Contents

Original Papers

Comparing Self-Reported Running Distance and Pace With a Commercial Fitness Watch Data: Reliability Study (e39211)
Garrett Bullock, Joanne Stocks, Benjamin Feakins, Zahra Alizadeh, Amelia Arundale, Stefan Kluzek. ................................. 24

Prevalence of Body Dysmorphic Disorder in the Spanish Population: Cross-Sectional Web-Based Questionnaire Study (e46515)
Álvaro Loewen, Hilario Blasco-Fontecilla, Chao Li, Marcos Bella-Fernández, Belén Ruiz-Antorán. ................................. 33

Methodological Insights on Recruitment and Retention From a Remote Randomized Controlled Trial Examining the Effectiveness of an Alcohol Reduction App: Descriptive Analysis Study (e51839)
Melissa Oldham, Larisa Dinu, Gemma Loebenberg, Matt Field, Matthew Hickman, Susan Michie, Jamie Brown, Claire Garnett. ................................. 48

Remote Delivery of the Cuidándome Telehealth Intervention for Self-Management of Depression and Anxiety Among Latina Immigrant Women: Randomized Controlled Trial (e52969)
Carmen Alvarez, Subhash Aryal, Elizabeth Vrany, Maria Sanchez R, Rosalpie Quiles, Lia Escobar-Acosta, Felicia Hill-Briggs. ................................. 61

Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care After Surgical Abortion (the FACTS Study Phase 3): Mixed Methods Prospective Pilot Study (e46284)
Roopan Gill, Gina Ogilvie, Wendy Norman, Brian Fitzsimmons, Ciana Maher, Regina Renner. ................................. 77

The Sukaribit Smartphone App for Better Self-Management of Type 2 Diabetes: Randomized Controlled Feasibility Study (e46222)
Cecilia Josefsson, Thea Liljeroos, Margareta Hellgren, Ulrika Pöder, Mariann Hedström, Erik Olsson. ................................. 91

Office Workers’ Views About the Uses, Concerns, and Acceptance of Hand Hygiene Data Collected From Smart Sanitizers: Exploratory Qualitative Interview Study (e47308)
Sophie Rutter, Sally Sanger, Andrew Madden, Sukaina Ehdeed, Catherine Stones. ................................. 106

Population-Level Portal-Based Anxiety and Depression Screening Perspectives in HIV Care Clinicians: Qualitative Study Using the Consolidated Framework for Implementation Research (e48935)
Daniela Zimmer, Erin Staab, Jessica Ridgway, Jessica Schmitt, Melissa Franco, Scott Hunter, Darnell Motley, Neda Laiteerapong. ................................. 119

Dentists’ Information Needs and Opinions on Accessing Patient Information via Health Information Exchange: Survey Study (e51200)
Shuning Li, Grace Felix Gomez, Huiping Xu, Anushri Rajapuri, Brian Dixon, Thankam Thyvalikakath. ................................. 131
Promoting Self-Efficacy of Individuals With Autism in Practicing Social Skills in the Workplace Using Virtual Reality and Physiological Sensors: Mixed Methods Study (e52157)
Sung-In Kim, So-young Jang, Taewan Kim, Bogaan Kim, Dayoung Jeong, Taehyung Noh, Mingon Jeong, Kaely Hall, Meelim Kim, Hee Yoo, Kyungsiik Han, Hwajung Hong, Jennifer Kim. ................................................................. 143

Mapping the Cardiometabolic Patient Experience and Self-Care Behaviors to Inform Design, Implementation, and Persistent Use of Digital Health Care Solutions: Mixed Methods Study (e43683)
Jan Liska, Marie Mical, Christophe Maillard, Cécile Dessapt, Europa Bendig, Daniel Mai, John Piette, Sabina De Geest, Guillaume Fontaine. 1

Linguistic Variables and Gender Differences Within a Messenger-Based Psychosocial Chat Counseling Service for Children and Adolescents: Cross-Sectional Study (e51795)
Zeki Ele, Sabrina Baldofski, Elisabeth Kohls, Melanie Eckert, Shadi Saee, Julia Thomas, Richard Wundrack, Christine Rummel-Kluge. ................................................................. 168

Rosie, a Health Education Question-and-Answer Chatbot for New Mothers: Randomized Pilot Study (e51361)
Quynh Nguyen, Elizabeth Aparicio, Michelle Jasczynski, Amara Channell Doig, Xiaohe Yue, Heran Mane, Neha Srikanth, Francis Gutierrez, Nataly Delcid, Xin He, Jordan Boyd-Graber. ................................................................. 183

Clinical Needs Assessment of a Machine Learning–Based Asthma Management Tool: User-Centered Design Approach (e45391)
Lu Zheng, Joshua Ohde, Shauna Overgaard, Tracey Brereton, Kristelle Jose, Chung-II Wi, Kevin Peterson, Young Juhn. ................................................................. 196

Exploring Adolescents’ Attitudes Toward Mental Health Apps: Concurrent Mixed Methods Study (e50222)
Helene Høgsdal, Henriette Kyrrstad, Marte Rye, Sabine Kaiser. ................................................................. 207

The Impact of Social Media Use on Mental Health and Family Functioning Within Web-Based Communities in Saudi Arabia: Ethnographic Correlational Study (e44923)
Bdour Alwuqaysi, Allie Abdul-Rahman, Rita Borgo. ................................................................. 217

Enabling Personalization for Digital Cognitive Stimulation to Support Communication With People With Dementia: Pilot Intervention Study as a Prelude to AI Development (e51732)
Nick Hird, Tohmi Osaki, Samik Ghosh, Sucheendra Palaniappan, Kiyoshi Maeda. ................................................................. 232

Designing and Developing a Mobile App for Management and Treatment of Gestational Diabetes in Nepal: User-Centered Design Study (e50823)
Aarthi Shanmugavel, Prabin Shakya, Archana Shrestha, Jyoti Nepal, Abha Shrestha, Jean-Francois Daneault, Shristi Rawal. ................................................................. 250

Preoperative Virtual Reality to Expose Patients With Breast Cancer to the Operating Room Environment: Feasibility and Pilot Case Series Study (e46367)
Jordana Sommer, Kristin Reynolds, Pamela Hebbard, Michael Smith, Natalie Mota, W Mutch, Jessica Maples-Keller, Leslie Roos, Renée El-Gabalawy. ................................................................. 267

Educational Video Intervention to Improve Health Misinformation Identification on WhatsApp Among Saudi Arabian Population: Pre-Post Intervention Study (e50211)
Ebtihal Alsaaal, Sharifah AlDossary. ................................................................. 284

A Blended Intervention Targeting Emotion Dysregulation in Adults With Attention-Deficit/Hyperactivity Disorder: Development and Feasibility Study (e53931)
Emilie Nordby, Frode Guribye, Viktor Schenning, Sander Andersen, Jonna Kunts, Astrid Lundervold. ................................................................. 299

Designing a Web-Based Navigation Tool to Support Access to Youth Mental Health Services: Qualitative Study (e48945)
Alison Calear, Philip Batterham, Sonia McCallum, Michelle Banfield, Elizabeth Moore, Natalie Johnson, Alyssa Morse. ................................................................. 315
Development and Usability Testing of an mHealth Tool for Trauma-Informed Prevention of Substance Use, HIV Acquisition, and Risky Sexual Behaviors Among Adolescents: Mixed Methods Study (e52835)
Carla Danielson, Angela Moreland, Austin Hahn, Devin Banks, Kenneth Ruggiero. 326

Virtual Exercise in Medicine: A Proof of Concept in a Healthy Population (e45637)
Barbara Le Roy, Charles Martin-Krumm, Charlotte Poupon, Raphaëlle Richieri, Eric Malbos, Fanny Barthélémy, Eric Guedj, Marion Trousselard. 342

Experiences Among Patients With Cystic Fibrosis in the MucoExocet Study of Using Connected Devices for the Management of Pulmonary Exacerbations: Grounded Theory Qualitative Research (e38064)
Maxime Morsa, Amélie Perrin, Valérie David, Gilles Rault, Enora Le Roux, Corinne Alberti, Rémi Gagnayre, Dominique Pougeon Bertrand. 358

Culturally Adapting the World Health Organization Digital Intervention for Family Caregivers of People With Dementia (iSupport): Community-Based Participatory Approach (e46941)
Anna Messina, Rebecca Amati, Anna Annoni, Beatrice Bano, Emiliano Albanese, Maddalena Fiordelli. 376

A Bluetooth-Enabled Device for Real-Time Detection of Sitting, Standing, and Walking: Cross-Sectional Validation Study (e47157)
Reza Daryabegi-Khotbehsara, Jonathan Rawstorn, David Dunstan, Sheikh Shariful Islam, Mohamed Abdelrazek, Abbas Kouzani, Poojith Thummala, Jenna McVicar, Ralph Maddison. 391

Demographics and Social Factors Associated With Persistent Nonuse of Video Appointments at a Multisite Health Care Institution: Cross-Sectional Study (e50572)
Pravesh Sharma, Celia Kamath, Tabetha Brockman, Anne Roche, Pamela Sincirope, Ruoxiang Jiang, Paul Decker, Vanessa Pazdernik, Christi Patten. 405

Transfer of Knowledge on Pneumoconiosis Care Among Rural-Based Members of a Digital Community of Practice: Cross-Sectional Study (e52414)
Brian Soller, Orrin Myers, Akshay Sood. 420

Improving Medical Photography in a Level 1 Trauma Center by Implementing a Specialized Smartphone-Based App in Comparison to the Usage of Digital Cameras: Prospective Panel Study (e47572)
Jan El Barbari, Maxim Fikuart, Nils Beisemann, Michael Müller, Hannah Syrek, Paul Grützner, Jochen Franke, Sven Vetter. 431

Machine Learning and Symptom Patterns in Degenerative Cervical Myelopathy: Web-Based Survey Study (e54747)
Alvaro Touzet, Tanzil Ruijeedawa, Colin Munro, Konstantinos Margetis, Benjamin Davies. 441

Testing a Behavioral Activation Gaming App for Depression During Pregnancy: Multimethod Pilot Study (e44029)
Rachel Vanderkruik, Craig Ferguson, Lauren Kobylski, Joseph Locascio, Gabriella Hamlett, Parker Killenber, Robert Lewis, Noah Jones, Ella Rossa, Hannah Dineen, Rosalind Picard, Lee Cohen. 449

Telephone-Based Training Intervention for Using Digital Communication Technologies for Social Housing Residents During the COVID-19 Pandemic: Mixed Methods Feasibility and Acceptability Evaluation (e45506)
Tim Walker, Sarah Buckingham, Ria Poole, Lewis Elliott, Tamaryn Menneer, Gengyang Tu, Karyn Morrissey. 468

Acceptance of a Web-Based Intervention in Individuals Who Committed Sexual Offenses Against Children: Cross-Sectional Study (e48880)
Sonja Schröder, Claudia Buntrock, Louisa Neumann, Jürgen Müller, Peter Fromberger. 485

Development and Implementation of Digital Diagnostic Algorithms for Neonatal Units in Zimbabwe and Malawi: Development and Usability Study (e54274)
Hannah Gannon, Leyla Larsson, Simbarashe Chimhuya, Marcia Mangiza, Emma Wilson, Erin Kesler, Gwendoline Chimhini, Felicity Fitzgerald, Gloria Zailani, Caroline Crehan, Nushrat Khan, Tim Hull-Bailey, Yali Sassoon, Morris Baradza, Michelle Heyes, Msandeni Chiulume. 497
Patient Phenotyping for Atopic Dermatitis With Transformers and Machine Learning: Algorithm Development and Validation Study (e52200)
Andrew Wang, Rachel Fulton, Sy Hwang, David Margolis, Danielle Mowery. 512

Examining a Remote Group-Based Type 2 Diabetes Self-Management Education Program in the COVID-19 Era Using the ORBIT Model: Small 6-Week Feasibility Study (e46418)
Madison Hiemstra, Sonja Reichert, Marc Mitchell. 527

Home Blood Pressure Telemonitoring Technology for Patients With Asymptomatic Elevated Blood Pressure Discharged From the Emergency Department: Pilot Study (e49592)
Karen Tran, Meagan Mak, Laura Kuyper, Jesse Bittman, Birinder Mangat, Heather Lindsay, Chad Kim Sing, Liang Xu, Hubert Wong, Martin Dawes, Nadia Khan, Kendall Ho. 535

Acceptability of a Self-Led Mindfulness-Based Intervention for Teens with Type 1 Diabetes: Pilot Randomized Controlled Trial (e45659)
Tori Humiston, Caroline Cummings, Stephen Suss, Laura Cohen, Holly Hazlett-Stevens, Amy Hughes Lansing. 549

“I Just Wanted a Dentist in My Phone”—Designing Evidence-Based mHealth Prototype to Improve Preschool Children’s Oral and Dental Health: Multimethod Study of the Codevelopment of an App for Children’s Teeth (e49561)
Waraf Al-yaseen, Daniela Raggio, Mariana Araujo, Nicola Innes. 558

Administrative Dashboard for Monitoring Use of a Web-Based Parent Training Intervention: Usability Study (e53439)
Susan Breitenstein, Julia Berteletti, Shea Smoske, Charles Barger, Kyrie Tipps, Nathan Helsabeck. 575

Web-Based Mindfulness Meditation as an Adjunct to Internet-Delivered Cognitive Behavioral Therapy for Public Safety Personnel: Mixed Methods Feasibility Evaluation Study (e54132)
Caeleigh Landry, Hugh McCall, Janine Beahm, Nickolai Titov, Blake Dear, R Carleton, Heather Hadjistavropoulos. 585

Formative Evaluation of a Comprehensive Self-Management Intervention for Irritable Bowel Syndrome, Comorbid Anxiety, and Depression: Mixed Methods Study (e43286)
Kendra Kamp, Pei-Lin Yang, Emily Friedman, Alejandra Lopez, Sarah Irribarren, Pamela Barney, Sean Munson, Margaret Heitkemper, Rona Levy. 603

Investigating the Potential of a Conversational Agent (Phyllis) to Support Adolescent Health and Overcome Barriers to Physical Activity: Co-Design Study (e51571)
Richard Moore, Abdel-Karim Al-Tamimi, Elizabeth Freeman. 615

Improvements in Adolescents' Disordered Eating Behaviors in a Collaborative Care Digital Mental Health Intervention: Retrospective Observational Study (e54253)
Landry Huffman, Darian Lawrence-Sidebottom, Aislinn Beam, Amit Parikh, Rachael Guerra, Monika Roots, Jennifer Huberty. 634

Data Representation Structure to Support Clinical Decision-Making in the Pediatric Intensive Care Unit: Interview Study and Preliminary Decision Support Interface Design (e49497)
Najia Yakob, Sandrine Laliberté, Philippe Doyon-Poulin, Philippe Jouvet, Rita Noumeir. 646

Comparison of Blended Learning With Traditional Dermatology Learning for Medical Students: Prospective Evaluation Study (e49616)
Cristiana Silva, Cidia Vasconcellos, Murilo Souza, Juliana Fernandes, Vitoria Rego. 667
Support for Chronic Pain Management for Breast Cancer Survivors Through Novel Digital Health Ecosystems: Pilot Usability Study of the PainRELife Mobile App (e51021)
Marianna Masiero, Chiara Filipponi, Elisa Fragale, Silvia Pizzoli, Elisabetta Munzone, Alessandra Milani, Luca Guido, Vittorio Guardamagna, Sara Marcegglia, Roberto Prandin, Marco Prenassi, Annamaria Caruso, Vania Manzelli, Chiara Savino, Costanza Conti, Federica Rizzi, Alice Casalino, Giulia Candiani, Francesca Memini, Luca Chiveri, Andrea Vitali, Massimo Corbo, Roberto Grasso, Florence Didier, Roberta Ferrucci, Gabriella Pravettoni.

Mental Health and Well-Being in Racial or Ethnic Minority Individuals After Using a Faith and Prayer Mobile App (Pray.com): Feasibility and Preliminary Efficacy Trial (e52560)
Breanne Laird, Sabrina Zuniga, Joshua Hook, Daryl Van Tongeren, Lynda Joeman, Jennifer Huberty.

Attitudes Toward School-Based Surveillance of Adolescents' Social Media Activity: Convergent Parallel Mixed Methods Survey (e46746)
Colin Burke, Cynthia Triplett, Caryn Rubanovich, Melissa Karnaze, Cinnamon Bloss.

Internet Use and Effects on Mental Well-being During the Lockdown Phase of the COVID-19 Pandemic in Younger Versus Older Adults: Observational Cross-Sectional Study (e48824)
Chou Yu, Nien Tou, James Low.

Explainable Machine Learning Classification to Identify Vulnerable Groups Among Parenting Mothers: Web-Based Cross-Sectional Questionnaire Study (e47372)
Akiko Hanao, Tetsuo Ishikawa, Shoko Sugao, Makoto Fujii, Kei Hirai, Hiroko Watanabe, Masayo Matsuzaki, Goji Nakamoto, Toshihiro Takeda, Yasuji Kitabatake, Yuichi Itoh, Masayuki Endo, Tadashi Kimura, Eiryo Kawakami.

Tumor Immunotherapy–Related Information on Internet-Based Videos Commonly Used by the Chinese Population: Content Quality Analysis (e50561)
Chen-xu Ni, Yi-bo Fei, Ran Wu, Wen-xiang Cao, Wenhao Liu, Fang Huang, Fu-ming Shen, Dong-jie Li.

Remote Self-Administration of Cognitive Screeners for Older Adults Prior to a Primary Care Visit: Pilot Cross-Sectional Study of the Reliability and Usability of the MyCog Mobile Screening App (e54299)
Stephanie Young, Elizabeth Dworak, Greg Byrne, Callie Jones, Lihua Yao, Julia Yoshino Benavente, Maria Diaz, Laura Curtis, Richard Gershon, Michael Wolf, Cindy Nowinski.

Vision-Language Model for Generating Textual Descriptions From Clinical Images: Model Development and Validation Study (e32690)
Jia Ji, Yongshuai Hou, Xinyu Chen, Youcheng Pan, Yang Xiang.

A Digitally Enabled Combined Lifestyle Intervention for Weight Loss: Pilot Study in a Dutch General Population Cohort (e38891)
Rahul Gannamani, José Castela Forte, Pytrik Folkertsmia, Sven Hermans, Sridhar Kumaraswamy, Sipko van Dam, Niels Chavannes, Hendrikus van Os, Hanno Pijl, Bruce Wolfenbuttel.

Predictors of Mental Health Literacy in a Sample of Health Care Major Students: Pilot Evaluation Study (e43770)
Pia Tohme, Nour Ali Fadel, Nour Yaktine, Rudy Abi-Habib.

Exploring User Perspectives on Brief Reflective Questioning Activities for Stress Management: Mixed Methods Study (e47360)
Ananya Bhattacharjee, Pan Chen, Abhijoy Mandal, Anne Hsu, Katie O'Leary, Alex Mariakakis, Joseph Williams.

Investigation of the Impact of Wellinks on the Quality of Life and Clinical Outcomes in Patients With Chronic Obstructive Pulmonary Disease: Interventional Research Study (e47555)
Kerry Pierz, Nicholas Locantore, Gretchen McCreary, Robert Calvey, Nickole Hackney, Pooja Doshi, John Linnell, Abrammy Sundaramoorthy, Carol Reed, Julie Yates.
The Impact of Intervention Design on User Engagement in Digital Therapeutics Research: Factorial Experiment With a Mixed Methods Study (e51225)

Hyerim Lee, Eung Choi, Jung Shin, Tae-Gyun Kim, Jooyoung Oh, Bokyoung Shin, Jung Sim, Jaeyong Shin, Meelim Kim. 829


Zahra Zandesh. 844

Evaluating the Intensity of Exposure to MTV Shuga, an Edutainment Program for HIV Prevention: Cross-Sectional Study in Eastern Cape, South Africa (e44111)


Nonuse of Blended Web-Based and Face-To-Face Cognitive Behavioral Therapy for Alcohol Use Disorder: Qualitative Study (e45471)

Kristine Tarp, Regina Christiansen, Randi Bilberg, Simone Borkner, Caroline Dalsgaard, Marie Folker, Anette Nielsen. 880

The Effect of Web-Based Culinary Medicine to Enhance Protein Intake on Muscle Quality in Older Adults: Randomized Controlled Trial (e49322)

Emily Salas-Groves, Michelle Alcorn, Allison Childress, Shannon Galyean. 889

Using #ActuallyAutistic on Twitter for Precision Diagnosis of Autism Spectrum Disorder: Machine Learning Study (e52660)

Aditi Jaiswal, Peter Washington. 905

Determining the Right Levels of Health Coaching and Heart Rate Variability Biofeedback in a Workplace Behavior Change Intervention: Multiphase Optimization Strategy Preparation Study (e47181)

Sean Locke, Jenna Osborne. 918

Measurement of Head Circumference Using a Smartphone: Feasibility Cohort Study (e54194)

Stefan Yordanov, Kalsoom Akhter, Jye Quan Teh, Jawad Naushahi, Ibrahim Jalloh. 936

Use Intention and User Expectations of Human-Supported and Self-Help eHealth Interventions: Internet-Based Randomized Controlled Trial (e38803)

Talia Cohen Rodrigues, Thomas Reijnders, Linda Breeman, Veronica Janssen, Roderik Kraaijenhagen, Douwe Atsma, Andrea Evers. 947

Effects of Embodiment in Virtual Reality for Treatment of Chronic Pain: Pilot Open-Label Study (e34162)

Adam Saby, Anthony Alvarez, David Smolins, James Petros, Lincoln Nguyen, Michael Trujillo, Oytun Aygun. 957

Contactless Monitoring System Versus Gold Standard for Respiratory Rate Monitoring in Emergency Department Patients: Pilot Comparison Study (e44717)

Charlotte Goldfine, Md Oshim, Brittany Chapman, Deepak Ganesan, Tauhidur Rahman, Stephanie Carreiro. 967

An Intelligent Customer-Driven Digital Solution to Improve Perioperative Health Outcomes Among Children Undergoing Circumcision and Their Parents: Development and Evaluation (e52337)

Zhi Kwa, Jinqui Li, Dale Loh, Yang Lee, Guangyu Liu, Lixia Zhu, Minna Pikkarainen, Honggu He, Vidyadhar Mali. 978

Development of Cost-Effective Fatty Liver Disease Prediction Models in a Chinese Population: Statistical and Machine Learning Approaches (e53654)

Li Zhang, Yueqing Huang, Min Huang, Chun-Hua Zhao, Yan-Jun Zhang, Yi Wang. 993
Patient and Therapist Perceptions of a Publicly Funded Internet-Based Cognitive Behavioral Therapy (iCBT) Program for Ontario Adults During the COVID-19 Pandemic: Qualitative Study (e50113)
Serena Thapar, Megan Nguyen, Bilal Khan, Roz Fanaeyan, Vanessa Kishimoto, Rebecca Liu, Blanca Bolea-Alamañac, Marisa Leon-Carlyle, Anne O’Riordan, Maggie Keresteci, Onil Bhattacharyya. ................................................................. 1012

Assessment of a Daily Diary Study Including Biospecimen Collections in a Sample of Sexual and Gender Minority Young Adults: Feasibility and Acceptability Study (e52195)
Stephanie Cook, Erica Wood, Mariana Rodrigues, Janice Jachero Caldas, Maxline Delorme. ................................................................. 1025

Acceptance of Telemedicine by Specialists and General Practitioners in Cardiology Care: Cross-Sectional Survey Study (e49526)
Felix Muehlensiepen, Marie Hoffmann, Jonathan Nübel, Yury Ignatyev, Martin Heinze, Christian Butter, Anja Haase-Fielitz. ................................................................. 1042

Enhancing Health Care Accessibility and Equity Through a Geoprocessing Toolbox for Spatial Accessibility Analysis: Development and Case Study (e51727)
Soheil Hashtrakhani, David Schwartz, Arash Shaban-Nejad. ................................................................. 1055

Group, Blended and Individual, Unguided Online Delivery of Mindfulness-Based Cognitive Therapy for People With Cancer: Feasibility Uncontrolled Trial (e52338)
Nasim Badaghi, Mette van Kruijssbergen, Anne Speckens, Joëlle Vilé, Judith Prins, Saskia Kelders, Linda Kwakkenbos. ................................................................. 1066

Comprehensive Assessment and Early Prediction of Gross Motor Performance in Toddlers With Graph Convolutional Networks-Based Deep Learning: Development and Validation Study (e51996)
Sulim Chun, Sooyoung Jang, Jin Kim, Chanyoung Ko, Joohyun Lee, JaeSeong Hong, Yu Park, ................................................................. 1081

Perception of Apps for Mental Health Assessment With Recommendations for Future Design: United Kingdom Semistructured Interview Study (e48881)
Erin Funnell, Benedetta Spadaro, Nayra Martin-Key, Jiri Benacek, Sabine Bahn. ................................................................. 1094

Co-Production of a Flexibly Delivered Relapse Prevention Tool to Support the Self-Management of Long-Term Mental Health Conditions: Co-Design and User Testing Study (e49110)

Help-Seeking, Support, and Engagement in Gestational Diabetes Mellitus Online Communities on Facebook: Content Analysis (e49494)
Sheila Pham, Kate Churruca, Louise Ellis, Jeffrey Braithwaite. ................................................................. 1130

How to Identify e-Cigarette Brands Available in the United States During 2020-2022: Development and Usability Study (e47570)
Shaoying Ma, Adeeba Kaareen, Hojin Park, Yanyun He, Shuning Jiang, Zefeng Qiu, Zidian Xie, Dongmei Li, Jian Chen, Richard O’Connor, Geoffrey Fong, Ce Shang. ................................................................. 1136

Using a Novel Gameplay Intervention to Target Intrusive Memories After Work-Related Trauma: Iterative Qualitative Analysis of Intensive Care Unit Staff Experiences (e47458)
Priya Patel, Susan Brown, Bolian Guo, Emily Holmes, Lalitha Iyadurai, Jonathan Kingslake, Julie Highfield, Richard Morriss. ................................................................. 1145

A Mobile Applet for Assessing Medication Adherence and Managing Adverse Drug Reactions Among Patients With Cancer: Usability and Utility Study (e50528)
Chenxu Ni, Yi-fu Wang, Yun-ting Zhang, Min Yuan, Qing Xu, Fu-ming Shen, Dong-Jie Li, Fang Huang. ................................................................. 1167

Using mHealth to Improve Communication in Adult Day Services Around the Needs of People With Dementia: Mixed Methods Assessment of Acceptability and Feasibility (e49492)
Amy Zheng, Marissa Bergh, Komal Patel Murali, Tina Sadarangani. ................................................................. 1177
The Impact of a Web-Based Restorative Dentistry Course on the Learning Outcomes of Dental Graduates: Pre-Experimental Study (e51141)
Rasha Al-Sbei, Jawdat Ataya, Issam Jamous, Mayssoon Dashash .................................................. 1190

Comparison of the Discrimination Performance of AI Scoring and the Brixia Score in Predicting COVID-19 Severity on Chest X-Ray Imaging: Diagnostic Accuracy Study (e46817)
Eric Tenda, Reyhan Yunus, Benny Zulkarnaen, Muhammad Yugo, Ceva Pitoyo, Moses Asaf, Tiara Islamiyati, Arierta Pujitresnani, Andry Setiadharma, Joshua Hennrina, Cleopas Rumende, Vally Wulani, Kuntjoro Harimurti, Aida Lydia, Hamzah Shatri, Pradana Soewondo, Prasandhya Yusuf. . . 1 1 9 9

Developing a Novel Web-Based Self-Management Support Intervention for Polycystic Ovary Syndrome: Mixed Methods Study With Patients and Health Care Professionals (e52427)
Carol Percy, Andrew Turner, Charys Orr .......................................................... 1214

A Deep Learning–Based Approach for Prediction of Vancomycin Treatment Monitoring: Retrospective Study Among Patients With Critical Illness (e45202)
Dohyun Kim, Hyun-Soo Choi, Dong-Hoon Lee, Minkyu Kim, Yoon Kim, Seon-Sook Han, Yeonjeong Heo, Ju-Hee Park, Jinkyeong Park .................................................. 1236

Optimization of Using Multiple Machine Learning Approaches in Atrial Fibrillation Detection Based on a Large-Scale Data Set of 12-Lead Electrocardiograms: Cross-Sectional Study (e47803)
Beau Chuang, Albert Yang .......................................................... 1250

Designing an App to Support Measurement-Based Peer Supervision of Frontline Health Workers Delivering Brief Psychosocial Interventions in Texas: Multimethod Study (e55205)
Anubhuti Poudyal, Delta-Marie Lewis, Sarah Tah, Alyssa Martinez, Lauren Magoun, Y Ho, Natali Carmio, John Naslund, Katherine Sanchez, Neal Lesh, Vikram Patel .......................................................... 1259

Improving Behavioral-Based Safety Training in Using Verbal Commands Through a Theory-Driven and Feedback-Based Nonimmersive Virtual Reality Game: Development and Usability Study (e48080)
Chukwudebube Atagbuzia, Ean Ng, Ganapathy Natarajan .......................................................... 1276

Implementing a Sodium-Glucose Cotransporter 2 Inhibitor Module With a Software Tool (Future Health Today): Qualitative Study (e50737)
Matthew Suen, Jo-Anne Manski-Nankervis, Caroline McBride, Natalie Lumsden, Barbara Hunter .......................................................... 1286

Characteristics of Users of a Digital Hypnotherapy Intervention for Hot Flashes: Retrospective Study (e53555)
Morgan Snyder, Gary Elkins .......................................................... 1294

Adapting the Number of Questions Based on Detected Psychological Distress for Cognitive Behavioral Therapy With an Embodied Conversational Agent: Comparative Study (e50056)
Kazuhiro Shidara, Hiroki Tanaka, Hiroyoshi Adachi, Daisuke Kanayama, Takashi Kudo, Satoshi Nakamura .......................................................... 1302

A Closed-Loop Digital Health Tool to Improve Depression Care in Multiple Sclerosis: Iterative Design and Cross-Sectional Pilot Randomized Controlled Trial and its Impact on Depression Care (e52809)
Kyra Henderson, Jennifer Reihm, Kanisha Koshal, Jaeleene Wijangco, Narender Sara, Nicolette Miller, Marianne Doyle, Alicia Mallory, Judith Sheridan, Chu-Yueh Guo, Lauren Oommen, Katherine Rankin, Stephan Sanders, Anthony Feinstein, Christina Mangurian, Riley Bove .......................................................... 1320

Assessing a GPS-Based 6-Minute Walk Test for People With Persistent Pain: Validation Study (e46820)
Joshua Simmich, Nicole Andrews, Andrew Claus, Megan Murdoch, Trevor Russell .......................................................... 1334

Using Principles of Digital Development for a Smartphone App to Support Data Collection in Patients With Acute Myocardial Infarction and Physical Activity Intolerance: Case Study (e33868)
Diana Cáceres Rivera, Luz Rojas, Lyda Rojas, Diana Gomez, David Castro Ruiz, Luis López Romero .......................................................... 1348
Preliminary Efficacy of a Cognitive Behavioral Therapy–Based Smartphone App for Smoking Cessation in China: Randomized Controlled Pilot Trial (e48050)
Shanshan Chen, Jingsong Tang, Congyang Wu, Ge Zhang, Jing Zhang, Yanhui Liao. .................................................. 1356

Use of Digital COVID-19 Exposure Notifications at a Large Gathering: Survey Analysis of Public Health Conference Attendees (e50716)
Caitlin Drover, Adam Elder, Brandon Guthrie, Debra Revere, Nicole Briggs, Laura West, Amanda Higgins, William Lober, Bryant Karras, Janet Baseman. ................................................................. 1369

Developing a Text Messaging Intervention to Prevent Binge and Heavy Drinking in a Military Population: Mixed Methods Development Study (e55041)
Chase Aycock, Indika Mallawaarachchi, Xin-Gun Wang, Daniel Cassidy, Jordan Ellis, Robert Klesges, G Talcott, Kara Wiseman. ................................................................. 1382

Association of Digital Engagement With Relaxation Tools and Stress Level Reduction: Retrospective Cohort Study (e50506)
Inbar Breuer-Asher, Marilyn Ritholz, David Horwitz, Omar Manejwala, Ephraim Behar, Yifat Fundoiano-Hershcovitz. ................................................................. 1396

An Ecological Mobile Momentary Intervention to Support Dynamic Goal Pursuit: Feasibility and Acceptability Study (e49857)
Ciarán O’Driscoll, Aneesha Singh, Iya Chichua, Joachim Cloidic, Anjali Desai, Dara Nikolova, Alex Yap, Irene Zhou, Stephen Pilling. ................................................................. 1408

Effectiveness of a Web-based and Mobile Therapy Chatbot on Anxiety and Depressive Symptoms in Subclinical Young Adults: Randomized Controlled Trial (e47960)
Stanisław Karkosz, Robert Szymański, Katarzyna Sanna, Jarosław Michalowski. ................................................................. 1421

Parents’ User Experience Accessing and Using a Web-Based Map of COVID-19 Recommendations for Health Decision-Making: Qualitative Descriptive Study (e53593)
Samantha Cyrkot, Lisa Hartling, Shannon Scott, Sarah Elliott. ................................................................. 1437

Assessing Knowledge, Competence, and Performance Following Web-Based Education on Early Breast Cancer Management: Health Care Professional Questionnaire Study and Anonymized Patient Records Analysis (e50931)
Michael Gnant, Khatijah Abdullah, Frances Boyle, Chiun-Sheng Huang, Katie Bickford, Sola Neunie, Alexander Noble, Anne Nunn, Caroline Sproat, Nadia Harbeck, Carlos Barrios. ................................................................. 1449

Development of Therapeutic Alliance and Social Presence in a Digital Intervention for Pediatric Concussion: Qualitative Exploratory Study (e49133)
Kiarah O’Kane, Thalia Otamendi, Noah Silverberg, Esther Choi, Veronik Sicard, Roger Zemek, Katherine Healey, Olivier Brown, Lauren Butterfield, Anda Smith, Gary Goldfield, Rachel Kardash, Bechara Saab, Andrée-Anne Ledoux, Molly Cairncross. ................................................................. 1464

Low-Fidelity Prototype of a Sensor-Dependent Interaction Platform: Formative Evaluation With Informal Caregivers of Older Adults With Cognitive Impairment (e53402)
Nikita Sharma, Karen Grotenshuys, J Gemert-Pijnen, Harri Oinas-Kukkonen, L Braakman-Jansen. ................................................................. 1478

Community Members’ Perceptions of a Resource-Rich Well-Being Website in California During the COVID-19 Pandemic: Qualitative Thematic Analysis (e55517)
MarySue Heilemann, Jianchao Lai, Madonna Cadiz, Jocelyn Meza, Daniela Flores Romero, Kenneth Wells. ................................................................. 1499

Best Practices in Evolving Privacy Frameworks for Patient Age Data: Census Data Study (e47248)
Colin Moffatt, Jonah Leshin. ................................................................. 1515

Adapting mHealth Interventions (PrEPmate and DOT Diary) to Support PrEP Retention in Care and Adherence Among English and Spanish-Speaking Men Who Have Sex With Men and Transgender Women in the United States: Formative Work and Pilot Randomized Trial (e54073)
Albert Liu, Cat-Dancing Alleyne, Susanne Dobblecki-Lewis, Kimberly Koester, Rafael Gonzalez, Janie Vinson, Hyman Scott, Susan Buchbinder, Thiago Torres. ................................................................. 1523
User Friendliness and Perioperative Guidance Benefits of a Cataract Surgery Education App: Randomized Controlled Trial (e55742)
Rokas Gerbutavicius, David Merle, Armin Wolf, Spyridon Dimopoulos, Karsten Kortuem, Friederike Kortuem. 1541

Factors Explaining the Use of Web-Based Consultations With Physicians by Young and Middle-Aged Individuals in China: Qualitative Comparative Analysis (e50036)
Chunyu Zhang, Ning Hu, Rui Li, Aiping Zhu, Zhongguang Yu. 1555

A Novel Electronic Record System for Documentation and Efficient Workflow for Community Health Workers: Development and Usability Study (e52920)
Harshdeep Acharya, Kevin Sykes, Ton Neira, Angela Scott, Christina Pacheco, Matthew Sanner, Elizabeth Abliah, Kevin Oyowe, Edward Ellerbeck, K Greiner, Erin Corrineau, Sarah Finocchiaro-Kessler. 1565

Web-Based Emotion Regulation Training for Sexual Health: Randomized Controlled Trial (e50850)
Vinicius Jobim Fischer, Maila Rossato Holz, Joë Billieux, Gerhard Andersson, Claus Vögele. 1575

Using Smartphones to Enhance Vision Screening in Rural Areas: Pilot Study (e55270)
Zheng Wang, John Kempen, Gang Luo. 1590

The Role of Mental Health Stigma in University Students’ Satisfaction With Web-Based Stress Management Resources: Intervention Study (e50018)
Sohyun Cho, Laurianne Bastien, Julia Petrovic, Bilun Böke, Nancy Heath. 1597

Development and Evaluation of a Clinician-Vetted Dementia Caregiver Resources Website: Mixed Methods Approach (e54168)
Jaye McLaren, Dat Hoang-Gia, Eugenia Dorisca, Stephanie Hartz, Stuti Dang, Lauren Moo. 1609

Evaluation of an e-Learning Program for Community Pharmacists for Dispensing Emicizumab (Hemlibra) in France: Nationwide Cross-Sectional Study (e54656)
Valérie Chamouard, Julie Freysenge, Béatrice Clairaz-Mahiou, Felicia Ferrera Bibas, Laurie Fricicelli. 1620

Acceptability of a Self-Guided Lifestyle Intervention Among Young Men: Mixed Methods Analysis of Pilot Findings (e53841)
Jean Reading, Melissa Crane, Justin Guan, Ronston Jackman, Maria Thomson, Jessica LaRose. 1630

Fast Healthcare Interoperability Resources–Based Support System for Predicting Delivery Type: Model Development and Evaluation Study (e54109)
João Coutinho-Almeida, Alexandrina Cardoso, Ricardo Cruz-Correia, Pedro Pereira-Rodrigues. 1642

Mobile Technology Use in Clinical Research Examining Challenges and Implications for Health Promotion in South Africa: Mixed Methods Study (e48144)
Khuthala Mabetha, Larske Soepnel, Gugulethu Mabena, Molebogeng Motlhatlhedi, Lukhanyo Nyati, Shane Norris, Catherine Draper. 1652

Development and Implementation of an eHealth Oncohematonootric Program: Descriptive, Observational, Prospective Cohort Pilot Study (e49574)
Beatriz Sánchez-Quiliones, Cristina Antón-Maldonado, Nataly Ibarra Vega, Isabel Martorell Mariné, Amparo Santamaria. 1669

Implementation of Video-Based Care in Interdisciplinary Primary Care Settings at the Veterans Health Administration: Qualitative Study (e52830)
Claudia Der-Martirosian, Cynthia Hou, Sona Hovsepian, Maia Diarra Carter, Leonie Heyworth, Aram Dobalian, Lucinda Leung. 1681
Nudges and Prompts Increase Engagement in Self-Guided Digital Health Treatment for Depression and Anxiety: Results From a 3-Arm Randomized Controlled Trial (e52558)

Trevor van Mierlo, Renante Rondina, Rachel Fournier. ................................................................. 1695

Time Efficiency, Reliability, and User Satisfaction of the Tooth Memo App for Recording Oral Health Information: Cross-Sectional Questionnaire Study (e56143)

Palinee Detsomboonrat, Pagaporn Pisaratratik. ............................................................................. 1718

Race and Socioeconomic Status as Predictors of Willingness to Use Digital Mental Health Interventions or One-On-One Psychotherapy: National Survey Study (e49780)

Lorenzo Lorenzo-Luaces, Akash Wasil, Corinne Kacmarek, Robert DeRubeis. ......................... 1730

Usability and Feasibility Evaluation of a Web-Based and Offline Cybersecurity Resource for Health Care Organizations (The Essentials of Cybersecurity in Health Care Organizations Framework Resource): Mixed Methods Study (e50968)

Niki O’Brien, Roberto Fernandez Crespo, Fiona O’Driscoll, Mabel Prendergast, Deeph Chana, Ara Darzi, Saira Ghafur. .......................................................... 1743

Mental Distress, Label Avoidance, and Use of a Mental Health Chatbot: Results From a US Survey (e45959)

Kristin Kosyluk, Tanner Baeder, Karah Greene, Jennifer Tran, Cassidy Bolton, Nele Loecher, Daniel DiEva, Jerome Galea. ............................................................. 1755

Promoting a Patient-Centered Understanding of Safety in Acute Mental Health Wards: A User-Centered Design Approach to Develop a Real-Time Digital Monitoring Tool (e53726)

Gemma Louch, Kathryn Berzins, Lauren Walker, Gemma Wormald, Kristin Blackwell, Michael Stephens, Mark Brown, John Baker. ......................................................... 1772

The Effects of a Single-Session Virtual Rumination Intervention to Enhance Cognitive Functioning in Veterans With Subjective Cognitive Symptoms: Multimethod Pilot Study (e48525)

Tara Austin, Jennifer Smith, Borsika Rabin, Laurie Lindamer, James Pittman, Staley Justice, Elizabeth Twamley, Crystal Lantrip. ......................................................... 1786

Novel Approach for Detecting Respiratory Syncytial Virus in Pediatric Patients Using Machine Learning Models Based on Patient-Reported Symptoms: Model Development and Validation Study (e52412)

Shota Kawamoto, Yoshihiko Morikawa, Naohisa Yahagi. ............................................................ 1796

Developing a Multiprofessional Mobile App to Enhance Health Habits in Older Adults: User-Centered Approach (e54214)

Andressa Sobrinho, Grace Gomes, Carlos Bueno Júnior. ............................................................ 1806

Clinical Decision Support System for Guidelines-Based Treatment of Gonococcal Infections, Screening for HIV, and Prescription of Pre-Exposure Prophylaxis: Design and Implementation Study (e53000)

Saugat Karki, Sarah Shaw, Michael Lieberman, Alejandro Pérez, Jonathan Pincus, Priya Jakhmola, Amrita Tailor, Qyinkansola Ogunrinde, Danielle Sill, Shane Morgan, Miguel Alvarez, Jonathan Todd, Dawn Smith, Ninad Mishra. ......................................................... 1822

Assessing the Accuracy of Generative Conversational Artificial Intelligence in Debunking Sleep Health Myths: Mixed Methods Comparative Study With Expert Analysis (e55762)

Nicola Bragazzi, Sergio Garbarino. .................................................................................................. 1836

Experiences, Lessons, and Challenges With Adapting REDCap for COVID-19 Laboratory Data Management in a Resource-Limited Country: Descriptive Study (e50897)

Kagiso Ndlovu, Kabelo Gaumo, Onalenna Makhura, Robin Hu, Nkwezi Motlhogola, Audrey Masizana, Emily Lo, Thongbotho Mphoyakgosi, Sihle Moyob. ......................................................... 1848

Landscape of Digital Technologies Used in the National Health Service in England: Content Analysis (e51859)

Jake Allcock, Mengdie Zhuang, Shuyang Li, Xin Zhao. .................................................................. 1863
AI-Led Mental Health Support (Wysa) for Health Care Workers During COVID-19: Service Evaluation (e51858)
Christel Chang, Chaitali Sinha, Madhavi Roy, John Wong .......................................................... 1875

Remote Symptom Monitoring Using Patient-Reported Outcomes in Patients With Chronic Kidney Disease: Process Evaluation of a Randomized Controlled Trial (e48173)
Birgith Grove, Annette de Thurah, Per Iversen, Ann Kvisgaard, Niels Hjollund, Regine Grytnes, Liv Schougaard .......................................................... 1884

Exploring the Use of Customized Links to Improve Electronic Engagement With Sexual and Reproductive Health Care Among Young African American Male Individuals: Web-Based Survey Study (e48371)
Sandy Arena, Mackenzie Adams, Jade Burns .............................................................................. 1901

Cancer Care Supportive Text Messaging Program (Text4Hope) for People Living With Cancer and Their Caregivers During the COVID-19 Pandemic: Longitudinal Observational Study (e53668)
Reham Shalaby, Wesley Vuong, Belinda Agyapong, April Gusnowski, Shireen Surood, Vincent Agyapong .......................................................... 1910

Objective Assessment of Physical Activity at Home Using a Novel Floor-Vibration Monitoring System: Validation and Comparison With Wearable Activity Trackers and Indirect Calorimetry Measurements (e51874)
Yuki Nakajima, Asami Kitayama, Yuji Ohta, Nobuhisa Motooka, Mayumi Kuno-Mizumura, Motohiko Miyachi, Shigeho Tanaka, Kazuko Ishikawa-Takata, Julien Tripette .......................................................... 1925

Developing Social Enhancements for a Web-Based, Positive Emotion Intervention for Alzheimer Disease Caregivers: Qualitative Focus Group and Interview Study (e50234)
Ian Kwok, Emily Lattie, Dershung Yang, Amanda Summers, Paul Cotten, Caroline Leong, Judith Moskowitz .......................................................... 1940

Characterizing Technology Use and Preferences for Health Communication in South Asian Immigrants With Prediabetes or Diabetes: Cross-Sectional Descriptive Study (e52687)
Lu Hu, Laura Wyatt, Farhan Mohsin, Sahnah Lim, Jennifer Zanowiak, Shinh Mammen, Sarah Hussain, Shahmir Ali, Deborah Onakomaiya, Hayley Belli, Angela Alfa, Nadia Islam ........................................................................... 1953

Acceptance, Satisfaction, and Preference With Telemedicine During the COVID-19 Pandemic in 2021-2022: Survey Among Patients With Chronic Pain (e53154)
Michael Harnik, Alina Scheidegger, Larissa Blätter, Zdenek Nemecek, Thomas Sauter, Andreas Limacher, Florian Reisig, Martin grosse Holtforth, Konrad Streitberger .......................................................... 1965

Attributes, Quality, and Downloads of Dementia-Related Mobile Apps for Patients With Dementia and Their Caregivers: App Review and Evaluation Study (e51076)
Tzu Chen, Shin-Da Lee, Wei-Fen Ma ......................................................................................... 1977

Precision Assessment of Real-World Associations Between Stress and Sleep Duration Using Actigraphy Data Collected Continuously for an Academic Year: Individual-Level Modeling Study (e53441)
Constanza Vidal Bustamante, Garth Coombs III, Habiballah Rahimi-Eichi, Patrick Mair, Jukka-Pekka Onnela, Justin Baker, Randy Buckner .......................................................... 2016

Effectiveness of a Smartphone App (Heia Meg) in Improving Decisions About Nutrition and Physical Activity: Prospective Longitudinal Study (e48185)
Christine Olsen, Daniel Lungu .................................................................................................. 2029

Community-Dwelling Older Adults’ Readiness for Adopting Digital Health Technologies: Cross-Sectional Survey Study (e54120)
Dietmar Assuerhofer, Giuliano Piccoliori, Adolf Engl, Angelika Mahnknecht, Barbara Piagg, Verena Barbieri, Nicoletta Colletti, Stefano Lombardo, Timon Gärtner, Waltraud Tappeiner, Heike Wieser, Christian Wiedermann ..........................................................
An Exploration of the Goodness of Fit of Web-Based Tools for M ori: Qualitative Study Using Interviews and Focus Groups (e50385)
Liesje Donkin, Marie-Claire Bidois-Putt, Holly Wilson, Penelope Hayward, Amy Chan. .......................................................... 2040

Assessment of Stress and Well-Being of Japanese Employees Using Wearable Devices for Sleep Monitoring Combined With Ecological Momentary Assessment: Pilot Observational Study (e49396)
Shotaro Kinoshita, Sayaka Hanashiro, Shiori Tsutsumi, Kiko Shiga, Momoko Kitazawa, Yasuyo Wada, Jun Inaishi, Kazuhiro Kashiwagi, Toshikazu Fukami, Yasumasa Mashimo, Kazumichi Minato, Taishiro Kishimoto. .......................................................... 2052

Group Cohesion and Necessary Adaptations in Online Hearing Voices Peer Support Groups: Qualitative Study With Group Facilitators (e51694)
Alison Branitsky, Eleanor Longden, Sandra Bucci, Anthony Morrison, Filippo Varese. .......................................................... 2069

Accuracy of the Apple Watch Series 4 and Fitbit Versa for Assessing Energy Expenditure and Heart Rate of Wheelchair Users During Treadmill Wheelchair Propulsion: Cross-sectional Study (e52312)
Marius Danielsson, Melanie Vergeer, Guy Plasqui, Julia Baumgart. .......................................................... 2081

Needs for Successful Engagement in Telemedicine Among Rural Older US Veterans and Their Caregivers: Qualitative Study (e50507)
Jacqueline Boudreau, Lauren Moo, Meaghan Kennedy, Jennifer Conti, Chitra Anwar, Camilla Pimentel, Kathryn Nearing, William Hung, Eileen Dryden. .......................................................... 2094

Patient and Provider Perspectives About the Use of Patient-Generated Health Data During Pregnancy: Qualitative Exploratory Study (e52397)
Sarah MacEwan, Ramona Olvera, Pallavi Jonnalagadda, Naleef Fareed, Ann McAlearney. .......................................................... 2106

Relationship Between Product Features and the Prices of e-Cigarette Devices Sold in Web-Based Vape Shops: Comparison Study Using a Linear Regression Model (e49276)
Yanyun He, Qian Yang, Yousef Alish, Shaoying Ma, Zefeng Qiu, Jian Chen, Theodore Wagener, Ce Shang. .......................................................... 2119

Decision Support for Managing Common Musculoskeletal Pain Disorders: Development of a Case-Based Reasoning Application (e44805)
Fredrik Granviken, Ottar Vasseljen, Kerstin Bach, Amar Jaiswal, Ingebrigt Meisingset. .......................................................... 2130

Medical Students' Perceptions on Identifying and Addressing Emotional Responses in Emergency Medicine: Pilot Investigation (e50827)
Anish Agarwal, Rachel Gonzales, Cory Munden, DaCarla Albright, Suzana Tsao. .......................................................... 2145

Factors Associated With Perception of Stigma Among Parents of Children With Cleft Lip and Palate: Cross-Sectional Study (e53353)
Yanan Zhang, Xinwen Zhang, Jinzhuo Jiang, Wanhua Xie, Daoman Xiang. .......................................................... 2152

Health Behavior Change Intervention Preferences Expressed by American Indian Cancer Survivors From a Southwest Tribal Community: Semistructured Interview Study (e51669)
Jennifer Erdrich, Felina Cordova-Marks, William Carson, Jennifer Bea, William Montfort, Cynthia Thomson. .......................................................... 2161

Exploring the Implementation of Shared Decision-Making Involving Health Coaches for Diabetes and Hypertension Self-Management: Qualitative Study (e51848)
Sungwon Yoon, Chao Tan, Jie Phang, Venice Liu, Wee Tan, Yu Kwan, Lian Low. .......................................................... 2171

The Effect of a Combined Intermittent Fasting Healthy Plate Intervention on Anthropometric Outcomes and Body Composition Among Adults With Overweight and Obesity: Nonrandomized Controlled Trial (e51542)
Effect of the COVID-19 Pandemic on Gambling Behavior in Mainland Chinese Gamblers in Macau: Cross-Sectional Survey Study (e45700)
Jinquan Zhou, Hong-Wai Ho, ChiBiu Chan ................................................................. 2196

Patient Satisfaction With the Health Care Services of a Government-Financed Health Protection Scheme in Bangladesh: Cross-Sectional Study (e49815)
Md Hasan, Md Rabbani, Orin Akter, Gazi Mehdi, Mohammad Ahmed, Sayem Ahmed, Mahbub Chowdhury .................................................. 2215

Dashboards to Support Implementation of the Quebec Alzheimer Plan: Evaluation Study With Regional and Professional Considerations (e55064)
GenevieveArsenault-Lapiere, Alexandra Lernay-Compagnat, Maxime Guillette, Yves Couturier, Victoria Massamba, Isabelle Dufour, Éric Maubert, ChristineFournier, Julie Denis, Caroline Morin, Isabelle Vedel .................................................. 2231

Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction: Retrospective Cohort Study (e48487)
Pin Zhang, Lei Wu, Ting-Ting Zou, ZiXuan Zou, JiaXin Tu, Ren Gong, Jie Kuang ................................................................. 2245

Differences in Psychological Inflexibility Among Men With Erectile Dysfunction Younger and Older Than 40 Years: Web-Based Cross-Sectional Study (e45998)
Junichi Saito, Hiroaki Kumano, Mohammad Ghazizadeh, Chigusa Shimokawa, Hideki Tanemura .................................................. 2256

Determining Distinct Suicide Attempts From Recurrent Electronic Health Record Codes: Classification Study (e46364)
Kate Bentley, Emily Madsen, Eugene Song, Yu Zhou, Victor Castro, Hyunjoon Lee, Younga Lee, Jordan Smoller .................................................. 2264

Synergizing the Behavior Change Wheel and a Cocreative Approach to Design a Physical Activity Intervention for Adolescents and Young Adults With Intellectual Disabilities: Development Study (e51693)
Laura Maenhout, Julie Latomme, Greet Cardon, Geert Crombez, Geert Van Hove, Sofie Compernolle .................................................. 2274

Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study (e47091)
Stephen Bonett, Willey Lin, Patrina Sexton Topper, James Wolfe, Jesse Golinkoff, Aayushi Deshpande, Antonia Villarruel, José Bauermeister .................................................. 2307

Virtual and Interprofessional Objective Structured Clinical Examination in Dentistry and Dental Technology: Development and User Evaluations (e44653)
MengWei Pang, YanLing Dong, XiaoHan Zhao, JiaWu Wan, Li Jiang, JinLin Song, Ping Ji, Lin Jiang .................................................. 2331

Use of Machine Learning Tools in Evidence Synthesis of Tobacco Use Among Sexual and Gender Diverse Populations: Algorithm Development and Validation (e49031)
Shaoying Ma, Shuning Jiang, Olivia Yang, XuanZhi Zhang, Yu Fu, Yusen Zhang, Aadeeba Kaareen, Meng Ling, Jian Chen, Ce Shang .................................................. 2343

Designing and Developing Online Training for Diabetes Prevention Program Coaches Using an Integrated Knowledge Translation Approach: Development and Usability Study (e50942)
Kaela Cranston, Natalie Grieve, Tineke Dineen, Mary Jung .................................................. 2352

Evaluation of Eligibility Criteria Relevance for the Purpose of IT-Supported Trial Recruitment: Descriptive Quantitative Analysis (e49347)
RominaBlasini, Cosima Strantz, Christian Gulden, Sven Helfer, Jakub Lidke, Hans-Ulrich Prokosch, Keywan Sohrabi, Henning Schneider .................................................. 2365

User Requirements in Developing a Novel Dietary Assessment Tool for Children: Mixed Methods Study (e47850)
Zoë van der Heijden, Femke de Gooijer, Guido Camps, Desiree Lucassen, Edith Feskens, Marlou Lasschuijt, Elske Brouwer-Brolsma .................................................. 2372
The Development and Use of a New Visual Tool (REVISIT) to Support Participant Recall: Web-Based Interview Study Among Older Adults (e52096)
Eileen Dryden, Chitra Anwar, Jennifer Conti, Jacqueline Boudreau, Meaghan Kennedy, William Hung, Kathryn Nearing, Camilla Pimentel, Lauren Moo.

Testing Behavioral Messages to Increase Recruitment to Health Research When Embedded Within Social Media Campaigns on Twitter: Web-Based Experimental Study (e48538)
Sandro Stoffel, Jing Law, Robert Kerrison, Hannah Brewer, James Flanagan, Yasemin Hirst.

Clinical Informatics Team Members' Perspectives on Health Information Technology Safety After Experiential Learning and Safety Process Development: Qualitative Descriptive Study (e53302)
Chantelle Recsky, Kathy Rush, Maura MacPhee, Megan Stowe, Lorraine Blackburn, Allison Muniak, Leanne Currie.

Attitudes, Barriers, and Motivators Toward Daily Walking and a Mobile App to Increase Walking Among Women: Web-Based Anonymous Survey (e48668)
Catherine Jones, Shikha Chandarana, Amita Vyas, Melissa Napolitano.

Brief Intervention as a Method to Reduce Z-Hypnotic Use by Older Adults: Feasibility Case Series (e51862)
Maria Bjelkarøy, Tone Simonsen, Tahreem Siddiqi, Sigrid Halset, Socheat Cheng, Ramune Grambaite, J rat Benth, Jennifer Gerwing, Espen Kristoffersen, Christofer Lundqvist.

Investigating the Impact of Prompt Engineering on the Performance of Large Language Models for Standardizing Obstetric Diagnosis Text: Comparative Study (e53216)
Lei Wang, Wenshuai Bi, Suling Zhao, Yinyao Ma, Longting Lv, Chenwei Meng, Jingru Fu, Hanlin Lv.

Efficacy of Digital Outreach Strategies for Collecting Smoking Data: Pragmatic Randomized Trial (e50465)
Lauren Kearney, Emily Jansen, Hasmeena Kathuria, Katrina Steiling, Kayla Jones, Allan Walkey, Nicholas Cordella.

mHealth App Usability Questionnaire for Stand-Alone mHealth Apps Used by Health Care Providers: Canadian French Translation, Cross-Cultural Adaptation, and Validation (Part 1) (e50839)
Julie Gagnon, Sebastian Probst, Julie Chartrand, Michelle Lalonde.

Predictive Criterion Validity of the Parsley Symptom Index Against the Patient-Reported Outcomes Measurement Information System-10 in a Chronic Disease Cohort: Retrospective Cohort Study (e53316)
Hants Williams, Sarah Steinberg, Kendall Leon, Ryan Vingum, Mengyao Hu, Robin Berzin, Heather Hagg, Patrick Hanaway.

Daily Activity Lifelogs of People With Heart Failure: Observational Study (e51248)
Teketo Tegegne, Ly-Duyen Tran, Rebecca Nourse, Cathal Gurrin, Ralph Maddison.

Designing and Validating a Novel Method for Assessing Delay Discounting Associated With Health Behaviors: Ecological Momentary Assessment Study (e48954)
Amanda Luken, Jill Rabinowitz, Jonathan Wells, David Sosnowski, Justin Strickland, Johannes Thurl, Gregory Kirk, Brion Maher.

The Influence of Human Connections and Collaboration on Research Grant Success at Various Career Stages: Regression Analysis (e49905)
Akiko Hashiguchi, Makoto Asashima, Satoru Takahashi.

Identifying Unmet Needs in Major Depressive Disorder Using a Computer-Assisted Alternative to Conventional Thematic Analysis: Qualitative Interview Study With Psychiatrists (e48894)
Michelle Worthington, Richard Christie, Aaron Masino, Sarah Kark.

Risk Identification in Perinatal Health Care Settings via Technology-Based Recruitment Methods: Comparative Study (e48823)
Jessica Beaty, Logan Zelenak, Spencer Gillon, Lucy McGoron, Gregory Goyert, Steven Ondersma.
Machine Learning–Based Approach for Identifying Research Gaps: COVID-19 as a Case Study (e49411)

Feasibility and Acceptability of Web-Based Structured Oral Examinations for Postgraduate Certification: Mixed Methods Preliminary Evaluation (e40868)
Vanessa Burch, Jessica McGuire, Eric Buch, Mike Sathekge, Francis M'bouaflou, Flavia Senkubuge, Johannes Fagan. ................................................................. 2559

Development of a Social Risk Score in the Electronic Health Record to Identify Social Needs Among Underserved Populations: Retrospective Study (e54732)
Elham Hatef, Hsien-Yen Chang, Thomas Richards, Christopher Kitchen, Janya Budaraju, Iman Foroughmand, Elyse Lasser, Jonathan Weiner. 2 5 7 2

Patient-Centered Approaches for Designing Destigmatizing Sexual Pain-Related Web-Based Platforms: Qualitative Study (e53742)
Abdul-Fatawu Abdulai, Hasti Naghdali, Heather Noga, Paul Yong. ........................................................................................................... 2583

Investigating the Roles and Responsibilities of Institutional Signing Officials After Data Sharing Policy Reform for Federally Funded Research in the United States: National Survey (e49822)
Jinyoung Baek, Jonathan Lawson, Vasiliki Rahimzadeh. ........................................................................................................... 2593

Automated Category and Trend Analysis of Scientific Articles on Ophthalmology Using Large Language Models: Development and Usability Study (e52462)
Hina Raja, Asim Munawar, Nikolaos Mylonas, Mohammad Delaaz, Yeganeh Madadi, Muhammad Elahi, Amr Hassan, Hashem Abu Serhan, Onur Inam, Luis Hernandez, Hao Chen, Sang Tran, Wuqaas Munir, Alaa Abd-Alrazaq, Siamak Yousefi. ........................................................................................................... 2603

Comparison of Self-Tracking Health Practices, eHealth Literacy, and Subjective Well-Being Between College Students With and Without Disabilities: Cross-Sectional Survey (e48783)
Soyoung Choi. ........................................................................................................... 2619

Examining and Comparing the Validity and Reproducibility of Scales to Determine the Variety of Vegetables Consumed: Validation Study (e55795)
Kaya Ominami, Osamu Kushida. ........................................................................................................... 2631

Assessing Priorities in a Statewide Cardiovascular and Diabetes Health Collaborative Based on the Results of a Needs Assessment: Cross-Sectional Survey Study (e55285)
Elizabeth Beverly, Sarah Koopman-Gonzalez, Jackson Wright, Kathleen Dungan, Harini Pallerla, Rose Gubitosi-Klug, Kristin Baughman, Michael Konstan, Shari Bolen. ........................................................................................................... 2641

Integrating Explainable Machine Learning in Clinical Decision Support Systems: Study Involving a Modified Design Thinking Approach (e50475)
Michael Shulha, Jordan Hovdebo, Vinita D’Souza, Francis Thibault, Rola Harmouche. ........................................................................................................... 2655

Recruitment Strategies in the Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization Study (meSH): Multicenter Survey Study (e48767)
Chinonyelum Nwosu, Hamda Khan, Rita Masese, Judith Nocek, Siera Gollan, Taniya Varughese, Sarah Bourne, Cindy Clesca, Sara Jacobs, Ana Baumann, Lisa Kiesges, Nirmish Shah, Jane Hankins, Matthew Smeltzer. ........................................................................................................... 2668

Development of a Management App for Postviral Fibromyalgia-Like Symptoms: Patient Preference-Guided Approach (e50832)
Marc Blanchard, Cinja Koller, Pedro Azevedo, Tiffany Prétat, Thomas Hugle. ........................................................................................................... 2678

Creating High-Quality Synthetic Health Data: Framework for Model Development and Validation (e53241)
Elaz Karimian Sichani, Aaron Smith, Khaled El Emam, Lucy Mosquera. ........................................................................................................... 2687
Momentary Factors and Study Characteristics Associated With Participant Burden and Protocol Adherence: Ecological Momentary Assessment (e49512)
Allan Tate, Angela Fertig, Junia de Brito, Emile Ellis, Christopher Carr, Amanda Trofholz, Jerica Berge. 2706

Assessing Electronic Health Literacy in Individuals With the Post–COVID-19 Condition Using the German Revised eHealth Literacy Scale: Validation Study (e52189)
Matthias Marsall, Hannah Dinse, Julia Schröder, Eva-Maria Skoda, Martin Teufel, Alexander Bäuerle. 2718

Proactive Identification of Patients with Diabetes at Risk of Uncontrolled Outcomes during a Diabetes Management Program: Conceptualization and Development Study Using Machine Learning (e54373)
Arash Khalilnejad, Ruo-Ting Sun, Tejaswi Kompala, Stefanie Painter, Roberta James, Yajuan Wang. 2732

Expanding Youth-Friendly HIV Self-Testing Services During the COVID-19 Pandemic: Qualitative Analysis of a Crowdsourcing Open Call in Nigeria (e46945)
Onyekachukwu Anikamadu, Oliver Ezechi, Alexis Engelhart, Ucheoma Nwaozuru, Chisom Obiez-Umeh, Ponmile Ogunjemite, Babatunde Bale, Daniel Nwachukwu, Titilola Gbaja-biamila, David Oladele, Adesola Musa, Stacey Mason, Temitope Ojo, Joseph Tucker, Juliet Iwelunmor. 2745

Real-World Gait Detection Using a Wrist-Worn Inertial Sensor: Validation Study (e50035)

Regulatory Issues in Electronic Health Records for Adolescent HIV Research: Strategies and Lessons Learned (e46420)
Sara Green, Sung-Jae Lee, Samantha Chahin, Meardith Pooler-Burgess, Monique Green-Jones, Sitaji Gurung, Angulique Outlaw, Sylvie Naar. 2773

ChatGPT as a Tool for Medical Education and Clinical Decision-Making on the Wards: Case Study (e51346)
Anthony Skryd, Katharine Lawrence. 2781

Acceptance of Medical Artificial Intelligence in Skin Cancer Screening: Choice-Based Conjoint Survey (e46402)
Inga Jagemann, Ole Wensing, Manuel Stegemann, Gerrit Hirschfeld. 2789

Digital Health Needs and Preferences During Pregnancy and the Postpartum Period: Mixed Methods Study (e48960)
Natalie Henrich, Alison Brinson, Alyssa Arnold, Hannah Jahnke. 2799

Acceptability and Feasibility of a Smartphone-Based Real-Time Assessment of Suicide Among Black Men: Mixed Methods Pilot Study (e48992)
Leslie Adams, Thomasina Watts, Aubrey Devinney, Emily Haroz, Johannes Thurl, Jasmin Stephens, Mia Campbell, Denis Antoine, Benjamin Lé Cook, Sean Joe, Roland Thorpe Jr. 2817

An Initial Validation of Community-Based Air-Conduction Audiometry in Adults With Simulated Hearing Impairment Using a New Web App, DigiBel: Validation Study (e51770)
Anna Sienko, Arun Thirunavukarasu, Tanya Kuzmich, Louise Allen. 2829

Sleep Patterns of Premedical Undergraduate Students: Pilot Study and Protocol Evaluation (e45910)
Gargi Rajput, Andy Gao, Tzu-Chun Wu, Ching-Tzu Tsai, Jennifer Molano, Danny Wu. 2838

Acceptability and Utility of a Digital Group Intervention to Prevent Perinatal Depression in Youths via Interactive Maternal Group for Information and Emotional Support (IMAGINE): Pilot Cohort Study (e51066)
Keshet Ronen, Anupa Gewali, Kristin Dachelet, Erica White, Marimirica Jean-Baptiste, Yolanda Evans, Jennifer Unger, S Tandon, Amritha Bhat. 2852
Benefits, Recruitment, Dropout, and Acceptability of the Strength Back Digital Health Intervention for Patients Undergoing Spinal Surgery: Nonrandomized, Qualitative, and Quantitative Pilot Feasibility Study (e54600)

Annemieke van der Horst, Laura Meijer, Harmieke van Os - Medendorp, Jan-Jukema, Ernst Bohlmeijer, Karlein Schreurs, Saskia Kelders.

Web-Based Screening, Brief Intervention, and Referral to Treatment for Traumatic Stress and Alcohol Misuse Among Survivors of Sexual Assault and Intimate Partner Violence: Usability and Acceptability Study (e49557)

Christine Hahn, Emily Tilstra-Ferrell, Selime Salim, Nada Goodrum, Alyssa Rheingold, Amanda Gilmore, Sara Barber, Angela Moreland.

Sodium Intake Estimation in Hospital Patients Using AI-Based Imaging: Prospective Pilot Study (e48690)


Human-Written vs AI-Generated Texts in Orthopedic Academic Literature: Comparative Qualitative Analysis (e52164)

Hassan Hakam, Robert Prill, Lisa Korte, Bruno Lovreković, Marko Ostojić, Nikolai Ramadanov, Felix Muehlensiepen.

Investigating the Feasibility of Using a Wearable Device to Measure Physiologic Health Data in Emergency Nurses and Residents: Observational Cohort Study (e51569)

Anish Agarwal, Rachel Gonzales, Kevin Scott, Raina Merchant.

Feasibility of Guided Internet-Based Cognitive Behavioral Therapy for Panic Disorder and Social Anxiety Disorder in Japan: Pilot Single-Arm Trial (e53659)

Seina Shinno, Kazuki Matsumoto, Sayo Hamatani, Yosuke Inaba, Yoshihito Ozawa, Yohei Kawasaki, Tomoki Ikai, Chihiro Sutoh, Hiroyuki Hayashi, Eiji Shimitzu.

Six-Month Pilot Testing of a Digital Health Tool to Support Effective Self-Care in People With Heart Failure: Mixed Methods Study (e52442)

Alison Keogh, Carol Brennan, William Johnston, Jane Dickson, Stephen Leslie, David Burke, Peter Megyesi, Brian Caulfield.

Pharmacy Students’ Attitudes Toward Distance Learning After the COVID-19 Pandemic: Cross-Sectional Study From Saudi Arabia (e54500)

Saud Alsauali, Salim Almutairi, Saleem Almutairi, Saleh Almofadhi, Mohammed Anaam, Mohammed Alshammar, Suhaj Abdulmalik, Yasser Almogbel.

Nurse-Led Brief Intervention for Enhancing Safe Sex Practice Among Emerging Adults in Hong Kong Using Instant Messaging: Feasibility Study (e52695)

Sharon Pak, Man Wang, Anne Teitelman, Janet Wong, Daniel Fong, Edmond Choi.

Efficient Machine Reading Comprehension for Health Care Applications: Algorithm Development and Validation of a Context Extraction Approach (e52482)

Duy-Anh Nguyen, Minyi Li, Gavin Lambert, Ryszard Kowalczyk, Rachael McDonald, Quoc Vo.

Performance of ChatGPT on the India Undergraduate Community Medicine Examination: Cross-Sectional Study (e49964)

Aravind Gandhi, Felista Josaph, Vineeth Rajagopal, P Aparnavi, Sushma Karkuri, Sonal Dayama, Prakash Satapathy, Mahalaqta Khatib, Shilpa Gaidhane, Quazi Zahiruddin, Ashish Behera.

Development and Evaluation of a Digital App for Patient Self-Management of Opioid Use Disorder: Usability, Acceptability, and Utility Study (e48068)

Van King Jr, Gregg Siegel, Henry Priesmeyer, Leslie Siegel, Jennifer Potter.
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral Activation Mobile App to Motivate Smokers to Quit: Feasibility and Pilot Randomized Controlled Trial</td>
<td>Belinda Borrelli, Y Bartlett, Daniel Fulford, Greg Frasco, Christopher Armitage, Alison Wearden</td>
<td>3016</td>
</tr>
<tr>
<td>Evaluation of a Pilot Program to Prevent the Misuse of Prescribed Opioids Among Health Care Workers: Repeated Measures Survey Study</td>
<td>Stephen Hebard, GracieLee Weaver, William Hansen, Scarlett Ruppert</td>
<td>3028</td>
</tr>
<tr>
<td>Lessons From the Field From a Volunteer Telehealth Ambassador Program to Enhance Video Visits Among Low-Income Patients: Qualitative Improvement Study</td>
<td>Delphine Tuot, Aarya Mukherjee, Amanda Churape, Triveni DeFries, George Su, Elaine Khoong, Courtney Lyles</td>
<td>3037</td>
</tr>
<tr>
<td>Mi Sleep Coach Mobile App to Address Insomnia Symptoms Among Cancer Survivors: Single-Arm Feasibility Study</td>
<td>Noel Arring, Debra Barton, Carolyn Lafferty, Bryana Cox, Deirdre Conroy, Lawrence An</td>
<td>3044</td>
</tr>
<tr>
<td>Integrating Virtual Mindfulness-Based Stress Reduction Into Inflammatory Bowel Disease Care: Mixed Methods Feasibility Trial</td>
<td>Kaitlyn Chappell, Diana Meakins, Melanie Marsh-Joyal, Allison Bihari, Karen Goodman, Jean-Michel Le Melledo, Allen Lim, Farhad Peerani, Karen Kroeker</td>
<td>3056</td>
</tr>
<tr>
<td>Impact of Electronic Patient-Reported Outcomes on Unplanned Consultations and Hospitalizations in Patients With Cancer Undergoing Systemic Therapy: Results of a Patient-Reported Outcome Study Compared With Matched Retrospective Data</td>
<td>Andreas Trojan, Christian Kühne, Michael Kiessling, Stefan Dröse, Christian Singer, Christian Jackisch, Christoph Thomssen, Gerd Kullak-Ublick</td>
<td>3073</td>
</tr>
<tr>
<td>Feasibility of Deploying Home-Based Digital Technology, Environmental Sensors, and Web-Based Surveys for Assessing Behavioral Symptoms and Identifying Their Precipitants in Older Adults: Longitudinal, Observational Study</td>
<td>Wan-Tai Au-Yeung, Yan Liu, Remonda Hanna, Sarah Gothard, Nathaniel Rodrigues, Cierra Leon Guerrero, Zachary Beattie, Jeffrey Kaye</td>
<td>3109</td>
</tr>
<tr>
<td>The Primary Care and Environmental Health e-Learning Course to Integrate Environmental Health in General Practice: Before-and-After Feasibility Study</td>
<td>Jean-Baptiste Tostain, Marina Mathieu, Agnès Oude Engberink, Bernard Clary, Michel Amouyal, Béatrice Lognos, Pascal Demoly, Isabella Annesi-Maesano, Grégory Ninot, Nicolas Molinari, Arnaud Richard, Maha Badreddine, Claire Dullos, Francois Carbonnel</td>
<td>3097</td>
</tr>
<tr>
<td>Changepoint Detection in Heart Rate Variability Indices in Older Patients Without Cancer at End of Life Using Ballistocardiography Signals: Preliminary Retrospective Study</td>
<td>Naotake Yanagisawa, Yuji Nishizaki, Bingwei Yao, Jianting Zhang, Takatoshi Kasai</td>
<td>3109</td>
</tr>
<tr>
<td>Engaging Cancer Care Physicians in Off-Label Drug Clinical Trials: Human-Centered Design Approach</td>
<td>Maren Parsell, Morgan Greenleaf, Greeshma Kombara, Vikas Sukhatme, Wilbur Lam</td>
<td>3120</td>
</tr>
<tr>
<td>Implementation Documentation and Process Assessment of the PharmNet Intervention: Observational Report</td>
<td>Lori Eldridge, Beth Meyerson, Jon Agley</td>
<td>3131</td>
</tr>
<tr>
<td>Machine Learning–Based Prediction of Changes in the Clinical Condition of Patients With Complex Chronic Diseases: 2-Phase Pilot Prospective Single-Center Observational Study</td>
<td>Celia Alvarez-Romero, Alejandro Polo-Molina, Eugenio Sánchez-Ubeda, Carlos Jimenez-De-Juan, Maria Cuadrín-Benitez, Jose Rivas-Gonzalez, Jose Portela, Rafael Palacios, Carlos Rodriguez-Morcillo, Antonio Muñoz, Carlos Parra-Calderon, Maria Nieto-Martín, Manuel Ollero-Baturone, Carlos Hernández-Quilless</td>
<td>3146</td>
</tr>
</tbody>
</table>
Privacy Concerns About Sharing General and Specific Health Information on Twitter: Quantitative Study (e45573)
Pouyan Esmaeilzadeh. ................................................................. 3159

Evaluation of the Effectiveness of Suicide.ca, Quebec's Digital Suicide Prevention Strategy Platform: Cross-Sectional Descriptive Study (e46195)
Louis-Philippe Côté, Julie Lane. .................................................. 3182

Professional Social Media Use Among Orthopedic and Trauma Surgeons in Germany: Cross-Sectional Questionnaire-Based Study (e53336)
Yasmin Youssef, Tobias Gehlen, Jörg Ansorg, David Back, Julian Scherer. ................................................................. 3202

Assessment of Qatar's Health Care Community Call Center Efficacy in Addressing COVID-19 Pandemic Health Care Challenges: Cross-Sectional Study (e42753)
Muhammad Waheed, Lolwa Al Mannai, Hanan Khudadad, Jamil Alenbawi, Mariama Mansaray, Samya Al Abdulla. .................................................. 3233

Jim Swift, Noel O'Kelly, Chris Barker, Alex Woodward, Sudip Ghosh. ................................................................. 3245

User-Friendly Chatbot to Mitigate the Psychological Stress of Older Adults During the COVID-19 Pandemic: Development and Usability Study (e49462)
Ya-Hsin Chou, Chemin Lin, Shwu-Hua Lee, Yen-Fen Lee, Li-Chen Cheng. ................................................................. 3255

Novel Web-Based Drop-In Mindfulness Sessions (Pause-4-Providers) to Enhance Well-Being Among Health Care Workers During the COVID-19 Pandemic: Descriptive and Qualitative Study (e43875)
Mary Elliott, Camille Khalilouf, Jennifer Hirsch, Diane de Camps Meschino, Orit Zamir, Paula Ravitz. ................................................................. 3269

Latino Parents’ Reactions to and Engagement With a Facebook Group–Based COVID-19 Vaccine Promotion Intervention: Mixed Methods Pilot Study (e51331)
Anna González-Salinas, Elizabeth Andrade, Lorrain Abroms, Kaitlyn Gómez, Carla Favetto, Valeria Gómez, Karen Collins. ................................................................. 3284

Barriers to COVID-19 Vaccination in a Troop of Fleet Antiterrorism Security Team Marines: Observational Study (e50181)
E Blazek, Amy Bucher. ................................................................. 3301

A Snapshot of COVID-19 Vaccine Discourse Related to Ethnic Minority Communities in the United Kingdom Between January and April 2022: Mixed Methods Analysis (e51152)
Nazifa Ullah, Sam Martin, Shoba Poduval. ................................................................. 3311

Factors Impacting Chinese Older Adults’ Intention to Prevent COVID-19 in the Post–COVID-19 Pandemic Era: Survey Study (e53608)
Huixin Guan, Wei Wang. ................................................................. 3322

Exploring the Prevalence of Tinnitus and Ear-Related Symptoms in China After the COVID-19 Pandemic: Online Cross-Sectional Survey (e54326)
Di Wang, Peifan Li, Xiaoling Huang, Yixuan Liu, Shihang Mao, Haoning Yin, Na Wang, Yan Luo, Shan Sun. ................................................................. 3338

Changes in the Clinical Practice of Mental Health Service Providers Throughout the COVID-19 Pandemic: Longitudinal Questionnaire Study (e50303)
Milena Gotra, Katharine Lindberg, Nicholas Jasinski, David Scarisbrick, Shannon Reilly, Jonathan Perle, Liv Miller, James Mahoney III. ................................................................. 3352
Web-Based Search Volume for HIV Tests and HIV-Testing Preferences During the COVID-19 Pandemic in Japan: Infodemiology Study (e52306)
Rie Kanamori, Futaba Umemura, Kosuke Uemura, Taiju Miyagami, Simon Valenti, Nobuyuki Fukui, Mayumi Yuda, Mizue Saita, Hirotake Mori, Toshio Naito. 3377

A Novel Approach for the Early Detection of Medical Resource Demand Surges During Health Care Emergencies: Infodemiology Study of Tweets (e46087)
Mahakprit Kaur, Taylor Cargill, Kevin Hui, Minh Vu, Nicola Bragazzi, Jude Kong. 3388

Exploring the Perspectives of Patients Living With Lupus: Retrospective Social Listening Study (e52768)
Erica Spies, Thomas Andreu, Matthias Hartung, Josephine Park, Paul Kamudoni. 3412

Patient Experiences and Insights on Chronic Ocular Pain: Social Media Listening Study (e47245)

Identification of Myths and Misinformation About Treatment for Opioid Use Disorder on Social Media: Infodemiology Study (e44726)
Mai ElSherief, Stephen Sumner, Vikram Krishnasamy, Christopher Jones, Royal Law, Akadia Kacha-Ochana, Lyna Schieber, Munmun De Choudhury. 3436

Arabic Web-Based Information on Oral Lichen Planus: Content Analysis (e49198)
Azzam AlMeshrafi, Arwa AlHamad, Hamoud AlKuraidees, Lubna AlNasser. 3446

Social Network Analysis of e-Cigarette–Related Social Media Influencers on Twitter/X: Observational Study (e53666)
Runtao Zhou, Zidian Xie, Qihang Dongmei Li. 3457

Preferences on Governance Models for Mental Health Data: Qualitative Study With Young People (e50368)

Media Discourse Regarding COVID-19 Vaccinations for Children Aged 5 to 11 Years in Australia, Canada, the United Kingdom, and the United States: Comparative Analysis Using the Narrative Policy Framework (e38761)
Verity Chadwick, Freya Saich, Joseph Freeman, Alexandra Martiniuk. 3483

Examining the Gateway Hypothesis and Mapping Substance Use Pathways on Social Media: Machine Learning Approach (e54433)
Yunhao Yuan, Erin Kasson, Jordan Taylor, Patricia Cavazos-Rehg, Munmun De Choudhury, Talayeh Aledavood. 3501

Viewpoints

Architecture and Development Framework for a Web-Based Risk Assessment and Management Platform Developed on WordPress to Address Opioid Overdose (e49759)
Alireza Kazemi, Marisha Boyd, Fiona Choi, Andy Tai, Vivian Tsang, Tam To, Jane Kim, Kerry Jang, Farhud Shams, Stefanie Schreiter, Maurice Cabants, Reinhard Krausz. 3210

Decentralizing Health Care: History and Opportunities of Web3 (e52740)
Aditya Narayan, Kydo Weng, Nirav Shah. 3221

AI Analysis of General Medicine in Japan: Present and Future Considerations (e52566)
Nozomi Aoki, Taiju Miyagami, Mizue Saita, Toshio Naito. 3228
Corrigenda and Addendas

Alessandra Bazzano, Lesley-Ann Noel, Tejal Patel, C Dominique, Catherine Haywood, Shenitta Moore, Andrea Mantsios, Patricia Davis.

Correction: Implementation Documentation and Process Assessment of the PharmNet Intervention: Observational Report (e59427)
Lori Eldridge, Beth Meyerson, Jon Agley.

Correction: Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study (e56570)
Christina Popescu, Grace Golden, David Benrimoh, Myriam Tanguay-Sela, Dominque Slowey, Eryn Lundrigan, Jérôme Williams, Bennet Desormeau, Divyesh Kardani, Tamara Perez, Colleen Rollins, Sonia Israel, Kelly Perlman, Caitrin Armstrong, Jacob Baxter, Kate Whitmore, Marie-Jeanne Fradette, Kaelan Felcarek-Hope, Ghassen Soufi, Robert Fratila, Joseph Mehlertter, Karl Looper, Warren Steiner, Soham Rej, Jordan Karp, Katherine Heller, Sagar Panikh, Rebecca McGuire-Snieckus, Manuela Ferrari, Howard Margolese, Gustavo Turecki.

Correction: The Influence of Human Connections and Collaboration on Research Grant Success at Various Career Stages: Regression Analysis (e58733)
Akiko Hashiguchi, Makoto Asashima, Satoru Takahashi.

Correction: Tumor Immunotherapy–Related Information on Internet-Based Videos Commonly Used by the Chinese Population: Content Quality Analysis (e59671)
Chen-xu Ni, Yi-bo Fei, Ran Wu, Wen-xiang Cao, Wenhao Liu, Fang Huang, Fu-ming Shen, Dong-jie Li.

Correction: Development of Therapeutic Alliance and Social Presence in a Digital Intervention for Pediatric Concussion: Qualitative Exploratory Study (e59722)
Kiarah O'Kane, Thalia Otamendi, Noah Silverberg, Esther Choi, Veronik Sicard, Roger Zemek, Katherine Healey, Olivier Brown, Lauren Butterfield, Andra Smith, Gary Goldfield, Rachel Kardish, Bechara Saab, Andrée-Anne Ledoux, Molly Cairncross.

Research Letters

The Relationship Between Changes in Mindfulness and Subsequent Changes in Well-Being Following Psychedelic Use: Prospective Cohort Study (e54632)
Grant Jones, Felipe Herrmann, Adam Bear, Robin Carhart-Harris, Hannes Kettner.

Smartwatch Versus Routine Tremor Documentation: Descriptive Comparison (e51249)
Catharina van Alen, Alexander Brenner, Tobias Warnecke, Julian Varghese.

A Health Information Technology Protocol to Enhance Colorectal Cancer Screening (e55202)
Adam Baus, Dannell Boatman, Andrea Calkins, Cecil Pollard, Mary Conn, Sujha Subramanian, Stephenie Kennedy-Rea.

Case Report

Examining Passively Collected Smartphone-Based Data in the Days Prior to Psychiatric Hospitalization for a Suicidal Crisis: Comparative Case Analysis (e55999)
Ross Jacobucci, Brooke Ammerman, Nilam Ram.
Short Paper

Tailoring of Health-Promotion Video Messaging for Reproductive-Aged Women at Risk for Developing Cardiometabolic Disease: Qualitative Focus-Groups Study (e52583)

Jacqueline Kent-Marvick, Bryan Gibson, Alycia Bristol, Stephanie St Clair, Sara Simonsen.
Comparing Self-Reported Running Distance and Pace With a Commercial Fitness Watch Data: Reliability Study

Garrett Bullock1*, DPhil, DPT; Joanne Stocks2*, PhD; Benjamin Feakins3, DPhil; Zahra Alizadeh4, MD; Amelia Arundale5, PhD, DPT; Stefan Kluzek2, MD, DPhil

1Wake Forest School of Medicine, Winston-Salem, NC, United States
2University of Nottingham, Nottingham, United Kingdom
3University of Oxford, Oxford, United Kingdom
4Department of Sports Science and Exercise Medicine, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran
5Red Bull Athlete Performance Center, Thalgau, Austria
*these authors contributed equally

Corresponding Author:
Garrett Bullock, DPhil, DPT
Wake Forest School of Medicine
475 Vine St
Winston-Salem, NC, 27411
United States
Phone: 1 3367144264
Email: garrettbullock@gmail.com

Abstract

Background: There is substantial evidence exploring the reliability of running distance self-reporting and GPS wearable technology, but there are currently no studies investigating the reliability of participant self-reporting in comparison to GPS wearable technology. There is also a critical sports science and medical research gap due to a paucity of reliability studies assessing self-reported running pace.

Objective: The purpose of this study was to assess the reliability of weekly self-reported running distance and pace compared to a commercial GPS fitness watch, stratified by sex and age. These data will give clinicians and sports researchers insights into the reliability of runners’ self-reported pace, which may improve training designs and rehabilitation prescriptions.

Methods: A prospective study of recreational runners was performed. Weekly running distance and average running pace were captured through self-report and a fitness watch. Baseline characteristics collected included age and sex. Intraclass correlational coefficients were calculated for weekly running distance and running pace for self-report and watch data. Bland-Altman plots assessed any systemic measurement error. Analyses were then stratified by sex and age.

Results: Younger runners reported improved weekly distance reliability (median 0.93, IQR 0.92-0.94). All ages demonstrated similar running pace reliability. Results exhibited no discernable systematic bias.

Conclusions: Weekly self-report demonstrated good reliability for running distance and moderate reliability for running pace in comparison to the watch data. Similar reliability was observed for male and female participants. Younger runners demonstrated improved running distance reliability, but all age groups exhibited similar pace reliability. Running pace potentially should be monitored through technological means to increase precision.

(JMIR Form Res 2024;8:e39211) doi:10.2196/39211

KEYWORDS

GPS; Garmin; training load; running; exercise; fitness; wearables; running; running distance; pace; pace distance

Introduction

Physical activity is an essential component of a healthy lifestyle [1]. There is a substantial body of evidence highlighting the physical, social, and psychological health benefits of regular physical activity [1-3]. Sustainable physical activity interventions are needed, given that 31% of the global population is sedentary [4]. The World Health Organization’s physical activity action plan [5] identifies sport as an underused yet significant contributor to physical activity.
One widely popular sport globally is running [6]. Over the past 40 years, running has become one of the most popular physical leisure activities [7,8]. An estimated 50 million people in Europe participate in running as a way to stay healthy [9]. Due to high running participation prevalence [9], researchers have attempted to quantify running habits and training load, most notably through self-report [10]. Running load or workload is the distance run in 1 session. A training session is 1 running bout. Running speed is the intensity at which one runs for 1 running session [9,10]. However, there are potential inaccuracies from over–self-reporting due to recall bias [11] and social desirability of higher levels of physical activity [12], with potential differences by sex and age groups [13]. Further, the reliability of self-reported running pace has not been investigated, which is an important factor in quantifying running training intensity [14]. Due to these issues, research has investigated the reliability of wearables in quantifying running load [15]. Wearables, such as accelerometers, have demonstrated excellent reliability in assessing gait patterns, acceleration, and velocity [15]. Although wearable accelerometers are ubiquitously used in the general population [16] and are reliably used in research to measure physical activity levels [17], they are rarely used by running populations to track running load and training [18]. Runners opt for wearable GPS watches to track running training [19], with up to 90% of regular runners using some form of GPS monitoring when running [18]. GPS wearable technology quantifies running workload and speed [20]. A systematic review determined that there is excellent reliability for step counting and moderate validity for energy expenditure and distance run [21]. The most popular GPS wearable technology used by runners is the Garmin watch, as indicated by a previous survey where 44% of GPS and sports watches were Garmin, compared to 27% for Polar and 7% for Nike watches [22]. There has been previous related work in evaluating the reliability of running self-reports in large samples. In a sample of 92 endurance runners, followed for a 52-week (ie, 1 year) period, 93% of the runners participated in the entire follow-up period [10]. In a study of 53 running participants over 18 weeks, the response rate was 73% over the reporting period [23]. Another study surveyed 228 runners for at least 6 months, with a 2.2% attrition rate [24]. Although there is substantial evidence investigating the reliability of running distance self-reporting [10] and GPS wearable technology [20], there are currently no studies investigating the reliability of participant self-reporting in comparison to GPS wearable technology in running populations. There is a critical sports science and medical research gap due to a paucity of reliability studies assessing self-report running pace. Further, as GPS wearable technology is expensive [25], there may be a barrier for some recreational runners, decreasing the efficacy of using GPS monitoring alone to assess running workload [22]. Therefore, this study aimed to assess the reliability of weekly self-reported running distance and pace compared to a commercial GPS fitness watch, stratified by sex and age. These data will give clinicians and sports researchers insights into the reliability of runners’ self-reported pace, which may improve training designs and rehabilitation prescriptions.

Methods

Study Design

A prospective cohort study was conducted using a mobile-based app. Participants accessed a dynamic digital consent form through the app or the recruitment website during the spring of 2021 over a 4-month period. During consent, participants could select different levels of study engagement. All levels of engagement involved the following: (1) an acknowledgement, understanding, and consent to participate in the study; (2) a baseline questionnaire collecting information on demographics, previous and current injury and illness history, footwear and foot posture, knee symptoms, lifestyle, and previous year’s training load; and (3) weekly reports on training load and each participant’s perceptions of cardiorespiratory symptoms, mood, and incidence of illness and injury in the last week. More advanced participation involved connecting participants’ Garmin Connect (Garmin Ltd) data, which included sharing data on running distance, running speed, and heart rate during each training or racing session. Participants added their Garmin Connect information at study recruitment. Garmin data were then automatically uploaded every week when the participant was within the study. Once the participant reaches the end of the study data collection or voluntarily leaves the study, the Garmin data collection link is terminated, ending data upload (Figure 1). Participants could opt out of the study at any time.
Ethical Considerations

This study received a favorable ethical review from the University of Nottingham (FMHS 113-1120). All methods were performed following the relevant guidelines and regulations of the Declaration of Helsinki. Before study inclusion, all participants were detailed about the risks and benefits of participation. All participants provided informed consent to participate.

Population and Recruitment Strategy

This study aimed to examine recreational runners. The inclusion criteria of this subgroup of the larger “Running Through” [26] study consisted of the following: (1) age ≥18 years; (2) performing running activities; and (3) connecting their Garmin Connect data to the weekly reports. Exclusion criteria consisted of individuals meeting the following conditions: (1) not able or willing to use the internet regularly; (2) diagnosed with an immunocompromised disorder; (3) diagnosed with memory impairment; (4) diagnosed with a neurodegenerative disorder; (5) diagnosed with inflammatory osteoarthritis; and (6) undergone trunk or lower extremity orthopedic surgery in the last 6 months. The larger “Running Through” study consisted of both Garmin and self-report data. Participants were recruited through email, the study website, social media, and word of mouth. Participants resided in the United Kingdom or Europe. All recruitment was performed in English. Participants did not receive compensation for participating in this study. Watch ownership was not known by the research team.

Weekly Survey

Participants were sent an encrypted text message or email weekly to report weekly running distance, pace, and incidence of illness and injury [10]. Garmin Connect data also monitored running distance and pace. Garmin monitoring has demonstrated excellent reliability and validity [20].

Data Storage

The University of Nottingham’s secure server hosted the research survey tool through the RedCap (Research Electronic Data Capture) service [27]. Data were queried from the secure database using a unique randomized and encrypted identification number.

Data Reduction

Watch data were downloaded to an encrypted SQL database using Garmin Connect software (Garmin Ltd). For convenience, these data were combined with the RedCap survey data and tables containing key variables that could be used to link these data. Custom functions were written in R using the DBI and MariaDB packages to interface with the database. The rjson and bit64 packages were additionally used to facilitate the extraction of JavaScript Object Notation format activity data and provide necessary extensions to R’s base data classes. Once data were downloaded, they were aggregated, cleaned, and checked for quality assurance. Data checks were performed through automation and manually.

Statistical Analyses

Participant statistics were described using mean (SD) or median (IQR) for continuous variables and frequencies (percentages)
for categorical variables. Overall running exposure was calculated in person kilometers.

To assess reliability, intraclass correlation coefficients (ICC_{2,1}) were calculated for weekly running distance and running pace between self-reports and weekly reports generated by the Garmin Connect data. Reliability was rated as poor (<0.50), moderate (0.50-0.75), good (0.75-0.90), and excellent (>0.90) [28]. Correlation and Bland-Altman plots were also calculated to assess any systematic measurement error. Analyses were then stratified by sex and age strata (18-40, 41-60, and ≥61 years). All analyses were performed in R 4.1.2 (R Foundation for Statistical Computing) [29], with the psych package for ICC calculations and BlandAltmanLeh for Bland-Altman plots.

### Results

A total of 485 participants linked their Garmin Connect data to the study, with 475 participants included for a total of 3602 participant weeks. Of these, 3 participants were excluded due to lack of follow-up, and another 7 did not run during the collection period (Table 1; the flow chart is available in Multimedia Appendix 1). Participants self-reported running a weekly median of 26.2 (IQR 12.8-39.7) km at a median pace of 6.0 (IQR 5.4-6.7) min/km compared to 25.9 (IQR 4.7-41.8) km running distance at a median pace of 6.1 (IQR 5.2-7.0) min/km (Table 2) recorded by the Garmin watch.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (n=475)</th>
<th>Female participants (n=248)</th>
<th>Male participants (n=227)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>49.5 (12.2)</td>
<td>51.0 (13.1)</td>
<td>47.8 (10.9)</td>
</tr>
<tr>
<td>BMI (kg/m^2), mean (SD)</td>
<td>24.0 (3.7)</td>
<td>24.2 (3.7)</td>
<td>23.8 (3.8)</td>
</tr>
<tr>
<td>Number of weeks followed, median (IQR)</td>
<td>17 (11-23)</td>
<td>15 (10-20)</td>
<td>17 (11-24)</td>
</tr>
</tbody>
</table>

**Smoking, n (%)**
- Current smoker: 10 (2) Female 6 (2) Male 4 (1)
- Ex-smoker: 62 (13) Female 30 (13) Male 32 (13)

Cigarettes per day, median (IQR):
- Male participants (n=227)
- Female participants (n=248)
- Overall (n=475)

Years smoked, median (IQR):
- Male participants (n=227)
- Female participants (n=248)
- Overall (n=475)

Diabetes, n (%): 6 (1) Female 1 (1) Male 5 (3)

Heart disease, n (%): 5 (1) Female 2 (1) Male 3 (1)

Cancer, n (%): 14 (3) Female 10 (4) Male 4 (2)

Asthma, n (%): 62 (13) Female 40 (16) Male 32 (11)

Hay fever (pollen allergies), n (%): 180 (38) Female 99 (40) Male 81 (35)

Days of running per week, mean (SD): 2 (1) Female 2 (1) Male 2 (1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (n=475)</th>
<th>Female participants (n=248)</th>
<th>Male participants (n=227)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time running (min), median (IQR)</td>
<td>144 (48-238)</td>
<td>139 (43-226)</td>
<td>153 (45-230)</td>
</tr>
<tr>
<td>Calories burned (kcal), median (IQR)</td>
<td>1450 (547-2353)</td>
<td>1212 (461-1963)</td>
<td>1755 (671-2840)</td>
</tr>
<tr>
<td>Kilometers, median (IQR)</td>
<td>25.9 (4.7-41.8)</td>
<td>22.7 (4.5-40.9)</td>
<td>29.6 (4.9-40.6)</td>
</tr>
<tr>
<td>Running pace (min/km), median (IQR)</td>
<td>6.1 (5.2-7.0)</td>
<td>6.7 (5.7-7.7)</td>
<td>5.7 (4.9-6.4)</td>
</tr>
<tr>
<td>Average heart rate (bpm^a), mean (SD)</td>
<td>130 (26)</td>
<td>127 (26)</td>
<td>130 (24)</td>
</tr>
<tr>
<td>Maximum heart rate (bpm^a), mean (SD)</td>
<td>163 (24)</td>
<td>162 (24)</td>
<td>163 (24)</td>
</tr>
</tbody>
</table>

^aBpm: beats per minute.

**Reliability**

Weekly distance and pace reliability were rated as good and moderate, respectively, for both sexes and for runners aged 41-60 and ≥61 years. Furthermore, weekly distance reliability was rated as excellent and moderate in runners aged 18-40 years. All results exhibited no discernable systematic bias (Figure 2; Table 3; Multimedia Appendix 1).
**Figure 2.** Correlation and Bland-Altman Plots of the Reliability of Self-Report and Garmin Connect Weekly Running Distance and Running Pace. A. Weekly Running Distance (km) B. Weekly Running Pace (min/km).
Table 3. Reliability of weekly self-report and Garmin watch data for running distance and pace.

<table>
<thead>
<tr>
<th>Group</th>
<th>Self-report</th>
<th>Garmin watch</th>
<th>ICC(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weekly distance run (km), median (IQR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall (n=475)</td>
<td>26.2 (12.8, 39.7)</td>
<td>25.9 (4.7-41.8)</td>
<td>0.88 (0.87-0.89)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (n=248)</td>
<td>24.8 (10.0-39.7)</td>
<td>22.7 (4.5-40.9)</td>
<td>0.86 (0.85-0.87)</td>
</tr>
<tr>
<td>Male (n=227)</td>
<td>29.0 (16.9-41.1)</td>
<td>29.6 (4.9-40.6)</td>
<td>0.89 (0.88-0.90)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-40 (n=113)</td>
<td>27.0 (10.9-43.0)</td>
<td>26.8 (7.2-44.6)</td>
<td>0.93 (0.92-0.94)</td>
</tr>
<tr>
<td>41-60 (n=262)</td>
<td>27.0 (16.0-38.0)</td>
<td>25.9 (11.5-41.2)</td>
<td>0.87 (0.85-0.88)</td>
</tr>
<tr>
<td>≥61 (n=100)</td>
<td>25.0 (13.8-36.2)</td>
<td>24.5 (9.7-39.3)</td>
<td>0.83 (0.80-0.85)</td>
</tr>
<tr>
<td>Average weekly running pace (min/km), mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall (n=475)</td>
<td>6.0 (1.1)</td>
<td>6.1 (2.0)</td>
<td>0.72 (0.69-0.75)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (n=248)</td>
<td>6.4 (1.2)</td>
<td>6.7 (1.9)</td>
<td>0.67 (0.62-0.72)</td>
</tr>
<tr>
<td>Male (n=227)</td>
<td>5.7 (0.9)</td>
<td>5.7 (2.1)</td>
<td>0.68 (0.65-0.71)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-40 years (n=113)</td>
<td>5.8 (1.1)</td>
<td>5.7 (2.5)</td>
<td>0.69 (0.65-0.73)</td>
</tr>
<tr>
<td>41-60 years (n=262)</td>
<td>6.0 (1.0)</td>
<td>6.1 (2.5)</td>
<td>0.70 (0.66-0.73)</td>
</tr>
<tr>
<td>≥61 years (n=100)</td>
<td>6.2 (1.2)</td>
<td>6.5 (2.5)</td>
<td>0.74 (0.68-0.78)</td>
</tr>
</tbody>
</table>

\(^a\)ICC: intraclass correlation coefficient.

**Discussion**

**Principal Findings**

The overall findings of this study indicate that the weekly self-report of running distance by runners wearing a Garmin watch demonstrated good reliability compared to the Garmin watch data. Distance reliability was similar between female and male participants and across age strata, except for participants aged 18-40 years, who demonstrated excellent reliability. Weekly self-report of running pace demonstrated moderate reliability compared to Garmin watch data, with similar reliability observed between sex and age strata. There were no discernable patterns or systematic biases concerning self-reported running distance or pace.

**Comparison to Previous Work**

Self-reported running distance exhibited good reliability compared to Garmin data. The reliability is higher in this study compared to a previous study on physical activity (ICC 0.67-0.81) [30]. However, the previous study examined multiple countries and recorded all physical activity beyond running. The homogenous country sample and the focus on running in our study may affect the comparison of these results [31]. Younger adults (aged 18-40 years) demonstrated increased running distance reliability reporting compared to the older age strata (aged 41-60 and ≥61 years). This is comparable with previous research, in which younger adults displayed improved self-report reliability [30]. Younger adults may have a greater aptitude to monitor their running through technology [32]. However, this is only speculative, and further research is required.

Self-reported running pace demonstrated moderate reliability compared to Garmin Connect data. There are currently no studies investigating the reliability of self-reported running pace. However, recreational runners usually train at one pace, with little change at different distances [33]. The moderate reliability observed in this sample may be due to these runners reporting their perceived running pace, with little fluctuation between sessions or weeks. However, specific variances may have occurred in the actual running pace, decreasing the reliability of these data. Previous literature has suggested that instant feedback through the use of heart rate or step cadence can increase a recreational runner’s ability to self-report running pace [33]. However, further work is needed to investigate the efficacy of this approach.

These findings necessitate future research. Participants were recreational runners, and most of them were older than 40 years. Future work is needed to assess the reliability of self-report in comparison to GPS monitoring data in elite runners of all ages and younger populations across different skill or competition levels. All runners in this study already owned a Garmin watch before the study enrollment. Understanding how self-reporting changes among new GPS activity monitor users is needed. Running pace demonstrated moderate reliability in this recreational runner sample. Future research is required to investigate the effectiveness of running pace training on self-report reliability.
Limitations
As with all studies, there are limitations to this study. First, there is the risk of recall bias due to the weekly intervals for self-reporting, which decreases the precision of these findings. Participants may have not worn or activated their Garmin watch for specific runs, decreasing the reliability of these data. GPS monitoring devices are expensive, causing a high barrier to entry. Such a barrier may add selection bias to this reliability study, decreasing the generalizability of these results to all running populations. Further, the sample in this study comprised recreational runners; therefore, the results are not generalizable to elite runners or populations that solely engage in walking for exercise. Finally, participants used different versions of the Garmin watch. As different watch versions may exhibit different reliability, there is a potential for decreased data precision.

Practical Applications
Physical activity monitors have effectively enhanced physical activity levels by providing user feedback, facilitating behavior change—following prescribed training—and preventing injuries [34]. The good to excellent reliability of self-reported weekly running distance observed in this cohort of recreational runners across all adult age groups supports previous research indicating that runners can effectively report running loads [30]. These findings strengthen the notion that self-report can be used to reliably monitor runners as they begin or maintain an exercise regimen or return to running following an injury. However, the moderate reliability exhibited for running pace suggests that recreational runners of all ages are not as adept at monitoring their running pace. Incorporating technological monitoring for running pace may be pertinent to maintain prescribed running paces either for specific training regimens (ie, preparing for a race) or when returning to running following an injury.

Conclusions
Weekly self-report demonstrated good reliability for running distance compared to the Garmin watch data, with similar reliability between sex and age groups. However, the weekly self-report demonstrated only moderate reliability for running pace compared to the Garmin data, with similar reliability between sex and age groups. Sports researchers and scientists can use weekly self-reported running distance in conjunction with Garmin data when quantifying weekly training load. However, caution should be exercised when relying on self-reported running pace to evaluate running intensity in recreational runners. Running pace potentially should be monitored through technological means to increase precision.

Data Availability
Data and corresponding codes are available within the Open Science Framework [35].

Acknowledgments
This study was funded through Technopolis Consulting Group Belgium (TGB).

Authors' Contributions
GB, JS, BF, and SK conceived the study idea. JS, ZA, and SK collected the data. GS, BF, JS, and SK extracted and analyzed the data. GB, JS, BF, AA, and SK wrote the original manuscript. GB, JS, BF, ZA, AA, and SK edited the manuscript. GB, JS, BF, ZA, AA, and SK approved the final draft of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Study flowchart.
[DOCX File, 25 KB - formative_v811e39211_app1.docx ]

References


20. Mallula C, Comparing Garmin Forerunner 405CX GPG and Nike + iPod to Accurately Measure Energy Expenditure, Distance, and Speed of Overground Running. MSL Academic Endeavours. Full text available at: https://engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=1726&context=etdarchive [accessed 2023-12-14]


## Abbreviations

**ACC:** intraclass correlational coefficient  
**RedCAP:** Research Electronic Data Capture
Prevalence of Body Dysmorphic Disorder in the Spanish Population: Cross-Sectional Web-Based Questionnaire Study

Álvaro Loewen¹, MD; Hilario Blasco-Fontecilla¹,²,³,⁴, MD, PhD; Chao Li¹, MD; Marcos Bella-Fernández²,⁵,⁶, MSc; Belén Ruiz-Antorán⁷, MD, PhD

¹Facultad de Medicina, Universidad Autónoma de Madrid, Madrid, Spain
²Servicio de Psiquiatría Infantil y Juvenil, Hospital Universitario Puerta de Hierro Majadahonda, Majadahonda, Spain
³Centro de Investigación Biomédica en Red - Salud Mental, Madrid, Spain
⁴Facultad de Ciencias de la Salud y Centro Médico, Universidad Internacional de La Rioja, Madrid, Spain
⁵Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain
⁶Departamento de Psicología, Universidad Pontificia de Comillas, Madrid, Spain
⁷Servicio de Farmacología Clínica, Hospital Universitario Puerta de Hierro Majadahonda, Majadahonda, Spain

Corresponding Author:
Hilario Blasco-Fontecilla, MD, PhD
Servicio de Psiquiatría Infantil y Juvenil
Hospital Universitario Puerta de Hierro Majadahonda
Instituto de Investigación Sanitaria Puerta de Hierro - Segovia de Arana
C/ Manuel de Falla, 1
Majadahonda, 28222
Spain
Phone: 34 911911690
Email: hmblasco@yahoo.es

Abstract

Background: Body dysmorphic disorder (BDD) is defined as excessive concern with mild or nonexistent defects in personal physical appearance, which are not perceived by others. The worldwide prevalence of BDD ranges between 0.5% and 3.2%, with no differences across genders. The mean age of onset of BDD is 16.9 years. BDD is typically associated with young age, psychiatric disorders, and dermatological procedures. Patients with BDD typically display poorer mental health status than patients diagnosed with other mental disorders.

Objective: The aim of this study was to estimate the prevalence of BDD in Spain and to identify the variables associated with BDD.

Methods: We performed a cross-sectional descriptive study by collecting data through an anonymous web-based survey targeting the Spanish population aged 18 years or older. The measures in this study were (1) sociodemographic variables, (2) variables associated with dermatological and psychiatric disorders and cosmetic procedures, (3) scales measuring quality of life (12-item Short Form health survey, version 2) and (4) BDD (BDD Questionnaire). Statistical analysis was performed with SPSS software version 21. P values less than .05 were considered significant.

Results: Of the 2091 participants who took the survey, 322 (15.2%) met the criteria of having BDD. The mean age of the participants with BDD was 23.5 (SD 9.6) years. In terms of BDD prevalence, women accounted for 19.9% (284/1421), men accounted for 5.2% (34/653), and students accounted for 25.2% (263/1043). Approximately 46.6% (150/322) of the participants with BDD reported a history of psychiatric comorbidities, including anxiety disorders, depressive disorders, and eating disorders. BDD was significantly associated with female gender, younger age (18-24 years), students, monthly income of less than €500 (€1=US $1.11), and the presence of dermatological and some psychiatric disorders such as depression, anxiety, and eating disorders (P<.05). The number of body parts of concern in participants with BDD was significantly higher than that in those without BDD (4.6 vs 2.2, respectively; P<.001). Regarding the body parts of concern, body fat was the most common concern for both groups with BDD and without BDD, followed by thighs, face, hips, and skin in the BDD group and thighs, teeth, and hair in the non-BDD group. Participants with BDD showed a significantly poorer self-perception of their mental health, irrespective of the presence of any mental disorder (P<.001).
Conclusions: Our findings showed that the prevalence of BDD in Spain was higher than expected. Further, BDD is frequently associated with other psychiatric disorders, particularly depressive disorder, anxiety disorder, and eating disorder. Participants with BDD had a poorer perception of quality of life associated with mental but not physical health problems. Finally, the perception of quality of mental health life in participants with BDD was independent of diagnosis of any mental disorder.

JMIR Form Res 2024;8:e46515 doi: 10.2196/46515

KEYWORDS
body dysmorphic disorder; prevalence; adults; Spain; comorbidities; mental health; depression; anxiety; OCD; obsessive-compulsive disorder

Introduction

Body dysmorphic disorder (BDD) is a common psychiatric disorder affecting 0.5%-3.2% of the general population worldwide [1]. A multicentric study in Spain showed that the prevalence of BDD was higher among patients with acne (10.6%) [2]. The prevalence of BDD across genders is debatable. One study showed similar prevalence of BDD across both genders (49% females and 51% males) [3]. Another study reported higher prevalence among females (68.5%) [4]. The average age of onset of BDD is 16.9 years [5]. One study reported an inverse relationship between the prevalence of BDD and age: 78.6% of participants with BDD were aged 18-27 years, 14.3% were aged 28-37 years, and 7.1% were aged 38 years or older [6]. Furthermore, BDD is closely associated with other mental disorders. A recent systematic review [1] showed that the highest prevalence of BDD was among psychiatric inpatients (5.8%-37.78%). The psychiatric disorders most frequently associated with BDD are depressive disorder (47%-56.3%), borderline personality disorder (54.3%), and eating disorders (12%-45%), whereas obsessive-compulsive disorder (3%-15.3%) and schizophrenia are less closely associated with BDD. Participants who had cosmetic procedures (2.9%-57%) slightly overlay with BDD prevalence in the general population. The prevalence of BDD in both psychiatric outpatients (0.3%-2%) and students (1.3%-5.8%) partially overlapped with that observed in the general population. Among dermatologic patients, the prevalence of BDD was reported to be 2.1%-36% [1].

The etiology of BDD is multifactorial, encompassing biological, psychological, and sociocultural factors. BDD has been associated with parental rejection, as well as physical, emotional, or sexual abuse during adolescence [7]. Studies have shown a possible genetic association in first-degree relatives, with affected patients being up to 3-8 times more likely to develop BDD than the general population [7]. Shy, anxious, and perfectionistic individuals may also have a greater predisposition to develop this disorder [8].

The most important symptom of BDD is distortion of body perception, which leads to low self-esteem, anxiety, depression, social isolation, and obsessive-compulsive behaviors [9]. The clinical profile of BDD is characterized by repetitive actions such as constant checking in mirrors, applying excessive makeup to cover defects, dermatillomania, comparing one’s appearance with that of others, and excessive exercise, taking up an average of 3-8 hours daily [10]. Patients with BDD are usually preoccupied with 5-7 different parts of their body [11], the most common being skin (53.8%), nose (38.5%), and hair (34.6%). The other body parts that are of frequent concern are weight and muscle mass (30.8%), face (30.8%), chest and trunk (19.2%), and teeth (19.2%) [1]. The mental and physical health status perceived by patients with BDD are lower than that perceived by the general population [12].

Given that the primary concern of patients with BDD revolves around their external appearance, it is common for people with BDD to predominantly seek dermatological and cosmetic procedures over seeking professional help for the treatment of their underlying psychiatric pathology [13]. Moreover, patients diagnosed with BDD often do not seek help for various reasons: they feel ashamed or lack insight [14]. Furthermore, BDD is likely to be underdiagnosed, given the large number of barriers to diagnosis such as the absence of appropriate tools, lack of information and awareness among health care professionals, and professionals’ refusal to lose a patient or inability to diagnose it properly [15]. In addition, not all health care professionals are familiar with this disorder [9]. This leads to poor identification in psychiatric and cosmetic settings where BDD cases are notoriously prevalent [14]. If professionals do not perform a detailed anamnesis, it is difficult for patients to disclose their concerns, given the shyness underlying this disorder [16]. To overcome this shortcoming, clinicians may use the BDD Questionnaire (BDDQ), a validated diagnostic tool for BDD, with sensitivity of 100% [17-19], specificity of 89%-93% for psychiatric inpatients [18], and 93% for dermatologic patients [19]. The BDDQ is a brief questionnaire that assesses different items of the patient’s body perception [17]. Finally, treatment is based on pharmacological and nonpharmacological measures. The former includes the use of fluoxetine, a serotonin reuptake inhibitor antidepressant. The latter is based on psychotherapy—most notably cognitive behavioral therapy [20]. Randomized clinical trials have shown a rate of 50%-80% improvement in patients as well as a lower relapse rate following pharmacological treatment [21].

In summary, BDD is a poorly studied and underdiagnosed psychiatric disorder. This may be because BDD is not perceived as a disorder by aesthetic practitioners, who may think that they are merely offering a “service” [22-24]. The main objective of our study was to estimate the prevalence of BDD in Spain. Additionally, we explored the association of BDD with sociodemographic variables, presence of dermatological or psychiatric disorders and cosmetic procedures, and quality of life.
Methods

Design and Scope of This Study
A cross-sectional descriptive study was conducted. The general population aged 18 years and older in Spain was invited to access the study protocol via a link to Google Forms. The security and lawful use of personal data collected on the website is guaranteed by accepting the data privacy policy included in the survey. The measures in the protocol consisted of (1) sociodemographic variables, (2) variables associated with dermatological and psychiatric disorders and cosmetic procedures, and (3) perception of health and quality of life, measured using the 12-item Short Form version 2 (SF-12v2) health survey, and (4) a validated scale for the diagnosis of BDD, using the BDD questionnaire screening test [17].

Study Sample Population
Participants 18 years and older from the general population residing and registered in Spain at the time the survey was performed and who voluntarily completed the study’s web-based questionnaire were included. The sample size was calculated using Epidat 4.0 (Dirección Xeral de Saúde Pública da Consellería de Sanidade da Xunta de Galicia) based on the following estimate: population size, 40,000,000; expected proportion, 5%; accuracy, 1%; confidence level, 95%; and design effect, 1.0. The minimum total number of responses required for 1% precision with 95% confidence level was 1825 participants.

Variables
The following variables were collected: sociodemographic variables such as gender (female, male, or other, ie, participants who do not identify themselves as male or female), year of birth, region of residence, race, educational level, employment status, and range of monthly income. Variables related to other comorbidities associated with BDD were comorbidity with dermatologic and psychiatric disorders and cosmetic procedures. Regarding the year of birth, the participants were classified into 4 age groups: 18-24 years, 25-44 years, 45-64 years, and 65 years or older. This classification was based on previous studies related to mental health and the use of these services according to age group [25].

Quality of Life Assessment
Data on quality of life and perception of their state of health were collected through the SF-12v2 health survey, which is validated as a psychometric instrument for numerous diseases and conditions, including mental illness. It assesses the participant’s physical and mental state through 8 health domains: 4 related to physical health, that is, general health, physical function, role-physical, and bodily pain; and 4 related to mental health, that is, vitality, social function, role-emotional, and mental health [26].

BDD Assessment
BDD assessment was performed using the BDDQ [17]. A version adapted and validated in Spanish was used [27]. The BDDQ is a brief (7-item) self-report measure derived from the Diagnostic and Statistical Manual of Mental Disorders, fourth edition diagnostic criteria. The first 6 items require a dichotomous answer (yes/no) and the last one is multiple choice. The test will be positive if the patient answers “yes” to questions 1 and 2, “yes” to question 3, 4, 5, or 6, and if in question 7, the time indicated is more than 1 hour per day [17].

Data Handling
The Google Forms platform was used for the web-based survey. The questionnaire was distributed telematically, both through a link via mobile phone and a printed QR code. The data obtained were extracted and sorted in Microsoft Excel. Subsequently, we used SPSS to create the database and perform the corresponding analyses. A license was obtained for the SF-12v2 health survey, which together with the use of the QualityMetric program provided, allowed for its correct interpretation. The database was generated in an anonymized form guaranteeing exclusive access by the principal investigator, thereby allowing respect, privacy, and confidentiality of the data.

Statistical Analysis
Statistical analysis was performed with SPSS software (version 21; IBM Corp). All statistical analyses were performed at .05 level of significance. A descriptive study was conducted for all the variables included in this study. Quantitative variables were expressed as mean and standard deviation (SD). Qualitative variables were expressed as absolute value (n) and percentage with an estimated 95% CI. Comparison of means was performed using 2-sided Student t test or Mann-Whitney U test as appropriate after checking normality with the Kolmogorov-Smirnov test. The association of the qualitative variables was estimated by means of the chi-square statistic. Multiple logistic regression models were used to determine the association of different variables with each other. A univariate analysis was performed where a significantly higher risk ratio for BDD diagnosis was obtained for some of the variables studied. The significant variables obtained in the univariate model were subsequently included in the multivariate analysis. Thereafter, given the heterogeneous conditions of the population, a subgroup analysis was performed with a multivariate model considering the gender and age variables.

Ethics Approval
This study was conducted in accordance with the requirements expressed in the Declaration of Helsinki 2013. Participants were invited to participate online by clicking on the link to the survey. Information about the purpose of the survey and its anonymous and voluntary nature was included in the survey header. Participants were identified by a numerical code in order to respect the confidentiality of the participants’ personal data. The automatic code is assigned directly by Google Forms at the time of download through a time stamp. This project was approved by the ethics committee of the Hospital Universitario Puerta de Hierro Majadahonda in Madrid (promoter protocol code PI 206/21, December 20, 2021).
Results

A total of 2091 participants were included in this study. The sociodemographic and clinical characteristics of the participants are described in Table 1.

The prevalence of BDD in the population assessed in this study was 15.2% (284/1421, 19.9% in females vs 34/653, 5.2% in males). Regarding age, the prevalence of BDD was higher in the youngest age group (18-24 years; 267/1091, 24.5%), followed by the 25-44 years age group (30/279, 10.8%) (Figure 1). The number of body parts of concern in participants with BDD was significantly higher than that in participants without BDD (4.6 vs 2.2, respectively; P<.001). Regarding the body parts of concern, body fat was the most common concern in both groups, followed by thighs, face, hips, and skin in the BDD group and thighs, teeth, and hair in the non-BDD group (Figure 2). Regarding the sociodemographic characteristics of the BDD population (n=322), the majority were females (284/322, 88.2%), with a mean age of 23.5 (SD 9.6) years, and Caucasians (243/722, 75.5%). Approximately 81.7% (1533/2091) were students, and 76.6% (247/322) of them had a university education level. Among the participants diagnosed with BDD, 63.4% (204/322) had a history of dermatologic disease, the most frequent being acne (115/204, 56.5%) and dermatitis (99/204, 48.5%). Approximately 46.6% (150/322) of the participants with BDD reported a history of psychiatric comorbidities, and the most frequent were anxiety disorders (108/150, 72%), depressive disorders (84/150, 56%), eating disorders (72/150, 48%), and attention-deficit/hyperactivity disorder (18/150, 12%). Approximately 17.7% (57/322) of the population with BDD had previously undergone a cosmetic procedure, the most frequent being laser treatment for acne, blemishes, and other skin lesions (17/57, 29.8%); mesotherapy (9/57, 15.8%); and rhinoplasty (7/57, 12.3%). The factors related to BDD are reported in univariate and multivariate analyses in Table 2. BDD diagnosis was significantly associated with female gender, other genders, age 18-24 years, students, monthly income level of less than €500 (€1=US $1.11), and participants with dermatologic and psychiatric comorbidities (P<.001). All these variables were included in the multivariate model, with gender (female and other), age, student occupation, depressive disorder, eating disorders, and anxiety disorders remaining at <.05 significance as diagnostic predictors of BDD.
### Table 1. Sociodemographic and clinical characteristics of the participants.

<table>
<thead>
<tr>
<th></th>
<th>All (N=2091)</th>
<th>With BDD(^a) diagnosis (n=322)</th>
<th>Without BDD diagnosis (n=1769)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender (female), n (%)</strong></td>
<td>1421 (68)</td>
<td>284 (88.2)</td>
<td>1137 (64.3)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>37.7 (16.6)</td>
<td>23.5 (9.6)</td>
<td>35.5 (16.9)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1091 (52.2)</td>
<td>267 (82.9)</td>
<td>824 (46.6)</td>
</tr>
<tr>
<td>25-44</td>
<td>279 (13.3)</td>
<td>30 (9.3)</td>
<td>249 (14.1)</td>
</tr>
<tr>
<td>45-64</td>
<td>654 (31.3)</td>
<td>23 (7.1)</td>
<td>631 (35.7)</td>
</tr>
<tr>
<td>&gt;64 years</td>
<td>66 (3.2)</td>
<td>2 (0.6)</td>
<td>64 (3.6)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African</td>
<td>11 (0.5)</td>
<td>3 (0.9)</td>
<td>8 (0.5)</td>
</tr>
<tr>
<td>American</td>
<td>6 (0.3)</td>
<td>0 (0)</td>
<td>6 (0.3)</td>
</tr>
<tr>
<td>Asian</td>
<td>26 (1.2)</td>
<td>7 (2.2)</td>
<td>19 (1.1)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>1533 (73.3)</td>
<td>243 (75.5)</td>
<td>1290 (72.9)</td>
</tr>
<tr>
<td>Latin</td>
<td>306 (14.6)</td>
<td>23 (7.1)</td>
<td>283 (16)</td>
</tr>
<tr>
<td>Other</td>
<td>209 (10)</td>
<td>46 (14.3)</td>
<td>163 (9.2)</td>
</tr>
<tr>
<td><strong>Educational level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>11 (0.5)</td>
<td>1 (0.3)</td>
<td>10 (0.6)</td>
</tr>
<tr>
<td>Middle school</td>
<td>8 (0.4)</td>
<td>0 (0)</td>
<td>8 (0.5)</td>
</tr>
<tr>
<td>Professional education</td>
<td>100 (4.8)</td>
<td>15 (4.7)</td>
<td>85 (4.8)</td>
</tr>
<tr>
<td>High school</td>
<td>265 (12.7)</td>
<td>59 (18.3)</td>
<td>206 (11.6)</td>
</tr>
<tr>
<td>University</td>
<td>1707 (81.6)</td>
<td>247 (76.7)</td>
<td>1460 (82.5)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1043 (49.9)</td>
<td>263 (81.7)</td>
<td>780 (44.1)</td>
</tr>
<tr>
<td>Worker</td>
<td>865 (41.4)</td>
<td>50 (15.5)</td>
<td>815 (46.1)</td>
</tr>
<tr>
<td>Other</td>
<td>183 (8.8)</td>
<td>9 (2.8)</td>
<td>174 (9.8)</td>
</tr>
<tr>
<td><strong>Monthly income (€)(^b), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;500</td>
<td>861 (41.2)</td>
<td>205 (63.7)</td>
<td>656 (37.1)</td>
</tr>
<tr>
<td>500-999</td>
<td>138 (6.6)</td>
<td>34 (10.6)</td>
<td>104 (5.9)</td>
</tr>
<tr>
<td>1000-1999</td>
<td>343 (16.4)</td>
<td>39 (12.1)</td>
<td>304 (17.2)</td>
</tr>
<tr>
<td>2000-2999</td>
<td>291 (13.9)</td>
<td>9 (2.8)</td>
<td>282 (15.9)</td>
</tr>
<tr>
<td>&gt;3000</td>
<td>358 (17.1)</td>
<td>11 (3.4)</td>
<td>347 (19.6)</td>
</tr>
<tr>
<td>Not known</td>
<td>100 (4.8)</td>
<td>24 (7.5)</td>
<td>76 (4.3)</td>
</tr>
<tr>
<td><strong>Dermatologic disease, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acne</td>
<td>1072 (51.3)</td>
<td>204 (63.4)</td>
<td>868 (49.1)</td>
</tr>
<tr>
<td>Atopic dermatitis</td>
<td>532 (49.6)</td>
<td>115 (56.5)</td>
<td>417 (48)</td>
</tr>
<tr>
<td>Other dermatitis</td>
<td>434 (40.5)</td>
<td>99 (48.5)</td>
<td>335 (38.6)</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>35 (3.3)</td>
<td>10 (4.9)</td>
<td>25 (2.9)</td>
</tr>
<tr>
<td>Rosacea</td>
<td>50 (4.7)</td>
<td>2 (1)</td>
<td>48 (5.5)</td>
</tr>
<tr>
<td>Urticaria</td>
<td>64 (6)</td>
<td>5 (2.5)</td>
<td>59 (6.8)</td>
</tr>
<tr>
<td>Pityriasis</td>
<td>15 (1.4)</td>
<td>1 (0.5)</td>
<td>14 (1.6)</td>
</tr>
<tr>
<td>Eczema</td>
<td>37 (3.5)</td>
<td>9 (4.4)</td>
<td>28 (3.2)</td>
</tr>
<tr>
<td>Skin infections</td>
<td>32 (3)</td>
<td>6 (2.9)</td>
<td>26 (3)</td>
</tr>
<tr>
<td>Category</td>
<td>All (N=2091)</td>
<td>With BDD diagnosis (n=322)</td>
<td>Without BDD diagnosis (n=1769)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------------</td>
<td>----------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>11 (1)</td>
<td>0 (0)</td>
<td>11 (1.3)</td>
</tr>
<tr>
<td>Other</td>
<td>65 (6.1)</td>
<td>7 (3.4)</td>
<td>58 (6.7)</td>
</tr>
<tr>
<td><strong>Psychiatric disorder, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td>493 (23.6)</td>
<td>150 (46.6)</td>
<td>343 (19.4)</td>
</tr>
<tr>
<td>Borderline personality disorder</td>
<td>244 (49.5)</td>
<td>84 (56)</td>
<td>160 (46.6)</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>9 (1.8)</td>
<td>3 (2)</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>33 (6.7)</td>
<td>12 (8)</td>
<td>21 (6.1)</td>
</tr>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>59 (12)</td>
<td>18 (12)</td>
<td>41 (12)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>7 (1.4)</td>
<td>1 (0.7)</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>292 (59.2)</td>
<td>108 (72)</td>
<td>184 (53.6)</td>
</tr>
<tr>
<td>Schizophrenia/psychosis</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>12 (2.4)</td>
<td>3 (2)</td>
<td>9 (2.6)</td>
</tr>
<tr>
<td>Other</td>
<td>18 (3.7)</td>
<td>3 (2)</td>
<td>15 (4.4)</td>
</tr>
<tr>
<td>History of plastic surgery procedures, n (%)</td>
<td>355 (17)</td>
<td>57 (17.7)</td>
<td>298 (16.8)</td>
</tr>
</tbody>
</table>

*aBDD: body dysmorphic disorder.

b€1=US $1.11.

**Figure 1.** Prevalence of body dysmorphic disorder among subsamples. BDD: body dysmorphic disorder.
Figure 2. Body parts of concern in our study population.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Univariate risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>284/1421 (20)</td>
<td>4.55 (3.14-6.58)</td>
<td>.001</td>
<td>3.02 (2.01-4.53)</td>
</tr>
<tr>
<td>Male</td>
<td>34/653 (5.2)</td>
<td>Ref</td>
<td>_b</td>
<td>.024</td>
</tr>
<tr>
<td>Other</td>
<td>4/17 (23.5)</td>
<td>5.60 (1.73-18.09)</td>
<td>.004</td>
<td>0.97 (0.95-0.99)</td>
</tr>
<tr>
<td>Mean age</td>
<td></td>
<td>0.93 (0.92-0.95)</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td><strong>Age categories (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>267/1091 (24.5)</td>
<td>2.69 (1.79-4.02)</td>
<td>.001</td>
<td>_b</td>
</tr>
<tr>
<td>25-44</td>
<td>30/279 (10.8)</td>
<td>Ref</td>
<td>—</td>
<td>.050</td>
</tr>
<tr>
<td>45-64</td>
<td>23/654 (3.5)</td>
<td>0.30 (0.17-0.53)</td>
<td>.001</td>
<td>0.93 (0.92-0.95)</td>
</tr>
<tr>
<td>&gt;64</td>
<td>2/66 (3)</td>
<td>0.26 (0.060-1.11)</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>263/1043 (25.2)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Worker</td>
<td>50/865 (5.8)</td>
<td>0.18 (0.13-0.25)</td>
<td>.001</td>
<td>0.51 (0.28-0.94)</td>
</tr>
<tr>
<td>Other</td>
<td>9/183 (4.9)</td>
<td>0.15 (0.07-0.30)</td>
<td>.001</td>
<td>0.35 (0.14-0.88)</td>
</tr>
<tr>
<td><strong>Dermatologic disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>204/1072 (19)</td>
<td>1.79 (1.40-2.29)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>118/1019 (11.6)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Acne</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>115/532 (21.6)</td>
<td>1.80 (1.40-2.32)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>207/1559 (13.3)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Atopic dermatitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>99/434 (22.8)</td>
<td>1.90 (1.46-2.48)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>223/1657 (13.5)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Other dermatitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10/35 (28.6)</td>
<td>2.23 (1.06-4.70)</td>
<td>.03</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>312/2056 (15.2)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Psoriasis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2/50 (4)</td>
<td>0.22 (0.05-0.92)</td>
<td>.02</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>320/1041 (15.7)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Psychiatric disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>150/493 (30.4)</td>
<td>3.63 (2.83-4.65)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>172/1598 (10.8)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Depressive Disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>84/244 (34.4)</td>
<td>3.55 (2.64-4.78)</td>
<td>.001</td>
<td>1.68 (1.06-2.69)</td>
</tr>
<tr>
<td>No</td>
<td>238/1847 (12.9)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Eating behavior disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>72/136 (52.9)</td>
<td>7.67 (5.34-11.02)</td>
<td>.001</td>
<td>3.77 (2.26-6.30)</td>
</tr>
<tr>
<td>No</td>
<td>250/1955 (12.8)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Obsessive-compulsive disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>12/33 (36.4)</td>
<td>3.22 (1.57-6.62)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>310/2058 (15.1)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Given the heterogeneity of the population, most of whom were women aged 18-24 years, we performed a multivariate analysis considering gender and age (Table 3). On the one hand, after multivariate analysis by gender in the group of women, students, eating disorder, and anxiety disorder remained with a significance at <.05 as diagnostic predictors of BDD (Table 3). In the male group after multivariate analysis, none of the factors analyzed showed statistical significance. On the other hand, in the multivariate analysis by age (Table 3), the following factors remained with significance at <.05 as diagnostic predictors of BDD. In the 18-24 years analysis, the diagnostic predictors were female gender and other gender, students, depressive disorder, eating disorder, and anxiety disorder. In the 25-44 years analysis, the diagnostic predictor was income level between €2000 and €2999. In the 45-64 years analysis, the diagnostic predictor was female gender, and the >64 years analysis was not performed due to the small sample size.
<table>
<thead>
<tr>
<th>Female gender</th>
<th>BDD(^a) prevalence, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>233/752 (30.9)</td>
<td>Ref(^b)</td>
<td>—(^c)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Worker</td>
<td>43/541 (7.9)</td>
<td>0.19 (0.14-0.27)</td>
<td>.001</td>
<td>0.46 (0.22-0.96)</td>
<td>.04</td>
</tr>
<tr>
<td>Other</td>
<td>9/128 (7)</td>
<td>0.17 (0.08-0.34)</td>
<td>.001</td>
<td>0.26 (0.09-0.72)</td>
<td>.01</td>
</tr>
<tr>
<td>Monthly income (€)(^d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;500</td>
<td>181/617 (29.3)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>500-999</td>
<td>30/107 (28)</td>
<td>0.94 (0.59-1.48)</td>
<td>.79</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1000-1999</td>
<td>34/255 (13.3)</td>
<td>0.37 (0.25-0.55)</td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2000-2999</td>
<td>8/192 (4.2)</td>
<td>0.10 (0.05-0.22)</td>
<td>.001</td>
<td>0.37 (0.15-0.93)</td>
<td>.04</td>
</tr>
<tr>
<td>&gt;3000</td>
<td>91/173 (5.2)</td>
<td>0.13 (0.07-0.26)</td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Eating behavior disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>69/125 (55.2)</td>
<td>6.195 (4.23-9.07)</td>
<td>.001</td>
<td>4.03 (2.32-6.98)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>215/1296 (16.6)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>103/250 (41.2)</td>
<td>3.83 (2.85-5.16)</td>
<td>.001</td>
<td>2.94 (1.65-5.25)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>181/1171 (15.5)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>18-24 years age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>236/787 (30)</td>
<td>4.20 (2.75-6.42)</td>
<td>.001</td>
<td>3.19 (2.05-4.97)</td>
<td>.001</td>
</tr>
<tr>
<td>Male</td>
<td>27/292 (9.2)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Other</td>
<td>4/12 (33.3)</td>
<td>4.91 (1.39-17.37)</td>
<td>.01</td>
<td>6.38 (1.72-23.67)</td>
<td>.006</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>255/1005 (25.4)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Worker</td>
<td>8/67 (11.9)</td>
<td>0.40 (0.19-0.85)</td>
<td>.02</td>
<td>0.39 (0.17-0.86)</td>
<td>.02</td>
</tr>
<tr>
<td>Dermatologic disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>177/640 (27.7)</td>
<td>1.53 (1.15-2.05)</td>
<td>.004</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>90/451 (20)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Psychiatric disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>130/301 (43.2)</td>
<td>3.62 (2.70-4.86)</td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>137/790 (17.3)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>73/152 (48)</td>
<td>3.55 (2.49-5.06)</td>
<td>.001</td>
<td>1.71 (1.04-2.84)</td>
<td>.04</td>
</tr>
<tr>
<td>No</td>
<td>194/930 (20.7)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Eating disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>66/107 (61.7)</td>
<td>6.27 (4.12-9.54)</td>
<td>.001</td>
<td>3.70 (2.13-6.42)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>201/984 (20.4)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>93/194 (47.9)</td>
<td>3.83 (2.76-5.30)</td>
<td>.001</td>
<td>2.14 (1.22-3.78)</td>
<td>.008</td>
</tr>
<tr>
<td>No</td>
<td>174/897 (19.4)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

25-44 years age group
<table>
<thead>
<tr>
<th>Monthly income (€)</th>
<th>BDD prevalence, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;500</td>
<td>9/35 (25.7)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>500-999</td>
<td>3/17 (17.6)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1000-1999</td>
<td>15/109 (13.8)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2000-2999</td>
<td>2/74 (2.7)</td>
<td>0.08 (0.02-0.40)</td>
<td>.002</td>
<td>0.12 (0.02-0.86)</td>
<td>.04</td>
</tr>
<tr>
<td>&gt;3000</td>
<td>1/43 (2.3)</td>
<td>0.07 (0.01-0.57)</td>
<td>.01</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

45-65 years age group

Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>BDD prevalence, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>21/403 (5.2)</td>
<td>6.76 (1.57-29.10)</td>
<td>.01</td>
<td>5.17 (1.14-23.45)</td>
<td>.03</td>
</tr>
<tr>
<td>Male</td>
<td>2/248 (0.8)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Dermatologic disease

<table>
<thead>
<tr>
<th>Dermatologic disease</th>
<th>BDD prevalence, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>16/273 (5.9)</td>
<td>3.33 (1.35-8.20)</td>
<td>.006</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>7/381 (1.8)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Other dermatitis conditions

<table>
<thead>
<tr>
<th>Other dermatitis conditions</th>
<th>BDD prevalence, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>3/13 (23.1)</td>
<td>9.31 (2.38-36.47)</td>
<td>.001</td>
<td>6.66 (1.44-30.87)</td>
<td>.02</td>
</tr>
<tr>
<td>No</td>
<td>20/641 (3.1)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

aBDD: body dysmorphic disorder.
bRef: reference value.
cNot applicable.
d€1=US $1.11.

Regarding quality of life, BDD was not statistically associated with physical health status. However, those diagnosed with BDD showed significantly lower levels of mental health than those without BDD (Figure 3). Additionally, quality of life analysis was performed considering psychiatric comorbidity as a factor affecting quality of life. Differences regarding mental health status remained statistically significant (P < .05) for patients with BDD, irrespective of the presence of psychiatric pathology (Figure 4).

Figure 3. Association of body dysmorphic disorder with (A) physical and (B) mental health status. BDD: body dysmorphic disorder.
Figure 4. Perceived quality of life. (A) Physical health quality of life among patients without psychiatric comorbidities. (B) Physical health quality of life among patients with psychiatric comorbidities. (C) Mental health quality of life among patients without psychiatric comorbidities. (D) Mental health quality of life among patients with psychiatric comorbidities. BDD: body dysmorphic disorder.

Discussion

Principal Findings

Our study reveals many interesting aspects of BDD. Most of our findings are consistent with those reported previously [1,4,6,12,13]. For example, BDD is particularly prevalent in the young adult population [5], and patients with BDD are particularly concerned about an average of 4.6 different body parts. However, some findings are novel in our study. For instance, the prevalence of BDD in our sample population (15.2%) was higher than expected (0.5%-3.2%) [1]. Further, in addition to the described association between BDD and eating and depressive disorders, we report that BDD is closely associated with anxiety disorders. The most relevant finding of our study is the perception of quality of life: participants with BDD had a poorer perception of quality of life associated with mental but not physical health problems. Moreover, the perception of quality of mental health in patients with BDD was independent of diagnosis of any mental disorder.

Our study reports that the prevalence of BDD in adults is 15.2% in Spain, which is higher than that reported in the general population (0.5%-3.2%) in another study [1] and higher than that reported in a Spanish multicentric study in patients with acne (10.6%) [2]. Our findings may be explained by 2 factors. First, our sample population was particularly young, with more than half of the participants being in the age group of 18-24 years. As shown in a previous study [6], the younger the age of the patient, the greater is the possibility of BDD diagnosis. Second, our data reflect more of a screening diagnosis as compared with a definitive diagnosis established with BDDQ via an interview by a health professional, which is more demanding [17].

Our study shows that the sociodemographic characteristics most associated with the diagnosis of BDD are gender (female and other), age group of 18-24 years, students, income level of less than €500/month, and a diagnosis of previous dermatologic or psychiatric disease. Regarding gender, females showed a statistically significant (P<.001) higher prevalence of BDD than males (284/1421, 19.9% vs 34/653, 5.2%, respectively). Previous studies have reported similar prevalence between men and women [3] or increased prevalence in women (68.5%) compared to that in men [4]. In our study, almost 67.9% (1421/2091) of the participants were women, which may suggest the need for future studies controlled by sex to clarify the differences. Regarding age, the mean age at diagnosis of BDD in the participants in our study was 23.5 years, which is higher than the mean age of 16.9 years described elsewhere [5]. This is probably because our study only included populations 18
Participants with BDD in our study were concerned about an average of 4.6 different body parts, which is in line with that reported by previous studies (5-7 body parts) [13]. The body parts that were of the most concern were body fat (248/322, 77%), thighs (191/322, 59.3%), face (166/322, 51.6%), hip (144/322, 44.7%), and skin (126/322, 39.1%). Skin (14/26, 54%) and face (8/26, 31%) were the body parts of the greatest concern in previous studies [1]. The prevalence of BDD in participants with dermatologic conditions (204/1072, 19%) in our study falls within the range (4.9%-21.1%) reported in the literature [1]. However, it is necessary to emphasize that previous studies did not distinguish between dermatologic patients per se and those undergoing aesthetic procedures [1].

Regarding the association between previous dermatologic disease and BDD, we found no significant relationships. The association between BDD and acne (115/532, 21.6%) in our study was slightly higher than that previously described (8.8%-21.1%) [1]. The association of BDD with having undergone rhinoplasty was 12.1% (39/322), which was lower than that previously documented (20.1%) [1]. However, the sample size limited our capability to extract meaningful conclusions on this issue. Among participants with previous psychiatric pathology and BDD, there was a significant association with eating disorders (72/136, 52.9%) and depressive disorders (84/244, 34.4%), similar to the findings of 12%-45% and 47%-56.3%, respectively, reported in a previous systematic review [1]. In addition, our study showed a significant comorbidity with anxiety disorders (108/292, 36.9%), making it necessary to conduct future studies in this subgroup.

Regarding quality of life, participants diagnosed with BDD had a perception of having a poorer mental health status than those without BDD. In contrast, there were no significant differences in the physical status between participants with and without BDD [12]. Ultimately, the diagnosis of BDD was associated with a perception of reduced quality of life that is not subsidiary to the presence of mental health disorders. In other words, our study suggests that BDD could be used as a marker or predictor of an individual’s perception of quality of life, which is independent of the presence of mental health problems.

**Limitations and Strengths**

Our study was based on the use of a questionnaire that was disseminated telematically, which is why we obtained a heterogeneous participant base, being represented mostly by female students in an age range of 18-24 years. The higher percentage of prevalence obtained may be linked to the specific population in this study. However, we must bear in mind that, despite obtaining the first diagnostic approximation with BDDQ, its confirmation must subsequently be backed up by an interview with a health professional [17]. Furthermore, the mean age at BDD diagnosis (23.5 years) in our study was higher than that (16.9 years) described previously [5] because this study was limited to participants aged 18 years and older. Additionally, the number of participants in our study who had undergone previous plastic surgeries was too low to achieve a proper statistical power.

With regard to the strengths of this study, we increased the number of variables related to BDD compared to the number of variables used in previous studies [2-6,12,13], which, together with the total number of participants, resulted in a large database. Further, we introduced the quality of life indicators through the SF-12v2 health survey.

**Conclusion**

Patients with BDD experience serious biopsychosocial repercussions. This study provides the first approximation of the prevalence of BDD in the Spanish population, which was found to be higher than expected, although our results should be interpreted cautiously. BDD was particularly prevalent in participants aged 18-24 years, students, and women. BDD is associated with psychiatric conditions such as eating disorder, anxiety disorder, and depressive disorder, and with dermatologic conditions such as acne and dermatitis. No significant associations were found between BDD and the performance of previous aesthetic procedures, which could be due to our small sample size. Finally, BDD could be a marker of an individual’s perception of quality of health, irrespective of psychiatric diagnoses. Future studies should confirm our preliminary findings.

**Acknowledgments**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Authors’ Contributions**

HBF and BRA conceptualized this study, contributed to the methodology, supervised this study, and contributed to the resources. AL and CL curated the data and contributed to visualization. CL and MBF performed a formal analysis of this study. AL, HBF, and BRA contributed to the investigations. HBF contributed to project administration. AL and MBF wrote the original draft. HBF performed reviewing and editing.
Conflicts of Interest

In the last 24 months, HBF has received lecture fees from Takeda, BIAL, laboratorios Rubio, and laboratorios Rovi. He has also been granted 3 prizes for the development of a serious videogame for treating attention-deficit/hyperactivity disorder (The Secret Trail of Moon) called as the Shibuya Prize by Takeda, first prize of the College of Psychologists of Madrid, and a prize for the Best Innovative Health Initiative within Healthstart. He is the Principal Investigator of an iPfIS research contract (accessed on August 12, 2022; IF16/00039), Coprincipal Investigator of a MINECO research grant (RTI2018-101857-B-I00), and Principal Investigator of a research of the Sincronia project, funded by the start-up Bitsphi; recipient of a Fund for the Improvement of Postsecondary Education grant and an IDIHP/PSA intensification grant; involved in 2 clinical trials (Mensia Koala, Newrofeed Study; ESKETSUI2002); and a cofounder of Haglaia Solutions. He is also an employee and member of the advisory board of Ita Salud Mental (Korian).

References


Abbreviations

BDD: body dysmorphic disorder
BDDQ: Body Dysmorphic Disorder Questionnaire
SF-12v2: 12-item Short Form version 2
Original Paper

Methodological Insights on Recruitment and Retention From a Remote Randomized Controlled Trial Examining the Effectiveness of an Alcohol Reduction App: Descriptive Analysis Study

Melissa Oldham¹, PhD; Larisa Dinu¹, MSc; Gemma Loebenberg¹, MSc; Matt Field², PhD; Matthew Hickman³, PhD; Susan Michie⁴, PhD; Jamie Brown¹, PhD; Claire Garnett¹, PhD

¹University College London, London, London, United Kingdom
²Department of Psychology, University of Sheffield, Sheffield, United Kingdom
³Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
⁴Centre for Behaviour Change, University College London, London, United Kingdom

Corresponding Author:
Melissa Oldham, PhD
University College London
London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 20 7679 883
Email: m.oldham@ucl.ac.uk

Abstract

Background: Randomized controlled trials (RCTs) with no in-person contact (ie, remote) between researchers and participants offer savings in terms of cost and time but present unique challenges.

Objective: The goal of this study is to examine the differences between different forms of remote recruitment (eg, National Health Service [NHS] website, social media, and radio advertising) in the proportion of participants recruited, demographic diversity, follow-up rates, and cost. We also examine the cost per participant of sequential methods of follow-up (emails, phone calls, postal surveys, and postcards). Finally, our experience with broader issues around study advertising and participant deception is discussed.

Methods: We conducted a descriptive analysis of 5602 increasing-and-higher-risk drinkers (Alcohol Use Disorders Identification Test score ≥8), taking part in a 2-arm, parallel group, remote RCT with a 1:1 allocation, comparing the intervention (Drink Less app) with usual digital care (NHS alcohol advice web page). Participants were recruited between July 2020 and March 2022 and compensated with gift vouchers of up to £36 (a currency exchange rate of £1=US $1.26988 is applicable) for completing follow-up surveys, with 4 stages of follow-up: email reminders, phone calls, postal survey, and postcard.

Results: The three main recruitment methods were advertisements on (1) social media (2483/5602, 44.32%), (2) the NHS website (1961/5602, 35.01%), and (3) radio and newspapers (745/5602, 13.3%), with the remaining methods of recruitment accounting 7.37% (413/5602) of the sample. The overall recruitment cost per participant varied from £0 to £11.01. Costs were greater when recruiting participants who were men (£0-£28.85), from an ethnic minority group (£0-£303.81), and more disadvantaged (£0-£49.12). Targeted approaches were useful for recruiting more men but less useful in achieving diversity in ethnicity and socioeconomic status. Follow-up at 6 months was 79.58% (4458/5602). Of those who responded, 92.4% (4119/4458) responded by email. Each additional stage of follow-up resulted in an additional 2-3 percentage points of the overall sample being followed up, although phone calls, postal surveys, and postcards were more resource intensive than email reminders.

Conclusions: For remote RCTs, researchers could benefit from using a range of recruitment methods and cost-targeted approaches to achieve demographic diversity. Automated emails with substantial financial incentives for prompt completion can achieve good follow-up rates, and sequential, offline follow-up options, such as phone calls and postal surveys, can further increase follow-up rates but are comparatively expensive. We also make broader recommendations focused on striking the right balance when designing remote RCTs. Careful planning, ongoing maintenance, and dynamic decision-making are required throughout a trial to balance the competing demands of participation among those eligible, deceptive participation among those who are not eligible, and ensuring no postrandomization bias is introduced by data-checking protocols.
Introduction

Randomized controlled trials (RCTs) are used to examine the efficacy of interventions on a wide range of health-related behaviors and outcomes [1-4]. RCTs examining the efficacy of digital interventions are increasingly taking place on the web or remotely. Web-based trials feature no in-person contact between researchers and participants, with the administration of the intervention and all measures completed on the web. Remote trials also have no in-person contact between researcher and participant but may involve some offline follow-up options, such as completing surveys over the phone or by post. Web-based and remote trials can be cheaper and less labor-intensive than in-person trials, although they present some unique challenges around recruitment, retention, and participant deception. Here, we present methodological insights from a large-scale (n=5602) remote RCT examining the effectiveness of a digital intervention, the “Drink Less” app [5], in helping increasing-and-higher-risk drinkers (Alcohol Use Disorders Identification Test [AUDIT] score ≥8) to reduce their alcohol consumption.

Digital interventions, such as websites and apps, are increasingly being used for a wide range of health behaviors [6] and can offer benefits over in-person interventions in terms of cost, convenience, and anonymity [7]. RCTs aiming to evaluate digital interventions can be conducted on the web or remotely and may have several advantages relative to trials requiring in-person contact. First, web-based and remote settings could increase the external validity of the trial, as having to travel to in-person appointments for baseline or follow-up assessments does not reflect real-world implementation or how users access digital interventions [8]. Second, in theory, participants can be recruited from throughout nations or even globally, giving a larger and potentially more generalizable sampling frame [9]. Third, the cost of web-based or remote trials is likely to be much less as they can be partly automated, reducing demands on researcher time, and could potentially reduce researcher bias through double blind [9].

However, there are also significant challenges with web-based or remote RCTs beyond those conducted in person. First, it may be harder to recruit participants or to recruit a broadly representative sample [10,11], as some groups, such as older adults and people from less advantaged communities, may be less likely to engage with research conducted remotely [12]. Second, researchers have less control over who signs up, and it is possible that motivated individuals may sign up multiple times for financial incentives [8,13]. Third, once recruited, researchers may have less control over how participants engage with the intervention [14] or respond to follow-ups [8]. This could be particularly problematic with groups who may have low digital literacy and may not understand how to use the intervention, although this may be reflective of how people would engage with digital interventions in real-world settings. There are other challenges that are present in both remote and in-person trials. Contamination occurs when the comparator group finds the intervention being tested outside of the trial [9]. This could be particularly likely if the comparator group receives an intervention they do not deem acceptable and seeks out alternatives. These biases could introduce bias into RCTs, which could obscure the effect of the intervention.

Here, we draw on data from a large-scale remote RCT, evaluating the effectiveness of the Drink Less app [5] compared with usual digital care (the National Health Service [NHS] alcohol advice web page). Drink Less is a theory- and evidence-informed, app-based intervention designed by researchers [15,16] to help increasing-and-higher-risk drinkers reduce their alcohol consumption. To mitigate some of the potential challenges outlined above, the trial used a multipronged recruitment strategy, including an advertisement on the NHS website and social media advertising [5]. In line with previous research [11], and to maximize follow-up rates, we offered substantial financial incentives to complete follow-up surveys, including an additional amount for completing the primary outcome within the first 24 hours, and undertook a comprehensive follow-up approach by sequentially sending follow-up reminders through email, SMS text messages, and telephone and by post. These strategies and broader methodological issues will be discussed ahead.

This study aims to:

1. Compare different remote recruitment methods in terms of cost per recruited participant, retention rates, participant deception, and sociodemographic diversity.
2. Compare the proportion of returned responses using different strategies for follow-up at each time point, and compare the cost and time associated with each follow-up stage.
3. Consider broader methodological issues pertaining to recruitment, retention, and participant deception, and discuss the success of strategies to mitigate these issues throughout the trial.

Methods

The protocol and analysis plan were preregistered on the Open Science Framework [17]. The trial was registered (ISRCTN64052601). The main trial findings are reported elsewhere [18].

Design

Participants
A total of 5602 participants were randomized in the RCT evaluating Drink Less. Participants were eligible if they were aged 18 years or older, lived in the United Kingdom, were increasing-and-higher-risk drinkers (AUDIT score ≥8) [19], had access to an iOS device (iPhone, iPod touch, or iPad), and wanted to drink less alcohol. Recruitment ran from July 2020 to March 2022 and included an advertisement on the NHS website, a mail-out to a database of UK-based users of the smoking cessation app “Smoke Free”, radio and social media advertising, press releases, and local advertising through health care providers. Advertisements were codeveloped with public representatives.

Informed consent was sought at baseline to participate in 3 web-based follow-up surveys at 1, 3, and 6 months. Surveys were completed on the web through Qualtrics (Silver Lake), although at the 6-month follow-up, offline options (eg, phone and post) were available. The 6-month follow-up survey assessed primary and secondary outcomes relating to alcohol use and a range of related measures. The 1- and 3-month follow-up surveys only assessed secondary outcome measures relating to alcohol use. We attempted to contact participants within 30 days of their first invitation to complete each follow-up survey. To maximize data retention and to allow for time taken for answers to be posted at 6-month follow-up, data provided up to 2 weeks after the 30-day period were accepted.

Initially, as well as through 3 emails (days 0, 5, and 9) and (from January 15, 2022) a total of 2 SMS text messages (days 5 and 9), we had planned that at the 1-, 3-, and 6-month follow-up, all participants would also be sequentially offered opportunities to complete follow-up through phone (called twice from days 10 to 17), a mailed survey (from day 18), and a mailed postcard (from day 30). However, due to resource constraints, from November 2020 on, we only used automated emails on days 0, 5, 9, and 11 to contact participants at the 1- and 3-month follow-up; we no longer called or sent postal surveys. With the aim of improving these follow-up rates with less resource, we added SMS text messaging follow-ups. Phone calls, mailed surveys, and postcard follow-ups were retained for the 6-month follow-up survey (when the primary outcome was measured).

Measures
Recruitment Method
At baseline, participants were asked to specify where they saw the study advertised, with the following response options: NHS website, social media (eg, Facebook and Twitter [subsequently rebranded X]), other media (eg, radio and newspapers), emailed by the Smoke Free app, local health care provider, word of mouth, Google, general practitioner (GP) surgery, or other. If they selected “other,” free-text responses that fell within one of the response options were recoded (eg, Facebook would be social media). The response options “local health care provider” and “GP surgery” were collapsed. Throughout the study, both untargeted and targeted (eg, at men) social media advertisements were used. These were analyzed separately.

Participant Deception
We experienced 3 distinct subgroups of participant deception throughout the trial: duplicates, manual fraud, and bots. Duplicate responses, where individuals signed up more than once with identical names and phone numbers, were the least prevalent (n=49) and easiest to detect. Data checks were undertaken each month to search for duplicate values. Manual fraud was a more prevalent form of participant deception (n=297), defined as individuals who signed up multiple times with false information, such as phone numbers linked to businesses where they were not known or addresses that did not exist. To identify manual fraud, monthly checks were made on all addresses and telephone numbers provided to ensure street names matched the postcode and that numbers were mobile phone numbers. Any suspicious responses were flagged, and the participants were contacted and asked to confirm their details over the phone. Where individuals were not known at the phone number provided, they were removed from the study. To make it easier to automatically screen out those engaging in manual fraud, we added attention checks, whereby individuals were asked to select a certain response option. Participants were also asked to confirm their age at 2 different points in the baseline survey to ensure they were consistent. Individuals failing either of these attention checks were screened out of the survey before randomization. The most prevalent type of fraud were “bot” responses (n=863). These were fraudulent responses similar to manual fraud, but they occurred in batches of 20-30 at a time when contact information was given in noticeably similar formats (eg, firstname123@emailaddress.com), often with American street addresses (being UK-based was an inclusion criteria of the trial). These responses seemed to be automated and were identified using the same process of address checking as above (individuals not known at the phone number provided were removed from the study). Adding a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) to the survey eliminated this issue. A more detailed discussion on participant deception is described elsewhere [20].

Sociodemographic Characteristics
Sociodemographic measures were assessed at baseline. This study focuses on gender, ethnicity, and occupation (to derive socioeconomic status [SES]: ABC1 [managerial, professional, and intermediate occupations] versus C2DE [skilled, semiskilled, unskilled manual, and lowest-grade worked or unemployed]).

Analysis
Aim 1: Methods of Recruitment
Each recruitment method is compared in terms of the proportion of enrolled participants, the proportion of participants who were men, from a minority ethnic group, or from a more disadvantaged background (C2DE), and the proportion of participant deception. Cost-per-recruited participant citing each recruitment method (eg, total spend on recruitment method divided by the number of participants citing recruitment method) is reported. As well as the overall cost per participant, we also present the cost per participant stratified by gender (eg, for each
man recruited), ethnicity (eg, for those from ethnic minority individuals), and SES (eg, for those from more disadvantaged backgrounds). Finally, we present follow-up rates at 1-, 3-, and 6-months for each method of recruitment.

**Aim 2: Follow-Up**

The proportion of the sample responding at each sequential stage of follow-up (ie, emails, phone calls, postal surveys, and postcards) is reported. The cost of each follow-up stage per participant responding at each stage is also reported. This was derived by dividing the estimated researcher time and other relevant costs by the number of follow-ups completed at each stage.

**Aim 3: Broader Methodological Issues**

Broader methodological issues such as advertising, participant deception, technical support, contamination, and boosting retention are discussed. We describe and briefly discuss the strategies we used throughout the trial to mitigate issues.

**Ethical Considerations**

Ethical approval for the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone app) trial was granted by the University College London (UCL) Ethics Committee (16799/001). Participants provided informed consent before participating in the trial. Study data were pseudonymized and stored on a secure university drive. Participants were compensated with gift vouchers of up to £36 (a currency exchange rate of £1=US $1.26988 is applicable) for completing the 3 surveys: £6 for the survey at 1 and 3 months and £12 at 6 months, with an additional £12 if the 6-month survey was completed within 24 hours.

**Results**

**Sample Characteristics**

A total of 5602 participants completed the baseline survey between July 2020 and March 2022: 65.78% (3685/5602) responded at 1-month follow-up, 63.80% (3574/5602) at 3-month follow-up, and 79.58% (4458/5602) at 6-month follow-up. Over half (3207/5602, 57.25%) of the sample were women, 42.22% (2365/5602) were men, 0.46% (26/5602) were “other,” and 0.07% (4/5602) preferred not to say. Most of the sample were White (5296/5602, 94.54%) and earned above-average income (4151/5602, 74.01%). The sample characteristics were similar at each follow-up. Table 1 reports the sociodemographic characteristics of the sample at baseline and those responding at each stage of follow-up.

Table 1. Sample characteristics at baseline and among those who responded at 1-month, 3-month, and 6-month follow-up for increasing-and-higher-risk drinkers participating in the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial (RCT).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline</th>
<th>1-month follow-up</th>
<th>3-month follow-up</th>
<th>6-month follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>3207 (57.25)</td>
<td>2046 (55.52)</td>
<td>1992 (55.74)</td>
<td>2534 (56.84)</td>
</tr>
<tr>
<td>Women</td>
<td>2365 (42.22)</td>
<td>1620 (43.96)</td>
<td>1565 (43.79)</td>
<td>1903 (42.69)</td>
</tr>
<tr>
<td>Other</td>
<td>26 (0.46)</td>
<td>16 (0.43)</td>
<td>14 (0.39)</td>
<td>17 (0.38)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>4 (0.07)</td>
<td>3 (0.08)</td>
<td>3 (0.08)</td>
<td>4 (0.09)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>96 (1.71)</td>
<td>68 (1.85)</td>
<td>68 (1.9)</td>
<td>83 (1.86)</td>
</tr>
<tr>
<td>Black</td>
<td>47 (0.84)</td>
<td>35 (0.95)</td>
<td>39 (1.09)</td>
<td>41 (0.92)</td>
</tr>
<tr>
<td>Chinese</td>
<td>9 (0.16)</td>
<td>9 (0.24)</td>
<td>9 (0.25)</td>
<td>9 (0.2)</td>
</tr>
<tr>
<td>White</td>
<td>5296 (94.54)</td>
<td>3474 (94.27)</td>
<td>3361 (94.04)</td>
<td>4206 (94.35)</td>
</tr>
<tr>
<td>Mixed</td>
<td>113 (2.02)</td>
<td>75 (2.03)</td>
<td>71 (1.99)</td>
<td>84 (1.88)</td>
</tr>
<tr>
<td>Other</td>
<td>21 (0.37)</td>
<td>15 (0.41)</td>
<td>15 (0.42)</td>
<td>18 (0.4)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>19 (0.34)</td>
<td>9 (0.24)</td>
<td>11 (0.31)</td>
<td>16 (0.36)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (0.02)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0.02)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC1 b</td>
<td>4151 (74.01)</td>
<td>2759 (74.87)</td>
<td>2688 (75.21)</td>
<td>3337 (74.85)</td>
</tr>
<tr>
<td>C2DE c</td>
<td>1451 (25.9)</td>
<td>926 (25.13)</td>
<td>886 (24.79)</td>
<td>1121 (25.15)</td>
</tr>
</tbody>
</table>

aThe data is also reported in the main trial paper [18].
bABC1: managerial, professional, and intermediate occupations.
cC2DE: skilled, semiskilled, unskilled manual, and lowest-grade worked or unemployed.
Aim 1: Recruitment Methods, Demographic Diversity, and Cost Per Participant

Most participants recruited for this trial reported seeing it advertised on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), or through radio or newspapers (745/5602, 13.3%), with all other recruitment methods accounting for 7.37% (413/5602) of the sample (Table 2).

Table 2. Total recruitment and proportion of recruited sample of iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial who were men, of minority ethnic groups, had lower socioeconomic status (SES), and identified as a fraudulent response by recruitment method.

<table>
<thead>
<tr>
<th>Recruitment method</th>
<th>Included sample (N=5602), n (%)</th>
<th>Men, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ethnic minority group, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Low SES, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Fraudulent response, n/N (%)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untargeted social media</td>
<td>2119 (37.82)</td>
<td>650 (30.67)</td>
<td>147 (6.94)</td>
<td>507 (23.93)</td>
<td>1020/3139 (32.49)</td>
</tr>
<tr>
<td>Targeted social media</td>
<td>364 (6.5)</td>
<td>353 (96.98)</td>
<td>13 (3.57)</td>
<td>90 (24.73)</td>
<td>8/372 (2.15)</td>
</tr>
<tr>
<td>National Health Service website</td>
<td>1961 (35.01)</td>
<td>628 (32.02)</td>
<td>76 (3.88)</td>
<td>570 (29.07)</td>
<td>123/2084 (5.9)</td>
</tr>
<tr>
<td>Radio or newspapers</td>
<td>745 (13.3)</td>
<td>591 (79.33)</td>
<td>27 (3.62)</td>
<td>167 (22.42)</td>
<td>19/764 (2.49)</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>142 (2.53)</td>
<td>74 (52.11)</td>
<td>9 (6.34)</td>
<td>41 (28.87)</td>
<td>11/153 (7.19)</td>
</tr>
<tr>
<td>Google</td>
<td>159 (2.84)</td>
<td>44 (27.67)</td>
<td>7 (4.4)</td>
<td>50 (31.45)</td>
<td>9/168 (5.36)</td>
</tr>
<tr>
<td>Smoke Free email</td>
<td>55 (0.98)</td>
<td>13 (23.64)</td>
<td>3 (5.45)</td>
<td>11 (20)</td>
<td>10/65 (15.38)</td>
</tr>
<tr>
<td>Health care provider or general practitioner</td>
<td>15 (0.27)</td>
<td>5 (33.33)</td>
<td>3 (20)</td>
<td>4 (26.67)</td>
<td>16/31 (51.61)</td>
</tr>
<tr>
<td>Other</td>
<td>42 (0.75)</td>
<td>7 (16.67)</td>
<td>2 (4.76)</td>
<td>11 (26.19)</td>
<td>0/42 (0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The percentage of participants recruited from each method (ie, N=Included sample value).

<sup>b</sup>The percentage of the overall sample from each recruitment method including those removed after participant deception checks.

Ongoing sociodemographic tracking throughout the study revealed that women, White, and advantaged participants were being overrecruited. In response, strategies targeted at a more diverse sample in terms of gender, SES, and ethnicity were introduced with mixed success. This included targeted social media advertisements aimed at men and radio advertisements on Talk Radio, Asian Sounds (in English and Urdu), and Punjabi Radio (in English and Punjabi).

Recruitment methods differed in the proportion of men (range 17%-97%), with targeted approaches including social media advertising (353/364, 97% men) and radio advertising (591/745, 79.3%) being the most successful in recruiting a sample of men. Word of mouth was most effective in terms of recruiting a balanced sample in terms of gender (74/142, 52.1% men) but recruited a small proportion (142/5602, 2.53%) of the sample overall.

Recruiting through GP surgeries and local health care providers resulted in the highest proportion of participants from minority ethnic groups (3/15, 20%) but recruited a small proportion of participants in total (15/5602, 0.27%). Untargeted social media advertisements and word of mouth were the next best, with 6.94% (147/2119) and 6.3% (9/142) of the sample coming from ethnic minority individuals, respectively.

The NHS website, word of mouth, and Google all recruited around a third of participants who were more disadvantaged. However, both Google (159/5602, 2.84%) and word of mouth (142/5602, 2.53%) recruited a small proportion of participants in total.

The final column of Table 2 refers to the proportion (and number) of participants who were removed from the study due to participant deception, citing each recruitment method. A total of 84.54% (1028/1216) of participants identified as fraudulent cited social media as the place they saw the advertisement. It should be noted here that these participants may not have been honest in terms of where they saw the study advertisement and may have been deliberately misreporting where they found the study or responding at random.

Money spent on each of the recruitment methods varied from £0 for the NHS advertisement and word of mouth to £8203 for radio or newspaper advertisements (Table 3; a currency exchange rate of £1=US $1.26988 is applicable). Of the paid forms of recruitment, social media advertising and advertising through health care providers were the cheapest ways of recruiting participants who were men, of ethnic minorities, or from more disadvantaged backgrounds.

Although the overall number of participants recruited from health care settings was low, this was impeded by the COVID-19 pandemic. The initial recruitment plan was to have posters in primary care surgeries throughout the United Kingdom; however, due to the pandemic and associated lockdowns for most of the recruitment period, many people received health care on the web and were not visiting GP surgeries. We only started advertising in GP surgeries for the last 5 months of trial recruitment (in November 2021).

Those recruited from health care providers (15/15, 100%), Smoke Free email (51/55, 93%), and word of mouth (126/142, 88.7%) appeared to have the highest response rates and those recruited through advertisements on Google (109/159, 69%), and the NHS website (1513/1961, 77%) appeared among the lowest. Table 4 presents the follow-up rates at 1-, 3-, and 6-month follow-up.

https://formative.jmir.org/2024/1/e51839
Table 3. Total cost per participant and cost per participant who were men, of ethnic minority groups, and lower socioeconomic status (SES) by recruitment method for participants of the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial. A currency exchange rate of £1=US $1.26988 is applicable.

<table>
<thead>
<tr>
<th>Recruitment method</th>
<th>Recruited, n</th>
<th>Total cost (£)</th>
<th>Cost per participant (£)</th>
<th>Cost per man (£)</th>
<th>Cost per ethnic minority participant (£)</th>
<th>Cost per low SES participant (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untargeted social media</td>
<td>2119</td>
<td>6750.00</td>
<td>3.19</td>
<td>10.38</td>
<td>45.92</td>
<td>13.31</td>
</tr>
<tr>
<td>Targeted social media</td>
<td>364</td>
<td>690.00</td>
<td>1.90</td>
<td>1.95</td>
<td>53.08</td>
<td>7.67</td>
</tr>
<tr>
<td>National Health Service website</td>
<td>1961</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Radio or newspapers</td>
<td>745</td>
<td>8203.00</td>
<td>11.01</td>
<td>13.88</td>
<td>303.81</td>
<td>49.12</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>142</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Google</td>
<td>159</td>
<td>1247.00</td>
<td>7.84</td>
<td>28.34</td>
<td>178.14</td>
<td>138.56</td>
</tr>
<tr>
<td>Smoke Free email</td>
<td>55</td>
<td>375.00</td>
<td>6.82</td>
<td>28.85</td>
<td>125.00</td>
<td>34.09</td>
</tr>
<tr>
<td>Health care provider or general practitioner</td>
<td>15</td>
<td>61.00</td>
<td>4.07</td>
<td>12.20</td>
<td>20.33</td>
<td>15.24</td>
</tr>
<tr>
<td>Other</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4. Follow-up rates at 1-, 3-, and 6-month follow-up among increasing-and-higher-risk drinkers participating in the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial by recruitment method.

<table>
<thead>
<tr>
<th>Recruitment method</th>
<th>Recruited, n</th>
<th>Follow-up rate at 1 month, n (%)</th>
<th>Follow-up rate at 3 months, n (%)</th>
<th>Follow-up rate at 6 months, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untargeted social media</td>
<td>2119</td>
<td>1376 (64.94)</td>
<td>1340 (63.24)</td>
<td>1708 (80.6)</td>
</tr>
<tr>
<td>Targeted social media</td>
<td>364</td>
<td>277 (76.1)</td>
<td>256 (70.33)</td>
<td>295 (81.04)</td>
</tr>
<tr>
<td>National Health Service website</td>
<td>1961</td>
<td>1237 (63.08)</td>
<td>1210 (61.7)</td>
<td>1513 (77.15)</td>
</tr>
<tr>
<td>Radio or newspapers</td>
<td>745</td>
<td>520 (69.8)</td>
<td>495 (66.44)</td>
<td>603 (80.94)</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>142</td>
<td>107 (75.35)</td>
<td>106 (74.65)</td>
<td>126 (88.73)</td>
</tr>
<tr>
<td>Google</td>
<td>159</td>
<td>87 (54.72)</td>
<td>81 (50.94)</td>
<td>109 (68.55)</td>
</tr>
<tr>
<td>Smoke Free email</td>
<td>55</td>
<td>38 (69.09)</td>
<td>43 (78.18)</td>
<td>51 (92.73)</td>
</tr>
<tr>
<td>Health care provider or general practitioner</td>
<td>15</td>
<td>12 (80)</td>
<td>10 (66.66)</td>
<td>15 (100)</td>
</tr>
<tr>
<td>Other</td>
<td>42</td>
<td>31 (73.81)</td>
<td>33 (78.57)</td>
<td>38 (90.48)</td>
</tr>
</tbody>
</table>

Aim 2: Retention During Sequential Follow-Up
At 6-month follow-up, 92.4% (4119/4458) of those who responded did so in response to 1 of the 3 email notifications. An additional 2.02% (90/4458) responded following 2 phone calls from the research team, and 3.25% (145/4458) responded following a postal survey. The final stage of recruitment, a postcard sent through mail to participants featuring just the key outcome measure for the trial (AUDIT-C), yielded a further 2.33% (104/4458) of the followed-up sample. The estimated costs of each sequential stage of follow-up are presented in Tables 5 and 6.
Table 5. A summary of time spent and cost on different stages of contacting participants of the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trials at 1- and 3-month follow-up. A currency exchange rate of £1=US $1.26988 is applicable.

<table>
<thead>
<tr>
<th>Time point and method</th>
<th>Total follow-up sent, n</th>
<th>Follow-up per month, mean (range)</th>
<th>Responded</th>
<th>Hours spent sending follow-up&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Hours spent sending vouchers</th>
<th>Total research hours&lt;sup&gt;b&lt;/sup&gt; (£)</th>
<th>Cost research hours&lt;sup&gt;b&lt;/sup&gt; (£)</th>
<th>Other costs (£)</th>
<th>Total cost (£)</th>
<th>Cost per participant (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1-month follow-up</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated email</td>
<td>5602</td>
<td>267 (65-621)</td>
<td>1874</td>
<td>0</td>
<td>34</td>
<td>34</td>
<td>672</td>
<td>0</td>
<td>672</td>
<td>0.36</td>
</tr>
<tr>
<td>First manual email and SMS text message&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3800 and 1057</td>
<td>181 (34-448)</td>
<td>1130</td>
<td>130</td>
<td>20</td>
<td>150</td>
<td>2966</td>
<td>106&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3072</td>
<td>2.72</td>
</tr>
<tr>
<td>Second and third manual email and SMS text message&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3291 and 648</td>
<td>175 (0-462)</td>
<td>643</td>
<td>112</td>
<td>12</td>
<td>124</td>
<td>2452</td>
<td>65&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2517</td>
<td>3.91</td>
</tr>
<tr>
<td><strong>3-month follow-up</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated email</td>
<td>5602</td>
<td>267 (65-621)</td>
<td>2053</td>
<td>0</td>
<td>37</td>
<td>37</td>
<td>732</td>
<td>0</td>
<td>732</td>
<td>0.36</td>
</tr>
<tr>
<td>First manual email and SMS text message&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3610 and 1282</td>
<td>172 (26-419)</td>
<td>1056</td>
<td>123</td>
<td>19</td>
<td>142</td>
<td>2807</td>
<td>128&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2935</td>
<td>2.78</td>
</tr>
<tr>
<td>Second and third manual email and SMS text message&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3698 and 874</td>
<td>176 (0-511)</td>
<td>460</td>
<td>126</td>
<td>8</td>
<td>134</td>
<td>2649</td>
<td>87&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2736</td>
<td>5.95</td>
</tr>
</tbody>
</table>

<sup>a</sup>Time spent sending manual reminders and SMS text messages. On average, an email and SMS text message reminder took 2 minutes and 5 seconds to send, and a voucher email took 1 minute and 8 seconds to send.

<sup>b</sup>The cost here is the average of 2 research staff salaries (£19.77) × research hours.

<sup>c</sup>For the first 3 months of follow-up, we contacted participants twice manually by email, followed sequentially by phone calls, a written survey, and a postcard with the primary outcomes. However, this was not sustainable, so the subsequent follow-up stages were dropped at 1 and 3 months and are not presented below but are included in this total. 1-month phone completions=22, and 1-month postcard completions=16. 3-month phone completions=4, and 3-month postcard completions=1. SMS text messages were added 18 months into recruitment and sent at the same time as the first and second emails, so individual effects cannot be differentiated. SMS text messages did not add significantly to the time spent sending them, as they were also sent through mail merge at the same time.

<sup>d</sup>Based on 10 pence (US $0.12) per SMS text message.
Table 6. A summary of time spent and cost on different stages of contacting participants of the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial (RCT) at 6-month follow-up. A currency exchange rate of £1=US $1.26988 is applicable.

<table>
<thead>
<tr>
<th>Method</th>
<th>Total follow-up sent, n</th>
<th>Follow-up per month, mean (range)</th>
<th>Responded (n=4458)</th>
<th>Follow-Up hours(^a)</th>
<th>Voucher hours(^b)</th>
<th>Data entry hours(^c)</th>
<th>Total hours</th>
<th>Cost hours(^d) (£)</th>
<th>Other costs (£)</th>
<th>Total cost (£)</th>
<th>Cost per participant (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email through Qualtrics</td>
<td>5602</td>
<td>266 (64-621)</td>
<td>2358</td>
<td>0</td>
<td>42</td>
<td>_e</td>
<td>42</td>
<td>830</td>
<td>0</td>
<td>830</td>
<td>0.35</td>
</tr>
<tr>
<td>First manual follow-up email and SMS text message(^e)</td>
<td>1886 and 505</td>
<td>52 (4-132)</td>
<td>948</td>
<td>64</td>
<td>17</td>
<td>—</td>
<td>81</td>
<td>1601</td>
<td>51(^f)</td>
<td>1652</td>
<td>1.74</td>
</tr>
<tr>
<td>Second manual follow-up email and SMS text message(^e)</td>
<td>1077 and 450</td>
<td>51 (4-132)</td>
<td>813</td>
<td>37</td>
<td>15</td>
<td>—</td>
<td>52</td>
<td>1028</td>
<td>45</td>
<td>1073</td>
<td>1.32</td>
</tr>
<tr>
<td>Phone calls</td>
<td>2118</td>
<td>101 (8-260)</td>
<td>90</td>
<td>117</td>
<td>2</td>
<td>—</td>
<td>119</td>
<td>2353</td>
<td>0</td>
<td>2353</td>
<td>26.14</td>
</tr>
<tr>
<td>Posted survey</td>
<td>1378</td>
<td>66 (2-167)</td>
<td>145</td>
<td>68</td>
<td>3</td>
<td>24</td>
<td>95</td>
<td>1878</td>
<td>2384(^b)</td>
<td>4262</td>
<td>29.39</td>
</tr>
<tr>
<td>Postcard</td>
<td>1161</td>
<td>55 (2-156)</td>
<td>104</td>
<td>59</td>
<td>2</td>
<td>9</td>
<td>70</td>
<td>1384</td>
<td>1080(^i)</td>
<td>2464</td>
<td>23.69</td>
</tr>
</tbody>
</table>

\(^a\)Based on average times of 2.05 minutes per email or SMS text message, 3.31 minutes per phone call, 2.94 minutes per survey, and 3.07 minutes per postcard.

\(^b\)A voucher email took 1.08 seconds to send.

\(^c\)Based on 10 minutes to input a survey and 5 minutes to input a postcard.

\(^d\)The cost here is the average of 2 research staff salaries (£19.77) × research hours.

\(^e\)Not available.

\(^f\)Text messages were added 18 months (from January 15, 2022) into recruitment and sent at the same time as the first and second emails, so individual effects cannot be differentiated.

\(^g\)Based on 10 pence (US $0.12) per SMS text message.

\(^h\)Based on estimated stationary and postage costs of £1.73 per survey.

\(^i\)Based on estimated stationary and postage costs of £0.93 per postcard.

Aim 3: Broader Methodological Insights

Retention
Each SMS text message cost 10 pence (US $0.12) to send and required minimal researcher time as texts were sent to participants through mail merge at the same time as email reminders were sent. This was relatively low cost and low effort, and there was an increase in the average follow-up rate at 1- and 3-month follow-up in the 3 months before and after the introduction of the SMS text messages (from 58.0% (221/381) to 71.43% (830/1162) at 1 month and 58.5% (223/381) to 64.80% (753/1162) at 3 months).

Recruitment
Remote trials may unintentionally exclude participants with less experience using web-based surveys and digital interventions or with lower digital literacy. To mitigate this risk, in the recommendation email and at the end of the baseline survey, we included a link to a pictorial step-by-step guide to downloading and using the app [21] and encouraged participants to contact the research team if they needed technical support. Less than 10 participants contacted the research team for technical support throughout the trial.

Advertisement Development
Advertising any research study involves balancing incentivizing the target audience to participate while avoiding incentivizing those outside of the target market to falsify information to gain reimbursement. This is particularly true of remote research, where there is no face-to-face contact with researchers and therefore fewer barriers to participant deception. Below, we outline the process of developing the study advertisement, involving feedback from public and patient involvement (PPI) groups and dynamic changes throughout the trial in response to higher rates of participant deception.

PPI Feedback on Advertising
To improve the clarity and appeal of the advertisement, we attended meetings with 2 PPI groups (the Sheffield Addiction Recovery Research Panel and the Alcohol and Food Discussion Group at the University of Stirling) and asked for feedback on an advertisement we had designed (Figure 1). The PPI group highlighted language (eg, “Researchers at UCL” and “trial”) that they felt was too formal and would make the study sound frightening or labor-intensive. Furthermore, they did not like the phrase “digital support tools,” which they felt was unclear, and instead suggested we use the phrase “online support tools.” The group also suggested that to make the advertisement more appealing, we should make it clear that people would get support...
to drink less alcohol, highlight the financial incentives in a more prominent position, and include pictures.

**Figure 1.** The original study advertisement designed by the research team to recruit participants to the iDEAS randomized controlled trial (left), advertisement following public and patient involvement feedback (middle), and advertisement following issues with participant deception (right). iDEAS: iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone; UCL: University College London.

### Advertising and Participant Deception

Following issues with participant deception, edits were made to the advertisement to disincentivize those who did not meet the inclusion criteria from signing up for financial reimbursement. The mention of the vouchers was removed from the heading and moved to the body of the advertisement. The specific amount was removed, and the text was updated to make it clear that there was no immediate financial incentive to participate in the study; rather, vouchers were sent after 1-, 3-, and 6-month follow-up surveys were completed.

### Negative Engagement With Advertising

Throughout the study, we also experienced negative engagement with our social media advertising, particularly on Facebook. Unhelpful comments included those joking about wanting to drink more (eg, “I need support with drinking MORE alcohol”), leaving negative messages about the research team (eg, “killjoy weirdos”), highlighting the reimbursement amount (eg, “vouchers sound good”), or telling other users reasons they had been screened out (eg, “people who use android rather than apple ones are not wanted”). We decided against disabling comments on advertising posts, as other people used them to engage positively with the study and to tag friends. Rather than respond to or delete posts, which may have further antagonized people, we used the “hide” feature on negative comments on a weekly basis, meaning these comments could not be seen by others but that the original poster was not notified. A total of 46.6% (210/451) of comments were hidden throughout the study.

### Contamination

This was a pragmatic trial, as we were testing the effect of the recommendation rather than the use of the Drink Less app. Nevertheless, we took steps to minimize contamination. We were careful not to mention the name of the app or the trial in any advertising. We also included 2 sensitivity analyses to try and capture the extent of contamination in the trial. One focused on those who followed the recommendation determined by self-report (at 1-month and 6-month follow-ups). The second was an instrumental variable analysis that accounted for nonuse in the intervention group and contamination in the comparator group by operationalizing the difference in app use between the 2 groups.

These recommendations are summarized in Textbox 1.
Textbox 1. Methodological recommendations for remote randomized controlled trials (RCTs).

**Recruitment**
- Use a range of recruitment methods.
- Monitor the demographic composition of the sample during trials and have targeted methods for underrecruited groups.
- Targeted advertising on social media or radio can be successful in recruiting men and can yield large numbers of responses. Having advertisements run consecutively for weeks seemed to result in cumulative benefits.
- General practitioner (GP) surgeries and word of mouth were good for recruiting a more balanced sample in terms of gender, ethnicity, and socioeconomic status (SES) but overall yielded lower numbers of participants. However, these methods were likely impacted by the COVID-19 pandemic and may be more effective with an increased investment of time or money in future trials.
- Offer technical support for online surveys and intervention use, ideally in different forms such as through pictorial step-by-step guides or through phone or email to ensure recruitment and engagement are inclusive.

**Follow-up**
- Offline follow-up options, such as phone calls and postal surveys, are more resource intensive but can increase follow-up rates.
- SMS text messaging services can be a relatively low-cost and low-effort way of boosting follow-up rates.

**Advertising and incentives**
- Avoid overly formal language, which may alienate participants, and use pictures.
- Highlight benefits to participants other than financial incentives (eg, support for alcohol reduction).
- Tailor advertising strategies to ensure the right balance of incentivization across different platforms. For example, if advertising on social media or where barriers to sign up are low, mentioning incentives could result in motivated individuals falsifying information. However, where there are more barriers to sign up, for example, through a radio advertisement where participants must find the study link independently, it may be necessary to highlight incentives more explicitly.

**Participant deception**
- Be aware of different types of fraud and the best ways to detect them, and continuously monitor data as strategies are likely to evolve in response to checks and barriers introduced. These may include address checks, phone calls, or requiring participants to submit ID.
- When creating online surveys, researchers should use fraud detection software if it is offered (eg, CAPTCHAS [Completely Automated Public Turing test to tell Computers and Humans Apart]) and check licenses to see if additional fraud detection software is available.
- Include attention-check questions where participants are asked to give stable information at different points in a survey or where participants are asked to select a particular response option.
- Ensure costing is included for the data monitoring resources required.

**Contamination**
- Consider the inclusion of sensitivity analyses, such as instrumental variable analysis, to capture the extent of contamination in remote randomized controlled trials.

**Discussion**

**Summary of Findings**
In this remote RCT, the 3 main participant recruitment methods were through advertisements on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), and through radio or newspapers (745/5602, 13.3%), with all other methods of recruitment accounting for 7.37% (413/5602) of the sample. More participants who were women, White, and from more advantaged backgrounds responded to the initial recruitment. Targeted approaches through social media and radio advertising were successful in recruiting men but less successful in appealing to a more diverse demographic in terms of ethnicity and SES. The most effective methods for recruiting more balanced samples (health care providers and word of mouth) were often responsible for a relatively small proportion of the overall sample, suggesting greater investment in these methods could be a positive strategy in future trials. The costs associated with different recruitment methods varied. There was an increase in cost per participant when recruiting participants who were men, from ethnic minorities, and from more disadvantaged backgrounds across all recruitment methods.

There was evidence that the sequential approach taken to 6-month follow-up was successful, with 79.58% (4458/5602) follow-up rates at 6-months. Most participants responded following automated emails and substantial financial incentives, including an additional incentive to respond to the primary outcome within the first 24 hours, but each additional stage of follow-up resulted in an additional 2% to 3% of the sample following up. The advantage of the sequential approach is also evidenced by the greater follow-up rate (4458/5602, 79.58%) at 6-month follow-up when this process was followed, relative to the follow-up rates at 1- and 3-months (3685/5602, 65.78% and 3574/5602, 63.8%, respectively), where only email or SMS
text reminders were sent and less financial incentive was offered. However, each of the offline stages of follow-up was considerably more resource intensive than email reminders, so this is a practical consideration to be made at the costing stage. It would be of great interest to compare, across trials, the sociodemographic characteristics associated with the sample captured at each stage of follow-up. For example, it may be possible that offline stages of follow-up may be effective in retaining less digitally literate or less engaged participants.

**Implications**

When making methodological decisions about remote RCTs, there is rarely a right answer that is applicable to every study or circumstance. It is important to be aware of balancing forces, which often pull in different directions. For example, when considering advertising, it is important to balance making the study appealing to the target market with not making the study so appealing that it yields a high rate of participants who sign up with false information or who respond multiple times to gain financial reimbursement. There is a similar trade-off when considering processes aimed at reducing participant deception in the data. It is important that processes that aim to ensure participants are real and eligible do not add postrandomization bias to remote RCTs by removing “real” participants in potentially nonrandom ways. Part of navigating this balance is as well as to monitor and learn from decisions made throughout a trial.

**Previous Research**

The findings of this study are in line with other studies that have focused on methodological issues in remote studies and RCTs [11,13]. The recruitment strategy undertaken was informed by a previous smoking cessation trial, which recommended using a range of sources but also monitoring the success of strategies throughout to recruit a large, diverse sample [11]. We have reported on the success of each strategy here to inform the planning of future trials. An additional potential strategy that we did not use here to improve ethnic diversity in trial participation is geotargeting of social media advertisements in geographic areas with an ethnically diverse population [22].

The multistage follow-up strategy and stepped approach to incentives (eg, an additional £12 if completed within 24 hours at 6 months) undertaken throughout the iDEAS trial were also informed by previous research [11]. The need to have ongoing strategies to detect participant deception in web-based studies and trials is also supported in other studies, and other strategies recommended beyond those we used are to check participant IDs during onboarding and undertake IP address checks [13].

**Limitations**

This study offers valuable insights for researchers conducting web-based or remote RCTs, but it is not without limitations. The cost per participant is calculated for different sociodemographic groups to demonstrate the relative increase in costs required to recruit a balanced sample. However, this stepped increase in costs is conflated by narrowing the focus to smaller groups in the population. For example, we would expect that each participant from ethnic minority groups would cost more than each participant overall when simply dividing the cost by the number of participants, because there are proportionately fewer of them. Regardless, our estimates of comparative costs for different demographic groups across different recruitment methods may help other researchers who are planning future trials. Furthermore, this study does not consider costs related to setting up the trial, developing automation, designing materials for data collection and recruitment, and engaging with stakeholders to promote recruitment. These are additional upfront and ongoing costs that should be considered when costing RCTs. There are also 2 limitations related to the generalizability of these findings. Due to the very small numbers of some ethnic minorities, ethnicity was treated as White versus ethnic minority. Grouping all ethnic minority participants together in this way does not allow examination of different methods of recruitment for attracting different ethnic minorities. Furthermore, the Drink Less app is currently only available to those with an iOS device, and as such, iOS device ownership was an entry requirement for the trial. There are some sociodemographic differences in iPhone ownership: relative to Android devices, iPhone owners are younger, more likely to be women [23], and have higher average incomes [24].

**Conclusion**

Most participants in this remote RCT were recruited through advertisements on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), and through radio or newspapers (745/5602, 13.3%). Most recruitment methods oversampled participants who were more advantaged, women, and White. Targeted approaches through social media and radio advertising were successful in recruiting men but less successful in appealing to a more diverse demographic in terms of ethnicity and SES. There was evidence that the sequential approach taken to 6-month follow-up was successful, with 79.58% (4458/5602) follow-up rates at 6 months. This study offers recommendations for achieving balance in methodological challenges when conducting remote RCTs. Recruitment methods should be broad and targeted to achieve sociodemographic diversity. Automated emails with substantial financial incentives can achieve excellent follow-up rates of approximately 70%, but sequential offline follow-up can further boost retention by nearly 10% overall. SMS text messages can be a low-cost, low-effort way to improve follow-up rates. An important and broader takeaway is the importance of continuously monitoring, identifying, reacting to, and documenting new methodological challenges as they appear over the course of a trial. This is necessary not only to improve individual trials but also because pooling shared experiential learning can help research teams who are planning future trials.
Acknowledgments

This study is funded by the National Institute for Health and Care Research (NIHR; Public Health Research Programme, #127651). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. MO’s salary is funded by NIHR (NIHR127651) and Medical Research Council (MRC; MR/W026430/1). CG is funded by NIHR (NIHR302923).

Data Availability

The data sets generated during and analyzed during this study are available on Open Science Framework [25].

Conflicts of Interest

MH, GL, LD, MF, and SM declare no conflicts of interest. JB has received unrestricted funding related to smoking cessation research and sits on the scientific advisory board for the Smoke Free app. CG and MO have done paid consultancy work for the behavior change and lifestyle organization “One Year No Beer,” providing fact-checking for blog posts.

References

iDEAS methods paper protocol 1.4.docx. OSFHOME. URL: https://osf.io/aswmf [accessed 2023-11-30]


iDEAS trial imputed datasets. OSFHOME. URL: https://osf.io/2j9df/ [accessed 2023-12-04]

Abbreviations

ABC1: managerial, professional, and intermediate occupations
AUDIT: Alcohol Use Disorders Identification Test
C2DE: skilled, semiskilled, unskilled manual, and lowest-grade worked or unemployed
CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
GP: general practitioner
iDEAS: iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone
NHS: National Health Service
PPI: public and patient involvement
RCT: randomized controlled trial
SES: socioeconomic status
UCL: University College London

Edited by A Mavragani; submitted 14.08.23; peer-reviewed by D Leightley, J Roca; comments to author 14.10.23; revised version received 07.11.23; accepted 24.11.23; published 05.01.24.

Please cite as:

©Melissa Oldham, Larisa Dinu, Gemma Loebenberg, Matt Field, Matthew Hickman, Susan Michie, Jamie Brown, Claire Garnett. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Remote Delivery of the Cuidándome Telehealth Intervention for Self-Management of Depression and Anxiety Among Latina Immigrant Women: Randomized Controlled Trial

Carmen Alvarez¹, MSN, PhD; Subhash Aryal¹, MS, PhD; Elizabeth Vrany², PhD; Maria Jose Sanchez R³, MD, MPH; Rosalphie Quiles¹, PhD; Lia Escobar-Acosta⁴, BS; Felicia Hill-Briggs²†, PhD

¹School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
²Institute of Health System Science, Feinstein Institutes for Medical Research, Department of Medicine, Zucker School of Medicine, New York, NY, United States
³Department of Health Promotion, College of Public Health, University of Nebraska Medical Center, Omaha, NE, United States
⁴Division of Geriatric Medicine and Gerontology, Johns Hopkins University School of Medicine, Baltimore, MD, United States
⁵Johns Hopkins School of Nursing, Baltimore, MD, United States

†deceased

Corresponding Author:
Carmen Alvarez, MSN, PhD
School of Nursing
University of Pennsylvania
418 Curie Boulevard
Philadelphia, PA, 19104
United States
Phone: 1 2158980715
Email: alcarmen@nursing.upenn.edu

Abstract

Background:  Growing evidence suggests that Latina immigrant survivors of adverse childhood experiences (ACEs) are at increased risk for developing and remaining with either depression or anxiety or both symptoms. This study examined the feasibility and acceptability of a telehealth intervention—Cuidándome (quee-DAN-doh-meh, “taking care of myself”). Cuidándome is a 10-week, patient-centered, trauma-informed intervention delivered by a trained facilitator that promotes self-management of depression and anxiety symptoms through improved problem-solving skills and strategies.

Objective:  The aim of this study was to examine the feasibility and acceptability of Cuidándome delivered remotely (via Zoom) with Latina immigrant ACE survivors with either depression or anxiety or both symptoms. We also estimated the effect sizes associated with the intervention on decreasing depression and anxiety symptoms and improving social problem-solving styles.

Methods:  We evaluated Cuidándome using a randomized controlled trial design. Latina immigrants (N=47) who had experienced at least 1 ACE and had at least mild depression or anxiety symptoms were randomized to Cuidándome or a comparison group delivered by trained facilitators. We assessed for changes in depression and anxiety symptoms as well as social problem-solving styles at baseline, post intervention, and 3- and 6-month follow-up.

Results:  Analyses indicated significant decreases over time within both Cuidándome and comparison groups for depression and anxiety symptoms and maladaptive problem-solving. The intervention effect was largest for anxiety; at 6-month follow-up, Cuidándome participants had significantly lower anxiety scores than the comparison group. In addition, we observed a greater average point reduction in depression symptoms at 6 months among Cuidándome participants (5.7 points) than in the comparison group (3.7 points).

Conclusions:  A mental health program delivered via Zoom by a trained facilitator was feasible and acceptable to Latina immigrant women and can be beneficial for reducing anxiety and depression symptoms. More research is needed to assess the effectiveness of Cuidándome among a powered sample size of Latina immigrants.

Trial Registration:  ISRCTN Registry ISRCTN16668518; https://www.isrctn.com/ISRCTN16668518

(JMIR Form Res 2024;8:e52969)  doi:10.2196/52969
KEYWORDS
Latina immigrant; mental health; depression; anxiety; problem-solving; intervention study; trauma-informed; depressive; Latinx; Latino; Latina; Hispanic; Spanish; immigrant; immigrants; survivor; child; children; childhood; trauma; traumatic; adverse; telehealth; telemedicine; eHealth; digital health; feasibility; acceptability; randomized; controlled trial; controlled trials; mobile phone

Introduction

Background
Latina immigrant survivors of adverse childhood experiences (ACEs) are at increased risk of poor long-term health outcomes, as mental health disorders often go untreated in this population [1]. ACEs are a spectrum of adversities that occur in 18 years and younger of age and include physical and sexual abuse, family dysfunction such as living with an adult with mental illness, experiencing or witnessing community violence (eg, stabbing and shooting), and experiencing or witnessing violence perpetrated by an organized group (eg, gang violence and police or military brutality) [2]. These types of experiences are established risk factors for anxiety and depressive disorders in adulthood [1,3]. The United States has seen historic levels of immigrants from Central America—many fleeing from different types of adversity, trauma (eg, natural disasters, pervasive community, and political violence), and limited opportunity for socioeconomic advancement. Growing evidence suggests that foreign-born Latinos, particularly those from countries with high risks of community and political violence, experience high rates of early childhood adversity that are associated with poorer mental health outcomes [3]. When compared to the general US population and Latino immigrant men, Latina immigrants report significantly higher rates of multiple types of ACEs [1].

Despite the high burden of adversity and depression and anxiety symptoms, multiple barriers impede Latina immigrants’ access to mental health services. System-level barriers such as lack of health insurance and lack of language-concordant services are common barriers to accessing mental health services [4]. Despite the growth of the Latino population in the United States, there has been a decline in mental health services offered in Spanish [5]. In addition, while evidence-based psychological treatments are the recommended first-line treatment for mild to moderate depression [6], they are becoming less available in primary care settings [7]. Implementation of evidence-based psychological treatment services in Latino-serving health care settings also remains a challenge. These limitations make it difficult for Latino immigrants who prefer psychotherapy to pharmacotherapy to access the mental health care they need [8,9]. Studies targeting Latinos in primary care settings have often used licensed personnel as interventionists; however, the sustainability of providing such services in low-resource settings is questionable. To address these barriers, mental health experts recommend expanding access to behavioral health services by providing them outside of specialized settings, using telehealth services [10], and rigorously training and supervising unlicensed personnel (such as community health workers) to deliver high-quality services [11].

Teletherapy and the use of paraprofessionals both show promise in increasing acceptability and engagement in the treatment of Latino immigrants experiencing depression. Among Latinos, adherence to teletherapy sessions was higher (>80%) compared to in-person sessions (42%–80%) [12,13]. Despite the methodological limitations, paraprofessional-led interventions have demonstrated improvements in depression symptoms among Latinos [14,15]. Given the shortage of behavioral health professionals, delivering mental health care through paraprofessionals or community health workers is a promising strategy for increasing access for underserved populations [11,16].

Community health workers are also attuned to the needs of the communities they serve and have feasible solutions. Community health workers who serve Latino populations acknowledge (1) the need for more mental health services, (2) the training for community health workers to better meet this need, and (3) the use of teleservices to make care more accessible [17]. Community health workers have also proposed group support for addressing mental health needs. Indeed, the group format helps to reduce feelings of isolation and shame as participants hear from others who have similar life experiences with trauma and depression and anxiety symptoms [18]. Further, group support maximizes the community health workers’ reach as multiple individuals can be served and supported by each other.

Problem-solving is an established evidence-based approach for managing depression and anxiety symptoms. Social problem-solving refers to the cognitive behavioral process used to cope with life stressors [19]. According to problem-solving theory, coping with stressors involves two independent components: (1) problem orientation and (2) problem-solving style [20]. Problem orientation refers to one’s general cognition and attitudes when faced with a problem; this process is also framed by past experiences and self-appraisal about problem-solving ability. Problem-solving style refers to cognitive behavioral activities people use to cope with or manage stressful situations and include rational problem-solving (RPS; systematic and deliberate application of problem-solving skills), impulsive-careless style (ICS; impulsive approaches to problems), and avoidance style (AS; procrastination and avoiding addressing the problem) [21,22]. Effective social problem-solving involves identifying barriers to practicing recommended behaviors and brainstorming strategies to overcome barriers [23].

Among Latina immigrant women, ACE survivors had lower self-confidence in stress management compared to women who did not report ACEs [3]. In addition, experiencing more types of adversity was negatively associated with overall social problem-solving skills and positively associated with negative problem orientation (NPO) and AS [3]. Understanding and overcoming barriers through problem-solving underscores the importance of trauma-informed care, in which trauma survivors are supported in understanding how childhood adversities...
contribute to mental and physical health and reducing negative self-evaluations that impact problem-solving styles [24]. To date, the most widely used and evaluated psychological intervention among Latino immigrants is cognitive behavioral therapy, and the established benefit to Latina immigrants is based on 3 randomized controlled trials with limited generalizability [25]. Randomized controlled trials testing problem-solving therapy for decreasing depression symptoms among Latina immigrants showed clinically significant reductions in symptoms up to a year postintervention when compared to pharmacotherapy [26,27]. In summary, there remains a need to expand the portfolio of effective mental health interventions to maximize reach and enhance responsiveness to diverse needs among Latina immigrants.

**This Study**

Given the lack of mental health services for Latina immigrants and the evidence for problem-solving and trauma-informed care, we developed Cuidánde (quee-DAN-doh-meh, “taking care of myself”). Cuidánde is a 10-week, culturally appropriate, trauma-informed, group-based intervention delivered once a week by a trained facilitator that promotes self-management of depression and anxiety symptoms through improved problem-solving skills and strategies. Multiple strategies were used to provide a trauma-informed intervention, including training the research team in therapeutic communication, screening and education about ACEs and their impact on health, and creating a safe and trusting environment for participants to work through their barriers for implementing useful strategies for depression and anxiety symptom management. Details of the adaptation and development process for Cuidánde are documented elsewhere [28]. The aims of this study were to (1) examine the feasibility and acceptability of Cuidánde delivered remotely (via Zoom; Zoom Video Communications) with Latina immigrant ACE survivors with either depression or anxiety or both symptoms and (2) estimate the effect sizes associated with the Cuidánde intervention on decreasing depression and anxiety symptoms and improving social problem-solving styles. We hypothesized that compared to the comparison group, the intervention group would report lower depression and anxiety symptoms, higher positive problem orientation (PPO) and RPS, and lower NPO, AS, and ICS at postintervention and at 3- and 6-month follow-up.

**Methods**

**Recruitment**

We recruited participants over 2 weeks in July 2021. Both active and passive strategies were used to recruit participants. Actively, we developed a database of Latina immigrants with a prior study [3] and contacted these women to assess for eligibility and participate in this study if interested. We also shared the study flyer with community health workers in the area, who distributed the flyer within their networks, including placing study flyers inside bags of food that were being donated during a food drive. Our passive strategies included posting flyers at laundromats and grocery stores in Latino-concentrated neighborhoods. Women who were interested in participating in the study texted or called the research phone. A bilingual research assistant obtained informed consent and established eligibility over the phone for all women verbalizing interest in participating in the study.

Establishing eligibility included the completion of a baseline study questionnaire (including demographic information and assessments for depression and anxiety symptoms) to verify eligibility for the study. Eligibility criteria included (1) being ≥18 years, (2) foreign-born (or born on the island of Puerto Rico), (3) self-identify as a Latina, (4) self-report of ≥1 ACE, (5) ability to understand and speak Spanish, and (6) have a score of ≥5 on the Patient Health Questionnaire-8 (PHQ-8)—an assessment for depression symptoms [29] or ≥5 on the Generalized Anxiety Disorder-7 (GAD-7)—an assessment for anxiety symptoms [30]. We excluded women currently enrolled in another study about mental health (to limit potential confounding or carryover effects), and we excluded women who reported being pregnant (given that pregnancy can contribute to depression symptoms). Figure 1 displays the CONSORT (Consolidated Standards of Reporting Trials) diagram, participant enrollment, and retention (Multimedia Appendix 1).
**Ethical Considerations**

All study procedures were approved by the Johns Hopkins University Ethics Review Board (IRB00287200). Oral consent was obtained from all participants in Spanish over the phone by the bilingual research assistants (native proficiency). To secure and protect all participant information, all data were collected and directly entered into REDCap (Research Electronic Data Capture) hosted at Johns Hopkins University. Only select research team members could access these data. All data were deidentified prior to export to SPSS (version 28; IBM Corp) for data analysis. Given the cost associated with data use, we compensated our participants up to US $190 for study participation (US $15 per session attended) and completion of all follow-up study questionnaires.

**Procedures**

Women who provided consent were found eligible for the study (based on the baseline study questionnaire), and women who agreed to enroll in the study were randomized to receive either Cuidándome or educational content from a health promotion manual designed in Mexico [31]. Randomization was stratified based on ACE score so that one group would not have more people with higher average ACEs than the other. After completion of the baseline questionnaire and randomization, participants were told when their group sessions would begin, and participants were mailed the corresponding workbook for
their group assignment. Based on input from community partners and the significant use of mobile phones and apps in the study population, we did not make computer or internet access a requirement for participation. For those who were not familiar with Zoom, a brief orientation was scheduled to explain how to use Zoom. All participants were encouraged to join the group sessions when they were scheduled; if a participant could not attend the group session, a make-up session would be scheduled with the participant where the facilitator would review content from the missed week prior to the next group session. This progression was particularly important because the sessions were designed to build on each other.

Study questionnaires were completed at baseline (T0) as part of the eligibility assessment and enrollment process, within 1 month post intervention (T1), and again at 3 months (T2) and 6 months (T3) post intervention. Trained bilingual research assistants who were not involved in the intervention delivery administered the study questionnaire (see “Study Questionnaire” section for descriptions of items and measures) to participants via phone and entered responses into a secure REDCap database. Based on our experience and prior evidence, many low-income Latino immigrants rely on their smartphones for internet access, particularly if they do not subscribe to broadband services [32].

Our retention efforts included 3 weekly reminders via SMS text message for joining the Zoom sessions, mailing participants a Cuidándome bookmark and a booklet of poems.

**Intervention**

Table 1 provides the content overview for the intervention and comparison groups. To summarize, Cuidándome facilitates the learning and practice of systematic problem-solving through identification of the problem, generation of potential solutions, selection of the best solution, and implementation of the identified plan. Given participants’ history of trauma, the intervention sessions start with content about mental health and how ACEs, as well as other types of adversity, can contribute to mental health symptoms and conditions in adulthood. The remainder of the sessions guides participants through 5 evidence-based self-management strategies for managing depression and anxiety symptoms and identifying solutions for the barriers (life activities and stressors) that get in the way of practicing the recommended strategies. The weekly sessions lasted approximately 1 hour. In addition to the facilitator, all participants had the Cuidándome workbook that provided structured templates for guiding participants through the session activities. The first session included a discussion about ground rules, including the importance of confidentiality and not sharing comments made within the group with people outside of the group. During each session, the facilitator encouraged group discussion and shared reflections and strategies for overcoming challenges. The first 2 modules focus on psychoeducation and allow for discussion throughout. In the remaining modules, group learning through participant discussion is the priority; therefore, the facilitator presents the activity, guides participants through the exercises, and encourages discussion using vignettes and the workbook.
Table 1. Brief description of modules for Cuidándome and comparison program.

<table>
<thead>
<tr>
<th>Cuidándome</th>
<th>Comparison (health promotion group)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1A: ACEs(^a), depression, anxiety, and PTSD(^b)</strong></td>
<td><strong>1: Physical activity</strong></td>
</tr>
<tr>
<td>• ACEs and their associations with health</td>
<td>• What is physical activity?</td>
</tr>
<tr>
<td>• Signs and symptoms of depression, anxiety, and PTSD</td>
<td>• Benefits of physical activity</td>
</tr>
<tr>
<td>• Mental health stigma</td>
<td>• Suggestions for remaining physically active</td>
</tr>
<tr>
<td><strong>1B: Mental health and self-management strategies</strong></td>
<td><strong>2A: Healthy eating</strong></td>
</tr>
<tr>
<td>• Signs and symptoms (continued)</td>
<td>• Review and discussion of food groups</td>
</tr>
<tr>
<td>• Review and discussion of personal profile</td>
<td>• Discussion of portion sizes</td>
</tr>
<tr>
<td>• Self-management strategies for depression and anxiety</td>
<td>• Recommendations for healthy eating</td>
</tr>
<tr>
<td><strong>2: Overview of problem-solving</strong></td>
<td><strong>2B: Healthy eating</strong></td>
</tr>
<tr>
<td>• Identify behaviors for self-management of mental health</td>
<td>• Recommendations for health eating</td>
</tr>
<tr>
<td>• Identify barriers to self-management</td>
<td>• My BMI</td>
</tr>
<tr>
<td>• Understand the steps in problem-solving approach</td>
<td></td>
</tr>
<tr>
<td><strong>3: Taking control of stress and emotions (problem orientation)</strong></td>
<td><strong>3: Mental health</strong></td>
</tr>
<tr>
<td>• Understand negative versus positive problem orientation and its impact on</td>
<td>• What is mental health?</td>
</tr>
<tr>
<td>problem-solving</td>
<td>• Why is it important?</td>
</tr>
<tr>
<td>• Understand the relationship between emotions and behavior</td>
<td>• What can influence your mental health?</td>
</tr>
<tr>
<td><strong>4: What makes a problem a problem? (problem identification)</strong></td>
<td><strong>4: Substance misuse</strong></td>
</tr>
<tr>
<td>• Identify external and individual barriers to self-management</td>
<td>• Addiction prevention</td>
</tr>
<tr>
<td>• Demonstrate knowledge of the problem-solving process</td>
<td>• Assessing alcohol consumption</td>
</tr>
<tr>
<td>• Under the importance of identifying problems for appropriate goal setting</td>
<td>• Smoking cessation</td>
</tr>
<tr>
<td><strong>5: Know thyself: set goals that fit your life (generating alternative solutions)</strong></td>
<td><strong>5: Chronic diseases</strong></td>
</tr>
<tr>
<td>• Demonstrate an understanding of effective goal setting</td>
<td>• Prediabetes and diabetes</td>
</tr>
<tr>
<td>• Understand the importance of identifying problems for appropriate goal setting</td>
<td>• Hypertension</td>
</tr>
<tr>
<td>• Understand one’s own values and priorities in decision-making and problem-solving</td>
<td>• Hyperlipidemia</td>
</tr>
<tr>
<td>• Demonstrate understanding of the 4 problem-solving styles and the impact on problem-solving</td>
<td></td>
</tr>
<tr>
<td>• Identify rational problem-solving as the effective approach for solving problems</td>
<td></td>
</tr>
<tr>
<td><strong>6: Different ways to reach health goals: knowing yourself</strong></td>
<td><strong>6: Cancer screenings</strong></td>
</tr>
<tr>
<td>• Understand the importance of exploring multiple options for problem-solving</td>
<td>• Breast cancer risk factors and screening</td>
</tr>
<tr>
<td>• Acquire skills for attempting alternative solutions for solving problems</td>
<td>• Cervical cancer screening</td>
</tr>
<tr>
<td>• Demonstrate awareness of signs that a solution is not working</td>
<td></td>
</tr>
<tr>
<td><strong>7: That sounds good but does it work for me?</strong></td>
<td><strong>7: Osteoporosis</strong></td>
</tr>
<tr>
<td>• Demonstrate mastery of the rational problem-solving approach</td>
<td>• What is osteoporosis?</td>
</tr>
<tr>
<td>• Articulate the problem-solving approach for the management of mood</td>
<td>• Prevention of osteoporosis</td>
</tr>
<tr>
<td><strong>8: Take action and know the signs</strong></td>
<td><strong>8: Respiratory illnesses</strong></td>
</tr>
<tr>
<td>• Identify rational problem-solving as the effective approach for solving problems</td>
<td>• Prevention and control of communicable respiratory infections</td>
</tr>
<tr>
<td>• Demonstrate mastery of the rational problem-solving approach</td>
<td>• Prevention and control of noncommunicable respiratory illnesses</td>
</tr>
<tr>
<td>• Articulate the problem-solving approach for the management of mood</td>
<td></td>
</tr>
<tr>
<td><strong>9: Putting it all together</strong></td>
<td><strong>9: Review</strong></td>
</tr>
<tr>
<td>• Demonstrate mastery of the rational problem-solving approach</td>
<td>• What did you learn?</td>
</tr>
<tr>
<td>• Articulate the problem-solving approach for the management of mood</td>
<td>• What has helped you?</td>
</tr>
<tr>
<td>• What will you continue to do for your well-being?</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ACE: adverse childhood experience.

\(^b\)PTSD: posttraumatic stress disorder.

**Comparison**

Given our focus on Latina immigrant women with either depressive or anxiety or both symptoms, we opted to offer the comparison group some generic health education content rather than be waitlisted. The content for the comparison group came from a family health promotion manual from Instituto Mexicano del Seguro Social—the Mexican Institute for Social Security [31]. We selected this manual given our focus on a Spanish-speaking, immigrant Latina population. The content from Instituto Mexicano del Seguro Social was already in Spanish and culturally appropriate—particularly the nutrition...
content that referenced traditional foods and diets common to our participants. The comparison content included 1 session about mental health conditions in general, which did not overlap with the more detailed Cuidándome training. Delivery of the comparison content mirrored the format of the intervention group: 1-hour weekly sessions delivered via Zoom.

**Group Facilitators and Intervention Fidelity**

The intervention facilitator is a bilingual, Latina immigrant with a bachelor degree, who was trained in problem-solving therapy by our expert clinical psychologist and who received ongoing support and guidance. The facilitator of the comparison group was an experienced registered nurse with a master in health education and expertise (over 15 years of experience) in facilitating health promotion groups for Latina immigrant women. Aside from training on human participants, the nurse for the comparison group did not receive any specific training but was oriented to the purpose of the comparison group and provided with the corresponding workbook content. All Zoom sessions were audio recorded and reviewed after the sessions by the principal investigator to assess for client-centeredness (eg, showing empathy and encouraging autonomy) in both groups and to determine if the Cuidándome facilitator followed the facilitator script and demonstrated a problem-solving approach (eg, guiding participants to targets for change and focusing on positive action).

**Study Questionnaire**

**Demographic Characteristics**

The study questionnaire included questions about demographic characteristics: age, relationship status, children, nativity, length of time in the United States, educational attainment, and employment status. Items about demographic characteristics were only administered at baseline.

**Adverse Childhood Experiences**

The Adverse Childhood Experiences-International Questionnaire was used, at baseline only, to assess for occurrence (eg, “Did you live with a household member who was a problem drinker or alcoholic, or misused street or prescription drugs?” “Yes” or “No”) and frequency (eg, “Did a parent, guardian or other household member hit or cut you with an object, such as a stick (or cane), bottle, club, knife, whip etc.” “Many times;” “A few times,” “Once,” or “Never”) of different types of adversities that occurred in the age of 18 years and younger [33]. In addition to items that inquired about the traditional ACEs (eg, physical and emotional neglect), the Adverse Childhood Experiences-International Questionnaire also assesses for the types of adversity such as bullying and experiencing or witnessing community violence. Items about child marriage were not included in our assessment because we have not identified this experience as a significant part of our population’s history. We dichotomized item responses based on the presence or nonzero frequency of an experience (yes=1 and no=0) and summed all dichotomized item responses for a total score; higher scores indicated experiencing more types of adversity. This tool has been validated with Latina immigrants [3,34].

**Primary Outcome Variables**

**Depression**

We used the PHQ-8 to assess the frequency (0=not at all to 3=nearly every day) of depression symptoms during the last 2 weeks [29]. Item responses are summed for a total score (range 0-24), with higher scores indicating greater severity of symptoms. The PHQ-8 has been validated among Latina immigrants and demonstrated good reliability with our sample (α=.83).

**Anxiety**

We used the GAD-7 to assess the frequency (0=not at all to 3=nearly every day) of anxiety symptoms during the last 2 weeks [30]. Item responses are summed for a total score (range 0-21), with higher scores indicating greater severity of symptoms. The GAD-7 has also been validated among Latina immigrants and demonstrated good reliability with our sample (α=.76).

**Social Problem-Solving**

We assessed social problem-solving styles using the Social Problem-Solving Inventory-Revised (SPSI-R) [20]. The items assess attitude toward challenges as well as one’s tendencies and approach for managing stressors in everyday life. Items present different styles of thinking and reactions to scenarios to which participants report how accurately the statement reflects their attitudes or behaviors to challenges (0=not at all true of me to 4=extremely true of me). The SPSI-R assesses for (1) problem orientation and (2) problem-solving style. Problem orientation refers to one’s disposition and attitude toward a problem. People with a PPO perceive problems as solvable challenges and are optimistic and confident in their ability to manage the problem; higher scores on the PPO subscale indicate greater confidence and optimism for solving problems. People with an NPO tend to perceive problems as a threat and are less confident in their ability to address the problem; higher scores on the NPO indicate less confidence in their ability to address problems.

Three problem-solving styles are assessed with the SPSI-R: RPS, ICS, and AS. When faced with challenges, people who practice RPS tend to think through multiple solutions and intentionally implement the optimal approach; higher scores on the RPS subscale indicate higher levels of RPS. The ICS is used to describe the tendency to act on the first option that comes to mind rather than consider multiple solutions; higher scores on the ICS subscale indicate greater impulsivity when addressing problems. The AS describes the practice of procrastination or avoiding addressing a problem; higher scores on the AS subscale indicate greater tendency for practicing avoidance for addressing problems. Each subscale was added for a sum score. To obtain a total social problem-solving score, the subscales are calculated (maladaptive styles negatively impact the total score social problem-solving score) using the prescribed formula [20]. These assessments have been used among Latina immigrants and demonstrated good reliability (α=.74).
**Statistical Analysis**

Descriptive statistics (frequencies, means, and SDs) for all participant demographics and outcome variables were calculated. We tracked the number of sessions completed for each participant as an indicator of acceptability and asked participants if and how the intervention helped them at the end of the 10 weeks. The proportion of interested participants who consented and were screened as eligible to be in the study was also computed to help inform feasibility. We conducted independent t tests to assess differences in outcome variables (depression, anxiety, and social problem-solving styles) between the intervention and comparison groups at baseline and the follow-up time points (data not shown). In addition, paired t tests were used to compare the differences in means for the outcome variables from postintervention to 3- and 6-month follow-up time points (data not shown). For our primary analyses, mixed between-within participants’ ANOVA was used to examine differences in outcome variables between the intervention and comparison group, over time, from baseline to 6-month postintervention. We calculated effect sizes (Cohen d: small <0.50, medium ≥0.50 to <.80, and large ≥0.80) using the difference in outcome means for the different groups divided by the pooled SDs.

**Results**

**Participant Demographics, ACEs, and Retention**

Our sample included 47 Latina immigrants at baseline (Table 2) and 41 participants at all follow-up assessments. There were no significant differences between the intervention and comparison group at baseline.
Table 2. Participant characteristics\textsuperscript{a}.

<table>
<thead>
<tr>
<th></th>
<th>Total sample\textsuperscript{a} (N=47)</th>
<th>Intervention group (n=23)</th>
<th>Comparison group (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>35.72 (8.4)</td>
<td>36.78 (9.2)</td>
<td>34.7 (7.6)</td>
</tr>
<tr>
<td><strong>Relationship status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>8 (17)</td>
<td>3 (6)</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Married</td>
<td>16 (34)</td>
<td>8 (17)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Living together (not married)</td>
<td>20 (43)</td>
<td>10 (21)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Living apart (not married)</td>
<td>2 (4)</td>
<td>1 (2.1)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>N/A\textsuperscript{b}</td>
</tr>
<tr>
<td><strong>Children, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>4 (8)</td>
<td>3 (6)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>1-3</td>
<td>35 (75)</td>
<td>19 (41)</td>
<td>16 (34)</td>
</tr>
<tr>
<td>4 or more</td>
<td>8 (17)</td>
<td>1 (2)</td>
<td>7 (15)</td>
</tr>
<tr>
<td><strong>Nativity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>11 (23)</td>
<td>5 (10)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>El Salvador</td>
<td>9 (19)</td>
<td>5 (11)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Guatemala</td>
<td>5 (10)</td>
<td>2 (4)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Honduras</td>
<td>12 (26)</td>
<td>4 (9)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Other (Caribbean and South America)</td>
<td>10 (21)</td>
<td>7 (15)</td>
<td>3 (6)</td>
</tr>
<tr>
<td><strong>Length of time in the United States, mean (SD)</strong></td>
<td>10.48 (6.4)</td>
<td>11.3 (6.2)</td>
<td>9.7 (6.6)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school or less</td>
<td>9 (19)</td>
<td>3 (6)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Some high school education</td>
<td>9 (19)</td>
<td>7 (15)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>High school graduate or more</td>
<td>29 (62)</td>
<td>13 (28)</td>
<td>16 (34)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>26 (55)</td>
<td>16 (34)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Employed full-time</td>
<td>12 (26)</td>
<td>3 (6)</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Employed part-time</td>
<td>9 (19)</td>
<td>4 (9)</td>
<td>5 (11)</td>
</tr>
<tr>
<td><strong>Adverse childhood experiences, mean (SD)</strong></td>
<td>11.26 (4.8)</td>
<td>11.1 (5)</td>
<td>11.4 (4.6)</td>
</tr>
<tr>
<td><strong>Depression symptoms, baseline, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (&lt;5)</td>
<td>6 (13)</td>
<td>3 (6)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Mild (5-9)</td>
<td>18 (38)</td>
<td>8 (17)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Moderate depression (10-14)</td>
<td>12 (26)</td>
<td>6 (13)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Major depression, moderately severe (15-19)</td>
<td>6 (13)</td>
<td>4 (9)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Major depression, severe (20-24)</td>
<td>5 (11)</td>
<td>2 (4)</td>
<td>3 (6)</td>
</tr>
<tr>
<td><strong>Anxiety symptoms, baseline, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (&lt;5)</td>
<td>2 (4)</td>
<td>N/A</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Mild (5-10)</td>
<td>23 (49)</td>
<td>12 (26)</td>
<td>11 (23)</td>
</tr>
<tr>
<td>Moderate (10-14)</td>
<td>14 (30)</td>
<td>9 (19)</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Severe (15-21)</td>
<td>8 (17)</td>
<td>2 (4)</td>
<td>6 (13)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Baseline sample.

\textsuperscript{b}N/A: not applicable.
The most common ACEs included community violence (n=39, 83%), witnessing violence in the home (n=39, 83%), emotional abuse (n=36, 77%), physical abuse (n=36, 77%), and being bullied (n=35, 75%). In addition, 55% (n=26) of the sample reported some form of sexual abuse (unwanted sex, fondling, and attempted sex).

Figure 1 displays participant enrollment, retention, and adherence to the group sessions. Of the participants we assessed for eligibility, most (n=38, 63%) were women who contacted the research team indicating their interest to participate. The other participants were selected from the database for a previous study. Attrition was low, with 6 women discontinuing participation primarily due to work schedules. All Cuidándome participants (n=20, 100%) completed at least 9 of the 10 total sessions, and 76% (n=16) of the comparison group completed 9 of the total 10 sessions.

### Depression and Anxiety and Social Problem-Solving Overview

In Tables 3 and 4, we present the mean scores for depression, anxiety, and social problem–solving styles by study group (intervention and comparison) and time (baseline, postintervention, and 3- and 6-month follow-up). In Table 5, we compared for differences of change in scores between the intervention and comparison groups for depression, anxiety, and social problem–solving styles (interaction effect); compared the change in depression, anxiety, and social problem–solving styles over time within the groups (time main effect); and compared the 2 programs for changing depression, anxiety, and social problem–solving styles (intervention main effect).

#### Table 3. Group mean scores for depression, anxiety, and social problem–solving styles at baseline, postintervention, and 3- and 6-month follow-up.

<table>
<thead>
<tr>
<th>Time point</th>
<th>Depression(^a)</th>
<th>Anxiety(^b)</th>
<th>Social problem–solving styles(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention, mean (SD)</td>
<td>Comparison, mean (SD)</td>
<td>Effect size(^d)</td>
</tr>
<tr>
<td>Baseline</td>
<td>10.75 (5.19)</td>
<td>10.19 (6.12)</td>
<td>N/A(^c)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>4.15 (3.04)</td>
<td>5.95 (4.26)</td>
<td>0.48</td>
</tr>
<tr>
<td>3-month follow-up</td>
<td>6.10 (4.66)</td>
<td>7.80 (6.16)</td>
<td>0.31</td>
</tr>
<tr>
<td>6-month follow-up</td>
<td>5.05 (2.95)</td>
<td>6.47 (4.58)</td>
<td>0.37</td>
</tr>
</tbody>
</table>

\(^a\)Patient Health Questionnaire-8 [29].

\(^b\)Generalized Anxiety Disorder-7 [30].

\(^c\)Social Problem-Solving Inventory-Revised [20].

\(^d\)Cohen d: difference in outcome means for the different groups divided by the pooled SDs.

\(^e\)N/A: not applicable.

#### Table 4. Group mean scores for social problem–solving styles at baseline, postintervention, and 3- and 6-month follow-up.

<table>
<thead>
<tr>
<th>Time point</th>
<th>Negative problem orientation(^a)</th>
<th>Avoidance style(^b)</th>
<th>Impulsive-careless style(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention, mean (SD)</td>
<td>Comparison, mean (SD)</td>
<td>Effect size(^b)</td>
</tr>
<tr>
<td>Baseline</td>
<td>9.3 (4.29)</td>
<td>10.62 (4.90)</td>
<td>N/A(^c)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>6.1 (2.4)</td>
<td>6.42 (4.13)</td>
<td>0.10</td>
</tr>
<tr>
<td>3-month follow-up</td>
<td>5.45 (3.15)</td>
<td>7.66 (4.82)</td>
<td>0.54</td>
</tr>
<tr>
<td>6-month follow-up</td>
<td>4.85 (3.20)</td>
<td>6.42 (4.72)</td>
<td>0.38</td>
</tr>
</tbody>
</table>

\(^a\)Social Problem-Solving Inventory-Revised [20].

\(^b\)Cohen d: Difference in outcome means for the different groups divided by the pooled SDs.

\(^c\)N/A: not applicable.
Table 5. Intervention and time effects on depression, anxiety, and social problem–solving styles.

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
<th>Anxiety</th>
<th>Social problem–solving</th>
<th>Negative problem orientation</th>
<th>Avoidance style</th>
<th>Impulsive-careless style</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention by time interaction</td>
<td>0.96</td>
<td>0.57 (3, 37)</td>
<td>0.96</td>
<td>0.48 (3, 37)</td>
<td>0.99</td>
<td>0.02 (3, 37)</td>
</tr>
<tr>
<td>Time main effect</td>
<td>0.53</td>
<td>11.1 (3, 37)</td>
<td>0.51</td>
<td>11.9 (3, 37)</td>
<td>0.51</td>
<td>11.9 (3, 37)</td>
</tr>
<tr>
<td>Intervention main effect</td>
<td>N/A&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.91 (1, 39)</td>
<td>N/A</td>
<td>4.3 (1, 39)</td>
<td>N/A</td>
<td>0.84 (1, 39)</td>
</tr>
</tbody>
</table>

<sup>a</sup>P<.001.<br><sup>b</sup>P<.005.<br><sup>c</sup>P=.02.<br><sup>d</sup>N/A: not applicable.

**Depression**

Based on the PHQ-8, depression levels decreased from baseline to post intervention for both groups and remained below baseline at 3 and 6 months (Table 3). There were small effect sizes (Cohen d) at each time point (postintervention (T1): d=0.48; 3-month follow-up (T2): d=0.31; and 6-month follow-up (T3): d=0.37; Table 3). Depression symptoms significantly decreased over time (time main effect) for both intervention and comparison groups (Wilks Λ=0.53; F<sub>1,37</sub>=11.1; P<.001; Table 5). However, when comparing the 2 groups, the change in depression symptoms over time was not significant (intervention by time interaction in Table 5). There was also no significant difference between the 2 programs in reducing depression symptoms (intervention main effect in Table 5); specifically, at each time point, there was no significant difference in depression symptoms between the groups.

**Anxiety**

Based on the GAD-7, anxiety levels also decreased from baseline to post intervention and remained below baseline through 6 months for both groups (Table 3). We estimated small effect sizes (Cohen d) at each time point (T1: d=0.36 and T2: d=0.36) and medium (T3: d=0.65) effect sizes for reduced anxiety symptoms (Table 3). The reduction in anxiety symptoms over time was significant, with both groups showing a reduction in symptoms across the follow-up time points (Wilks Λ=0.51; F<sub>3,37</sub>=11.9; P<.001; Table 5). There was also a significant difference in the reduction of symptoms between the 2 groups, where Cuidánde was demonstrated to be more effective than the comparison program for reducing anxiety symptoms (F<sub>1,36</sub>=4.3; P<.001).

**Social Problem–Solving Styles**

Overall social problem-solving increased from baseline to all 3 time points. Similar to depression and anxiety symptoms, the increase in social problem-solving over time was significant with both groups showing improvement (Wilks Λ=0.51; F<sub>3,37</sub>=11.9; P<.001). When comparing the 2 groups, the increase in social problem-solving scores over time was not significant (intervention by time interaction in Table 5). There was also no significant difference between the 2 programs for increasing social problem-solving scores (intervention main effect in Table 5). The ANOVA analyses indicated that neither program had a significant effect on PPO or RPS (data not shown).

**Negative Problem Orientation**

NPO decreased from baseline to all 3 time points (Table 4). The reduction in NPO over time was significant for both groups (Wilks Λ=0.44; F<sub>3,37</sub>=16; P<.001; Table 5). Although there was change over time for both groups, this change did not differ by group (intervention by time interaction), and there was no significant difference in the effect of the 2 programs for reducing NPO (intervention main effect). However, the difference in NPO at 3-month follow-up between the 2 groups almost reached significance (Cuidánde: mean 5.45, SD 3.15 vs comparison: mean 7.66, SD 4.82; P=.09), and we estimated small (T1: d=0.10) and medium (T2: d=0.54 and T3: d=0.38) effect sizes (Table 4).

**Avoidance Style**

ASs decreased from baseline to all 3 time points (Table 4). Similar to NPO, the reduction in AS over time was significant for both groups (Wilks Λ=0.68; F<sub>3,37</sub>=5.8; P<.005; Table 5). There was no difference in change over time between the 2 groups, and there was no significant difference in the effect of the 2 programs for reducing ASs; however, the difference in mean scores postintervention (Cuidánde: mean 2.1, SD 2.17 vs comparison: mean 4.42, SD 4.81; P=.06) and at 6-month follow-up (Cuidánde: mean 1.65, SD 2.5 vs comparison: mean 4.0, SD 4.80; P=.06) approached significance, where Cuidánde participants reported lower AS at these time points (Table 4). We estimated medium effect sizes (T1: d=0.62, T3: d=0.61) for the intervention.

**Impulsive-Careless Style**

ICS decreased from baseline to all follow-up time points (Table 4). ICS decreased over time for both groups (Wilks Λ=0.75; F<sub>3,37</sub>=4.2; P=.02); however, the change in scores did not differ significantly between the groups, and neither program was more effective at reducing ICS scores (Table 5). Nonetheless, we
observed lower mean scores for Cuidándome participants compared to the comparison group that approached significance at 3-month follow-up (Cuidándome: mean 4.55, SD 4.08 vs comparison: mean 7.1, SD 4.54; \( P=0.07 \); Table 4). We calculated small (T1: \( d=0.14 \)) and medium (T2: \( d=0.59 \) and T3: \( d=0.40 \)) effect sizes for the intervention.

**Discussion**

**Principal Findings**

This study is one of the first to assess the feasibility and acceptability of a trauma-informed, problem-solving-based, self-management program delivered remotely for Latina immigrant ACE survivors with at least mild depression and anxiety symptoms. Our rapid recruitment (completed in 2 weeks), high attendance, and retention indicated that participants desired the program and found it acceptable. Based on the PHQ-8 means, participants in both groups were experiencing moderate levels of depression symptoms at baseline. Depression symptoms significantly decreased for both groups, with no significance in change between the intervention and comparison group. However, we observed lower depression scores among Cuidándome participants compared to the comparison group suggesting Cuidándome participants experienced fewer days with depression symptoms. Although both groups experienced improvements, Cuidándome participants on average reported a greater reduction in depression symptoms (5.7 points) compared to the comparison group (3.7-point reduction). For anxiety symptoms, Cuidándome was significantly more effective at reducing anxiety symptoms. On average, Cuidándome participants reported minimal to no anxiety symptoms at all follow-up time points compared to comparison group participants who on average reported mild symptoms. The effect sizes for depression and anxiety were small to medium, further supporting the beneficial impact of Cuidándome for these symptoms. Further study with a powered sample is needed to rigorously test the effectiveness of Cuidándome in this Latina immigrant population.

Contrary to our hypothesis, we did not see improvements in PPO or RPS—the components of social problem-solving that we anticipated Cuidándome would increase. Instead, we found that among Cuidándome participants, we observed lower scores for NPO and maladaptive problem-solving styles (avoidance and impulsive-careless). The goal setting and learning the problem-solving steps may have helped Cuidándome participants feel more inspired and empowered to address daily life challenges in order to pursue their goals. When discussing the benefits of Cuidándome, our participants shared that Cuidándome provided them with the steps for “how” to achieve their goals; this may have helped women have a more positive outlook on addressing challenges [28]. Women also shared that they felt a greater sense of confidence managing daily challenges and thinking through options before reacting to a situation. Regarding the lack of findings with the positive subscales, although we did not include assessments of social desirability, social desirability may have influenced participant responses and minimized the scales’ sensitivity to change. Further, we were not powered to identify statistically significant changes with any of our outcomes.

Other studies that have examined the effectiveness of problem-solving therapy for depression among Latinos have also identified improvements in depression symptoms [27,35]. However, this study is the first to show promising findings on anxiety as well as social problem-solving styles, which are the potential mechanisms of action for improving mental health outcomes.

We unexpectedly observed significant reductions in depression and anxiety symptoms in both groups. On review of the session recordings, we learned that the nurse facilitator for this group used both goal-setting and problem-solving (particularly brainstorming solutions) strategies in her sessions—particularly for the nutrition and physical activity sessions. At the end of these sessions, participants were encouraged to set a goal based on the session topic, and they discussed strategies for achieving those goals. Participants also exchanged contact information with fellow participants in the chat feature of Zoom. We did not assess social support, but it is possible that through these sessions, participants were able to expand their social network and increase social support, which is associated with lower depression and anxiety symptoms [36]. In addition, given that there were 2 sessions about physical activity and nutrition, comparison group participants may have increased their physical activity; indeed, increased physical activity is associated with a reduction in depression symptoms among Latina women [37] and other populations [38]. Finally, for all participants, we made ourselves available to connect them with community resources. Participants often called for information about where they could find health care services, work, and food. In our future work, we will assess whether change in social needs is associated with improved mental health.

Based on prior work, we surmise that the trauma-informed content and care from our research team were important contributors to retention. During the development phase of Cuidándome, the review of ACEs and their association with mental health conditions were the most time-consuming sessions because of participant engagement. Similar to findings by Kaltman et al [39] who also examined the feasibility and acceptability of an in-person, trauma-informed intervention, Cuidándome participants had positive reviews about the discussions on trauma, and they found it validating to learn that their current depression and anxiety symptoms could be related to early life adversities. Participants who were mothers felt inspired to engage with their children in a more positive way to not perpetuate the cycle of ACEs. Qualitative analysis of participant discussions during these sessions may provide more insight into participant responses to the trauma-informed content.

The success of this feasibility study may also be attributed to the intervention being offered remotely. Multiple structural (eg, documentation status) [40] and system-level barriers (eg, accessibility, health insurance, and language concordant services) make mental health services and care inaccessible for Latina immigrant women [41]. Cuidándome eliminated several of these macrolevel barriers—there was no need for participants...
to present themselves in any establishment with government-issued identification in order to obtain services, participants did not have to travel to a physical location, health insurance was not required, and the program was offered in Spanish. Using telehealth and trained personnel eliminated barriers that prevent marginalized groups from accessing a program that may be beneficial for mental health. More research with stakeholders is needed to determine how programs such as Cuidándome can be made more accessible and sustainable in community-based settings.

Limitations
We acknowledge several limitations with this study. First, we sought to establish acceptability and feasibility and did not calculate a sample size a priori. Our relatively small sample size may explain the few statistically significant findings between Cuidándome and the comparison program. In addition, our sample represented women primarily in urban and suburban settings with access to broadband services. A larger sample size that includes some geographic diversity may yield more generalizable findings.

Despite the limitations, this work contributes to the body of literature highlighting specific useful strategies (telehealth and nonlicensed personnel) that can be used to expand access to mental health services for populations socially at risk and underserved populations. Nonlicensed personnel such as community health workers have successfully delivered mental health services in low-resource settings [16]. This work aligns with other studies demonstrating the acceptability and effectiveness of training nonlicensed personnel to deliver mental health interventions [39,42] as well as the use of a web-based platform for administering these programs.

Conclusions
Our findings indicate that the Cuidándome intervention can improve depression and anxiety symptoms among Latina immigrant ACE survivors. Further, Cuidándome may also be beneficial for decreasing maladaptive behaviors (avoidance and impulsivity) associated with depression and anxiety symptoms. As the Latina immigrant population continues to grow, so should community-based mental health resources. More methodologically rigorous study of Cuidándome is needed; however, this study shows the promise of an intervention that leverages nonlicensed personnel and uses a web-based platform to increase the availability of a beneficial mental health program.

Acknowledgments
FH-B developed the program on which Cuidándome is based. This study was supported by the Robert Wood Johnson Foundation Harold Amos Faculty Development Program.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to ethical considerations but are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT e-HEALTH (V.1.6.1) checklist.

References


33. Adverse Childhood Experiences International Questionnaire (ACE-IQ). World Health Organization. URL: https://www.who.int/publications/m/item/adverse-childhood-experiences-international-questionnaire-(ace-iq) [accessed 2020-08-01]


Abbreviations

ACE: adverse childhood experience
AS: avoidance style
CONSORT: Consolidated Standards of Reporting Trials
GAD-7: Generalized Anxiety Disorder-7
ICS: impulsive-careless style
NPO: negative problem orientation
PHQ-8: Patient Health Questionnaire-8
PPO: positive problem orientation
REDCap: Research Electronic Data Capture
RPS: rational problem-solving
SPSI-R: Social Problem-Solving Inventory-Revised
Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care After Surgical Abortion (the FACTS Study Phase 3): Mixed Methods Prospective Pilot Study

Roopan Gill¹²³, BSc, MD; Gina Ogilvie²⁴, MD, PHD; Wendy V Norman²⁴, MD, CCFP, DTM&H, MHSC; Brian Fitzsimmons²³, MD; Ciana Maher², BA; Regina Renner²³, MD, MPH

¹Department of Obstetrics and Gynecology, University of Toronto, Toronto, ON, Canada
²Women's Health Research Institute, University of British Columbia, Vancouver, BC, Canada
³Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, BC, Canada
⁴Department of Family Practice, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Roopan Gill, BSc, MD
Department of Obstetrics and Gynecology
University of Toronto
123 Edward Street
Suite 1200
Toronto, ON, M5G1E2
Canada
Phone: 1 4169782216
Email: roopan.gill@gmail.com

Abstract

Background: In Canada, 1 in 3 women and people of gestational age undergo an abortion in their lifetime. Despite the liberal legal context, barriers continue to exist for women and people who can become pregnant to access this service.

Objective: This study aims to (1) conduct a pilot study to demonstrate the feasibility and acceptability of myPostCare to support follow-up care after a procedural abortion; (2) use the findings to understand whether myPostCare has the potential to improve contraceptive behavior and knowledge, emotional well-being, and sexual health knowledge; and (3) develop a better understanding of how innovative mobile solutions can support integrative health programs in British Columbia with the goal of expanding to other sites across Canada.

Methods: People of gestational age (aged 14–45 y) who underwent a procedural abortion were recruited from 2 urban abortion facilities in British Columbia. The participants completed a baseline quantitative survey and were provided access to myPostCare for up to 30 days. A follow-up quantitative survey was sent via email on day 30. Qualitative interviews were conducted to explore user satisfaction and usability of myPostCare. Responses to the survey questions were summarized using descriptive statistics, and the system usability scale (SUS) was scored according to the instructions. A secure analytics platform was implemented to obtain data on the overall use of the website by users. Qualitative analysis was conducted with NVivo using a thematic analysis approach. This study was approved by the Women’s and Children’s Research Ethics Board.

Results: Overall, 62 participants were recruited (average age 30 y); 40% (25/62) of the participants completed the exit surveys, and 24% (6/25) consented to participate in the semistructured interviews; 40 participants had undergone an immediate postabortion intrauterine device (IUD) insertion, and 22 did not have an IUD inserted. Participants were satisfied with myPostCare. The SUS average score was 81.5 (SD 9.7; median 82.5, IQR 77.5–87.5), indicating high usability of the tool. Overall, 88% (22/25) of the participants changed their contraceptive method to an IUD. Web-based analytics demonstrated that there were 61 unique visitors to the site, and the top pages visited were Postprocedure Care, Emotional Well-Being, and Contraception Explorer. The longest time spent on the website was 56 minutes. The overall email open rate was 80%, with a click rate of 36%.

Conclusions: This study demonstrates that communities and individuals are important collaborators in developing a mobile innovation that facilitates access to high-quality patient-centered abortion care. Through the cocreation process, a digital platform such as myPostCare highlighted a gap in abortion care in Canada, particularly around follow-up support after a procedural abortion.
intervention was satisfactory, acceptable, and usable for women and people who can become pregnant to support and information available to women and people who can become pregnant after an abortion.

Three trials of mHealth interventions have aimed to study the role of mobile interventions in increasing the use of contraception [15-18]. Mobile for Reproductive Health and Mobile Alliance for Maternal Action have used best practices from health communication programs to systematically develop family planning text messages [18]. Furthermore, Smith et al [13,14] explored women’s needs in Cambodia to develop a mobile phone–based intervention to support postabortion family planning, specifically contraceptive adherence. In the United States, research on the acceptability and feasibility of remote technologies for follow-up after medication abortion suggested that women prefer either a telephone call or a text message over a clinic visit [19]. Most recently, researchers from University of San Francisco’s Program in Women-Centered Contraception developed a tablet-based contraceptive decision support tool for women [20]. This study used a multiphase approach that incorporated the end user throughout the entire design of the project. The tool has been designed in collaboration with key stakeholders and designers from Bedsider [21]. Using an iterative process informed by patient and provider input throughout, this family planning innovation demonstrated that including users in development led to a more patient-centered innovation [22]. Despite the development and implementation of these mHealth innovations for family planning, research is limited in understanding the follow-up needs of women and people who can become pregnant and undergo an abortion, and how they would perceive a tool to support them and to engage them as active participants in the design process.

Given the existing evidence in support of mHealth for family planning innovations, we aimed to determine whether a mobile technology intervention would be acceptable and feasible for women and people who can become pregnant to support follow-up care after first or second trimester procedural abortion. We developed a 3-phased study based on human-centered design and the “Development-evaluation-implementation” process from the Medical Research Council Frameworks for Complex Interventions [23] rooted in 2 theories: Technology Acceptance Model and Theory of Reasoned Action [24,25] phases 1 and 2 have been published previously [26,27]. This study was a prospective pilot that aimed to determine whether the intervention was satisfactory, acceptable, and usable for women and people who can become pregnant to support them in follow-up after a procedural abortion. Ultimately, this study is

KEYWORDS
mobile health; mHealth; digital health; abortion; human-centered design; sexual and reproductive health; qualitative research; mixed methods

Introduction

More than 30% of women and people who can become pregnant in Canada undergo an abortion. To support access to this essential service, Canadian federal legislation requires that abortion services be provided by each provincial and territorial health system [1,2]. Recent literature suggests that despite the prevalence of the procedure, stigma surrounding abortion in Canada leads to barriers for women and people who can become pregnant to access postabortion support and share their abortion experiences [3]. The provision of information is an essential part of good-quality abortion services, which include follow-up care after a procedural abortion [4,5]. According to the National Abortion Federation Clinical Policy Guidelines, “appropriate and accessible post-procedure and follow-up care is essential to patients’ wellbeing” [4]. There is little evidence to suggest that “mandatory” follow-up visits will detect conditions that women and people who can become pregnant cannot detect themselves; however, there is evidence to suggest that more novel methods of offering “follow-up” visits for postabortion support are desired [6]. This includes the use of innovative mobile health (mHealth) solutions defined by the World Health Organization as “the use of mobile and wireless technologies to support the achievement of health objectives” [6,7]. More importantly, given the potential reach of mHealth, evidence has pointed to its potential to provide remote support and follow-up, particularly for women and people who can become pregnant and live in rural and remote areas [8].

In Canada, geographic barriers impact abortion access, leading women and people who can become pregnant to travel long distances for services. Upon returning to their communities, they may face challenges in accessing minimal, ineffective, or nonexistent follow-up care [8,9]. A qualitative study further explored women’s abortion experiences in the Yukon territory, a remote Canadian service area, highlighting that “fragmented services left women unsatisfied, stressed, and upset about lack of information, multiple appointments, and lengthy wait times” [10]. Women further expressed frustration with lack of follow-up counseling and recommended that it be routinely offered as they feel contact with health care providers is cut off after the procedure [10]. In addition to access issues, barriers of cost, knowledge among the general public, and health care provider competence and attitudes have also been highlighted in the literature [9]. Another study explored women’s expressed desire for postabortion support services, highlighting the stigma surrounding abortion that exists in political and social contexts, preventing women from sharing their experiences [3]. This study specifically highlighted that although women may not necessarily need mandatory physical follow-up, they desire access to postabortion support for emotional well-being [3]. Furthermore, there is a great deal of inconsistency in the type of support and information available to women and people who could become pregnant after an abortion.

The New England Journal of Medicine published a special report on Telehealth in the United States, highlighting its utility and future. In 2016, Kaiser Permanente of Northern California reported that its virtual (email, telephone, and video) communications had exceeded in-person visits [11]. Similarly, research supports the safe and effective use of telehealth for the provision of medication abortion care globally [12-14].
the first to use mHealth and human-centered design in Canada as a novel approach to support follow-up care for women and people who can become pregnant and undergo procedural abortion.

Methods

Participants

Participants were recruited from 2 publicly funded abortion clinics in British Columbia, Canada. The eligibility criteria were as follows: (1) consent to undergo a first or second trimester procedural abortion, (2) ability to read and write English, (3) ability to participate in study procedures, and (4) aged ≥14 years. Participants were excluded if they were (1) attending the clinics because of fetal anomaly or miscarriage, (2) undergoing medication abortion, (3) in a situation where it may be dangerous to use a mobile intervention, and (4) unable to provide consent to participate. To elicit whether a woman was in a dangerous situation, counselors asked the patients as part of routine care if they felt safe in their current relationships. In cases where a risk is identified, counselors provided resources and would refer to the appropriate provider or service.

Study Design

The overall study design is a mixed methods user-centered design approach with 3 phases based on the “Development-evaluation-implementation” process from the Medical Research Council Frameworks for Complex Medical Interventions [28]. This is the final phase of the 3-phase study, with the findings from phases 1 and 2 already published [26,27]. Phase 3 is a prospective pilot mixed methods study conducted in 2 urban clinics in Vancouver, British Columbia, between March and June 2018 to test the acceptability and feasibility of myPostCare when implemented as part of clinical care. This study was approved by the Children’s and Women’s Research Ethics Board (H18-00036).

Eligible participants were screened by a primary investigator. They were then introduced to the study and consented under the supervision of the research coordinator. Participants consented to be contacted for a qualitative interview at 4 weeks. A baseline questionnaire that was adapted from validated survey tools was filled out to collect demographic information, contraception history, and levels of perceived well-being and distress in the past 2 months before the abortion [29-31]. The Arizona Integrative Outcomes Score was used and is a validated 1-item visual analog tool that allows self-rated global assessment of spiritual, social, mental, emotional, and physical well-being over the past 24 hours and 1 month [32].

The participants were registered on the website at the end of each recruitment day. Participants received 7 automatic email notifications that were timed with what would be expected after the procedure and prompted them to the website over the course of 30 days. At the end of 30 days, participants received a link to their email to complete a questionnaire adapted from the validated questionnaires [29-31,33,34]. This questionnaire specifically included questions about satisfaction with myPostCare, a system usability scale (SUS) comprising 10 questions, and an evaluation of the impact of various aspects of myPostCare including emotional well-being, contraceptive behavior, immediate postprocedural care, and sexual health. Data analytics were collected using a secure data analytic platform housed at the BC Children’s Hospital. Participants were compensated for their participation.

Participants who consented to the qualitative part of the study were contacted and invited to participate in semistructured interviews to explore their engagement with the mobile tool. This included system usability, experience of receiving email notifications, emotional well-being, contraceptive decision-making, immediate postprocedural concerns, and questions about sexual health. Questions explored experience with receiving timed email messages, feedback on the content of the notifications themselves, if they found the notifications helpful and why, did they follow the recommendations of the notifications, and did they visit the website after being prompted by the notifications. Participants received additional compensation for their participation in the interview.

Data Analysis

Descriptive analysis of each variable from the quantitative surveys and secure data analytic platforms was reported as mean (SD) or median for continuous variables and count (percentage) for categorical variables. All statistical analyses were performed in R (R Foundation for Statistical Computing). Using Pwkw, a secure web analytics through the BC Children’s Hospital Research Institute, specific user engagement data were gathered from February 20 to May 2, 2018. The semistructured interview transcripts were uploaded to NVivo 11 (Lumivero) and read by 2 researchers. Inductive analysis was performed to identify emerging themes that were further refined through collaborative analysis with the first author and coinvestigator [35]. The highlighted text was coded into nodes representing similar or repeated ideas. Some text was coded to >1 node, reflecting the number of ideas presented. The nodes were categorized into specific themes, forming a thematic map that was later discussed with the research team. To enhance the validity of the findings, a triangulation approach was used. This involved cross-referencing data from the quantitative survey and the subsequent 2 phases of this study.

Ethical Considerations

This study received ethics approval by the Children’s and Women’s Research Ethics Board (H18-00036). Informed consent was obtained from all the participants included in this study. The study data were anonymized and deidentified. All data were stored in an encrypted file only accessible to the research team involved in the analysis of the study. Compensation was not provided to those who had completed the survey. A CAD $25 (US $18.38) honorarium was provided to those who completed an interview.

Results

Participant Characteristics

Participants were recruited from 2 abortion clinics in Vancouver, British Columbia. A total of 62 participants were recruited and completed the baseline survey. Of the 62 participants recruited, 25 (40%) women and people who can become pregnant
responded to the follow-up survey. We investigated whether systematic differences existed between women who responded and those who did not. Table S1 in Multimedia Appendix 1 provides a summary of the demographic information from the baseline survey and a comparison between responders and nonresponders. There were no substantial differences between these 2 groups for any of the variables listed, although there was a nonsignificant trend for the responders to have a lower Arizona Integrative Outcomes Score. These results were not statistically significant ($P<.05$). All the participants identified as ciswomen.

For the qualitative interviews, of the 25 participants who completed the exit survey, 6 (24%) consented to participate in semistructured individual interviews. These were conducted via telephone.

**Quantitative**

**Change in Contraceptive Method**

Most of the respondents (22/25, 88%) indicated that they had changed their contraceptive method to an intrauterine device (IUD) at the time of their abortion, and 21 (95%) of the 25 respondents indicated that they had changed to a Mirena, whereas 1 (4%) of the 25 respondents indicated changing from a copper to Mirena. The contraceptive method of choice was not influenced by the website; however, the website and email notifications helped reassure participants about the signs, symptoms, and effectiveness of the IUD.

**System Usability Scale**

The SUS comprised 10 questions [36]. The average SUS was 81.5 (SD 9.7), and the median was 82.5 (IQR 77.5-87.5), which revealed that 75% (19/25) of the respondents indicated an SUS score $>77$, which is a very high score.

**Satisfaction**

Most of the respondents were satisfied with the website. Figure 1 graphically displays these results as percentages.

**Qualitative**

**Overview**

Qualitative analysis of the interviews was completed using thematic analysis including both inductive and deductive themes. Nine key themes were identified and are listed in Textbox 1.
Textbox 1. Key themes.

<table>
<thead>
<tr>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ease of use</td>
</tr>
<tr>
<td>2. Usefulness of myPostCare</td>
</tr>
<tr>
<td>3. Website</td>
</tr>
<tr>
<td>4. Frequency of use</td>
</tr>
<tr>
<td>5. Time spent on the website</td>
</tr>
<tr>
<td>6. Suggestions for improvement</td>
</tr>
<tr>
<td>7. Recommend to friend</td>
</tr>
<tr>
<td>8. Privacy and security</td>
</tr>
<tr>
<td>9. Design features</td>
</tr>
<tr>
<td>10. Overall impressions of myPostCare</td>
</tr>
</tbody>
</table>

**Ease of Use**

Overall, there was unanimous agreement that myPostCare was easy to use with an organized and easy-to-navigate design. One participant highlighted the following:

> I thought it was very easy to use, which I really liked. I felt the information was well laid out with the menu sidebar on the side. The writing was easy to interpret and was not overly scientific. It was easy to navigate throughout the whole website. It didn’t feel like I was reading a research article. It was nicely spaced out and got to the point very quickly. [participant 3]

Furthermore, the language was accessible and user friendly. The participants felt that the drop-down features were very effective. We added this feature after the usability testing from phase 2, and therefore, it was consistent among our participants to hear that they appreciated this feature. The participants were highly satisfied with the ease of use of myPostCare. When asked about ease of use as it pertained to the information, 1 participant shared the following:

> The information there was superb. It was very user-friendly. Anyone could use that and get what they were looking for, no problem. [participant 5]

**Usefulness of myPostCare**

We asked about the overall usefulness of myPostCare by asking separately about the website and email notifications. Participants were satisfied with the overall frequency and timing of the email notifications:

> I like there was one email per week, it was not overwhelming. It gave you time to go back to the website in increments, not getting overwhelmed and not having it constantly on your mind, but it was a good refresher every week. This is what I needed. [participant 3]

> The timing was impeccable when you would get these e-mails and what you would be feeling. When they would come, they were right on point. I always felt like someone was at my fingertips if I needed help. [participant 5]

They found that the emails helped to navigate the recovery process from immediate signs and symptoms, emotional well-being, and contraception decision-making to general sexual health, such as a better understanding of their menstrual cycle. They also found that the emails helped them feel supported and not alone. This was an important point that resonated with all participants interviewed. A few participants stated the following:

> When I would get the email it would say, “Okay, now you might be going through this and this and this,” it gave me a moment to be like, “Right, I am. I might be going through this. I’m still having some symptoms. How am I actually feeling?” It was a reminder to check in with myself and also to think about how I may be experiencing symptoms at that time. [participant 1]

> It was nice to feel as though there was “someone” checking up on you even though it wasn’t a person. There was new content with each e-mail and helped to direct you to different stages of recovery process. I found that helpful. [participant 1]

> I felt like I was cared for. It was amazing to get, “Hey, I hope you’re doing okay. Take care.” It just felt that someone was there for me and saying if you need to call or anything, you can at any time. [participant 2]

> I think this is a great resource. It was a really beneficial thing for me to have, for sure. [participant 6]

**Website**

Overall, all participants stated that they did not have a favorite page but that each category was helpful depending on the stage at which they were in the recovery process. Each participant mentioned that the Postprocedure care page and the Emotional Well-Being Support tool were the most effective:
The emotional well-being tool was helpful. I liked how each emotion had a little blurb about it. I liked the meditation. [participant 2]

Talking about various emotions that occur was important because I found that one week I felt one way but then all of a sudden I would feel different. It was nice to go back to the website, have those feelings identified and made me feel normal. [participant 3]

Most found that the emails were well timed with the website, and providing links embedded within the emails to direct participants to the website was appreciated. Participants stated that they did not click on the emotion “Good” but did use the suggestions provided such as the meditation, journaling, and going for a walk:

I wrote an entire journal entry one day, and that was really good and definitely got some crying out while I was doing that, so I think the website prompted me to do that day, yes. [participant 3]

Just going through and trying to be at one with this, checking in, using the tools. There was good days, bad days. I have a wonderful program at work as well but I didn’t have to reach out to it because there was stuff here about meditation and making sure that I am looking after myself and doing something nice for myself. [participant 5]

In addition, a participant commented that the website had credible information, which helped to answer questions that she would have seen her physician about and, therefore, kept her out of the office or emergency room. When further asked if the website helped her understand when to seek hospital care, she said:

It [myPostCare] kept my husband and I out of emergency rooms...Here we are, two weeks and three days, and all of a sudden there’s an email about IUDs being that you could have spotting for three to six months. I am like, “Okay. We are good.” Then the bleeding stopped. It was just very empowering to have that information. [participant 5]

Finally, some participants found the website useful to support them as they did not have anyone else to talk to about their abortion, and the website helped them not feel isolated:

I went every time I got an e-mail and then there were two times that I went on it by myself when I was feeling pretty emotional and the emotional support tool helped. [participant 2]

It was good for me because I didn’t tell anybody. I didn’t have anyone to talk to. [participant 4]

It [myPostCare] is so critical, and I hope it never goes away and that it’s there for as long as women need this procedure. I hope that this site is always there. It was truly instrumental to my whole well-being through this whole procedure, so I thank you. [participant 5]

**Frequency of Use**

Participants used the website on its own but also clicked on the links within the emails. Some participants saved all the emails so they could return to them. Using the website for 1 month seemed to be sufficient for all participants. One participant stated that she had visited the website 10 to 15 times:

I went very much every time I got a new e-mail and then there were two times that I went on it by myself when I was feeling pretty emotional and the emotional support tool helped. [participant 2]

**Time Spent on Website**

The time spent on the website, as expressed in the qualitative interviews, was consistent with the results from the web-based analytics presented in the **Web-Based Secure Analytics** section. Most participants stated that they were on the website for anywhere from 2 minutes to 1 hour:

I’d probably went all together 10 to 15 times. There was one time I was on it for probably an hour, but the other times, it was probably anywhere between two to five minutes. [participant 2]

I’d say around half an hour: [participant 3]

Maybe the first couple of weeks, I kind of looked at it. I looked through it for half an hour at dinner, 20 minutes, 15, half an hour. [participant 4]

...going to guess at least an hour going back through, making sure I didn’t miss anything going to the link, so at least an hour. [participant 5]

**Suggestions for Improvement of myPostCare**

There was a strong sentiment to include blogs and stories shared by women and people who can become pregnant and have undergone an abortion. This was available on the website, but these were found under the “Good” emotion. Many did not necessarily explore this section and mentioned that if they were feeling good, they did not necessarily explore the emotional support tool and would be keener on using the Contraception Explorer or Sexual Health pages. Furthermore, suggestions to add these on the main landing page or to have rotating articles that are specific to women and people who can become pregnant telling their stories would be very useful. This was highlighted by participants as a means of further enhancing the community feeling and not feeling alone in their experience. When further explored, this also highlighted that sharing stories was also a way to help destigmatize the experience that many women and people who can become pregnant and who participated in our interviews had internalized. The following excerpts highlight this theme:

I really wanted to hear someone’s story that was positive. I would have liked to listen to just having a couple of people’s stories and how it affected them just to compare myself to them. I don’t know. [participant 2]

Putting up videos or even having articles on different stories. [participant 2]
For me personally, if I’m feeling good or when I was starting to feel good about myself again, I wouldn’t have gone on the website to check that. [participant 2]

I think more testimonials and more quotes that you can use on that website from people who have been through the experience, the better because it gives validation for what women are going through and kind of makes us feel less alone. [participant 6]

The more testimonials and the more feedback you can get from women of all ages, all experiences, all the better. [participant 6]

**Recommend to a Friend**

All participants who were interviewed would recommend myPostCare to a friend. Some also suggested that this would be specifically good for friends who did not necessarily feel comfortable going to their physician or who did not have a family physician with whom they had a trusting relationship:

Yes, I think I totally would. I think it’d be very helpful to have. I don’t think it’s going to solve a friend’s problems or anything, but for me, it was helpful to have. [participant 2]

Yes, and it’s definitely one that I want to, like, if I ever know somebody that is going through that, I’m definitely going to recommend that to them. [participant 3]

I really would. This is a great resource for the person that doesn’t think that their issues always warrant a call to their doctor. [participant 6]

**Privacy and Security**

Our participants were satisfied with the level of privacy and security afforded by the emails and website. In particular, they noted that the emails were separate from the website, and some participants suggested that it would be essential to keep it this way when myPostCare would be made live. One participant stated the following:

It’s very discrete, and I liked that. The login is required to get in on the website, so to me, it was certainty sufficient. My name is not all over the website, so even if I left it open, it is what it is, who knows what I was in. It’s not too specific so I was never worried if I had it open in public. [participant 5]

**Design Features**

All participants stated that the design was professional and the language was unbiased. Many participants commented that the design of the website and emails was calming and supportive. They also enjoyed the consistency between the website and email notifications. Words such as “clearly thought out,” “pleasant and cool pictures,” “nice blues and greys,” and “positive and well-crafted” came up frequently among our interview participants. One participant commented the following:

I really liked the resources simply because it went beyond just what we went through. Yes, I think that was one of my favorite or one of the things that when I got to, I was like, “Okay, there is crisis lines, and there are counselors.” Yes, of course, that’s what I expected to be on there. It went past that. It went to sexuality. It went to LGBTQ, or it went through different topics, so I feel like it was good education beyond what I just went through. [participant 2]

Yes, I was really happy I signed up for it and I was getting those e-mails weekly. I was able to access it, once again, read about different perspectives. I think
there were some things that I felt like it was only me
or it wasn’t normal, and then it would say something
on the website that would make me feel better, more
calm. [participant 2]

I would just grab my phone and then just go, look at
the thing and, “Okay, this is normal to feel like this.”
I don’t know if I had a favourite part, but I just found
that everything was useful. [participant 4]

It was my other rock. My husband was my one rock,
and the other one was this. It knew when things were
going to happen, and when I was panicking about
things, all of the sudden, there will be an e-mail. It
was just perfect timing, and it was amazing. It truly
was. I felt like I wasn’t alone. I went through every
link. Even the links that were outside the website, I
checked out every one of them. I read stories. It
brought a sense of calm to me, I guess. It was truly,
I never had such a good experience off of a website
like this one. It was amazing. My husband went
through everything. You would be panicking. I don’t
know how many times we went back to this website
to make sure that something that was going on wasn’t
out of the ordinary, and of course, there would be,
that it wasn’t out of the ordinary, so it was amazing.
It truly was. [participant 5]

No, I think overall it was pretty straightforward. There
wasn’t anything that I was surprised to see, and there
wasn’t anything that I can remember that didn’t kind
of fit in with what was expected through the e-mails.
It all kind of made sense. [participant 6]

I think it was just a great experience to trial the
website. I have my own personal reasons for my
procedure and how I came about doing so, but I think
it’s a great source for people that want to have that
sense of community. I think it works really well for
the specific areas that you’re trying to find more
clarity. [participant 6]

Web-Based Secure Analytics

Table 1 presents the analytics results. Specifically, of the 62
participants, the number of unique visitors on the website was
61 (98%). Although only 25 participants completed the exit
survey, all participants except 1 (98%) visited the website at
least once. The number of returning visitors was 42. The average
daily page views were 5; the total number of page views through
the study period was 432; the highest number of hits at a single
visit was 35; and the top 3 pages were Postprocedure Care,
Emotional Well-Being, and Contraception Explorer. In total,
75% (47/62) of the participants were mobile users and 25%
(16/62) were desktop visitors. The most popular contraceptive
page visited was the IUD. The details of the number of page
views throughout myPostCare are presented in Table 2.
Table 1. myPostCare web-based analytics for user engagement from February 20, 2018, to May 2, 2018 (N=62).

<table>
<thead>
<tr>
<th>Web-based analytics data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique visitors to the website, n (%)</td>
<td>61 (98.4)</td>
</tr>
<tr>
<td>Average time spent on the website by visitors</td>
<td>1 min and 28 s</td>
</tr>
<tr>
<td>Longest visit on the website</td>
<td>35 hits</td>
</tr>
<tr>
<td>Total number of page views, n</td>
<td>432</td>
</tr>
<tr>
<td>Average daily page views, n</td>
<td>5</td>
</tr>
<tr>
<td>Participants who are mobile users, n (%)</td>
<td>47 (75)</td>
</tr>
<tr>
<td>Participants who are desktop visitors, n (%)</td>
<td>16 (25)</td>
</tr>
<tr>
<td>Participants who visited the Emotional Well-Being page, n (%)</td>
<td>28 (45)</td>
</tr>
<tr>
<td>Participants who visited the Contraception Explorer page, n (%)</td>
<td>27 (43)</td>
</tr>
<tr>
<td>Participants who visit the Postprocedure Care page, n (%)</td>
<td>46 (74)</td>
</tr>
<tr>
<td>Participants who visit the Sexual Health page, n (%)</td>
<td>21 (33)</td>
</tr>
</tbody>
</table>

Top 3 pages on the website
Postprocedure Care, Emotional Well-Being, and Contraception Explorer

Most popular contraceptives visited from the contraception tool in page views, n

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormonal IUD&lt;sup&gt;a&lt;/sup&gt;</td>
<td>13</td>
</tr>
<tr>
<td>Sterilization</td>
<td>7</td>
</tr>
<tr>
<td>Copper IUD</td>
<td>3</td>
</tr>
<tr>
<td>Vaginal ring</td>
<td>3</td>
</tr>
<tr>
<td>Fertility awareness</td>
<td>2</td>
</tr>
<tr>
<td>Patch</td>
<td>2</td>
</tr>
<tr>
<td>Abstinence</td>
<td>1</td>
</tr>
<tr>
<td>Depo shot</td>
<td>1</td>
</tr>
<tr>
<td>Female condom</td>
<td>1</td>
</tr>
<tr>
<td>Male condom</td>
<td>1</td>
</tr>
<tr>
<td>Withdrawal</td>
<td>1</td>
</tr>
</tbody>
</table>

Visits to given feelings (good, okay, and not so good) from the Emotional Well-Being tool<sup>b</sup>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Okay</td>
<td>18</td>
</tr>
<tr>
<td>Good</td>
<td>2</td>
</tr>
<tr>
<td>Not so good</td>
<td>5</td>
</tr>
</tbody>
</table>

Visits to given emotion from the Emotional Well-Being tool, n<sup>b</sup>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grief</td>
<td>6</td>
</tr>
<tr>
<td>Relief</td>
<td>6</td>
</tr>
<tr>
<td>Supported</td>
<td>4</td>
</tr>
<tr>
<td>Sadness</td>
<td>2</td>
</tr>
<tr>
<td>Guilt</td>
<td>3</td>
</tr>
<tr>
<td>Regret</td>
<td>1</td>
</tr>
<tr>
<td>Shame</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup>IUD: intrauterine device.
<sup>b</sup>Returning and 1-time visitors.
Table 2. Number of page views for myPostCare.

<table>
<thead>
<tr>
<th>myPostCare pages</th>
<th>Values, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postprocedure Care</td>
<td>46</td>
</tr>
<tr>
<td>Emotional Well-Being</td>
<td>28</td>
</tr>
<tr>
<td>Contraception Explorer</td>
<td>27</td>
</tr>
<tr>
<td>Postprocedure FAQs</td>
<td>25</td>
</tr>
<tr>
<td>Emotional Support Tool</td>
<td>21</td>
</tr>
<tr>
<td>Sexual Health</td>
<td>21</td>
</tr>
<tr>
<td>Abortion Myth or Fact Quiz</td>
<td>15</td>
</tr>
<tr>
<td>Resources</td>
<td>13</td>
</tr>
<tr>
<td>Menstrual Cycle 101</td>
<td>12</td>
</tr>
<tr>
<td>Meditation 101: Meditation for Beginners</td>
<td>10</td>
</tr>
<tr>
<td>Book an Appointment</td>
<td>6</td>
</tr>
<tr>
<td>Dealing with Difficult Feelings</td>
<td>6</td>
</tr>
<tr>
<td>About Us</td>
<td>5</td>
</tr>
</tbody>
</table>

Email Notifications Analytics

Among the 62 participants enrolled, 2 (3%) unsubscribed from email notifications after the “Welcome Message” on day 0, and 2 (3%) participants’ email address was not valid. The average open rate was 80%, and the click rate was 36%. The highest open and click rates were for Welcome Message at 73.1% and 31.3%, respectively. Interestingly, the desktop device was 57.4% and mobile was 42.6%, which is different from the device from which the website was viewed. The email open rates were higher than the click rates throughout. The open rates declined for both IUD and no IUD over time; however, they remained stable at an average of 53.7% and 53.8%, respectively. Figure 2 graphically represents data of the IUD versus no IUD streams open and click rates for given days.

Figure 2. Comparison of email open and click rates for intrauterine device (IUD) versus no IUD stream in percentage.

Discussion

Principal Findings

myPostCare is the first comprehensive web-based postabortion tool in Canada and has the potential to be integrated as part of family planning services. Integration of myPostCare into clinical practice provides an opportunity to consider a new approach to supplement follow-up care for abortion care specifically but women’s health generally. This study demonstrates the design and development of a comprehensive mobile intervention to facilitate care for women and people who can become pregnant and undergo a procedural abortion to support and normalize the emotional and physical aspects after abortion. We used a human-centered design methodology, an iterative development process that was informed by input from key stakeholders such as patients, family planning experts, and administrators involved in abortion care [22,37,38]. The results from the pilot evaluation of myPostCare demonstrated that it was feasible, acceptable, and satisfactory for women and people who can become pregnant.

Specifically, this 3-phase study demonstrates the importance of including the end users and key stakeholders in the design, development, and evaluation of a mobile intervention that services a population and health care issue that continue to be stigmatized. Formative research has provided important information regarding women’s interactions with technology, their needs and desires around follow-up and access to information, and feedback on design, which is essential for the success of myPostCare. A unique finding of this study that was
supported in the literature was the importance of including a component of emotional support as part of follow-up abortion care [39]. Furthermore, we learned that the success of myPostCare was not only owing to the interactive tools and information provided by the website but that the appropriately timed automatic email notifications that women received was an important aspect of their care throughout the 30 days after the procedure. An iterative design process was important to ensure that the research team was continually evaluating that myPostCare realized the needs of the target users.

We adopted a few theoretical frameworks, all of which use a comprehensive participatory approach to developing eHealth technologies. This was similarly performed by Gilbert et al [38] in the development of Get Checked Online, a web-based sexually transmitted infection testing resource. More specifically, integrating the Technology Acceptance Model and Theory of Reasoned action with the human-centered design methodology, we used a holistic approach to developing myPostCare. According to the Technology Acceptance Model, perceived ease of use and perceived usefulness of a system are the 2 predominant indicators of system adoption [27,32]. Participants in our study were accustomed to using some form of technology, either mobile phones or computers; did not require acquisition of new skills; and were keen on developing a technology-based tool to support follow-up care after an abortion. Importantly, myPostCare will not eliminate structural barriers to comprehensive abortion care, and although it may not directly affect health behavior and decision-making, it may assist in making the delivery of abortion care more efficient, convenient, patient centered, and accessible.

myPostCare is a unique addition to the literature because of its methodology and outcome. There is evidence to support eHealth technologies to improve health care; however, currently there is limited research on mobile interventions specifically to address postabortion care, although there are various interventions for contraception use. A randomized trial in Cambodia demonstrated that involving women in the design and testing of a mobile intervention to support postabortion contraception led to more women in the intervention group reporting use of effective contraception at 4 months; specifically, the use of long-acting contraceptives was higher in the intervention group at 4 and 12 months after the procedure [40]. Previous feasibility trials focused on usability and acceptability have highlighted the importance of conducting a pilot study first, which can then assist with the design of a larger randomized trial to measure effectiveness [41]. Finally, similar to studies on the development and testing of contraception tools, the integration of evaluation in real-time clinical care is essential to ascertain the barriers and challenges to implementation in the future [22].

The limitations of this study include overall generalizability to other populations, small convenience sample sizes for all 3 phases, loss to follow-up and low response rates in this challenging population, and recruitment bias. The sample size of 6 participants in the qualitative interviews was small, ideally requiring 20 participants to achieve meaningful saturation. Given that this study is an extension of 2 previous phases, researchers felt confident in the analysis being generalizable compared with the findings of the 2 previous phases and from previous studies highlighting the type of gaps that myPostCare fills as per the participants’ reflections. As it pertains to recruitment bias, those who consented to participate were likely individuals who are more engaged with technology, have higher socioeconomic demographics, and are more likely to be early adopters of a digital health intervention to support abortion care. In previous studies, this is referred to as a Digital Divide, which suggests that although many developers of technology-based health interventions are optimistic about their impact; this needs to be balanced by the fact that the pattern of adoption is along social gradients [38]. New technologies such as myPostCare may further reinforce these social divides. Furthermore, abortion continues to be a stigmatized issue, which can be a limitation for research, as this can be a sensitive topic for most and posed difficulties with recruitment and loss to follow-up in our study. We evolved throughout each phase of the study to consider the challenges faced with patient engagement. For instance, recruitment took longer than expected for the qualitative interviews. We assumed that lack of participant engagement may be associated with stigma about abortion. In addition, we recognized that conducting research immediately after the procedure might be a sensitive time for individuals. This will need to be taken into consideration for future studies, particularly when thinking about diversifying the participants recruited and obtaining robust response rates for analysis.

Balancing these limitations are the strengths of our study, including the successful development of human-centered design elements, wide stakeholder engagement, diverse expertise on the research team, a large proportion of our sample size that was from rural locations, rigorous research methodologies, iterative design process, and development of the first web-based postabortion tool in Canada.

Further research could involve evaluating the effectiveness of myPostCare.ca and the overall patient experience through a randomized controlled trial. In addition, as suggested in other web-based literature [38], a health equity impact assessment with expert consultation and literature review may also help identify ways in which myPostCare reinforces or alleviates health inequities in sexual health services.

Conclusions

myPostCare was found to be feasible and acceptable to women and people who can become pregnant to support follow-up care after a procedural abortion. There are obvious digital divides in health care specifically, as there are limited digital tools for women’s health in Canada. Thus, there is great potential for expansion of myPostCare. More specifically, since the introduction of Mifepristone in Canada, the first area of expansion will be for medication abortion. Generally, the expansion may then involve other aspects of women’s reproductive health.

We learned that key stakeholder engagement and understanding the organizational context are important. These factors are important for ongoing research initiatives and their implementation in clinical practice. Engaging stakeholders and potential users in a participatory process throughout the entire design and development of myPostCare was crucial to its...
success. Applying an iterative design and evaluation process that was flexible and dynamic, considering the factors of implementation at the outset, keeping in mind how myPostCare could change health care delivery, and the use of a multidisciplinary team were all unique and important aspects. This study demonstrated that a technology-based intervention for postabortion care is feasible and acceptable. The success of myPostCare was based on the incorporation of a multidisciplinary team; participatory user-centered design process; robust stakeholder engagement; and the provision of nonjudgmental, nondirective, and medically accurate information. This study provides an example of the ongoing development of technology-based family planning services and is aligned with a larger gender-equitable, evidence-based programmatic agenda in Canada.

Acknowledgments
The authors thank the clinics that participated in this study and offered time.

Data Availability
Qualitative and quantitative data are available upon request; however, the data will be destroyed 10 years after its collection for this study.

Conflicts of Interest
WVN declares funding support for work contributing to this article for a Chair in Family Planning Public Health Research from the Canadian Institutes of Heath Research and The Public Health Agency of Canada (2014-2024) and for a Scholar award from the Michael Smith Foundation for Health Research (2012 - 2020).

Multimedia Appendix 1
Demographic data.

[PDF File (Adobe PDF File), 146 KB - formative_v8i1e46284_app1.pdf ]

References
3. LaRoche KJ, Foster AM. “I kind of feel like sometimes I am shoving it under the carpet”: documenting women’s experiences with post-abortion support in Ontario. FACETS 2017 Sep 01;2(2):754-763 [FREE Full text] [doi: 10.1139/facets-2017-0059]


38. Sauro J. Measuring usability with the system usability scale (SUS). MeasuringU®. URL: https://measuringu.com/sus/ [accessed 2017-08-21]


Abbreviations

- **IUD**: intrauterine device
- **mHealth**: mobile health
- **SUS**: system usability scale

©Roopan Gill, Gina Ogilvie, Wendy V Norman, Brian Fitzsimmons, Ciana Maher, Regina Renner. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
The Sukaribit Smartphone App for Better Self-Management of Type 2 Diabetes: Randomized Controlled Feasibility Study

Cecilia Josefsson1, MSc; Thea Liljeroos2, MSc; Margareta Hellgren3, PhD; Ulrika Pöder1, PhD; Mariann Hedström1, PhD; Erik M G Olsson2, PhD

1Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden
2Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
3Skaraborg Institute, Skövde, Sweden

Corresponding Author:
Cecilia Josefsson, MSc
Department of Public Health and Caring Sciences
Uppsala University
BMC, Husargatan 3
Uppsala, 751 22
Sweden
Phone: 46 18 471 66 72
Email: cecilia.josefsson@pubcare.uu.se

Abstract

Background: A new app, Sukaribit, was designed to enable contact between the caregiver and the patient with the intent to improve self-care and glycemic control (hemoglobin A1c [HbA1c]).

Objective: This study investigated the feasibility of the study methodology and the intervention in preparation for a larger effectiveness study.

Methods: Adults with type 2 diabetes were recruited in this randomized controlled feasibility study with a mixed methods design. The intervention group (n=28) tried Sukaribit for 2 months. They were encouraged to report blood glucose levels and medications, and they received feedback from a physician. The control group (n=31) received standard care. Both groups were evaluated with pre and postmeasurements of glycemic control (HbA1c), diabetes distress, physical activity, and self-care. Feasibility was evaluated against 5 progression criteria regarding recruitment, study methods, and active participation.

Results: Of the 5 progression criteria, only 2 were met or partially met. The recruitment process exceeded expectations, and data collection worked well for self-reported data but not for HbA1c measured with a home testing kit. The participants were less active than anticipated, and the effect sizes were small. Only the number of blood glucose tests per day was positively affected by the intervention, with 0.6 more tests per day in the intervention group.

Conclusions: Recruitment of participants to a future fully powered study may work with minor adjustments. The collection of HbA1c using home testing constituted a major problem, and an alternative strategy is warranted. Finally, the app was not used as intended. In order to proceed with a larger study, the app and study procedures need improvement.

(JMIR Form Res 2024;8:e46222) doi:10.2196/46222

KEYWORDS

diabetes mellitus; type 2; health behavior; mobile health; mobile application; pilot study; mobile app; mHealth; diabetes; diabetic; RCT; randomized; glycemic; self care; self management; blood sugar; T2D; diabetes type 2; home-testing; digital health

Introduction

Background

Type 2 diabetes is a serious disease affecting the prognosis of many other diseases, including cardiovascular disease. Diabetes increases the risk of acute myocardial infarction, stroke, and heart failure [1,2]. To reduce the risk of both microvascular and macrovascular complications, it is important to control blood glucose levels [3] (ie, glycemic control), blood pressure, and lipid levels [4]. For people with type 2 diabetes, a prerequisite for good glycemic control is regular and frequent self-monitoring and knowledge of how blood glucose levels respond to food and physical activity. Many patients have
elevated levels of blood glucose, which suggests that self-management is often suboptimal [5].

**Diabetes and Mobile Apps**

Even if technical solutions to support diabetes self-management, such as smartphone apps, have become more common, they are used by a minority of patients [6]. It is not clear how many people use diabetes apps in Sweden, but in Australia, only 8% of people reportedly use diabetes mobile apps [6], despite almost unlimited availability with thousands of apps on the market. Reasons for people with diabetes to not use apps can be unawareness of their existence, technical literacy barriers, no need (the disease is not that bad or self-management is sufficient anyway), no recommendation from a health care professional, the resulting increased accountability for one’s own behaviors, or the time-consuming nature of some apps [7,8]. However, studies have shown that people with type 2 diabetes want to use smartphone apps, to reduce not only the practical burden but also the cognitive and emotional burden of diabetes self-management [9]. Studies also have shown that patients want to have more contact with their nurse or physician through digital media than is the case today [9,10]. The most effective app-based technical solutions, in terms of the potential to reduce hemoglobin A1c (HbA1c), are interactive and include components such as patient-generated health data, individualized feedback, 2-way communication, and tailored education [11,12]. These components are in line with the 2 behavior change techniques of “feedback on behavior” and “self-monitoring of behavior” that are associated with better glycemic control [13].

**Sukaribit Smartphone App**

The smartphone app Sukaribit (Beta version 1.1, Maishabit AB) was developed with a special focus on the interaction between the patient and caregiver. It has an intentionally basic design to be usable with more basic mobile phones, as it needs less capacity. The app stores and displays blood glucose measurements (patient-generated health data), enables digital 2-way patient-physician or nurse communication, provides individualized feedback, and delivers tailored education. For example, if the person with diabetes enters blood glucose measurements or steps (self-monitoring of behavior), the clinician can provide individualized feedback via the 2-way communication mechanism. The physician can give advice about medications or empower health-related behaviors (feedback on behaviors). The app aims to result in more frequent measurements, better blood glucose control, and better self-efficacy, which could be reflected in more optimal HbA1c (see Figure 1). Sukaribit aims to complement standard care by enabling feedback from the caregiver when patients are not at the clinic. There are several diabetes apps on the market. However, the American Diabetes Association requests longer-term clinical evidence, and clinical outcomes have been published in peer-reviewed literature for only a few diabetes smartphone apps [14]. In line with the British Medical Research Council guidelines for developing and evaluating complex interventions [15], this is the first scientific evaluation of the feasibility of the diabetes app Sukaribit.

**Aim**

The purpose of this study was to investigate the feasibility of the study methodology and the intervention before conducting a larger effectiveness study. Our research questions were as follows: (1) Are the study procedures feasible and effective? (2) Is the Sukaribit smartphone app (version 1.1) usable and accepted by people with type 2 diabetes? (3) How large are the effect sizes for the use of the Sukaribit smartphone app on HbA1c and other potential outcomes? In line with recommendations for feasibility evaluations, we developed predetermined progression criteria to decide whether to proceed to a full-scale randomized trial [16].

**Methods**

**Research Design**

The study was a randomized feasibility study with pre and postmeasurements from an intervention group and a control group. The control group received standard care. The report follows the Consolidated Standards of Reporting Trials (CONSORT) guidelines [17].

**Ethical Considerations**

The trial protocol was approved by the Swedish ethical review authority (diary number 2020-04894), and the participants provided written informed consent.

**Progression Criteria**

The aim was to study (1) the feasibility of study procedures and (2) the usability and acceptability of the intervention. This follows general recommendations for pilot studies by Avery et al [16]. In addition, we also studied (3) preliminary effect sizes (see the Preliminary Effect Sizes section). Aims (1) and (2) were evaluated against predetermined progression criteria (see Textbox 1) [16]. These progression criteria were set prospectively by the authors considering the possibility of finalizing recruitment of participants for a fully powered randomized controlled trial (RCT) within approximately 2 years and having an activity level in the intervention high enough to draw conclusions about its use. If the progression criteria were met, this indicated that a larger study is feasible using the procedure evaluated; otherwise, revisions should be considered.
Feasibility data were collected in a log by the research assistant, and the automated activity log from the Sukaribit app was shared with the researchers by Maishabit AB. To further explore if the app is usable and accepted by people with type 2 diabetes, an additional qualitative evaluation was conducted. Participants were asked open-ended questions in the portal about opinions and possible improvement of the app. The intervention group also participated in semistructured telephone interviews for further input about the acceptability of the intervention. The interview guide contained questions about the participant's diabetes, self-care, and study participation, as well as about the mobile app. The interviews were audio-recorded (average length: 25 minutes) and transcribed. The physician was also interviewed about participation with a separate but similar interview guide.

Participants and Procedures

The study included 59 adults (age >18 years) with type 2 diabetes. Exclusion criteria were other serious illnesses, HbA1c >70 mmol/mol, BMI <25 kg/m², no regular access to the internet, and not owning a blood glucose monitor. The following 2 initial exclusion criteria were abandoned as they were not that important and not feasible for effective recruitment: people with HbA1c <50 mmol/mol (4 were initially excluded) and an age >65 years.

Participants were recruited (between February 2021 and April 2021) at health care centers in Uppsala, through nationwide adverts in 3 major Swedish newspapers, and via advertising on the national Swedish Diabetes Federation’s web page and in diabetes-specific social media groups. People with type 2 diabetes reported their interest on a study-specific website hosted by Uppsala University or directly to the research assistant via email or telephone. Thereafter, they were contacted by the research assistant who informed them about the study. People who were still interested provided written consent to participate. Thereafter, the research assistant checked the inclusion and exclusion criteria preliminarily and ensured that the participant had a pedometer app on their smartphone or helped them install one.

Participants were sent a home testing kit for HbA1c, which meant that they took a blood sample at home and sent it to an accredited laboratory for analysis. As recruitment proceeded, the authors recognized that the wait time for baseline HbA1c test results could be long (mean 15.6, SD 6.4 days). Therefore, we decided to include and randomize participants before the HbA1c test results arrived and exclude them afterwards if necessary; 4 participants were excluded on this premise.

All questionnaires were administered using the Uppsala University Psychosocial Care Program (U-CARE) Portal (the portal). The participants answered the questionnaires at the time of randomization (1) and 2 months later (2) or delayed response of a maximum of 18 days was allowed. Randomization occurred in the portal (see the following paragraph), was totally automated, and occurred in blocks of 6 immediately after the completion of baseline questionnaires.

Those randomized to the intervention were supported in downloading the Sukaribit smartphone app and had a brief user education via telephone. They also received instructions in a PDF brochure. The intervention group was asked to share their blood glucose measurements in the app. Additional follow-up support was requested by 5 participants, as they were uncertain of particular features of the app (eg, input of medications). Those randomized to the control group received standard care [18]. All participants were contacted 2 months later for follow-up data. Those who participated in the intervention were also asked to participate in a semistructured interview about their experience with the intervention. Of those invited to the interview, 16 participants accepted (3 people declined) and were interviewed via telephone by 1 of the 2 research assistants.

Preliminary Effect Sizes

In addition to the feasibility of the app, the preliminary effect sizes of the Sukaribit smartphone app were also explored. They could be used to calculate the sample size for a fully powered study. Effect sizes were studied for (1) HbA1c, (2) number of blood glucose measurements reported the previous week, (3) physical activity, (4) general self-rated health (visual analogue scale from the EQ-5D) [19], (5) diabetes self-management, and (6) diabetes-related distress. This study only explored the changes in these measures, as the study was not sufficiently powered to detect efficacy.

HbA1c was analyzed from a home testing blood test at an accredited laboratory. The blood glucose measurements were recorded by the participants in their own diary of choice and reported in the portal as an outcome. The intervention group could use the app to record their measurements. Physical activity was measured as steps via pedometers on the participants’ smartphones, and the last 7 days were reported in the portal.

Textbox 1. Research questions 1 and 2 and their respective progression criteria.

<table>
<thead>
<tr>
<th>Research questions</th>
<th>Progression criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Are the study procedures feasible and effective?</td>
<td>At least 60 people reported interest in participating in the study within 3 months of recruitment.</td>
</tr>
<tr>
<td>2. Is the Sukaribit smartphone app usable and accepted by patients with type 2 diabetes?</td>
<td>At least 80% of those initially interested and eligible actually started participating.</td>
</tr>
<tr>
<td>3. At least 60 people reported interest in participating in the study within 3 months of recruitment.</td>
<td></td>
</tr>
<tr>
<td>4. At least 80% of those who initially reported interest and eligible actually started participating.</td>
<td></td>
</tr>
<tr>
<td>5. At least 75% of those randomized (to any of the groups) in the study completed the postmeasurements (ie, had complete data).</td>
<td></td>
</tr>
<tr>
<td>6. At least 80% of those who reported interest were eligible for inclusion in the study (ie, met the inclusion but not the exclusion criteria).</td>
<td></td>
</tr>
<tr>
<td>7. At least 75% of those randomized (to any of the groups) in the study completed the postmeasurements (ie, had complete data).</td>
<td></td>
</tr>
</tbody>
</table>

(1) Are the study procedures feasible and effective?

(2) Is the Sukaribit smartphone app usable and accepted by patients with type 2 diabetes?
Participants also reported the number of occasions per week over the last month they had exercised more than 30 minutes for fitness purposes. Diabetes self-management was measured using the Diabetes Self-Management Questionnaire (DSMQ) \[20,21\]. The DSMQ has 16 items divided into 4 subscales, namely (1) glucose management, (2) dietary control, (3) physical activity, and (4) health care use, with a maximum score of 64. A higher score indicates higher frequency of diabetes self-care behaviors. DSMQ has shown good psychometric properties in several contexts \[21\]. The Diabetes Distress Scale (DDS) was used to measure diabetes distress \[22\]. The DDS has 17 items divided into 4 subscales, namely (1) emotional burden, (2) physician-related distress, (3) regimen-related distress, and (4) diabetes-related interpersonal distress, with a maximum score of 102. A higher score indicates more distress. DDS has shown good psychometric properties in several contexts \[22\].

The Intervention

The intervention group used the smartphone app Sukaribit (version 1.1) for 2 months. In this app, participants entered their medication list, blood glucose levels, and (optionally) blood pressure levels. Participants could choose to send the recorded measurements to the study physician or not. They were encouraged to send blood glucose measurements at least once a week. The physician was a specialist in family medicine and an associate professor in general practice. She actively participated in the design of the study and evaluated and proposed changes to the app. When measurements were sent, the physician responded with feedback to the participant. All communication occurred through the Sukaribit app. The physician encouraged participants who did not send measurements on their own initiative to register and provide the requested information. This was done at least once for each participant at the start of intervention and regularly approximately once a week if no measurement was sent by the participant during that time. The physician checked messages and measures once a week and replied. There were 2 versions of the app: one for Android and one for iOS.

Data Analysis

The collected data on recruitment and intervention use were compared with the prespecified progression criteria to decide if they matched. Qualitative data were analyzed with quantitative content analysis \[23\]. Data from both the interviews and open-ended questions were analyzed together. Within and between-group effect sizes (Cohen $d$) were calculated for HbA$_1c$ and self-reported outcomes, dividing the mean differences with pooled SDs, with the aim of being the basis for statistical power and sample size calculations for a future study. The between-group effect sizes used the pooled baseline SDs as recommended by Morris \[24\]. A value of $d>0.8$ is classified as a large effect size, $d=0.5$ is classified as a medium effect size, and $d=0.2$ is classified as a small effect size according to Cohen \[25\]. Preliminary inference statistics were also performed utilizing linear regression analysis with the posttreatment value as the outcome and group allocation, baseline values, sex, and age included as covariates. The adjusted estimate can be interpreted as the adjusted mean difference for the treatment group when compared with the control group (the reference). $P<.05$ was considered significant.

Results

Participant Characteristics

Among the randomized participants (n=59), the majority were male (42/59, 71%), born in Sweden (54/49, 92%), and retired (32/59, 54%). The mean age was 61.1 (SD 10.3) years. Most participants (35/59, 59%) reported being lightly active at baseline (eg, practicing yoga, walking, and gardening), with main health issues including hypertension (39/59, 66%) and dyslipidemia (7/59, 46%). Diabetes complications, including eye disease, neuropathy, kidney disease, or sexual dysfunction, were reported by 29% (17/59). For a complete description of the participant characteristics, see Table 1.
Table 1. Participant characteristics at baseline (n=59).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total sample</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants randomized</td>
<td><em>a</em></td>
<td>28 (48)</td>
<td>31 (53)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>17 (29)</td>
<td>11 (39)</td>
<td>6 (19)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>61.8 (9)</td>
<td>60.2 (12)</td>
<td>61.8 (9)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>15 (25)</td>
<td>7 (25)</td>
<td>8 (26)</td>
</tr>
<tr>
<td>Cohabiting/married</td>
<td>41 (70)</td>
<td>20 (71)</td>
<td>21 (68)</td>
</tr>
<tr>
<td>Living alone but have a steady partner</td>
<td>3 (5)</td>
<td>1 (4)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Country of birth, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>54 (92)</td>
<td>25 (89)</td>
<td>29 (94)</td>
</tr>
<tr>
<td>Outside Sweden</td>
<td>5 (9)</td>
<td>3 (11)</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>8 (14)</td>
<td>1 (4)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Secondary</td>
<td>17 (29)</td>
<td>10 (36)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>University (≤3 years)</td>
<td>17 (29)</td>
<td>7 (25)</td>
<td>10 (32)</td>
</tr>
<tr>
<td>University (&gt;3 years)</td>
<td>17 (29)</td>
<td>10 (36)</td>
<td>7 (23)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unemployed</td>
<td>2 (3)</td>
<td>0</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Retired</td>
<td>32 (54)</td>
<td>14 (50)</td>
<td>18 (58)</td>
</tr>
<tr>
<td>Employed (any status)</td>
<td>25 (42)</td>
<td>14 (50)</td>
<td>11 (36)</td>
</tr>
<tr>
<td>Employed full time</td>
<td>23 (39)</td>
<td>12 (43)</td>
<td>11 (36)</td>
</tr>
<tr>
<td>Employed part time</td>
<td>2 (7)</td>
<td>2 (7)</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Exercise intensity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mostly sedentary</td>
<td>13 (22)</td>
<td>7 (25)</td>
<td>6 (19)</td>
</tr>
<tr>
<td>Lightly active</td>
<td>35 (59)</td>
<td>17 (60)</td>
<td>18 (58)</td>
</tr>
<tr>
<td>Moderately active</td>
<td>8 (14)</td>
<td>3 (10)</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Very active</td>
<td>3 (5)</td>
<td>1 (4)</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Days per week with ≥30 minutes of physical activity, mean (SD)</td>
<td>2.5 (2)</td>
<td>2.3 (2)</td>
<td>2.7 (2)</td>
</tr>
<tr>
<td>Steps per day in the last week, mean (SD)</td>
<td>4966 (3862)</td>
<td>4798 (3164)</td>
<td>5094.5 (4371)</td>
</tr>
<tr>
<td>Current smoker (Yes), n (%)</td>
<td>6 (10)</td>
<td>2 (7)</td>
<td>4 (13)</td>
</tr>
<tr>
<td><strong>Medical history, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>39 (66)</td>
<td>16 (57)</td>
<td>23 (74)</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>27 (46)</td>
<td>11 (39)</td>
<td>16 (52)</td>
</tr>
<tr>
<td>Stroke</td>
<td>1 (2)</td>
<td>0</td>
<td>1 (3)</td>
</tr>
<tr>
<td>History of mental illness</td>
<td>8 (14)</td>
<td>3 (11)</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>4 (7)</td>
<td>2 (7)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Other cardiovascular disease</td>
<td>8 (14)</td>
<td>6 (21)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Total sample</td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>--------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>Diabetes complications (Yes(^d), n (%))(^c)</td>
<td>17 (29)</td>
<td>4 (14)</td>
<td>13 (42)</td>
</tr>
<tr>
<td>Blood glucose tests per week, mean (SD)(^e)</td>
<td>7.5 (12)</td>
<td>8.1 (13)</td>
<td>6.9 (11)</td>
</tr>
</tbody>
</table>

\(^a\)Not applicable.
\(^b\)Missing data for 6 (10%) participants.
\(^c\)Missing data for 11 (19%) participants.
\(^d\)For example, eye disease, neuropathy, kidney disease, or sexual dysfunction.
\(^e\)Missing data for 1 (2%) participant.

### Feasibility of Study Procedures

Table 2 summarizes the progression criteria fulfilment. There were 182 people that reported interest in participating in the study; of this group, the majority (176/182, 96.7%) registered their interest on a web page. That met progression criterion 1 (n ≥ 60) by a good margin. Of the 182 people interested, 133 were reached and assessed for eligibility. However, a considerable proportion of the participants who registered their interest were ineligible or unable to be included in the study; hence, progression criterion 2 (50% inclusion rate) was not met. The main reason for exclusion at this stage was a BMI < 25 kg/m\(^2\) (n=26). In total, 55 people were excluded. Of those eligible, 19 people never logged into the portal even after being reminded. Finally, 59 (76%) of the 78 eligible participants were randomized in the study (treatment: n=28; control: n=31). For the complete recruitment flow, see Figure 2.

<table>
<thead>
<tr>
<th>Progression criteria</th>
<th>Value</th>
<th>Goal reached</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) At least 60 people reported interest in participating in the study within 3 months of recruitment.</td>
<td>• 182 people (in 2 months and 12 days)</td>
<td>Yes</td>
</tr>
<tr>
<td>(2) At least 50% of those who reported interest were eligible for inclusion in the study.</td>
<td>• 43% (78/182)</td>
<td>No</td>
</tr>
</tbody>
</table>
| (3) At least 75% of those randomized in the study completed the postmeasurements (ie, had complete and valid data). | • 64% (38/59) with complete questionnaire data and HbA\(_1c\) test results  
• 81% (48/59) with complete questionnaire data  
• 70% (41/59) with complete HbA\(_1c\) test results | Partially     |
| (4) At least 80% of those initially interested and eligible actually started participating. | • 76% (59/78)                                                      | No           |
| (5) At least 50% of those who participated in the intervention sent at least 8 blood glucose measurements during the 2-month intervention (about 1 per week). | • 11% (3/28; based on the “Number of sent diagnostic data”) | No           |
Figure 2. Recruitment flow. DDS: Diabetes Distress Scale; DSMQ: Diabetes Self-Management Questionnaire; HbA1c: hemoglobin A1c.

Progression criterion 3, at least 75% complete data at follow-up, was met regarding questionnaire data (81%). However, when also considering HbA1c tests, the completeness was 64%; thus, the criterion was not met. At baseline, 5 HbA1c test results were missing. At follow-up, 11 randomized participants did not complete their questionnaires, and there were 18 missing tests (Table 3). For baseline HbA1c, 41 manual reminders were sent in total; for the follow-up HbA1c, 63 manual reminders were sent. See Table 3 for details.
Table 3. Feasibility data (n=59).

<table>
<thead>
<tr>
<th>Data collected</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Self-reported instruments</strong></td>
<td></td>
</tr>
<tr>
<td>Time between inclusion and completion of the baseline instruments (days), mean (SD)</td>
<td>5.7 (6.0)</td>
</tr>
<tr>
<td>HbA1c&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Time between being sent the test kit and the test results (days), mean (SD)</td>
<td>15.6 (6.4)</td>
</tr>
<tr>
<td>Manual reminders for HbA1c, n</td>
<td>41</td>
</tr>
<tr>
<td><strong>Follow-up</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Self-reported instruments</strong></td>
<td></td>
</tr>
<tr>
<td>Participants with complete data, n</td>
<td>48</td>
</tr>
<tr>
<td>Time between the prompt and completion of the follow-up instruments (days), mean (SD)</td>
<td>4.4 (4.1)</td>
</tr>
<tr>
<td>Manual reminders, n</td>
<td>37</td>
</tr>
<tr>
<td>HbA1c</td>
<td></td>
</tr>
<tr>
<td>HbA1c test results&lt;sup&gt;b&lt;/sup&gt;, n</td>
<td>42</td>
</tr>
<tr>
<td>Time between being sent the test kit and the test results (days), mean (SD)</td>
<td>15.4 (8.7)</td>
</tr>
<tr>
<td>Manual reminders for HbA1c, n</td>
<td>63</td>
</tr>
<tr>
<td><strong>Sukaribit user data</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Participant activity</strong></td>
<td></td>
</tr>
<tr>
<td>Active participants, n</td>
<td>27</td>
</tr>
<tr>
<td>Number of messages sent per participant, mean (range)</td>
<td>1.0 (1-5)</td>
</tr>
<tr>
<td>Number of messages received from physicians per participant, mean (range)</td>
<td>3.0 (0-6)</td>
</tr>
<tr>
<td>Technical issues reported by participants to the developer, n</td>
<td>4</td>
</tr>
<tr>
<td><strong>Physician activity</strong></td>
<td></td>
</tr>
<tr>
<td>Time spent on all participant responses per week (hours)</td>
<td>2</td>
</tr>
<tr>
<td>Time spent on participant responses per week per participant (minutes)</td>
<td>5</td>
</tr>
<tr>
<td>Technical issues reported by the physician to the developer, n</td>
<td>5</td>
</tr>
</tbody>
</table>

<sup>a</sup>HbA1c: hemoglobin A1c.

<sup>b</sup>Missing tests + defective tests: n=18.

### Feasibility of the Intervention

Of the 28 participants in the treatment group who completed the study, 27 were active users of the app (ie, they completed 2299 data entries in total [blood glucose value, blood pressure value, and medications] in the app and sent 211 of the entries to the physician at some point). In addition, they sent 28 text messages to the physician (see Table 3).

For the 4 participants who requested technical support while using the app, the reasons for contact included difficulties logging in, issues with iOS graphic data, messages not being sent, or that the app had stopped working altogether.

Considering progression criterion 4, 76% of the eligible people actually started participating in the study. This was slightly lower than the criterion of 80%. Regarding criterion 5, only 11% of the participants sent diagnostic data 8 times in 2 months; thus, this criterion was not met.

### Client Satisfaction and the Physician’s Evaluation

A summary of the interviews is presented in Table 4. The findings show that smartphones with the iOS operating system were the most commonly used among the responding participants (15/20, 75%). Concerning the overall quality of the app, a majority of the participants reported the app was of fair quality, with only a few of their individual needs having been met. The 4 technical issues reported to the developer mainly concerned the iOS version of the app. The physician had technical problems but thought the contact was rewarding when it worked. She also experienced varying activity from the participants (Table 4).
<table>
<thead>
<tr>
<th>Questions and categories</th>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expectations for the app and study</strong></td>
<td></td>
</tr>
<tr>
<td>Want to have contact with a physician or health care professional (feedback)</td>
<td>8</td>
</tr>
<tr>
<td>Interest in diabetes apps</td>
<td>6</td>
</tr>
<tr>
<td>Contribute to research</td>
<td>4</td>
</tr>
<tr>
<td>Thinking that a diabetes app is part of the future for diabetes care</td>
<td>3</td>
</tr>
<tr>
<td>Help with more motivation to perform self-care</td>
<td>2</td>
</tr>
<tr>
<td>Want more knowledge</td>
<td>3</td>
</tr>
<tr>
<td>Ability to collect everything in the same place (though it is not working)/facilitate everyday life</td>
<td>2</td>
</tr>
<tr>
<td><strong>Thoughts about the app</strong></td>
<td></td>
</tr>
<tr>
<td>Technical problems</td>
<td>12</td>
</tr>
<tr>
<td>Difficult to add their medicine in the list</td>
<td>8</td>
</tr>
<tr>
<td>Technical problems when sending messages/values to the physician</td>
<td>3</td>
</tr>
<tr>
<td>Thought the app was difficult</td>
<td>6</td>
</tr>
<tr>
<td>Did not like the appearance of the app</td>
<td>2</td>
</tr>
<tr>
<td>Easy to navigate</td>
<td>8</td>
</tr>
<tr>
<td>Simple but functioning</td>
<td>2</td>
</tr>
<tr>
<td>Easier to manage more frequent blood glucose monitoring</td>
<td>3</td>
</tr>
<tr>
<td>Easy access to and communication with health care staff</td>
<td>3</td>
</tr>
<tr>
<td>Possible to get feedback on test results from physician</td>
<td>3</td>
</tr>
<tr>
<td>Increased motivation for self-care/increased awareness</td>
<td>3</td>
</tr>
<tr>
<td>Interesting to see how blood glucose is affected by food</td>
<td>1</td>
</tr>
<tr>
<td>Possibility to log data/follow data over time</td>
<td>3</td>
</tr>
<tr>
<td>Good support from the developer</td>
<td>2</td>
</tr>
<tr>
<td><strong>Contact with the physician</strong></td>
<td></td>
</tr>
<tr>
<td>Good and relevant replies</td>
<td>9</td>
</tr>
<tr>
<td>Good contact and fast communication</td>
<td>3</td>
</tr>
<tr>
<td>Some sort of miscommunication due to technical issues and maybe a lack of personal knowledge</td>
<td>1</td>
</tr>
<tr>
<td>No/very little communication with the physician</td>
<td>4</td>
</tr>
<tr>
<td><strong>Desired improvements</strong></td>
<td></td>
</tr>
<tr>
<td>Wish for an easier app</td>
<td>4</td>
</tr>
<tr>
<td>Improved design</td>
<td>1</td>
</tr>
<tr>
<td>Faster and more communication with caregiver</td>
<td>2</td>
</tr>
<tr>
<td>Direct communication between the app and blood glucose meter</td>
<td>5</td>
</tr>
<tr>
<td>Linked to other health applications</td>
<td>3</td>
</tr>
<tr>
<td>See old values and a graph function (to be able to learn)</td>
<td>4</td>
</tr>
<tr>
<td>Notifications when receiving message overview/table/graph</td>
<td>2</td>
</tr>
<tr>
<td>Wanted bigger text or a computer version</td>
<td>2</td>
</tr>
<tr>
<td>Information/news about diabetes in the app</td>
<td>3</td>
</tr>
<tr>
<td>Be able to send photos</td>
<td>1</td>
</tr>
<tr>
<td>Be able to log physical activities</td>
<td>1</td>
</tr>
<tr>
<td>Be able to set goals</td>
<td>2</td>
</tr>
</tbody>
</table>
Effect Sizes of Outcome Measures

The effects of treatment on a number of potential outcomes were analyzed based on complete data. No imputations were used. Both the within and between-group Cohen $d$ values suggested, at best, small effects. The largest between-group effect size ($d=0.36$) was achieved for the EQ-5D-VAS, and the effect was mainly dependent on the decrease in the control group. In the linear regression analysis, only the number of blood glucose tests per day was significant, indicating 0.57 more tests per day in the intervention group than in the control group (adjusted beta=0.57, 95% CI 0.09-1.06). This effect resulted from a reduction of tests per day in the control group, while the treatment group remained at a stable level. See Table 5.
Table 5. Complete case analyses of outcome measures.

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>Treatment group</th>
<th>Control group</th>
<th>Between group (post), d&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Linear regression analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline, mean (SD)</td>
<td>Post, mean (SD)</td>
<td>n</td>
<td>Baseline, mean (SD)</td>
</tr>
<tr>
<td>HbA&lt;sub&gt;1c&lt;/sub&gt; (mmol/mol)</td>
<td>50.1 (6.91)</td>
<td>49.3 (7.15)</td>
<td>20</td>
<td>−0.11</td>
</tr>
<tr>
<td>Blood glucose tests per day</td>
<td>1.27 (1.62)</td>
<td>1.23 (1.14)</td>
<td>20</td>
<td>−0.03</td>
</tr>
<tr>
<td>Physical activity per week&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.25 (1.8)</td>
<td>3.15 (1.8)</td>
<td>20</td>
<td>−0.06</td>
</tr>
<tr>
<td>Steps per day</td>
<td>4761 (3099)</td>
<td>5407 (3117)</td>
<td>18</td>
<td>0.21</td>
</tr>
<tr>
<td>EQ-5D-VAS (1-100)</td>
<td>55.1 (22.8)</td>
<td>57.8 (25.5)</td>
<td>18</td>
<td>0.11</td>
</tr>
<tr>
<td>Total DDS&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2.45 (0.83)</td>
<td>2.35 (0.76)</td>
<td>20</td>
<td>−0.12</td>
</tr>
<tr>
<td>Total DSMQ&lt;sup&gt;e&lt;/sup&gt;</td>
<td>6.62 (1.34)</td>
<td>6.59 (1.53)</td>
<td>19</td>
<td>−0.02</td>
</tr>
</tbody>
</table>

<sup>a</sup>The posttreatment between-group effect size was adjusted for baseline values.

<sup>b</sup>The difference between groups after treatment was adjusted for age, sex, and baseline values of the respective measure. The control group is the reference.

<sup>c</sup>Number of times, in the last month, the participant performed exercise for more than 30 minutes.

<sup>d</sup>DDS: Diabetes Distress Scale.

<sup>e</sup>DSMQ: Diabetes Self-Management Questionnaire.

**Discussion**

In this feasibility study, we explored the prerequisites for conducting a larger study (full-scale RCT) to investigate the effect of the smartphone app Sukaribit on glycemic control. Of the 5 prespecified progression criteria, only 1 was fully met, and 1 was partially met. This indicates that improvements should be considered both regarding study procedures and the intervention before further evaluations. The effect sizes were generally small. Given the low amount of participant activity, this was to be expected.

Another way to facilitate the recruitment process could be to add inclusion or exclusion questions on the study-specific website to better be able to reach the right target group.

Adding to the loss of potential participants in the early recruitment phases, the proportion of eligible and initially willing people who did not finally start participating was also slightly lower than that specified in criterion 2 (76% vs 80%). However, we could relatively easily compensate for these losses in recruitment with a longer and more aggressive recruitment campaign and by reconsidering the arguments for the BMI exclusion criteria.

Although data collection from self-reported questionnaires worked well, meaning that progression criterion 3 was partially met, the collection of HbA<sub>1c</sub> data through home testing kits did not work well. The first problem was the long administration time. The mean time from sending the kits out until the results were received by the project team was 16 days, with the main delay appearing to be at the participant’s home. There was also a large amount of missing data due to both defective tests and missing tests, even though several manual reminders were sent. Hence, there is a need to make the collection of HbA<sub>1c</sub> data more reliable and efficient. Previous studies have also reported difficulties using these test kits [26]. Better or additional instructions, more telephone reminders, another test kit brand, or another lab are things to consider. Most likely, the biggest advantage can be gained by improving the participant handling of the test and posting. Other ways to handle this could be to...
conduct this kind of trial within the health care system so that the blood test is managed by health care professionals and not by the participant. Another thing that we could have done differently is to not have HbA1c as an inclusion criterion. In a full-scale trial, the participant’s glycemic control at the start of the intervention could be a minor issue, since it is the effect of the intervention (the difference) that is measured.

**Feasibility and Acceptability of the Intervention**

Not all eligible participants who signed up for the trial started the intervention (progression criterion 4). We do not know the reasons for this; possibly, they just regretted the enrollment. When or if conducting a larger study, the possibility that not all who are accepting of study participation will actually join the study needs to be considered.

A few of the participants in the qualitative evaluation thought that the app improved self-care, but the majority did not think so. Many participants appeared to have been less active than anticipated, especially based on the amount of diagnostic data and messages sent to the physician. This was progression criterion 5, which was not met. Some participants described technical issues that interfered with the use of the app (eg, lack of access to pedometer data in the app as well as difficulties logging medications and viewing summary features). These problems could most often be related to the iOS version. This could have had an impact on user motivation leading to less activity. Participants also suggested improvements in the message function and added features when logging data (eg, in the calendar function, graph) in order to make the app more user-friendly. For the app to be beneficial, it is important that it is used. Previous studies [12] have shown that unsatisfied users will be less active and therefore will not benefit from using this kind of app. Multistep tasks, difficult system navigation, limited functionality, and limited interaction are generally the most common and important usability problems.

To improve user activity, the instructions given to the participants could be improved or routine follow-up telephone calls could be conducted with the participants in the treatment group. The intervention itself could have been more specific, with more guidelines for the participants to enhance their participation, and that might have led to more active self-care. However, this might have been perceived as a bigger effort. Nevertheless, the basic features of the app (ie, self-monitoring and facilitating patient-caregiver communication) appear to be valued by participants. For some, it facilitated a shift in routines toward more frequent blood glucose measurements and a larger understanding of the underlying causes of variations in their blood glucose levels. A feature that may enhance patient engagement is personalized content; for example, individual messaging between the caregiver and user seems to have positive effects in other studies. However, this is something that has not been adequately studied [27].

**Effect Sizes**

The effect sizes were small or not existing. Due to the feasibility concerns already raised, it would be premature to calculate a sample size for a full-scale RCT based on these results. If one still would, the only significant result was the number of blood tests, which had an effect size of 0.19. This would result in a necessary sample size of 870 (435 per study group; power= .80, \( \alpha = .05 \)). Based on the HbA1c results, the required sample size would be close to 1000. One could, based on the almost nonexistent effect sizes, reconsider the choice of self-rated outcome measures. However, with the low activity levels, it is difficult to say if the measures were not sensitive enough or if the intervention did not have a large enough impact.

**Clinical Significance**

The results of this study demonstrate the importance of conducting a feasibility trial in order to avoid unnecessary financial as well as study burden for those involved. In order to proceed with a larger clinical trial, a number of problems both in study design and the intervention, as described in the previous sections, need to be addressed. The next step then is to perform a sufficiently powered RCT. If the results are favorable, this will be the first step toward clinical evidence for the intervention, and a new digital treatment helping people to better manage their type 2 diabetes may be available shortly [28].

The participants who signed up for this study were particularly interested in mobile apps; therefore, the results from this study are applicable for patients with type 2 diabetes who want a digital aid. The app could complement standard care and possibly increase empowerment and self-care management. The main advantage of this app is that it enables a new and, maybe, faster way for communication between the person with diabetes and the diabetes nurse or general practitioner. This app, along with other available apps, could be suitable for health care now as well as in future, more digital health care [29].

**Limitations**

The smartphone app needed improvements during the trial period. Both the participants and study physician experienced development problems. This probably affected the participants’ experiences with the app. Another possible area of development is of the intervention itself, perhaps with a bigger focus on lifestyle and possibly with other professions involved such as a diettian, physiotherapist, or diabetes nurse. A possible bias in this study was that the study physician was part of the study team. Since she followed the study protocol and was not involved in the data collection, we believe this issue to be of minor importance. However, an independent physician or diabetes nurse would be preferable. The most preferable option would have been to involve the participants’ own family physician or diabetes nurse, who would have had personal knowledge of the patient. Another bias could be that the participants who signed up for this study were particularly interested in mobile apps. Therefore, the results from this study are applicable for people with type 2 diabetes who want a digital aid and not for the entire population.

**Strengths**

A strength of this feasibility study is that the trial was rather large and comprehensive for being a feasibility trial. Another strength is that the app and study methods have been evaluated in several ways with both quantitative and qualitative data, and the evaluation placed a lot of emphasis on the participants’
views. It is important to use different types of methods and validated instruments to get a more comprehensive evaluation of a diabetes app [30].

**Conclusion**

Recruitment of participants to a future fully powered study may work with adjustments. The collection of HbA1c using home testing constituted a major problem, and an alternative strategy for this measure is warranted. Finally, the app was not used by participants as intended, and further development is needed. In summary, in order to proceed with a larger randomized study, the app and study procedures need improvement.

**Acknowledgments**

The authors would like to thank the study participants. We also thank Michael Nahashon of Maishabit AB for making the Sukaribit app available and providing technical support.

The study was funded by Uppsala University Innovation and Uppsala University Project AB.

**Authors' Contributions**

EMGO and MH conceptualized the study. EMGO, UP, and M Hed supervised the study. TL performed project administration. EMGO and MH designed the methodology. EMGO curated the data. CJ and EMGO performed the formal analysis. CJ created the visualizations. CJ and TL wrote the original manuscript draft and performed the investigation. CJ, MH, UP, M Hed, and EMGO reviewed and edited the manuscript.

Generative AI was not used in any portion of the manuscript writing.

**Conflicts of Interest**

None declared.

**References**


**Abbreviations**

**CONSORT:** Consolidated Standards of Reporting Trials  
**DDS:** Diabetes Distress Scale  
**DSMQ:** Diabetes Self-Management Questionnaire  
**HbA1c:** hemoglobin A1c  
**RCT:** randomized controlled trial  
**U-CARE:** Uppsala University Psychosocial Care Program
Office Workers’ Views About the Uses, Concerns, and Acceptance of Hand Hygiene Data Collected From Smart Sanitizers: Exploratory Qualitative Interview Study

Sophie Rutter¹, PhD; Sally Sanger¹, PhD; Andrew D Madden¹, PhD; Sukaina Ehdeed¹, PhD; Catherine Stones², PhD

¹Information School, University of Sheffield, Sheffield, United Kingdom
²School of Design, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Sophie Rutter, PhD
Information School
University of Sheffield
The Wave 2, Whitham Road
Sheffield, S10 2AH
United Kingdom
Phone: 44 0114 222 2659
Email: s.rutter@sheffield.ac.uk

Abstract

Background: COVID-19 and the prospect of future pandemics have emphasized the need to reduce disease transmission in workplaces. Despite the well-established link between good hand hygiene (HH) and employee health, HH in nonclinical workplaces has received little attention. Smart sanitizers have been deployed in clinical settings to motivate and enforce HH. This study is part of a large project that explores the potential of smart sanitizers in office settings.

Objective: Our previous study found that for office workers to accept the deployment of smart sanitizers, they would need to find the data generated as useful and actionable. The objectives of this study were to identify (1) the potential uses and actions that could be taken from HH data collected by smart sanitizers (2) the concerns of office workers for the identified uses and actions and (3) the circumstances in which office workers accept HH monitoring.

Methods: An interview study was conducted with 18 office workers from various professions. Interview questions were developed using a framework from personal informatics. Transcripts were analyzed thematically.

Results: A wide range of uses of smart sanitizer data was identified including managing hygiene resources and workflows, finding operating sanitizers, communicating the (high) standard of organizational hygiene, promoting and enforcing organizational hygiene policies, improving workers’ own hygiene practices, executing more effective interventions, and identifying the causes of outbreaks. However, hygiene is mostly considered as a private matter, and it is also possible that no action would be taken. Office workers were also concerned about bullying, coercion, and use of hygiene data for unintended purposes. They were also worried that the data could be inaccurate or incomplete, leading to misrepresentation of hygiene practices. Office workers suggested that they would be more likely to accept monitoring in situations where hygiene is considered important, when there are clear benefits to data collection, if their privacy is respected, if they have some control over how their data are collected, and if the ways in which the data will be used are clearly communicated.

Conclusions: Smart sanitizers could have a valuable role in improving hygiene practices in offices and reducing disease transmission. Many actionable uses for data collected from smart systems were identified. However, office workers consider HH as a personal matter, and acceptance of smart systems is likely to be dynamic and will depend on the broad situation. Except when there are disease outbreaks, smart systems may need to be restricted to uses that do not require the sharing of personal data. Should organizations wish to implement smart sanitizers in offices, it would be advisable to consult widely with staff and develop systems that are customizable and personalizable.

(JMIR Form Res 2024;8:e47308) doi:10.2196/47308

KEYWORDS
hand hygiene; smart sanitizers; Internet of Things; IoT; offices; workplaces; smart systems
Introduction

Hand Hygiene in Workplaces

Recent pandemics (such as severe acute respiratory syndrome 1, H1N1 influenza, Middle East Respiratory Syndrome, and COVID-19) have demonstrated that public health threats are synonymous with occupational health threats [1]. Good hygiene protects people from acquiring and spreading gastrointestinal infections and respiratory infections [2]. When employees work in close proximity with others (colleagues, clients, and customers), share spaces (such as offices, kitchens, and break rooms) and share resources (such as computers, photocopiers, water coolers, and sinks), infectious diseases can be easily spread [3]. Hygiene practices such as using sanitizer, washing hands, and disinfecting surfaces have proved to be effective in reducing pathogen spread [4,5] and reducing illness in workplaces [3] and may be adopted more readily than other public health measures such as mask wearing [6].

Several studies have found that good hand hygiene (HH) reduces both absenteeism and presenteeism (attending work when ill), increases productivity [3,7], reduces the pressure on health services, and helps to tackle antimicrobial resistance through the reduced use of antibiotics [8]. How much time is lost to work owing to poor hygiene is difficult to quantify [9] as hygiene-related absences may be brief organizations tend not to report them to authorities, and it is often not possible to directly connect an acquired illness with poor hygiene practice (eg, without specific tests, what led to an employee’s stomach upset is speculative). Nonetheless, many of the studies cited previously indicate that much time is lost to work because of poor hygiene practices. Furthermore, before COVID-19, for many workers, attending work was rarely considered a risk to health, but now, employees may be highly anxious about infections in the workplace [10].

The COVID-19 pandemic led to a proliferation of guidance about reducing infection transmission in the workplace [7,11,12]. People were, and still are in some settings, encouraged to work from home where possible, keep socially distant, wear masks, wash and sanitize hands, and get vaccinated. Workplaces can also use 3 strategies to control infections [13]. First, they can try to prevent infections from entering the workplace through health screenings and by reducing or eliminating contact between workers (eg, working-from-home policies). Second, workplaces can help stop the transmission of infection through ventilation strategies and by erecting barriers and screens to prevent movement of aerosols. Finally, workplaces can help protect the worker from acquiring infections with personal protective equipment. However, despite many of these measures being introduced into workplaces, infectious disease transmission remains as a challenge [1]. This matters because it is vital that workplaces are in a position to adopt infection prevention and control strategies as and when required for current and future infectious diseases [1].

Smart Sanitizers

Smart sanitizers are already on the market and deployed in clinical settings. In this study, we consider the potential of smart HH systems in offices. Also known as automated hand hygiene monitoring systems (AHHMS) and electronic monitoring systems, smart sanitizers are Internet of Things devices. The device (the “thing”) stores and dispenses soap when activated by a sensor. Sensors within the device collect information about soap consumption (activation of the soap dispenser and fill level of the dispenser). When networked with other sensor data, such as movement of people (eg, entry to a room or building) and person tags (such as staff ID cards), the smart system can monitor a person’s HH based on where they are and what they are doing. The individual or aggregated data can then be shared on personal devices and apps, with sanitizer users and anyone else on the network. The system can also send messages and reminders and give feedback to registered users. The basic functionality has been established for >10 years and continues to develop and evolve. Recently, there has been a move to develop smart systems that can measure hand washing quality (correct technique and adequate time) [14].

Smart sanitizers have been adopted in clinical settings where HH is operationally crucial to help stop the spread of health care–associated infections, and there is a requirement to audit health care workers’ HH when caring for patients [15,16]. In clinical settings, smart sanitizers are generally considered to be effective in increasing HH, at least in the short term [17]. As HH is an important part of health care workers’ professional practice, many health care workers welcome the use of technology to improve hygiene adherence [18,19]. However, there are concerns about the loss of privacy and the potential for coercion, with many health care workers expressing a preference for systems that do not collect any personal data [14,19,20]. Health care workers are also concerned that the data collected may not accurately represent hygiene practices if the technology is prone to error, deliberately manipulated, or the context of HH (or rather, lack of HH) is not taken into account [14,18-20]. Furthermore, there are concerns about infrastructure costs and the potential for side effects of using systems that use radio frequency interference and UV light [14].

There has been little deployment of smart sanitizers outside clinical settings. Whether office workers would be as open as health care workers to adopting this technology needs further investigation, particularly because acceptance of smart sanitizers is dependent on organizational culture and how monitoring is implemented [17,18]. Moreover, as the professional concerns differ, how the technology is deployed and used in office settings may be different.

In 2021, Zivich et al [21] conducted a feasibility study for collecting HH data in offices and data about person-to-person contacts. Sensors were installed in soap and alcohol sanitizers in 2 US offices, and those participating (n=43) also carried sensors. From the data collected, first, the study authors found that office workers likely overestimate the frequency of their HH practices and, second, those with supervisory roles had fewer in-person interactions than those without supervisory responsibility. The authors also found that study participants were willing to carry sensors and have their interactions tracked. However, participation in the study was not obligatory and those participating were appropriately incentivized with a US $25 gift card. It is therefore not clear whether these office workers would be happy with such tracking as part of their usual working
practices, and in fact, some study participants suggested that they would need an increase in compensation to participate in a long study. Together, these findings suggest that smart sanitizers could be useful in (1) helping office workers identify their HH practices and (2) understanding and managing disease transmission in offices. However, it is not clear whether office workers would be willing in everyday life to have their hygiene data and contacts with other people collected.

Further to this, we investigated the attitudes toward the use of smart sanitizers in the workplace using a survey of workers in nonclinical settings (n=314), followed up with a qualitative questionnaire (n=12) and interview (n=3) [22]. Survey participants were generally in agreement that at work, high standards of HH is important and that smart sanitizers could usefully inform maintenance staff when to refill. However, there was little consensus with regards to the acceptance of collecting data that would give office workers an overview of their own HH practice, allow them to compare their own practices with those of others, provide them with personal messages, and give managers an anonymized view of HH practices. What was clear from the written responses and interviews is that participants thought it important that the data should only be collected if they can be acted upon, that is, the data should not just be collected because the technology allows it. This means that, before introducing smart sanitizers to the workplace, it is necessary to identify what actions could be usefully informed by the data. This led to our first research question (RQ), for which we adopted an exploratory approach to identify all the potential actions: RQ1—What actions could be taken from HH data collected by smart sanitizers?

However, whether these actions would be accepted by office workers requires further investigation, because survey participants were also concerned that HH data could be misused and misinterpreted. In particular, participants were concerned that collecting HH data could be an invasion of privacy, and the data collected may not be accurate. What HH data are needed will depend on how the data are to be used; therefore, it would be helpful to know the concerns associated with possible actions. This led to our second RQ: RQ2—For the actions identified, what, if any, are the data collection concerns of office workers?

Finally, the survey was conducted during the pandemic (July 2021 to August 2021), at a time when participants may have considered HH as particularly important. Concern for their health could have influenced the extent to which participants were willing to accept monitoring. When else, if ever, office workers would be more likely to accept smart sanitizers is not known. This led to our final RQ: RQ3—Under what circumstances would office workers accept HH monitoring?

**Methods**

**Overview**

This study was conducted as part of a large project to develop a smart hand sanitizer for the office environment. The project is a collaboration between the University of Sheffield (Information School) and the University of Leeds (School of Design), together with Savortex (a manufacturer of HH technology). The study reported in this paper, including data collection and analysis, was conducted solely by the universities.

**Recruitment**

This was a qualitative study to identify the potential uses of smart sanitizers from the perspective of those who work in offices all or most of the time. Interviews were conducted either using video link or via telephone, and they occurred between January 2022 and March 2022. The questions were pilot-tested with 2 participants known to the project team. Participants from a previous survey of attitudes toward the use of smart sanitizers in the workplace [22], who had expressed interest in further participation, were invited to participate in this study: 11 participants consented. To elicit a range of views, additional 7 participants were recruited using the research team’s networks. Although half of the participants (9/18, 50%) were from the education sector, sector did not account for differences in responses in the previous survey [22]. There were 18 participants in total, 3 (17%) of whom had some responsibility for hygiene within their organization (Table 1).
Table 1. Distribution of study participants based on role, sector, and responsibility for hygiene.

<table>
<thead>
<tr>
<th>Participant number</th>
<th>Role</th>
<th>Sector</th>
<th>Responsibility for hygiene</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Health care professional (office based)</td>
<td>Private health services</td>
<td>No</td>
</tr>
<tr>
<td>P2</td>
<td>Educator</td>
<td>Education</td>
<td>No</td>
</tr>
<tr>
<td>P3</td>
<td>Administrator</td>
<td>Local government</td>
<td>No</td>
</tr>
<tr>
<td>P4</td>
<td>Administrator</td>
<td>Health services</td>
<td>No</td>
</tr>
<tr>
<td>P5</td>
<td>Researcher</td>
<td>Education</td>
<td>No</td>
</tr>
<tr>
<td>P6</td>
<td>Social worker</td>
<td>Local government</td>
<td>No</td>
</tr>
<tr>
<td>P7</td>
<td>Educator</td>
<td>Education</td>
<td>No</td>
</tr>
<tr>
<td>P8</td>
<td>Disability liaison officer and educator</td>
<td>Education</td>
<td>No</td>
</tr>
<tr>
<td>P9</td>
<td>Conveyancer</td>
<td>Legal</td>
<td>No</td>
</tr>
<tr>
<td>P10</td>
<td>Director</td>
<td>Research and design</td>
<td>No</td>
</tr>
<tr>
<td>P11</td>
<td>Facilities manager</td>
<td>Education</td>
<td>Yes</td>
</tr>
<tr>
<td>P12</td>
<td>Hearing impairment teacher</td>
<td>Education</td>
<td>No</td>
</tr>
<tr>
<td>P13</td>
<td>Not known</td>
<td>Media and culture</td>
<td>No</td>
</tr>
<tr>
<td>P14</td>
<td>Deputy facilities manager</td>
<td>Education</td>
<td>Yes</td>
</tr>
<tr>
<td>P15</td>
<td>Facilities manager</td>
<td>Soft service industry</td>
<td>Yes</td>
</tr>
<tr>
<td>P16</td>
<td>Educator</td>
<td>Education</td>
<td>No</td>
</tr>
<tr>
<td>P17</td>
<td>Finance officer</td>
<td>Local government</td>
<td>No</td>
</tr>
<tr>
<td>P18</td>
<td>Educator</td>
<td>Education</td>
<td>No</td>
</tr>
</tbody>
</table>

Data Collection

To prepare for data collection, we turned to the field of personal informatics. Personal informatics systems “help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge” [23]. As such, smart HH systems can also be considered as personal informatics systems because individuals can use them to collect and track data about their HH practices. A semistructured interview guide was developed based on the stage-based model by Li et al [23]. This model is widely used in the design of personal informatics systems and holistically describes, from a user perspective, the stages of collecting and using personal data. A set of main questions (Table 2) relating to the 5 stages (preparation, collection, integration, reflection, and action) was prepared, together with several possible prompts. To allow for the identification of all the potential uses of HH data, we did not restrict the discussion to smart sanitizers that are currently on the market; at the beginning of the interview, participants were told that “By hand hygiene we mean using any kind of hand cleaning facility. This includes hand washing, using a wall sanitiser or using your own sanitiser from a bottle or a wipe.”

Table 2. Study interview guide based on the stage-based model of personal informatics by Li et al [23].

<table>
<thead>
<tr>
<th>Stage-based model of personal informatics</th>
<th>Main question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>How could hand hygiene data be used and what data should be collected?</td>
</tr>
<tr>
<td>Collection</td>
<td>How should hand hygiene data be collected?</td>
</tr>
<tr>
<td>Integration</td>
<td>How should the collected data be prepared and processed?</td>
</tr>
<tr>
<td>Reflection</td>
<td>Who should see the data and how should this be presented?</td>
</tr>
<tr>
<td>Action</td>
<td>What might you do as a result?</td>
</tr>
</tbody>
</table>

Data Analysis

Data were analyzed inductively using a “codebook” approach to thematic analysis [24], whereby a structured coding framework is used to analyze the data. Preliminary open coding was performed by the second author. At a follow-up meeting, developing codes were discussed with the first, third, and fourth authors and an initial codebook was compiled. The second author completed the coding of the remaining transcripts. The first author mapped the open codes to the RQs. This was then further reviewed by the second author.

Ethical Considerations

This study received ethics approval (038337) from the University of Sheffield Research Ethics Committee on February 16, 2021. All study participants received an information sheet about the project, and they were given opportunities to ask questions and advised that they could withdraw with no negative
consequences. All participants gave their informed consent. For confidentiality, their data are anonymized.

**Results**

RQ1: What Actions Could Be Taken From HH Data?

**Overview**

An exploratory approach was adopted for the first RQ, and all potential actions were identified. It was thought that HH data could be acted upon to (1) manage hygiene resources and workflows, (2) find operating sanitizers, (3) communicate the (high) standard of organizational hygiene, (4) improve own practice, (5) promote an organization’s hygiene policy, (6) enforce organizational policy, (7) target the training according to needs, (8) execute more effective interventions, and (9) identify the causes of outbreaks, and whether (10) any action would be taken was also considered.

**Manage Hygiene Resources and Workflows**

Facilities managers could find HH data useful when planning and maintaining hygiene facilities, including the purchase and decommissioning of sanitizers, purchase of soap and gel, optimal placement of sanitizers, and identification of when maintenance is required. Maintenance data could also make work processes and workflows more efficient:

> We then don’t have to send a cleaner every hour for no reason. [P15]

**Find Operating HH Facilities**

If stock fill level data were shared with everyone, building users could act upon the data to find operating HH facilities:

> It’s a bit like “FindmyPC” isn’t it?... If there’s nothing [soap and gel] on [place] I can go to [place] and I can get the stuff there. [P11]

**Communicate the (High) Standard of Organizational Hygiene**

Organizations could use HH data as tangible evidence to reassure employees and visitors that there is a high standard of HH in the building. This could be particularly useful for organizations that work with vulnerable people:

> This last calendar month we had 95% usage of all of our machines [that would communicate] we’re looking after our staff and the compliance of that. [P15]

**Improve Own HH Practice**

Through managing, tracking, and understanding their hygiene practices, including evaluating their HH technique, individuals could act upon the data to improve their HH practices:

> Like, let’s say you have a ring, it’s not very clean around the ring, I would then know and I would spend more time obviously. [P13]

If the system collected contextual data (including what the person is doing at the time, how they are feeling, and the current risk of catching an infection), the system could usefully identify trends, give insights, and make recommendations that would enable individuals to further act upon their HH data:

> Recommendations on my hand washing behaviour, like, you know: “In general, you do not seem to wash your hands very well on Tuesdays or on Wednesdays.” That might help me understand why that’s the case. Another thing that might be useful...then it might be nice to know if I’m washing hands when I’m meeting people more or if I have more meetings. So if it’s connected with the calendar then it might be able to give some more insight into why I think I’m not washing my hands more, when I’m washing them and where I’m going [next] so that might be good. [P2]

Participants thought that having access to other people’s HH data could enable people to benchmark and contextualize their own results:

> If I could compare my handwashing with somebody else’s, and if mine looked that I was hand-washing too often, then I’d have to look at if we’re all doing the same number of visits in a day, am I sanitising my hands too much, but then if I’m not getting infections at the same rate as other people then maybe I’m not hand sanitising my hands too much. [P12]

**Promote an Organizational Hygiene Policy**

Organizations can use smart systems to communicate their HH policy. Smart systems could help promote policies by sending reminders and keeping employees motivated through comparisons, competitions, and other incentives:

> I suppose they should have some sort of benchmark, you know, like “The rest of the organisation are all doing it really frequently and doing it for the right duration, but your team aren’t” so they have to have some sort of like benchmark as to where they fall on a scale, as it were. [P4]

It was thought that smart systems could be useful when new routines are introduced:

> If there were changes in expectations, such as more restrictions were put in place, if there was another outbreak. [P6]

This would also apply when new staff join the organization.

**Enforce Organizational HH Policy**

Organizations can use smart systems to identify compliance and changes in compliance. If a lack of compliance has been identified, organizations could target particular events (such as after using the toilet), individuals, teams, and departments to set hygiene goals that align with their policy:

> Showing trends, showing ups and downs, especially the downs, might highlight points to people to make them realise...you can use that to some effect then, can’t you, if you have a particular outbreak in a particular team or whatever. It might prompt people to take a bit more action to it maybe. [P9]
The data could also be synced with door entry systems to prevent people from entering spaces (such as food preparation areas), but none of the participants (0/18, 0%) thought this was a good idea.

**Target HH Training According to Needs**

An analysis of HH data could also help to identify who needs training and what their training needs are:

If they’re looking at training needs and compliance and safety and all those sorts of things, could use those to identify if there are any gaps. [P4]

**Execute More Effective Interventions**

Organizations and researchers could evaluate the effectiveness of interventions in real time and adapt them according to the results:

It doesn’t have to be Coronavirus, it could be the flu or something, it would be useful to see that, and to see how people responded to prompts and reminders. [P6]

**Identify the Causes of Outbreaks**

Participants also discussed the possibility that if HH data were combined with other health data, it could enable researchers to gain a better understanding of the impact of HH on health and the cause of infectious outbreaks:

Reporting that there’ve been a lot of stomach upsets, and that was linked in with the data on hand washing, which was very low, then you could put two and two together, and that could be useful. [P5]

**No Action**

Whether any action would be taken was also discussed. HH was often thought to be a personal matter and the responsibility of the individual. Several participants stated that they would not say anything or take any action if they knew their colleagues had inadequate HH practices:

I do think it’s pretty disgusting if people don’t wash their hands, [pause] but it’s not for me to tell them to...I’ve just got to be responsible for myself. [P17]

**RQ2: For the Actions Identified, What, if Any, Are the Concerns of Office Workers?**

Office workers are concerned that (1) intentions and messages could be misinterpreted and that data could be used for (2) bullying and coercion, (3) unintended purposes, (4) inaccurate representation of HH practices, and (5) incomplete representation of HH practices. Next, we have discussed which actions raise the concerns.

**Intentions and Messages Are Misinterpreted**

It was thought that messages generated by a smart sanitizer may not be received as intended. Using HH data to reassure building users about the status of HH in organizations could instead make them feel anxious:

Then again it could let people, like I say, who are socially anxious think “Oh my God, no-one’s cleaning their hands, it’s a really dirty place.” You will get people that will freak out about that. [P3]

Although organizations may install smart HH systems to reassure office workers, office workers may feel that monitoring could imply that a person is not able to manage on their own. The installation of devices that monitor HH could be construed as a message conveying lack of trust:

It felt like if we monitored something like that, then it would damage trust, it would make people less independent and capable of taking care of their health because it would set an expectation that someone else is going to monitor it. [P10]

**Bullying and Coercion**

Many interviewees felt that the data would be of particular interest to managers, but using the data to promote and enforce HH policy could lead to bullying, be divisive, and encourage rivalries:

I’d be concerned in some bits of the organisation that I worked in, that some managers would use it punitively to, not necessarily call out people publicly, but use it to...bully people or shame them or whatever. [P4]

Benchmarking one’s own HH against others was thought to be helpful in improving one’s own HH practice, but it was also thought that office workers may feel harshly and unfairly judged:

If my whole team does it, then if I don’t do it then I’m gonna surely [be] judged for, like, not cleaning my hands even though it [my reason for refusing] has nothing to do with that. [P13]

**Used for Unintended Purposes**

Participants thought that HH data could be (deliberately or inadvertently) used for purposes that do not benefit the organization or their employees. Moreover, HH data may reveal other personal information that would not be appropriate for organizations to know:

You might feel forced to say, “Oh, actually, I’ve got a bit of morning sickness. I think I might be pregnant,” and then you might have a miscarriage or something like that, so then it could – that might all – oh, dear, yeah. Or you might – say it could be an emotional reason why you’re going to the toilet. You might be going because you’re very upset about something. But I think, yeah, it could reveal all sorts of things about human behaviour, and actually, in an unintended way, reveal things about that person that are very private. [P5]

There was some concern that manufacturers of sanitizers and cleaning products would use the data to increase sales:

If the outcome is, how can we sell more hand sanitizer, what if we connected our hand sanitizer product to the internet...I don’t think that’s a good outcome, and I don’t think it comes from a good place. [P10]
Inaccurate Representation of HH Practices

Whether the data collected would be an accurate representation of HH practices and, therefore, whether any conclusions can be drawn from the data collected was also a concern. Participants worried that the use of smart systems could be manipulated, and therefore, the resulting data would be inaccurate and would misrepresent organizational HH practices:

*I know some people who are just going to go round and just put their hand under every time they walk past just so they’ve triggered it whether they’re washing their hands or not.* [P12]

Incomplete Representation of HH Practices

Participants expressed the concern that smart systems could not capture all the data necessary to represent all HH practices, and this also adds to concerns about whether any conclusions can be drawn from the data. First, smart systems alone cannot capture all HH events (such as an employee’s use of their own sanitizer and wipes, which may be a personal preference or a necessity, eg, if a person has allergies to a particular substance):

*Maybe you think, “I don’t want to touch the wall hand sanitizer because everybody else has touched that, so I’m going to stick to my personal sanitizer,” in which case, that wouldn’t capture any of that, so you would need both, for a true figure.* [P5]

Second, employees may work from home or in other locations outside the aegis of the organization, where it would be difficult for smart systems to capture HH events. Furthermore, for improving their own HH practices, people would want data beyond the work context:

*How long I’m spending washing my hands, gaps in between, but also if there’s any variation in days. So, I mean, Saturday and Sunday might not be different if I’m out and about, on a personal level, than the Thursday or Friday if I’m working. I would expect there to be, but if I was shopping, and I went in 20 shops on a Saturday, that might reflect that I was handwashing the same as I was in a working day.* [P12]

Finally, to fully interpret HH practices, it would also be necessary to collect data about what the employee was doing at the time and where. Otherwise, there is a danger that the system may incorrectly interpret HH practices as missing. Participants also questioned whether using HH data to draw comparisons between different departments would be meaningful, as different roles may have different HH requirements:

*If I’ve just come from the toilet and I’ve washed my hands and walked past a hand sanitiser, if I got a pop up on my machine...that says “You’ve been past a hand sanitiser and you haven’t used it” I would expect to be able to interact with it and explain to it why I haven’t used it.* [P4]

*RQ3: Under What Circumstances Would Office Workers Accept HH Monitoring?*

The concerns expressed previously suggest that there will likely be some resistance to many of the uses of HH data. However, office workers suggested that they are more likely to accept HH monitoring (1) if they or others could not be identified, (2) in situations where HH is considered important, (3) when events considered as private are not recorded, (4) when data collection can be customized, (5) when data are used for a beneficial purpose, and (6) when uses of the data are clearly communicated.

When Identity Is Protected

Participants expressed little or no concern about organizations accessing data from sensors in which no personal data are collected and were therefore generally accepting of the uses of HH data for resource planning purposes.

Most participants thought that data about individuals should not be shared with others. However, 11% (2/18) of the participants thought that attributed personal data should be seen by senior managers (P4 and P6) and another 11% (2/18) thought that attributed personal data could be seen by team managers also (P12 and P14):

*So with the data and reports, the only things that I think that someone else should be seeing about me are aggregate. So nothing where people can be identified.* [P2]

Senior management team or board level or Health and Safety Executive should have all the information by teams or by individuals but the individual managers...I wouldn’t want my manager of my department to have individualised data that makes them be able to say “[name] is not washing her hands often enough.” I would prefer that it’s anonymised at that level. [P4]

Although participants were generally uncomfortable with personal data being shared with others, they were mostly comfortable with personal data collection if they or their colleagues could not be identified in any reports:

*I think aggregated reports should be available to everyone, as a comparison purpose. I think maybe a little bit more detailed aggregated reports, for example, with a maximum, minimum, with a band, with a percentile band, with the longer period of change can be available to health and safety officer, can be a department manager, or what they call the senior manager group, steering committee.* [P8]

Situations Where HH Is Considered Important

Monitoring was thought to be more acceptable when the importance of HH is clear. Therefore, monitoring was seen as more acceptable in certain settings, notably, health care and food preparation, and for certain teams or roles, for example, food technicians and carers:

*If I worked in a food environment, it’d be very different.* [P17]
It was also thought more acceptable during infectious disease outbreaks:

Suppose there’s another virus outbreak and it’s demonstrated that hand washing is key to preventing its spread, and that you’re doing it for the public good...if it was, like, three years ago, I would’ve said this is ridiculous. Now, I think maybe, OK, in the right circumstances, I would go along with it, because the context seems to have changed. [P7]

It was thought to be acceptable at places in buildings where HH is important such as food preparation areas and toilets:

I would like to know that the people preparing my food wash their hands, that would be a good thing to know, because it’s crucial for there. The rest of them, I don’t need to know that, I don’t think...Although, I would prefer it if people washed their hands before they left the toilet – if an alarm went off there. [P5]

When Events Considered Private Are Not Recorded

Although it was thought helpful to capture HH data in locations where HH is important, data capture was felt to be more acceptable in some parts of the building than others. For example, monitoring HH on building entry was less controversial than monitoring outside a toilet:

Some people might think that it’s a bit of an invasion of privacy, being monitored in the toilet as well. Is there nowhere safe? Is there nowhere that I can just not be monitored? [P4]

When Data Collection Can Be Customized

It was thought that office workers would be more likely to accept monitoring if they can customize the system and control what data are collected and how they are presented:

I think perhaps like with the alerts, perhaps [they] could have the option to turn that on [recommendations, encouragement, advice] if you so desired, but it shouldn’t be a requirement. [P5]

The system should allow users to correct any errors in data collection and add explanations, so that managers do not unfairly target individuals:

So that you’ve got the chance to correct yourself if you need to, like, you see I would be going back to my computer and I would expect then to have a message on my computer that says “You’ve walked past a hand sanitiser and you didn’t use it. What was the reason?” [P4]

When Data Are Used for a Beneficial Purpose

Participants felt that data need to be collected for a purpose. The purpose needs to be justified, and the data should be retained only for as long as necessary. Furthermore, systems should be used to support individuals rather than punish them:

If there was a real, proper reason that they were collecting it for, then they could collect it for the relevant time period. So if there was some sort of disease outbreak and it lasted six months, then collect it for six months...it has to be justifiable...it’s not right to just collect it and hold that data. [P5]

It depends on what people perceive is the overall intention of whoever’s putting this policy in place. If it’s used--., if the perception is it’s used to beat people up about hygiene because it’s going to lead to a poor sick record or more transmissible covid than that is a different intention to “Well, I’m really bothered about how sore your fingers are becoming with all your hand washing.” [P12]

When Uses of the Collected Data Are Clearly Communicated

Participants recommended that the uses of the data need to be transparent and clearly communicated including how the data will be used and reported, who has access to what data, where and for how long the data are stored, and whether it is possible to opt out of data collection:

Why they’re doing it, who’s going to hold the data, who’s going to see it, how’s it going to be reported, who will it be shared with, yeah, where will the data go, how long will they hold the data for, and can I opt out, how do I opt out. [P2]

Discussion

Principal Findings

Our previous study found that office workers thought HH data should only be collected if they can be acted upon [22]. In this study, office workers were able to identify several actions that could usefully be informed by HH data (RQ1). These included using the data to manage hygiene resources and workflows, find operating sanitizers, communicate (high) organizational standards of hygiene, improve workers’ own practice, promote and enforce an organization’s hygiene policy, target the training according to needs, execute more effective interventions, and identify the causes of outbreaks. However, hygiene is mostly considered as a private matter, and it is possible that no action would be taken in practice. Furthermore, office workers expressed concerns (RQ2) that the data could be used to bully, to coerce and for unintended purposes. Moreover, the data could be misinterpreted, inaccurate, and an incomplete representation of hygiene practices. Office workers suggested that they would be more likely to accept monitoring for the identified uses (RQ3) when their privacy is respected, they have some control over how their data are collected, and how their data will be used is clearly communicated. Monitoring is also more likely to be accepted in situations where hygiene is considered important and there is a clear beneficial purpose for data collection.

HH Is (Mostly) a Personal Matter for Office Workers

Although the findings of this study suggest that facility managers, health and safety officers, departmental managers, building occupants and visitors, hygiene resource suppliers, researchers, and those interested in public health would find HH data useful, HH was thought to be a personal matter [25]. Several office workers reported that they would not take any action if they found that their colleagues’ HH practices were inadequate. For all the uses of HH data, there was a strong
preference for personal data to be anonymized or not collected at all. However, acceptance of HH monitoring is dynamic and dependent on the situation and the context within which the data are used. The findings of our study suggest that it is much more likely to be accepted during disease outbreaks, in certain locations (eg, entrance to buildings), and in sectors (eg, health and food) where HH is important to the ethos and culture of the organization (ie, will also influence office workers). Although none of the office workers (0/18, 0%) thought it acceptable to enforce HH by restricting access to areas (by syncing HH data with door entry system data), it is possible that under extreme circumstances and in certain locations, this could be acceptable.

Need for a Shared Understanding of “Good” Office HH

For those concerned about catching infectious diseases in the workplace [10], HH data could be used to reassure visitors to a building about the high standards of hygiene within the building and to promote and enforce organizational HH policies. In health care settings, smart sanitizers are already used to audit compliance and enforce the health sector’s policy of sanitizing hands before, during, and after patient care. Overall, 3 factors are likely to make smart sanitizers more acceptable in health care settings. First, they are used to enforce an HH policy that is considered important in professional practice [18]. Second, smart sanitizers can collect data that measure the compliance with policies such as the 5 moments for hygiene [26] that can be measured using the sanitizer supplied by the organization. Third, at least at a basic level, adherence to this policy can be monitored using room and sanitizer sensors, without the need for personal data collection.

For smart systems to be adopted in offices, it would help if there were an agreed-upon understanding of what is good hygiene practice, for example, how often and where (eg, entrances to buildings and exits from toilets) hygiene should be performed. Smart systems could then be used to reassure and promote HH in offices, if the policy can be complied with using office resources and without the need to collect personal data.

Office Workers Want Insights From All Their HH Practices

Health care workers may wish to track their HH practices around patients, as good HH is part of their professional identity [18], and having access to their personal HH data could help health care workers improve their HH practice [27,28]. Given that office workers are likely to be overestimating their HH practices [21], self-tracking could be beneficial. However, office workers in this study did not link self-tracking of HH with professional expectations and standards; rather, they were interested in gaining insights into their overall HH practices including in all locations (office, home, and when they are out of the office and across all facilities (sanitizer, soap, wipes, etc). No smart system (as yet) can automatically detect HH with such detail. This would only be possible if office workers were prepared to input data manually, and this would require considerable motivation.

Office Workers Share Health Care Workers’ Concerns

Many of the concerns that office workers expressed are similar to those of health care workers. Similar to health care workers [19,20], most office workers are concerned that personal data could be mishandled, exploited, or used to punish or bully employees, and use of HH data should be clearly and transparently communicated. It would be advisable to consult with office workers early in the system design process to engage them, explain what purposes the organization intends for the data, and identify what purposes they feel are acceptable and useful.

Another concern shared with health care workers is whether smart HH systems can accurately represent HH practices, because, first, systems can be gamed and deliberately misused, and second, systems may incorrectly interpret events as missed HH opportunities because they are not registering the wide context within which the event did or did not occur [18-20].

Given the shift to more hybrid and flexible working [29], it may also be necessary to analyze HH data alongside work patterns. More generally, studies of workplace tracking have found that systems that enable employees to customize and control what data are collected are more likely to be accepted [30].

Office workers were also concerned that the data could be used for purposes other than what it was intended for. This is understandable given that misuse of technologies is widely reported in the media; for example, AirTags designed to track property have been used to stalk individuals [31].

Useful for Health Researchers

Using sensors to remotely collect HH data resolves some of the challenges for health researchers who need to evaluate hygiene interventions. Researchers may evaluate interventions by observing HH practices, but the presence of an observer may change the behavior of the person being observed, particularly because HH is a social norm. Instead, researchers may use proxy measures such as changes in soap consumption. However, manually collecting soap consumption data from organizations is time consuming, and sensors can help in saving time [32].

Although it is thought likely that poor hygiene could contribute to disease transmission in workplaces, little data are available to support (or oppose) this point [9]. Connecting smart HH system data with other health data (such as data relating to employee absence) could help researchers understand the relationship.

Data collected from digital technologies (such as mobile phones, social media networks, and search engines) have been used to communicate public health messages and monitor and control outbreaks [33]. Smart HH systems could usefully be added to the arsenal of digital data sources that have been used to support health authorities’ response to COVID-19 and any future pandemics.

When Personal Data Are or Are Not Needed

To a large extent, how well a smart HH system is accepted depends on whether personal data are collected. We next consider what data smart HH systems can collect and what is needed for the identified actions (Table 3).
For some uses of HH data, there is no need for personal data to be collected. For resource and workflow planning and to find operating sanitizers, data about soap levels and soap consumption including date, time, and location of use can be collected from sensors in smart systems without the need to collect any personal data. Similarly, smart systems can give immediate feedback to users to improve their HH technique without collecting any personal data. It is also possible to communicate the overall standard of HH in an organization without collecting personal data. Smart system data combined with data from sensors that track movement in and out of spaces can be used to identify the extent to which all employees are practicing HH and whether they use HH facilities as they move around the building (eg, at the lift and after using the toilet). The data collected could also indicate overall compliance with organizational policy and be used to identify the overall education and training needs of the organizations. The same system could also give feedback to individuals at the point where they are using the facility.

Personal data are required for several of the uses of HH data identified in this study, particularly the uses where individuals are pinpointed such as identification of individual practices and training needs and enforcement of organizational policy. Sensors that track individuals (installed on staff ID cards, apps, or other personal devices) would be necessary to capture each person’s use of a hand sanitizer (date, time, technique, and location of use) and to send reminders. However, these data can be anonymized and aggregated to identify the uses of HH facilities by different groups (not individuals) within the organization. Good practice would be to offer a manual override that would allow employees to correct any system errors. For smart systems to fully represent a person’s HH practice, it would be necessary to allow users to manually input their use of any and all hygiene facilities such as wipes and their own sanitizer gel. To capture a person’s contextualized use of HH facility (including what the person had been doing and where they had been), the system would need to connect to other personal data such as calendars and mobile phones.

Whether personal data are needed to evaluate interventions will depend on the nature of the intervention and what needs to be evaluated. Identifying the causes of outbreaks will likely also require the collection of other personal data (eg, who is ill).

**Limitations**

Through this interview study, a wide range of applications for HH data collected via smart systems has been identified. However, this is an exploratory study; further investigation is needed to determine whether office workers would use smart sanitizers for the identified purposes. Such studies could build on these findings to further investigate the implementation and adoption of smart sanitizers, with trials in offices.

**Conclusions**

Smart sanitizers could, feasibly, make a contribution to the improvement of hygiene practices in offices [21], but for smart systems to be accepted, any data collected would need to be actionable [22]. This study contributes to knowledge by identifying the many potential uses for hygiene data collected from smart systems. As smart HH systems have not yet been introduced into offices, identification of constructive uses for data is important for their design and implementation.

Although smart sanitizers are widely deployed in clinical settings, health care workers recognize that HH is an important part of their professional practice [18,19]. Given that office workers consider HH to be a mostly personal matter, it seems less likely that they will want to adopt smart sanitizers. When there are disease outbreaks, office workers may consent to the sharing of personal data and the monitoring of their own and their colleagues’ HH. At other times, smart sanitizers may need

<table>
<thead>
<tr>
<th>Data to be collected</th>
<th>No personal data</th>
<th>Personal data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensors in dispensers of sanitizer, soap, etc</td>
<td>Sensors that track individuals</td>
</tr>
<tr>
<td>Soap levels and soap consumption including date, time, and location of use</td>
<td>✓</td>
<td>Manual input</td>
</tr>
<tr>
<td>Correct HH technique</td>
<td>✓</td>
<td>Personal devices</td>
</tr>
<tr>
<td>HH events or nonevents</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Date, time, duration, and location of a person’s use of a hand sanitizer</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Date, time, duration, and location of a person’s use of a hygiene facility</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Contextualized use of HH facility (including what the person had been doing and</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>where they had been)</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 3. Data collected by smart hand hygiene (HH) systems and whether these include personal data (from sensors that track individuals, manual input, and personal devices) or do not include personal data (from sensors in dispensers of sanitizer, soap, etc and sensors [including cameras] that track anonymized movement and location).
to be restricted to uses that do not require any personal data collection. Should organizations wish to implement smart sanitizers in offices, it would be advisable to consult widely with the staff and to develop systems that are customizable and personalizable. It should also be noted that office workers may find it more useful to have insights from all their HH practices, but these data cannot (yet) be automatically collected from smart systems.

In contrast to health care workers, as yet, there is no widely accepted HH policy for office workers. Future studies could usefully investigate what office workers would consider to be an appropriately high standard of hygiene and how often and where hygiene should be performed. A better understanding of what would be effective and acceptable HH policies in nonclinical settings would help to clarify how smart systems can be used and hence inform their design. Importantly, it could help workplaces adopt infection prevention and control strategies that are necessary for current and future infectious disease outbreaks [1].

Acknowledgments
This study was funded by Innovate UK (grant 8154) and Economic and Social Research Council’s (ESRC’s) Impact Acceleration Account (project 172894). This Innovate UK–funded study is part of a large project conducted in collaboration with Savortex, a manufacturer of hygiene technology. However, the authors of this paper have no financial or nonfinancial interests in this study.

Data Availability
This study involved human research participant data and could contain potentially identifying information. The data that support the findings of this study are available upon request from the corresponding author.

Authors’ Contributions
SR and CS were involved in conceptualization. SS, SR, and ADM contributed to the methodology. SS, SR, ADM, and SE were involved in formal analysis. SS and SE were involved in investigation. SR was involved in writing the original draft. SS, ADM, SE, and CS were involved in reviewing and editing the paper. SR contributed to project administration. SR and CS were involved in funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

References

https://formative.jmir.org/2024/1/e47308 JMIR Form Res 2024 | vol. 8 | e47308 | p.116 (page number not for citation purposes)


29. Matei A. ‘I was just really scared’: Apple AirTags lead to stalking complaints. Guardian. 2022. URL: https://www.theguardian.com/technology/2022/jan/20/apple-airtags-stalking-complaints-technology [accessed 2023-12-04]


Abbreviations

AHHMS: automated hand hygiene monitoring systems
HH: hand hygiene
RQ: research question

Edited by A Mavragani; submitted 15.03.23; peer-reviewed by E Toki, S Patel; comments to author 03.10.23; revised version received 13.11.23; accepted 04.12.23; published 11.01.24.

Please cite as:
Rutter S, Sanger S, Madden AD, Ehdeed S, Stones C
Office Workers’ Views About the Uses, Concerns, and Acceptance of Hand Hygiene Data Collected From Smart Sanitizers: Exploratory Qualitative Interview Study
JMIR Form Res 2024;8:e47308
URL: https://formative.jmir.org/2024/1/e47308
doi:10.2196/47308
PMID:38206674

©Sophie Rutter, Sally Sanger, Andrew D Madden, Sukaina Ehdeed, Catherine Stones. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Population-Level Portal-Based Anxiety and Depression Screening Perspectives in HIV Care Clinicians: Qualitative Study Using the Consolidated Framework for Implementation Research

Daniela Zimmer¹, MPH, MSW; Erin M Staab¹, MPH; Jessica P Ridgway¹, MD; Jessica Schmitt¹, MSW; Melissa Franco¹, MPH; Scott J Hunter¹,², PhD; Darnell Motley¹, PhD; Neda Laiteerapong¹, MS, MD

¹Section of General Internal Medicine, University of Chicago, Chicago, IL, United States
²Western Institutional Review Board- Copernicus Group, Princeton, NJ, United States

Corresponding Author:
Daniela Zimmer, MPH, MSW
Section of General Internal Medicine
University of Chicago
5841 S Maryland Ave
Chicago, IL, 60637
United States
Phone: 1 312 702 8847
Email: dazimmer@bsd.uchicago.edu

Abstract

Background: Depression and anxiety are common among people with HIV and are associated with inadequate viral suppression, disease progression, and increased mortality. However, depression and anxiety are underdiagnosed and undertreated in people with HIV owing to inadequate visit time and personnel availability. Conducting population-level depression and anxiety screening via the patient portal is a promising intervention that has not been studied in HIV care settings.

Objective: We aimed to explore facilitators of and barriers to implementing population-level portal-based depression and anxiety screening for people with HIV.

Methods: We conducted semistructured hour-long qualitative interviews based on the Consolidated Framework for Implementation Research with clinicians at an HIV clinic.

Results: A total of 10 clinicians participated in interviews. In total, 10 facilitators and 7 barriers were identified across 5 Consolidated Framework for Implementation Research domains. Facilitators included advantages of systematic screening outside clinic visits; the expectation that assessment frequency could be tailored to patient needs; evidence from the literature and previous experience in other settings; respect for patient privacy; empowering patients and facilitating communication about mental health; compatibility with clinic culture, workflows, and systems; staff beliefs about the importance of mental health screening and benefits for HIV care; engaging all clinic staff and leveraging their strengths; and clear planning and communication with staff. Barriers included difficulty in ensuring prompt response to suicidal ideation; patient access, experience, and comfort using the portal; limited availability of mental health services; variations in how providers use the electronic health record and communicate with patients; limited capacity to address mental health concerns during HIV visits; staff knowledge and self-efficacy regarding the management of mental health conditions; and the impersonal approach to a sensitive topic.

Conclusions: We proposed 13 strategies for implementing population-level portal-based screening for people with HIV. Before implementation, clinics can conduct local assessments of clinicians and clinic staff; engage clinicians and clinic staff with various roles and expertise to support the implementation; highlight advantages, relevance, and evidence for population-level portal-based mental health screening; make screening frequency adaptable based on patient history and symptoms; use user-centered design methods to refine results that are displayed and communicated in the electronic health record; make screening tools available for patients to use on demand in the portal; and create protocols for positive depression and anxiety screeners, including those indicating imminent risk. During implementation, clinics should communicate with clinicians and clinic staff and provide training on protocols; provide technical support and demonstrations for patients on how to use the portal; use multiple screening methods for broad reach; use patient-centered communication in portal messages; provide clinical decision support tools, training, and mentorship to help clinicians manage mental health concerns; and implement integrated behavioral health and increase mental health referral partnerships.
Introduction

Barriers to Depression and Anxiety Diagnosis
Depression and anxiety are common mental health conditions among people with HIV, with a prevalence of 20% to 45% [1-10]. People with HIV experiencing symptoms of depression or anxiety are more likely to miss appointments and have lower medication adherence, higher HIV viral loads, and higher mortality rates than those without depression or anxiety [1,2,5,10,11]. However, depression and anxiety are often underdiagnosed and undertreated in people with HIV, particularly among African Americans and Hispanics, because of the perceived stigma of mental health disorders, racial discrimination, HIV-related discrimination, and medical mistrust [1,3,5,12,13]. In the HIV Cost and Service Utilization survey of people with HIV identified as experiencing depression, only 45% had a formal depression diagnosis in their medical chart [14].

For people living with chronic conditions, such as HIV, specialty care clinics often serve as their primary source of health care [3,4]. Given the frequency of visits people with HIV have with their HIV care team, establishing mental health screening in HIV clinics is a key opportunity to address depression and anxiety underdiagnosis in people with HIV [3,4]. Patients and physicians have noted that depression screening in clinics is helpful in identifying, assessing, and treating depression [3]. However, competing demands and priorities during appointments, a lack of staff to complete assessments, and a shortage of resources to offer patients after diagnosis discourage clinicians from screening and treating depression [1,3,6,7,9,15,16].

Novel Mental Health Screening
A novel strategy to increase depression and anxiety screening in people with HIV is to perform screening at the population level using the patient portal. In recent years, health care systems have increased the adoption of electronic patient portals, and patients have increasingly used portals to facilitate their health care [8,9,17-20]. Studies in primary care settings have found that depression screening rates increased significantly when clinics adopted portal-based screening [3,7,18]. Notably, a population-level portal-based depression screening intervention, in which patients were invited to complete a depression screener regardless of having a scheduled appointment, also increased depression screening and diagnosis rates [21]. This population-level portal-based approach identified more patients with moderate to severe symptoms than screening during clinic appointments [22]. Moreover, portal-based screening increases the likelihood of discussing depression diagnosis and treatment during an appointment [7,19,20,22].

Population-level portal-based screening has been shown to be promising in primary care settings but has not been examined in HIV care settings. Guided by the Consolidated Framework for Implementation Research (CFIR), we conducted and analyzed qualitative interviews with clinicians at an urban HIV clinic [23]. This study explored clinicians’ perspectives on facilitators of and barriers to implementing population-level portal-based depression and anxiety screening for people with HIV. The objective of this formative study was to use the identified facilitators and barriers to develop implementation recommendations for HIV clinics.

Methods

Study Design
We completed a qualitative study to inform the design and implementation of population-level portal-based depression and anxiety screening at an HIV clinic. This paper reports the results of interviews conducted with clinicians and clinical staff.

Study Setting
The study was conducted in the Ryan White HIV Care Clinic at an academic medical center on the South Side of Chicago, the main provider of HIV care services for Chicago South Side residents. The South Side of Chicago is one of the communities most impacted by the HIV epidemic in the United States [24]. The clinic provides care for >630 people with HIV, most of whom are African Americans and publicly insured. Currently, staffed with 15 physicians, 6 fellows, a nurse practitioner, 2 pharmacists, and 2 licensed social workers, the clinic also provides mental health services.

In November 2020, the HIV clinic adopted a protocol for conducting depression and anxiety assessments during in-person clinic visits. Medical assistants were asked to complete the 2-item Patient Health Questionnaire (PHQ) and the 2-item Generalized Anxiety Disorder (GAD) scale with patients due for annual screening, as indicated by health maintenance topics and best practice advisories in the electronic health record (EHR) [25,26]. Scores of ≥3 were reflexed into the PHQ-9 and GAD-7, respectively. Medical assistants were also asked to complete the PHQ-9 and GAD-7 with patients with a history of depression or anxiety, respectively, who were due for ongoing symptom monitoring or surveillance. Physicians and advanced practice nurses were alerted via a critical, noninterruptive best practice advisory to scores of ≥3.

Concurrent with this study, a population-level portal-based depression screening intervention was tested in the primary care clinic at the academic medical center. Patients were invited to complete depression screening using the patient portal regardless of having a scheduled appointment [21]. The clinic saw an increase in screening and identification of depression [21]. These advances in screening in the primary care clinic at the institution motivated us to gauge the interest in and feasibility of integrating population-level portal-based screening in the HIV clinic.
Study Participants
All clinicians at the HIV clinic were eligible for study enrollment, including physicians, advanced practice nurses, pharmacists, nurses, and social workers. HIV clinicians were informed about the study at a clinic meeting and by email, and if interested in participating, they were instructed to contact the study project manager. The participants verbally consented before each interview.

Data Collection
Semistructured interviews were conducted one-on-one with each participant from January to April 2021, after the newly adopted in-clinic screening protocol was implemented. Demographic information was collected via electronic surveys in the REDCap (Research Electronic Data Capture; Vanderbilt University). Interviews were conducted over Zoom (Zoom Video Communications) and lasted for approximately 60 minutes. The interview questions were created by the research team using the CFIR interview guide tool for all 5 CFIR domains. The full interview guide is available in Multimedia Appendix 1. The following are example questions by domain:

- **Innovation characteristics**: “Do you think assessing anxiety and depression using the patient portal will be effective? Why or why not?”
- **Inner setting**: “What is the general level of receptivity in the clinic to using the patient portal?”
- **Outer setting**: “Do you think measuring anxiety and depression using the patient portal will meet the needs of the patients served by your clinic? Why or why not?”
- **Process**: “Who are other key influential individuals to get on board with assessing anxiety and depressive symptoms using the portal?”

Data Analysis
Descriptive analysis was used to summarize the characteristics of the participants interviewed. Initially, the original CFIR domains and constructs from the codebook were used for our interview analysis. During the initial coding phase, research team members identified additional themes and subthemes to be added to the interview analysis. Once consensus was achieved on the codebook for our analysis, 2 independent coders analyzed each interview transcript, and coding discrepancies were discussed until a consensus was reached. Analysis of coded transcripts was performed in the web-based software Dedoose (version 9.0.17; SocioCultural Research Consultants, LLC). A total of 2 research team members independently reviewed the coded excerpts to find common themes within each domain, identified each as a facilitator or a barrier, and discussed them to consensus. On the basis of these facilitators and barriers, the 2 research team members proposed implementation strategies and presented these strategies to the entire study team for validation and refinement.

Ethical Considerations
The study was reviewed and approved by the University of Chicago Biological Sciences Division Institutional Review Board (20-1313). The research team obtained oral consent from participants before the beginning of the interviews. Interview audio was recorded, and transcripts were deidentified before qualitative coding. The data were accessible to the research team only. Participants were given a US $40 e-gift card for interview completion.

Results
Participant Characteristics
Interviews with HIV clinicians continued until the team agreed that data saturation was met, as indicated by the lack of new themes emerging in the interviews. In total, 10 HIV clinicians completed the interviews. The participants ranged in age from 31 to 64 years. Most participants were identified as White (8/10, 80%) or male (6/10, 60%). As shown in Table 1, 70% (7/10) worked as physicians at the HIV clinic, and the remaining staff included a social worker (1/10, 10%), a nurse (1/10, 10%), and a pharmacist (1/10, 10%).
Table 1. Demographic information of interview participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants (N=10), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
</tr>
<tr>
<td>30-39</td>
<td>5 (50)</td>
</tr>
<tr>
<td>40-49</td>
<td>2 (20)</td>
</tr>
<tr>
<td>50-59</td>
<td>2 (20)</td>
</tr>
<tr>
<td>60-69</td>
<td>1 (10)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Female</td>
<td>4 (40)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1 (10)</td>
</tr>
<tr>
<td>White</td>
<td>8 (80)</td>
</tr>
<tr>
<td><strong>Clinical role</strong></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Social worker</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Nurse</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>1 (10)</td>
</tr>
<tr>
<td><strong>Caring for people with HIV (y)</strong></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>3 (30)</td>
</tr>
<tr>
<td>6-10</td>
<td>2 (20)</td>
</tr>
<tr>
<td>11-15</td>
<td>2 (20)</td>
</tr>
<tr>
<td>≥15</td>
<td>3 (30)</td>
</tr>
<tr>
<td><strong>Clinical experience (y)</strong></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>5 (50)</td>
</tr>
<tr>
<td>6-10</td>
<td>2 (20)</td>
</tr>
<tr>
<td>≥15</td>
<td>3 (30)</td>
</tr>
</tbody>
</table>

**Current Depression and Anxiety Screening Practice**

When asked about their current mental health screening practices, most participants mentioned informally screening patients by asking how they were feeling or if the patient was experiencing any thoughts of self-harm or suicidal ideation. As 1 participant said, “Within the review of systems [during patient intake], I often will ask if any depression and anxiety-type symptoms [were experienced] recently, but there is no standard way I approach every patient” (Participant 6, physician). Participants reported that patients expressing depression or anxiety would typically be screened with the PHQ and the GAD questionnaire. A few participants spoke of the newly implemented in-clinic mental health screening procedure put in effect before the interviews were conducted. However, they stated that the screening protocol was not regularly followed during the clinic visits.

Similarly, the participants mentioned that initiating mental health services relied on patients requesting services or bringing up suicidal ideation or self-harm. The clinic relies on HIV-trained clinical social workers to connect patients with mental health resources based on the individual’s insurance. One participant explained, “I’ll have our social worker call them and set them up with a resource and have her—obviously she’s more trained in that than I am, I believe—and have her assess them and provide appropriate resources” (Participant 9, physician). Before the COVID-19 pandemic, patients were regularly introduced to social workers during in-clinic appointments. The participants emphasized that these in-person interactions were valuable, as they built trust between the patient, the social worker, and the physicians to increase intervention uptake.

**Perceptions of Population-Level Portal-Based Screening**

**Overview**

Facilitators of and barriers to population-level portal-based depression and anxiety screening were identified within the 5 CFIR domains (Tables 2 and 3).
Table 2. Facilitators to population-level portal-based depression and anxiety screening from qualitative interviews with HIV clinicians.

<table>
<thead>
<tr>
<th>Domain and facilitators</th>
<th>Facilitator quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Innovation characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Advantages of systematic screening outside clinic visits</td>
<td>“I like the idea that it feels like it would make it a little less stressful on the clinic visit because it’s done independently of that. And then it might come up during the visit, but at least the initial screening and questions doesn’t get added to that clinic workflow. So from that standpoint, it’s a little bit relieving. Because we do want to do this, but you don’t want to have anyone having too much work or too many things on their plate. So from that standpoint, it feels like a good method to go about it.” (Participant 2, physician)</td>
</tr>
<tr>
<td>Expectation that assessment frequency could be tailored to patient needs</td>
<td>“I think if we did it this way, then we would have the information at the beginning of a visit, and could then walk into the visit knowing this. And maybe even have some additional background from our social worker, if they’ve reached out to them in the meantime, between the time they filled this out and got these results, and then the time we see it for an appointment.” (Participant 8, physician)</td>
</tr>
<tr>
<td><strong>Outer setting</strong></td>
<td></td>
</tr>
<tr>
<td>Greater respect for patient privacy</td>
<td>“I kind of think that’s where you get the most honest answers, in the patient’s environment. In clinic, the patient’s mental status is, I’m ready for clinic. So they have that person put on, their clinic person. And unless something is like really, really outstanding, they’re not forthcoming with their information, right?” (Participant 4, nurse)</td>
</tr>
<tr>
<td>Normalizing mental health screening</td>
<td>“…what I’m thinking is I really liked the way it is being approached as making it a routine part of HIV [care]...just destigmatizing and routinizing those questions for people. I think once that becomes routine as part of your, whatever, yearly check-in, I think that’s helpful.” (Participant 3, social worker)</td>
</tr>
<tr>
<td><strong>Inner setting</strong></td>
<td></td>
</tr>
<tr>
<td>Compatible with clinic culture, workflows, and systems</td>
<td>“Overall positive feelings towards it. I like the idea that it feels like it would make it a little less stressful on the clinic visit because it’s done independently of that. And then it might come up during the visit, but at least the initial screening and questions doesn’t get added to that clinic workflow. So from that standpoint, it’s a little bit relieving. Because we do want to do this, but you don’t want to have anyone having too much work or too many things on their plate.” (Participant 2, physician)</td>
</tr>
<tr>
<td>Protocol for addressing positive screening results</td>
<td>“Would be nice to have a pathway that’s somewhat predetermined. So, it’s like, ‘Okay, we identified this patient has this. We’re not sure if they’re going to be able to see a mental health provider because it might take 2 months to get in.... But in the meantime, this is the plan. This is our protocol for what we should do. These are first-line medications. This is the plan from our social work standpoint of how we’re going to follow up with them.’ So, things like that. That would make it easier once we do identify the need to take some of the guesswork out of what the next steps are.” (Participant 8, physician)</td>
</tr>
<tr>
<td><strong>Characteristics of individuals</strong></td>
<td></td>
</tr>
<tr>
<td>Participant beliefs about the importance of mental health screening and benefits HIV care</td>
<td>“Some of it could be a little bit more work within the appointment if you’re then talking about some of these issues and how they affect their other medical care, but I think it would be time likely worth spent and gratifying, and probably maybe more time spent in the front would help decrease time later needing to it if it were something that could be addressed and then would improve compliance, that would be very meaningful and worth discussion.” (Participant 6, physician)</td>
</tr>
<tr>
<td>Participant interest in evidence-based practices and desire to learn from prior implementation</td>
<td>“I think there’s a strong need to do it in general.... I haven’t read all the literature on it.... Most likely this needs to happen. We need to screen people. And then the question is just, ‘What’s the best way to screen?’ And looking at everything and talking through the pros and cons, it feels like this would probably be a good way to do it.” (Participant 2, physician)</td>
</tr>
<tr>
<td><strong>Process</strong></td>
<td></td>
</tr>
<tr>
<td>Team-based approach that leverages strengths of all clinic staff</td>
<td>“Our patient population is a bit delicate, which is why we have different levels to our team approach, because what patients wouldn’t share with their doctor they will share with me, because they easily identify with me. So they accept it on an extended family member kind of like basis. So their level of trust is greater. And we use that. It’s very effective.” (Participant 4, nurse)</td>
</tr>
<tr>
<td>Clear planning and communication with staff</td>
<td>“In general having something that’s standardized is good. Having something that doesn’t totally disrupt the workflow in clinic. So using the patient portal is excellent. And having a really clear plan for what the follow-up is for the patient. I think those are the really important things. And if those are well communicated to the clinic, to the section beforehand.... We have our Monday meetings at noon, something like that.... So that way everyone’s comfortable, I would be comfortable going forward with something like this, but making sure that everyone’s on the same page.” (Participant 2, physician)</td>
</tr>
<tr>
<td>System that empowers patients to communicate about their mental health</td>
<td>“But most of our patient population is a very secretive population. So I believe being able to have something on their own terms...[The social workers could] be like, ‘Hey, if you ever feel A, B, and C.... Hey just answer these questions. I get an alert and I will respond or someone will respond in a reasonable timeframe.’ Yeah. So if patients have the information that you can use MyChart to let us know if something is going on, I think that would be more successful than just screening patients as they check-in in clinic.” (Participant 4, nurse)</td>
</tr>
</tbody>
</table>
Facilitator: Expectation That Assessment Frequency Could be Tailored to Patient Needs

Participants generally thought that sending assessments via the portal every 6 months or once a year would be ideal; however, they mentioned that they would defer to the evidence on screening frequency best practices. The ability to send patients mental health screeners at a custom interval appealed to participants as it would keep them aware of the mental health concerns that arose. The participants also expressed interest in tailoring the screening frequency based on symptom severity.

Barrier: Difficulty Ensuring Prompt Response to Those in Imminent Risk to Themselves or Others

A common concern raised during the interviews was the complexity of responding to patients who indicated suicidal risk, self-harm, or homicidal ideation on the screener.
Participants were worried that patients with immediate mental health needs would not receive timely interventions if screening were performed via the portal.

**CFIR Domain 2: Outer Setting**

Codes within the outer setting domain focused on external factors that might affect the implementation of population-level portal-based screening, particularly the needs and resources of the patient population served by the clinic.

**Facilitator: Greater Respect for Patient Privacy**

Participants spoke about how patients seen at the HIV clinic value privacy. They thought that using the portal might increase screening uptake and encourage honest responses by allowing patients the flexibility to complete screening in environments where they are most comfortable.

**Barrier: Limited Patient Access, Experience, and Comfort Using the Portal**

The participants did not know if patients were familiar enough with the portal to complete the assessments electronically. The participants reported that several of their patients did not know how to access their laboratory work via the portal. Therefore, they were not confident that the patients could complete screeners via the portal without assistance or training. The participants were also concerned about usability issues regarding the small text and reading levels associated with using the portal.

**Barrier: Limited Availability of Mental Health Services**

Participants emphasized that accessible mental health referral pathways and resources were needed before the clinic could implement population-level portal-based screening. Otherwise, patients would be diagnosed without the proper resources to be treated. Although the clinic has existing partnerships with external mental health facilities, waitlists were long. Furthermore, transportation, insurance, and cost barriers limited patients’ access to mental health treatments. In addition, concerns about the capacity of the current referral network to handle an influx of newly diagnosed patients were expressed by participants.

**CFIR Domain 3: Inner Setting**

Codes within the inner setting domain focused on the clinic’s characteristics and readiness to implement population-level portal-based screening. Participants spoke about compatibility, available resources, access to knowledge and information, networks and communication, and culture.

**Facilitator: Compatibility With Clinic Culture, Workflows, and Systems**

Participants strongly expressed interest in implementing population-level portal-based screening for depression and anxiety through the portal and thought their colleagues would also be receptive. Participants stated that this would help create an open relationship with patients while prioritizing clinical values to provide holistic care to their patients. Participants said that using the portal for depression and anxiety screening would provide crucial information on patients’ mental health status without adding significant stress to clinical workflows.

**Facilitator: Protocol for Addressing Positive Screening Results**

Participants wanted a systematic process to manage patients who screened positive to avoid delays in connecting patients to resources. Specifically, participants expressed a desire for detailed guidance on available resources, referral pathways, and a follow-up plan for symptomatic patients. A few participants were interested in additional training or decision support tools to help them interpret screening results, connect patients to resources, and prescribe first-line medications.

**Barrier: Clinician Variation in the Use of EHRs**

Participants expressed concern that screening results in the EHR might be overlooked because clinicians did not always check their electronic in-baskets reliably because of variations in their proficiency and comfort with the EHR.

**Barrier: Limited Capacity to Address Mental Health Concerns During HIV Visits**

Participants raised concerns about having adequate personnel, time, and expertise to manage depression and anxiety. They reported limited time during appointments to address their patients’ health issues and social needs, and there may not be enough time to address depression and anxiety management.

**CFIR Domain 4: Characteristics of Individuals**

Codes within the individual characteristics domain focused on the participants’ knowledge, beliefs, and self-efficacy.

**Facilitator: Participant Beliefs That Mental Health Screening Is Important and Benefits HIV Care**

Most participants agreed that mental health screening was essential and valuable. They saw population-level portal-based screening as an opportunity to learn more about their patients and to address concerns that might not otherwise arise during appointments. Participants recognized the effects of mental health problems on HIV outcomes and were hopeful that addressing depression and anxiety would improve engagement with care and general health.

**Facilitator: Participant Interest in Evidence-Based Practices and Desire to Learn From Prior Implementations**

Participants indicated their willingness to abide by evidence-based mental health screening and management recommendations. They expressed a desire to learn more about how population-level portal-based depression and anxiety screening had been implemented in the primary care clinic so that the lessons learned could be applied to the HIV clinic.
Barrier: Participant Concerns About Limited Knowledge About Mental Health Treatments

Some participants were hesitant to implement population-level portal-based mental health screening because they thought that they lacked adequate expertise in mental health treatment and navigating mental health resources.

CFIR Domain 5: Process

Codes within the process domain focused on planning the intervention and engaging clinicians and patients.

Facilitator: Team-Based Approach That Leverages the Strengths of All Clinicians

Participants believed that a team approach would be crucial for successfully implementing population-level portal-based screening. Social workers were identified as key team members to provide knowledge on available resources and support connecting patients to care. Participants also indicated that engaging clinicians with strong relationships with patients would help lower patient hesitancy to engage with the portal.

Facilitator: Clear Planning and Communication With Staff

Participants emphasized the importance of having a standardized protocol that included details on which staff member was responsible for each step, especially in response to positive results, and training for all clinic personnel on this protocol before implementation. Participants highlighted the need for clear communication throughout the intervention’s preimplementation, implementation, and sustainability phases. They advised monitoring the intervention logistics and collecting iterative feedback from staff and patients throughout the intervention rollout.

Facilitator: A System That Empowers Patients to Communicate About Their Mental Health

Participants thought that population-level portal-based mental health screening could prompt patients to discuss their mental health with their care team. The participants believed that providing patients with the flexibility to complete screening assessments at their convenience and through their preferred screening method would empower them to inform their care team about their symptoms. Some suggested that screeners should always be readily available in the portal so patients could report their mental health symptoms as they feel them.

Barrier: Impersonal Approach to the Sensitive Topic of Mental Health

Participants expressed concerns that portal-based screening might be impersonal and that unexpected messages about mental health might seem invasive or cause anxiety in some patients. The participants emphasized that clear and patient-centered conversations would need to occur to explain the purpose of mental health screening. Otherwise, the participants feared that patients who did not understand the purpose or context of mental health screening would be unlikely to respond. The participants believed that if patients were informed about how these assessments pertained to their general health, they would be more likely to complete the screeners.

Discussion

Principal Findings

This qualitative study explored facilitators of and barriers to implementing population-level portal-based depression and anxiety screening in an HIV clinic. A total of 10 facilitators and 7 barriers were identified across 5 CFIR domains. Facilitators included the following: advantages of systematic screening outside clinic visits; the expectation that assessment frequency could be tailored to patient needs; evidence from the literature and previous experience in other settings; respect for patient privacy; empowering patients and facilitating communication about mental health; compatibility with clinic culture, workflows, and systems; staff beliefs about the importance of mental health screening and benefits for HIV care; engaging all clinic staff and leveraging their strengths; and clear planning and communication with staff. Barriers included difficulty in ensuring prompt response to suicidal ideation; patient access, experience, and comfort using the portal; limited availability of mental health services; variations in how providers use the EHR and communicate with patients; limited capacity to address mental health concerns during HIV visits; staff knowledge and self-efficacy regarding the management of mental health conditions; and the impersonal approach to a sensitive topic. Several barriers mentioned by participants, such as limited appointment times and limited access to mental health resources after diagnosis, are common challenges cited in similar implementation efforts [2,3,6,7,25,27].

Findings from our analysis have been used to compile a list of proposed implementation strategies to help integrate population-level portal-based depression and anxiety screening into practice within the HIV clinic setting.

Clinician-Focused Implementation Strategies

Strategy 1: Conduct a Local Assessment of Clinicians and Clinic Staff

Clinicians and clinic staff are essential to successfully implementing population-level portal-based depression and anxiety screening in the HIV clinic. To increase the feasibility and sustainability of the intervention, clinicians and clinic staff should be asked how the intervention would fit with their beliefs and values, the clinic culture, and its current clinical workflows. Clinicians’ and clinic staff’s thoughts should be incorporated into the implementation plan to assist in intervention compatibility and uptake.

Strategy 2: Engage Clinicians and Clinic Staff With Various Roles and Expertise to Support Implementation

The success of the intervention depends on clinicians’ engagement through the implementation process to inform the intervention using clinicians and clinic staff’s strengths. Clinicians and clinic staff in various roles may have different perspectives and ideas on implementing the intervention. Therefore, diversifying the staff perspective may provide crucial implementation strategies that might not be known by only interviewing clinicians.
Strategy 3: Highlight Advantages, Relevance, and Evidence for Population-Level Portal-Based Mental Health Screening

Before implementing population-level portal-based mental health screening, the advantages of depression and anxiety screening must be communicated and emphasized to all clinicians. These benefits should highlight how timely mental health discussions between patients and clinicians make efficient use of the limited appointment time. Information on relevant evidence and current clinical screening guidelines should also be provided to garner clinic support. Describing barriers encountered and lessons learned in other practices that have implemented similar interventions could ease concerns about implementation challenges.

Strategy 4: Communicate With Clinicians and Clinic Staff Throughout Implementation and Provide Training on Protocols

Training and involving clinicians throughout the rollout of the intervention will facilitate iterative feedback to troubleshoot any challenges that arise and help aid clinicians and clinic staff uptake. As clinicians and clinic staff tend to have established relationships with their patients, receiving their and their patients’ concerns will aid clinicians, clinic staff, and patient engagement throughout the intervention rollout.

Patient-Focused Implementation Strategies

Strategy 5: Provide Technical Support and Demonstrations for Patients on How to Use the Portal

Providing technical support and conducting training on using the portal might increase intervention uptake among patients. Demonstrations could decrease technology-related barriers and encourage patients to use the portal to complete assessments.

Strategy 6: Use Multiple Screening Methods for Broad Reach

Multiple screening approaches might be needed to reach all patients attending the clinic. For example, options could include completing depression and anxiety screening in the waiting room before an appointment, over the phone, or during an in-person appointment (eg, during intake before the clinician enters the room). Providing additional screening options for patients who are not technologically proficient or have limited access to technology may increase patient uptake of depression and anxiety screening.

Strategy 7: Use Patient-Centered Communication in Portal Messages

Patient-centered messages emphasizing privacy and framing mental health screening as part of routine care can provide a context for portal-based screeners and decrease patient hesitancy to answer questions about the potentially sensitive and stigmatized topic of mental health. Using the patient portal to send patient-centered messages will also allow patients to ask questions about population-level patient-based screening and address concerns.

IT-Focused Implementation Strategies

Strategy 8: Make Screening Frequency Adaptable Based on Patient History and Symptoms

Adaptability of screening frequency and leveraging the staff-patient relationship may improve intervention uptake. Clinicians could adjust the frequency of depression and anxiety screenings based on their relationship with the patient. By allowing staff to adjust the screening frequency, the clinic can check in on patients experiencing uncontrolled depression and anxiety symptoms. Likewise, the staff can lengthen the screening intervals when the patient is in remission for depression and anxiety. This adaptability will signal to patients that the clinic is prioritizing the patient’s health needs.

Strategy 9: Use User-Centered Design Methods to Refine How Results Are Displayed and Communicated in the EHR

When designing how portal-based results will be stored and displayed in the EHR, clinicians and clinic staff need to be engaged to ensure the utility of the screening information. Using a user-centered design with these essential stakeholders could increase the likelihood that portal-based depression and anxiety screening will be used in practice.

Strategy 10: Make Screening Tools Available for Patients to Use on Demand in the Portal

On-demand assessments would support patient autonomy and allow patients to signal when they are experiencing depression and anxiety symptoms. This patient-centered approach could enhance the clinic’s capacity to treat patients when needed. This differs from the traditional annual one-time screening, which aims to identify depression and anxiety in asymptomatic patients. Traditional screening may increase the demand for services and reduce the clinic’s ability to provide timely and appropriate care for symptomatic patients. Moreover, prioritizing on-demand assessments outside of appointment times could facilitate outreach between appointments and reduce the time to treatment.

Clinic-Focused Implementation Strategies

Strategy 11: Create Protocols for Positive Depression and Anxiety Screening Results, Including Those Indicating Imminent Risk

Establishing a standardized protocol for patients who are symptomatic of depression or anxiety may ease concerns about managing patients who are at imminent risk to themselves or others. For example, the protocol can describe who will contact the patient after the clinic has received a positive PHQ or GAD and how to assist patients in crisis. This will ease clinicians’ concerns about screening patients for depression and anxiety via the portal.

Strategy 12: Provide Clinical Decision Support Tools, Training, and Mentorship to Help Clinicians Manage Mental Health Concerns

Providing evidence-based information on treatment or referral strategies through decision support tools, ongoing training, and
clinician mentorship for managing mental health treatment would support clinicians’ confidence and ability to manage symptomatic mental health concerns in their patients. Through shared collaborations with mental health specialists and community mental health services, clinicians will be equipped to manage a potential influx of symptomatic patients via the portal.

**Strategy 13: Implement Integrated Behavioral Health and Increase Mental Health Referral Partnerships**

In concurrence with strategy 12, the clinic will need to invest and establish partnerships with local mental health sites to support the clinic’s capacity to treat newly diagnosed patients. Expanding the clinic’s referral network would create a safety net that the clinic can leverage to refer patients. This would prevent long wait times for treatment after a depression or anxiety diagnosis. By creating a behavioral health referral network, HIV clinicians can provide trusted resources to expand the clinic’s internal infrastructure, facilitate warm handoffs with community partners, and continue to support patient care.

**Limitations**

The study was conducted at a single academic HIV clinic. Therefore, the results of this qualitative analysis may not be generalizable to other HIV clinics with different patient populations, staffing needs, available resources, and portal uptake. The implementation strategies are merely recommendations from a single HIV clinic and may need to be adapted to fit the implementation setting. At the time of the interviews, clinic staff did not have experience with population-level portal-based mental health screening; therefore, their perspectives were based on how they perceived the intervention would be for patients and themselves once implemented. Although our study included perspectives from clinicians in various clinical roles, most interviewees were physicians, limiting available insight. Gaining patient perspectives through patient-focused interviews would provide further insight into facilitators, barriers, and intervention implementation strategies.

**Conclusions**

Our study provides information on clinicians’ views on population-level portal-based mental health screening within the HIV clinic setting. Participating clinicians expressed concerns about the accessibility of prompt mental health resources, patients’ perceptions of mental health screening, variation in clinician use of Epic (Epic Systems, Verona Wi), and limited clinician training on mental health management. Nevertheless, clinicians were interested in establishing population-level portal-based screening at the HIV clinic and were amenable to creating protocols for addressing positive mental health screening, to participating in training about available mental health resources and best practices, and to feeling it was compatible with the clinic. Others may build upon this work by exploring and identifying additional facilitators, barriers, and implementation strategies that were not found in our analysis.

**Acknowledgments**

This publication was made possible with support from the Third Coast Center for AIDS Research, a National Institutes of Health–funded center (P30 AI117943).

**Authors’ Contributions**

EMS, JPR, JS, MF, SJH, DM, and NL contributed to study concept and design. Data acquisition was performed by MF. Data analysis and interpretation were performed by DZ, EMS, and NL. DZ wrote the initial manuscript draft. Critical revision of the manuscript for intellectual content was performed by all the authors.

**Conflicts of Interest**

JPR has received fees for legal consulting from Gilead Sciences.

Multimedia Appendix 1

Interview guide.

[DOCX File, 123 KB - formative_v8i1e48935_app1.docx ]

**References**


Abbreviations

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
GAD: Generalized Anxiety Disorder
PHQ: Patient Health Questionnaire
REDCap: Research Electronic Data Capture
Original Paper

Dentists’ Information Needs and Opinions on Accessing Patient Information via Health Information Exchange: Survey Study

Shuning Li¹, PhD; Grace Gomez Felix Gomez¹,², BDS, PhD; Huiping Xu³, PhD; Anushri Singh Rajapuri¹, BDS, MS; Brian E Dixon²,⁴, PhD; Thankam Thyvalikakath¹,², DMD, MDS, PhD

¹Department of Dental Public Health and Dental Informatics, Indiana University School of Dentistry, Indianapolis, IN, United States
²Regenstrief Institute, Inc, Center for Biomedical Informatics, Indianapolis, IN, United States
³Department of Biostatistics and Health Data Sciences, Indiana University School of Medicine, Indianapolis, IN, United States
⁴Department of Epidemiology, Indiana University Richard M Fairbanks School of Public Health, Indianapolis, IN, United States

Corresponding Author:
Thankam Thyvalikakath, DMD, MDS, PhD
Department of Dental Public Health and Dental Informatics
Indiana University School of Dentistry
415 Lansing St
Indianapolis, IN, 46201
United States
Phone: 1 3172745460
Email: tpt@iu.edu

Abstract

Background: The integration of medical and dental records is gaining significance over the past 2 decades. However, few studies have evaluated the opinions of practicing dentists on patient medical histories. Questions remain on dentists’ information needs; their perception of the reliability of patient-reported medical history; satisfaction with the available information and the methods to gather this information; and their attitudes to other options, such as a health information exchange (HIE) network, to collect patient medical history.

Objective: This study aims to determine Indiana dentists’ information needs regarding patients’ medical information and their opinions about accessing it via an HIE.

Methods: We administered a web-based survey to Indiana Dental Association members to assess their current medical information-retrieval approaches, the information critical for dental care, and their willingness to access or share information via an HIE. We used descriptive statistics to summarize survey results and multivariable regression to examine the associations between survey respondents’ characteristics and responses.

Results: Of the 161 respondents (161/2148, 7.5% response rate), 99.5% (n=160) respondents considered patients’ medical histories essential to confirm no contraindications, including allergies or the need for antibiotic prophylaxis during dental care and other adverse drug events. The critical information required were medical conditions or diagnosis, current medications, and allergies, which were gathered from patient reports. Furthermore, 88.2% (n=142) of respondents considered patient-reported histories reliable; however, they experienced challenges obtaining information from patients and physicians. Additionally, 70.2% (n=113) of respondents, especially those who currently access an HIE or electronic health record, were willing to use an HIE to access or share their patient’s information, and 91.3% (n=147) shared varying interests in such a service. However, usability, data accuracy, data safety, and cost are the driving factors in adopting an HIE.

Conclusions: Patients’ medical histories are essential for dentists to optimize dental care, especially for those with chronic conditions. In addition, most dentists are interested in using an HIE to access patient medical histories. The findings from this study can provide an alternative option for improving communications between dental and medical professionals and help the health information technology system or tool developers identify critical requirements for more user-friendly designs.

(JMIR Form Res 2024;8:e51200) doi:10.2196/51200

KEYWORDS
dentistry; medical history; integrated medical and dental records; health information exchange; medical record; dental record; dental; medical information; dental care; adverse drug effect; medication; allergies; cost; data safety; data accuracy
Introduction

More than 20 years ago, the first US Surgeon General’s Report on Oral Health in America established oral health as an essential component of overall health and well-being [1]. In 2021, the National Institutes of Health (NIH) reemphasized the importance of establishing integrated medical and dental care in their updated report on Oral Health in America [2]. In addition, the NIH identified integrating medical and dental records as critical to enhancing medical and dental care [2]. The integration of medical and dental records is gaining significance for several reasons. First, increased evidence during the last 3 decades indicates strong associations and shared risk factors between oral and systemic diseases such as diabetes and heart diseases [2,3]. Second, the siloed systems of dental and medical data create challenges in information sharing [2,4,5], often resulting in incomplete or inaccurate patient medical information, which may cause significant patient care and safety issues in dental care [6-9]. Third, recent studies have revealed discrepancies in medical conditions and medications in the electronic dental record (EDR) versus electronic health record (EHR) [3,6,10-14] and demonstrated substantial delays when dentists are required to request additional medical information from physicians [15].

Other reasons for the increased significance of integrated medical and dental records include the rapid development of information technologies, which provides a solid base for integration, and the impact of the COVID-19 pandemic, which proves the urgency and importance of integrating medical and dental records [16-18].

The use of an integrated EDR-EHR system has been growing in large health care organizations (HCOs) such as the Veteran Affairs health care systems, Department of Defense, health maintenance organizations, and federally qualified health centers, where medical and dental practices are colocated and share patient care and records [19-21]. Numerous studies have reported physicians using integrated EDR-EHR systems to refer patients to dentists and vice versa for preventive and comprehensive care [19,21-23]. However, for most dentists who work in small independent practices, patients continue to be the primary source of their medical history and dentists’ communications with medical providers are limited [24]. Nevertheless, the solo and small-group dental practices, which constitute 50% of the dental workforce [25], cannot adopt such integrated systems without being credentialed to a major HCO. It is also not practical for dental practices to have separate interfaces to different EHR systems, which may interfere with their clinical workflows and business processes, such as billings and regulatory policies [3].

With the support of several federal policies, such as the Health Information Technology for Economic and Clinical Health Act and the 21st Century Cures Act [26,27], and the financial incentives established by the Centers for Medicare and Medicaid Services [28,29], community and regional health information exchanges (HIEs) have expanded significantly since 2009. HIE systems provide another option for integrating medical and dental records [24]. In an earlier study by the team, we modeled 3 methods for dentists to access their patient’s medical histories: the patient-reported medical history followed by the optional medical consults method, the integrated EDR-EHR, and the HIE approaches [24]. Our models showed that the HIE approach could provide benefits for reducing unnecessary medical consults, avoiding the delay of care, improving information quality, and cutting additional technical and financial overheads for small independent practices. In addition, a report published in 2021 indicated a decrease in dentists working solo [25]. Nevertheless, an HIE-based integrated solution can help small and large group practices improve data completeness and compliance by obtaining data from multiple HCOs and taking advantage of expert services provided by an HIE. However, efforts to connect dentists with an HIE are minimal compared to the extensive studies on integrated EDR-EHR [19-22].

Despite the widespread interest in integrating dental and medical care, few studies have evaluated the opinions of practicing dentists on patient medical histories [26,30]. For example, a recent study published by the American Dental Association Clinical Evaluators Panel reported that most dentists gathered their patients’ medical history and medication list via patients’ self-report and recorded vital signs during dental visits [30]. However, questions remain on dentists’ information needs; their perception of the reliability of patient-reported medical history; satisfaction with the available information and the methods to gather this information; and their attitudes to other options, such as an HIE, to collect patient medical history.

Given this knowledge gap, we surveyed dentists in 1 US state about their information needs and practices concerning retrieving patient medical history information. Our objectives were to determine their information needs regarding patients’ medical information and their opinions on accessing it via an HIE.

Methods

Recruitment

We administered a web-based survey to the Indiana Dental Association (IDA) members from March 19, 2021, to April 30, 2021. All participants are general dentists or specialists either currently or previously practicing in the State of Indiana. We only included dentists in this survey because they are responsible for diagnosing and planning treatments, which also involve ruling out contraindications. The survey was administrated through the Indiana University–approved Qualtrics Experience Management platform XM. We sent emails to 2148 IDA members over 6 weeks, including 1 initial invitation, 5 reminders, and 1 final thank you note.

Ethical Considerations

Participation in the survey was voluntary, responses were anonymous, and participants could only respond once (configurations blocked multiple responses in the web-based survey tool). Participants gave informed consent by accessing the link provided in the study invitation email. The patients were not compensated. This study received exemption approval from the Indiana University Institutional Review Board (Protocol #2012972646).
Survey Construction and Validation

The survey included 27 questions covering 3 topics: 12 on demographics, 11 on information needs and gathering, and 4 on exchanging patient medical information (Multimedia Appendix 1). Demographic information included sex, years in practice, primary practice information (type of practice, general practitioner or specialist, typical procedures, patient age groups), and EDR use. The information needs and data gathering section included questions related to dentists’ information needs, existing methods for collecting information, and challenges in these approaches. Finally, the exchange of patient medical information section included dentists’ opinions on using an HIE-based information platform to receive and share patients’ medical information. The survey had 23 multiple-choice questions, two 0-10 Likert-scale questions, and 2 open-ended questions. We administered the survey after assessing the face validity and content validity of the questionnaire. The face validity was assessed with research team members who were not involved in the development of the survey questions, and the content validity was assessed with 3 dentists—2 from the Indiana University School of Dentistry and 1 from private practice. These tests ensured that the survey was appropriate, understandable, and could be completed within a reasonable time.

Statistical Analysis

Data analysis included only completed responses. Partially answered surveys were eliminated from the final analysis. Data were summarized using frequencies and percentages for categorical variables and mean and SDs for continuous variables. Associations between characteristics of dentists (years in practice, dental professions, and current access to an HIE or a hospital or medical practice–based EHR [hereby referred to as HIE-EHR]) and their opinions on the importance and reliability of patients’ medical histories and perceptions of accessing patients’ medical histories via an HIE were examined using multivariable regression. The ordinal logistic regression model was used due to the ordinal nature of these variables. All statistical analyses were performed using SAS (version 9.4, The SAS Institute). \( P < .05 \) were considered statistically significant.

Results

At the end of 6 weeks, 219 (10.2%) out of 2148 IDA members accessed the survey, and 188 (8.8%) members responded to at least 1 question, of which 161 (7.5%) members reached the end of the survey.

Demographics

A total of 64.6% (102/158) of the respondents were male (Table 1). Their average years in practice was 25.72 (SD 13.52) years. A total of 74.5% (120/161) of the respondents were general practitioners, and the rest were dental specialists (Table 1). A total of 8 dental specialties were reported: oral and maxillofacial surgery (10/41, 24.2%), periodontics (8/41, 19%), orthodontics (7/41, 16.7%), pediatric dentistry (7/41, 16.7%), endodontics (4/41, 11.9%), operative dentistry (2/41, 4.8%), prosthodontics (2/41, 4.8%), and oral and maxillofacial pathology (1/41, 2.4%). The total percentage is more than 100% since some respondents reported more than 1 specialty.
Table 1. Characteristics of survey respondents.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (n=158), n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>56 (35.4)</td>
</tr>
<tr>
<td>Male</td>
<td>102 (64.6)</td>
</tr>
<tr>
<td>Dental profession (n=161), n (%)</td>
<td></td>
</tr>
<tr>
<td>General practitioner</td>
<td>120 (74.5)</td>
</tr>
<tr>
<td>Dental specialist</td>
<td>41 (25.5)</td>
</tr>
<tr>
<td>Years in practice (n=161), mean (SD)</td>
<td>25.72 (13.52)</td>
</tr>
<tr>
<td>Survey respondents’ primary dental practices</td>
<td></td>
</tr>
<tr>
<td>Type (n=160), n (%)</td>
<td></td>
</tr>
<tr>
<td>Private dental practice owner</td>
<td>100 (62.5)</td>
</tr>
<tr>
<td>Associate dentist of a private practice</td>
<td>26 (16.3)</td>
</tr>
<tr>
<td>Major dental care organizations</td>
<td>19 (11.9)</td>
</tr>
<tr>
<td>Public health practice, community health center, or</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>publicity-funded clinic (but not a federal facility)</td>
<td></td>
</tr>
<tr>
<td>Federal government facility (Veterans Affairs,</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Department of Defense, and Public Health Service)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>11 (6.9)</td>
</tr>
<tr>
<td>Number of dentists (including the respondent; n=160), n (%)</td>
<td>75 (46.9)</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2-5</td>
<td>66 (41.3)</td>
</tr>
<tr>
<td>6-10</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>16 (10)</td>
</tr>
<tr>
<td>Number of hygienists (n=161), n (%)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>31 (19.3)</td>
</tr>
<tr>
<td>1-5</td>
<td>110 (68.3)</td>
</tr>
<tr>
<td>6-10</td>
<td>15 (9.3)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>5 (3.1)</td>
</tr>
<tr>
<td>Use EDR(^a) to manage clinical data (n=161), n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>128 (79.5)</td>
</tr>
<tr>
<td>No</td>
<td>33 (20.5)</td>
</tr>
<tr>
<td>EDR system brands (n=126), n (%)</td>
<td></td>
</tr>
<tr>
<td>Dentrix</td>
<td>45 (35.7)</td>
</tr>
<tr>
<td>EagleSoft</td>
<td>19 (15.1)</td>
</tr>
<tr>
<td>axiUm</td>
<td>14 (11.1)</td>
</tr>
<tr>
<td>OpenDental</td>
<td>10 (7.9)</td>
</tr>
<tr>
<td>SoftDent</td>
<td>10 (7.9)</td>
</tr>
<tr>
<td>Practice Works</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Easy Dental</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Other(^b)</td>
<td>19 (15.1)</td>
</tr>
<tr>
<td>Whether or not have access to a state-based health information exchange, exchange capability between dental software and electronic medical record system, or integrated dental-medical record system? (n=161), n (%)</td>
<td>24 (14.9)</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>137 (85.1)</td>
</tr>
</tbody>
</table>
### Characteristics

<table>
<thead>
<tr>
<th>Patient age distribution (%; n=157), mean (SD)</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 years and younger</td>
<td>19.6 (23.3)</td>
</tr>
<tr>
<td>19-44 years</td>
<td>26.1 (13.8)</td>
</tr>
<tr>
<td>45-64 years</td>
<td>31.8 (15)</td>
</tr>
<tr>
<td>65 years and older</td>
<td>22.5 (13.1)</td>
</tr>
</tbody>
</table>

**EDR**: electronic dental record.

**Other EDR brands included**: Ascend (by Dentrix), Cloud9, Curve Dental, Denticon, DRM plus, DSN PerioExec, EPMS, MacPractice, Mconsent, Florida Probe, Mogo, NextGen, OMSvision, Ortho2 Edge, Practice Fusion, and Practice Web.

In all, 78.8% (126/160) of respondents reported working in private practices as owners or associate dentists (Table 1). Approximately half (75/160, 46.9%) of the respondents reported having 1 dentist in their primary dental practices, while 10% (16/160) of respondents reported their primary practice having more than 10 dentists. Most respondents’ (110/160, 68.3%) primary practices had 2-5 hygienists, while 19% (31/160) of respondents’ practices had no hygienists. The 3 most frequently performed procedures were diagnostic and preventive such as an examination, X-rays, scaling, prophylaxis, sealants, fluoride, etc (136/161, 84.4%); restorations or fillings (125/161, 78.1%); and tooth-supported or implant-supported crowns (105/161, 65.3%). The respondents served a diverse age group of the patient population, with an average of 19.6% (31/157) of patients 18 years or younger and 22.5% (35/157) of patients 65 years or older (Table 1).

About 4 in 5 (128/161, 79.5%) respondents reported using an EDR for not only billing or scheduling but also for clinical or patient data management. The top 3 brands of EDR were Dentrix (145/161, 35.7%), EagleSoft (19/161, 15.1%), and axiUm (14/161, 11.1%; Table 1).

### Dentists’ Opinions on the Importance of Medical Histories and Reliability of Patient-Reported Medical Histories

Almost all respondents (160/161, 99.5%) considered patients’ medical histories highly or moderately important during dental care (Table 2). They reviewed medical histories to (1) verify no contraindications exist to undergo a dental procedure (37/161, 23.2%), (2) confirm no need for antibiotic prophylaxis before the dental procedure (36/161, 22.7%), (3) rule out any allergies or adverse drug reactions (35/161, 22%), (4) assist with determining the prognosis of an oral disease or treatment outcomes (35/161, 21.7%), (5) detect normal and abnormal laboratory results (14/161, 8.4%), and (6) for other purposes (3/161, 1.9%). Only 1 respondent considered patients’ medical histories unimportant since they felt gathering medical history to be procedural and not essential for dental care. Regarding the reliability of patient-reported medical histories, 8% (n=13) of respondents considered them highly reliable, 79.5% (n=128) moderately reliable, and 12.4% (n=20) unreliable.
Table 2. Respondents’ opinions on the importance and reliability of patients’ medical histories and their perceptions of accessing patient history via an HIE a (n=161).

<table>
<thead>
<tr>
<th>Opinions to patients’ medical histories b</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How important is obtaining patient’s up-to-date medical history for you?</strong></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>7</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>8</td>
<td>13 (8.1)</td>
</tr>
<tr>
<td>9</td>
<td>18 (11.2)</td>
</tr>
<tr>
<td>10: Extremely important</td>
<td>126 (78.3)</td>
</tr>
<tr>
<td><strong>How reliable is the patient-reported medical history?</strong></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td>3</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>4</td>
<td>4 (2.5)</td>
</tr>
<tr>
<td>5</td>
<td>11 (6.8)</td>
</tr>
<tr>
<td>6</td>
<td>29 (18.0)</td>
</tr>
<tr>
<td>7</td>
<td>57 (35.4)</td>
</tr>
<tr>
<td>8</td>
<td>42 (26.1)</td>
</tr>
<tr>
<td>9</td>
<td>11 (6.8)</td>
</tr>
<tr>
<td>10: Extremely reliable</td>
<td>2 (1.2)</td>
</tr>
</tbody>
</table>

Perceptions of accessing patient history via an HIE

<table>
<thead>
<tr>
<th>Do you think access to such a system would be useful?</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: No</td>
<td>8 (5.0)</td>
</tr>
<tr>
<td>2: Maybe</td>
<td>40 (24.8)</td>
</tr>
<tr>
<td>3: Yes</td>
<td>113 (70.2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Would you consider using it to access your patient’s medical information?</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: No</td>
<td>9 (5.6)</td>
</tr>
<tr>
<td>2: Maybe</td>
<td>40 (24.8)</td>
</tr>
<tr>
<td>3: Yes</td>
<td>112 (69.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Would you allow other health care providers to access clinical information about your own patients?</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: No</td>
<td>15 (9.3)</td>
</tr>
<tr>
<td>2: Maybe</td>
<td>47 (29.2)</td>
</tr>
<tr>
<td>3: Yes</td>
<td>99 (61.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What is your interest to participate in a service to access such as a system?</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Not interested at all</td>
<td>15 (9.3)</td>
</tr>
<tr>
<td>2: Slightly interested</td>
<td>19 (11.8)</td>
</tr>
<tr>
<td>3: Moderately interested</td>
<td>58 (36.0)</td>
</tr>
<tr>
<td>4: Very interested</td>
<td>39 (24.2)</td>
</tr>
<tr>
<td>5: Extremely interested</td>
<td>30 (18.6)</td>
</tr>
</tbody>
</table>

aHIE: health information exchange.

b10-level Likert scale was used with ranges as follows: 1-5=not important or reliable, 6-8=moderately important or reliable, and 9-10=highly important or reliable.
Dentists’ Information Needs Regarding Their Patient’s Medical History

The 3 most needed information categories for new and existing patients were medical conditions or diagnosis, current medications, and allergies. Other categories included hospitalizations in the last 2 years, substance abuse, procedures in the previous 5 years, laboratory results from the last 6 months, and immunization records. The respondents evaluated the information needs of new and existing patients separately, and there were no significant differences in the results (Table 3).

<table>
<thead>
<tr>
<th>Patient medical information</th>
<th>Existing patients (n=161), n (%)</th>
<th>New patients (n=161), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical condition or diagnoses</td>
<td>150 (93.2)</td>
<td>153 (95)</td>
</tr>
<tr>
<td>Current medications</td>
<td>141 (87.6)</td>
<td>148 (91.9)</td>
</tr>
<tr>
<td>Allergies</td>
<td>138 (85.7)</td>
<td>143 (88.8)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>76 (47.2)</td>
<td>81 (50.3)</td>
</tr>
<tr>
<td>Hospitalization in the last 2 years</td>
<td>70 (43.5)</td>
<td>68 (42.2)</td>
</tr>
<tr>
<td>Procedures in the last 5 years</td>
<td>44 (27.3)</td>
<td>49 (30.4)</td>
</tr>
<tr>
<td>Laboratory results from the last 6 months</td>
<td>17 (10.6)</td>
<td>17 (10.6)</td>
</tr>
<tr>
<td>Immunization records</td>
<td>4 (2.5)</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>Others</td>
<td>3 (1.9)</td>
<td>5 (3.1)</td>
</tr>
</tbody>
</table>

Dentists’ Access to Their Patient’s Medical History

We also asked the dentists how they collected patient-reported medical histories and obtained additional information if needed. Paper-based health history forms constituted the most used method (127/161, 78.9%), followed by web-based health history forms (62/161, 38.5%) and electronic devices such as tablets (35/161, 21.7%). The total percentage is more than 100% since some respondents reported using more than 1 method. The top 3 challenges in collecting patient-reported medical history were as follows: (1) patients do not remember or recall medication names and dosage (156/161, 96.9%); (2) patients do not recall previous procedures and medical conditions (129/161, 80.1%); and (3) patients’ reluctance to share their medical history (84/161, 52.2%). When the respondents needed additional information, most (158/161, 98%) contacted physicians’ offices or health care providers directly via phone, fax, or email. Other communication methods included paper-based medical consult forms through the patient (46/161, 28.6%), patient’s pharmacy (39/161, 24.2%), state-based HIE (19/161, 11.8%), exchange capability between dental software and electronic medical record system (5/161, 3.1%), integrated dental-medical record system (4/161, 2.5%), and other (9/161, 5.6%). However, they experienced challenges such as the need for multiple attempts (97/161, 60.2%), not receiving information on time (80/161, 49.7%), physician offices being nonresponsive (66/161, 41.0%), need to contact numerous providers or specialists (55/161, 34.2%), need for patient intervention (44/161, 27.3%), and not receiving requested information (35/161, 21.7%).

Dentists’ Perceptions of Accessing Patient History via an HIE

A total of 69.6% (113/161) of respondents considered access to a regional HIE useful (Table 2). If such a system were available, 69.9% (n=113) of the respondents would consider using it to access their patient’s medical information, and 61.5% (n=99) would be willing to allow other health care providers to access their patients’ clinical information (Table 2). Furthermore, 91.3% (n=147) of the respondents expressed various interests in participating in a service to access an HIE (Table 2). However, they expressed concerns over the design and implementation of such a system, including data accuracy, data security and HIPAA (Health Insurance Portability and Accountability Act) compliance, cost of implementation (both time and money), and system usability.

The association between respondent characteristics (including dental profession, number of years in practice, and current access to an EHR or HIE) and their opinions on the importance and reliability of patients’ medical histories and perceptions of accessing patients’ medical histories via HIE based on multivariable ordinal logistic regression is displayed in Table 4. Dental profession type (general practitioner vs dental specialist) does not significantly affect one’s opinions toward the importance (P=.98) and reliability (P=.31) of patients’ medical history. However, respondents with more than 40 years in practice were less likely to consider obtaining up-to-date patient information important compared to those with less than 40 years in practice (odds ratio [OR] 0.351, 95% CI 0.139-0.889; P=.03) and more likely to think self-reported information to be reliable (OR 2.267, 95% CI 1.011-5.084; P=.047). In addition, respondents with access to an HIE-EHR were more likely to consider obtaining up-to-date patient information important compared to those who do not have access to an HIE-EHR (OR 2.590, 95% CI 1.080-6.209; P=.04). Regarding the respondents’ perceptions of using an HIE to access patients’ medical histories, we found that dental specialists were more interested than general practitioners in participating in service to access patient information via an HIE (OR 2.267, 95% CI 1.174-4.378; P=.02). Compared to respondents without current access to an HIE-EHR, those with access to an HIE-EHR were more likely to think it worthwhile to access such a system (OR 6.306, 95% CI 2.671-14.886; P<.001), more likely to consider using such a system to access their patient’s information (OR 5.538, 95% CI 2.379-12.892; P<.001), more likely to allow other providers to access their patient’s data (OR 2.943, 95% CI 1.342-6.456;


Table 4. Impact of respondents' demographics on their opinions on patient medical history and perceptions on accessing patient medical information via an HIE.²

<table>
<thead>
<tr>
<th>Patient medical information</th>
<th>Dental specialist practitioner OR² (95% CI)</th>
<th>P value</th>
<th>&gt;40 vs ≤40 years in practice OR (95% CI)</th>
<th>P value</th>
<th>Have access to HIE-EHRb vs no access OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>How important is obtaining a patient’s up-to-date medical history for you?</td>
<td>0.988 (0.401-2.437)</td>
<td>.98</td>
<td>0.351 (0.139-0.889)</td>
<td>.03d</td>
<td>2.590 (1.080-6.209)</td>
<td>.03d</td>
</tr>
<tr>
<td>How reliable is the patient-reported medical history?</td>
<td>0.713 (0.374-1.360)</td>
<td>.31</td>
<td>2.267 (1.011-5.084)</td>
<td>.047d</td>
<td>1.135 (0.554-2.327)</td>
<td>.73</td>
</tr>
<tr>
<td>Do you think access to such a system would be useful?</td>
<td>1.567 (0.674-3.643)</td>
<td>.30</td>
<td>2.435 (0.821-7.217)</td>
<td>.11</td>
<td>6.306 (2.671-14.866)</td>
<td>&lt;.001d</td>
</tr>
<tr>
<td>Would you consider using it to access your patient’s medical information?</td>
<td>2.187 (0.908-5.264)</td>
<td>.08</td>
<td>1.577 (0.577-4.309)</td>
<td>.37</td>
<td>5.538 (2.379-12.892)</td>
<td>&lt;.001d</td>
</tr>
<tr>
<td>Would you allow other providers to access clinical information about your own patients?</td>
<td>1.311 (0.623-2.759)</td>
<td>.48</td>
<td>1.517 (0.602-3.825)</td>
<td>.38</td>
<td>2.943 (1.342-6.456)</td>
<td>.007d</td>
</tr>
<tr>
<td>What is your interest in participating in a service to access such a system?</td>
<td>2.267 (1.174-4.378)</td>
<td>.02d</td>
<td>1.609 (0.722-3.585)</td>
<td>.24</td>
<td>3.894 (1.844-8.222)</td>
<td>&lt;.001d</td>
</tr>
</tbody>
</table>

aHIE: health information exchange.  
bEHR: electronic health record.  
cOR: odds ratio.  
dP<.05 were considered statistically significant.

Discussion

Principal Findings

We surveyed Indiana dentists to determine their information needs regarding patients’ medical histories and their opinions of accessing patient-specific medical information via a community or regional HIE. The survey respondents’ demographics distribution closely matched the dentists’ demographics in the 2020 Indiana oral health workforce data report [31]. In addition, the response rate of 7.5% (161/2148) is comparable to previous surveys of health care professionals, especially web-based surveys [32-34]. The results demonstrated dentists’ high priority in obtaining their patients’ medical diagnoses or conditions, medication histories, and allergies to provide optimum dental care. The survey respondents also reported challenges in getting medical information from patients and medical providers, although they considered patient-reported medical histories moderately or highly reliable. It is also significant that 70% (112/160) of surveyed dentists who work primarily in community practices (Table 1) expressed willingness to use and participate in a service to access and share their patients’ medical histories via an HIE.

Nevertheless, the participants commented that usability, data accuracy, data safety, and implementation costs would drive dental providers’ use of such services. Integration of dental and medical record data is critical to promote communication and care coordination between dental and medical providers and has gained tremendous attention in recent years [2]. However, existing studies only highlight case studies of integrating dental and medical care in large health care systems [19-21]. Through this study, we determined community practice dentists’ information needs and attitudes toward accessing patient medical information via an HIE. These study results contribute to dental professionals’ high-priority information needs and HIE functionalities for successfully using the expanding HIE network in the United States and other countries. In the sections below, we discuss the relevant findings in detail.

Dentists with <40 years of experience or having access to an HIE-EHR system felt patients’ medical histories were more critical than those with >40 years of experience, even though almost 90% (145/161) of the dentists considered patients’ medical histories essential (Table 2). This difference could be because, until 2 decades ago, only limited information technology existed for dentists to access their patient’s medical information except for patient-reported medical history and medical consults. This limited access to EHR data may explain why dentists with more than 40 years in practice were more likely to think patient-reported information as reliable (Table 4). Additionally, dentist respondents who already have access to an HIE-EHR system felt patients’ medical histories were more useful (Table 4). Our survey found that the most needed information categories were medical conditions or diagnosis, current medications, and allergies (Table 3), which was consistent with a previous survey [26]. Together these findings showed that some categories of patient medical information were more helpful to dentists during dental care. These findings can also be used to optimize the user interface design in either an EDR-EHR system or an HIE to avoid information overload. However, our team’s earlier studies on medical consults discovered that dentists’ most requested information categories were laboratory values and written diagnostic reports, followed by recommendations or medical
clearances [15]. The inconsistency of these results indicates that dentists’ information needs can evolve based on access to relevant information. As they gain access to EHR information, they can ask more specific and informed questions when consulting their medical colleagues, leading to increased responses from medical colleagues. This improved information access may enhance dentists’ patient management and treatment planning. The results also indicated dentists’ information needs for new and existing patients were almost identical (Table 3). Future studies should continue investigating dentists’ information needs as they gain direct access to patients’ up-to-date medical information via an EDR-EHR system or an HIE.

The survey respondents, especially those with access to an HIE-EHR, showed clear interest in using the HIE to optimize the information collection process (Tables 2 and 3). For instance, 11.8% (19/161) of the respondents reported access to a state-based HIE, which was higher than expected. This higher access rate could be attributed to dental providers’ access to state-wide information systems, such as Indiana’s Prescription Drug Monitoring Program, and may have mistaken it for an HIE. Nonetheless, several state-wide HIEs are promoting dentists’ use of HIEs to improve access to patient information [35-37]. However, the overall use during dental care remains low. For instance, a study of New York dentists’ use of the Rochester regional HIE demonstrated a 0.17% rate of use of the HIE during dental encounters [35]. This low use is not surprising given that the use of community HIEs, even by nondental providers, is still growing, with 1% to 5% use in all patient visits. In the New York dentists’ study, they accessed the HIE primarily for patients with chronic conditions, gingival and periodontal diagnosed diseases, and during the first dental visit [35]. The most frequently visited sections were the laboratory and radiology sections within the HIE, which is consistent with our earlier study results of dentists’ medical consult requests [15].

Although the emergence of community and vendor-supported HIEs has improved medical providers’ timely access to patient information [38-41], inefficient and cumbersome processes and poor user experiences are significant barriers to HIE use [42,43]. Previous studies in medical settings reported that some HIEs require users to have multiple logins; interrupt their workflow; and display overloaded and poorly arranged information [42,43]. Our study respondents expressed similar concerns about the usability of HIEs, such as difficulty accessing data, information overload, and nonintuitive interface designs that could prevent dentists’ use of HIEs. Therefore, future HIE tools’ design and development should focus on the accuracy and integration of the data (content) and the information display and navigation (presentation). Few respondents including those willing to use an HIE expressed concerns about accessing patients’ medical histories via an HIE due to data safety and HIPAA compliance concerns. This issue needs to be addressed both at the technical level with more new tools and methods to ensure safe data sharing and exchanging and at the regulatory level with new protocols and rules to support the use of HIEs. Furthermore, most respondents agreed that patients should be able to control the use of their health care information, and their consent must be received before any information exchange and sharing occur.

**Limitations**

This study only invited Indiana dentists who are IDA members. A more geographically diversified pool of participants may help improve the results’ validity and generalizability. In future studies, we also want to include other dental professionals, such as dental hygienists and dental assistants. We are aware of the relatively low response rate to the survey, which is not rare in surveys of health care professionals, especially web-based surveys [32-34]. This was an exploratory study and our initial step to determine dentists’ information needs and to help improve their information access. Based on the results of this survey, we will conduct key informant interviews and focus group studies to include a broader group of participants. Another limitation was that dentists may not be familiar with some of the terminologies used in the survey such as state-based HIEs. Terminology definitions and examples should be included in future survey designs.

**Conclusions**

Patients’ medical histories are essential for dentists to provide high-quality dental care. In addition, information such as medical conditions or diagnosis, current medications, and allergies are more relevant to dentists’ clinical decision-making. Paper-based health history forms and medical consults are still the most widely used methods to gather information. However, electronic forms and integrated systems are gaining attention to have direct access to information. Most dentists are interested in using an HIE to access patient medical histories. The findings from this study can provide an alternative option for improving communications between dental and medical professionals and help the health information technology system or tool developers identify critical requirements for more user-friendly designs.

**Acknowledgments**

This study was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under award 5R01DE031259-02. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Institute of Dental and Craniofacial Research.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request. Requests may need approval from Indiana University.
Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey on how essential patient medical history is for dental care.

References


Abbreviations

- **EDR**: electronic dental record
- **EHR**: electronic health record
- **HCO**: health care organization
- **HIE**: health information exchange
- **HIPAA**: Health Insurance Portability and Accountability Act
- **IDA**: Indiana Dental Association
- **NIH**: National Institutes of Health
- **OR**: odds ratio

Edited by A Mavragani; submitted 24.07.23; peer-reviewed by DB Rindal, K Williams; comments to author 11.10.23; revised version received 18.10.23; accepted 22.11.23; published 11.01.24.

Please cite as:
Li S, Felix Gomez GG, Xu H, Rajapuri AS, Dixon BE, Thyvalikakath T
Dentists' Information Needs and Opinions on Accessing Patient Information via Health Information Exchange: Survey Study
JMIR Form Res 2024;8:e51200
URL: https://formative.jmir.org/2024/1/e51200
doi:10.2196/51200
PMID:38206667

©Shuning Li, Grace Gomez Felix Gomez, Huiping Xu, Anushri Singh Rajapuri, Brian E Dixon, Thankam Thyvalikakath. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Promoting Self-Efficacy of Individuals With Autism in Practicing Social Skills in the Workplace Using Virtual Reality and Physiological Sensors: Mixed Methods Study

Sung-In Kim¹, MD; So-youn Jang², MS; Taewan Kim³, PhD; Bogoan Kim⁴, BA; Dayoung Jeong⁵, BS; Taehyung Noh⁵, BS; Mingon Jeong⁵, BS; Kaely Hall⁶, MS; Meelim Kim⁷,⁸,⁹,¹⁰, PhD; Hee Jeong Yoo¹¹,¹², MD, PhD; Kyungsik Han⁴,⁵, PhD; Hwajung Hong¹³*, PhD; Jennifer G Kim²*, PhD

¹Department of Psychiatry, Seoul National University Bundang Hospital, Bundang, Republic of Korea
²Georgia Institute of Technology, Atlanta, GA, United States
³Department of Industrial Design, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
⁴Department of Data Science, Hanyang University, Seoul, Republic of Korea
⁵Department of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea
⁶School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States
⁷Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
⁸Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, CA, United States
⁹The Design Lab, University of California San Diego, San Diego, CA, United States
¹⁰Center for Wireless & Population Health Systems Calit2's Qualcomm Institute, University of California San Diego, San Diego, CA, United States
¹¹Department of Psychiatry, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
¹²Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of Korea
¹³*these authors contributed equally

Corresponding Author:
Jennifer G Kim, PhD
Georgia Institute of Technology
225 North Ave
Atlanta, GA, 30332
United States
Phone: 1 404 894 2000
Email: jennifer.kim@cc.gatech.edu

Abstract

Background: Individuals with autism often experience heightened anxiety in workplace environments because of challenges in communication and sensory overload. As these experiences can result in negative self-image, promoting their self-efficacy in the workplace is crucial. Virtual reality (VR) systems have emerged as promising tools for enhancing the self-efficacy of individuals with autism in navigating social scenarios, aiding in the identification of anxiety-inducing situations, and preparing for real-world interactions. However, there is limited research exploring the potential of VR to enhance self-efficacy by facilitating an understanding of emotional and physiological states during social skills practice.

Objective: This study aims to develop and evaluate a VR system that enabled users to experience simulated work-related social scenarios and reflect on their behavioral and physiological data through data visualizations. We intended to investigate how these data, combined with the simulations, can support individuals with autism in building their self-efficacy in social skills.

Methods: We developed WorkplaceVR, a comprehensive VR system designed for engagement in simulated work-related social scenarios, supplemented with data-driven reflections of users’ behavioral and physiological responses. A within-subject deployment study was subsequently conducted with 14 young adults with autism to examine WorkplaceVR’s feasibility. A mixed methods approach was used, compassing pre- and postsystem use assessments of participants’ self-efficacy perceptions.

Results: The study results revealed WorkplaceVR’s effectiveness in enhancing social skills and self-efficacy among individuals with autism. First, participants exhibited a statistically significant increase in perceived self-efficacy following their engagement with the VR system (P=.02). Second, thematic analysis of the interview data confirmed that the VR system and reflections on the data fostered increased self-awareness among participants about social situations that trigger their anxiety, as well as the
behaviors they exhibit during anxious moments. This increased self-awareness prompted the participants to recollect their related experiences in the real world and articulate anxiety management strategies. Furthermore, the insights uncovered motivated participants to engage in self-advocacy, as they wanted to share the insights with others.

Conclusions: This study highlights the potential of VR simulations enriched with physiological and behavioral sensing as a valuable tool for augmenting self-efficacy in workplace social interactions for individuals with autism. Data reflection facilitated by physiological sensors helped participants with autism become more self-aware of their emotions and behaviors, advocate for their characteristics, and develop positive self-beliefs.

Key Words

autism; virtual reality; workplace; self-efficacy; social skills; data reflection

Introduction

Background

Approximately 40% of individuals with autism experience anxiety because of difficulties in socializing, sensory sensitivities, and other factors [1-3]. Specifically, workplace environments can amplify anxiety in individuals with autism, such as sensory overload, communication barriers, and unplanned interactions [4-8] as these environments are often designed with neurotypical expectations in mind. When people with autism constantly encounter situations that elicit anxiety, it can cause them to have negative self-beliefs about their skills and performance, which can lower their self-confidence in the workplace [9]. Therefore, self-efficacy—the personal judgment or belief in one’s ability to succeed in prospective situations—is crucial for people with autism because it can help individuals approach challenging workplace experiences from a positive perspective, as opposed to focusing on failure [10,11].

Virtual reality (VR) interventions have demonstrated potential effectiveness in promoting the self-efficacy of people with autism, as they provide a safe yet realistic environment to master specific social skills by offering opportunities to repetitively practice them [12-18]. Although this exposure and repetition are useful for mastering skills, gaining a deeper understanding of specific social situations that trigger anxiety can better prepare users to effectively manage those situations in the real world. An individual’s understanding of anxiety-inducing situations and crafting self-beliefs that they have the capabilities to succeed in those situations align with 2 key constructs of the self-efficacy theory by Bandura [19,20]—physiological states and verbal persuasion. Notably, VR can also support physiological state awareness and verbal persuasion. It can identify the VR situations that prompt significant changes in a user’s physiological data, enabling the user to reflect on these experiences and formulate strategies to respond effectively [18]. This process can empower the development of self-efficacy [19]. However, existing VR interventions often limit the scope of self-efficacy to mastery experiences only [21] and fail to provide the extended theoretical rationale or background of self-efficacy on the design or outcomes of the system.

In this study, we expanded the current VR system by incorporating a feedback model in which the user’s behavioral and physiological sensor data can be reviewed immediately following the experience simulation, allowing for situational reflection.

Our research aimed to increase self-efficacy through a VR-based social skill training system for individuals with autism. Furthermore, we investigated how the facilitation of self-understanding through incorporating reflection of physiological and behavioral data immediately after the social skills practice can impact self-efficacy in preparation for real-life scenarios.

Objectives

This study had 2 main objectives. First, we developed WorkplaceVR, a VR application that allows people with autism to engage in simulated work-related scenarios to help understand their performance through data-driven reflection, with users’ behavioral and physiological data collected during the VR experience. Second, we evaluated the feasibility of the WorkplaceVR by conducting a deployment study with 14 young adults with autism. Using a mixed methods approach, we investigated the changes in self-efficacy among users with autism using pre- and postsurvey questionnaires. In addition, we conducted interviews to identify how participants with autism use, expect, and encounter challenges in the VR experience.

Methods

In this study, we designed and implemented WorkplaceVR, a VR-based system that offers simulations of work-related social situations and data-driven reflection of users’ behavioral and physiological responses (Figure 1).
**Phase 1: Development of the VR System (WorkplaceVR)**

*Inclusive and Iterative Design for the VR Program Development*

Our user-centered design approach to developing the VR system is to ensure inclusivity; technology must empower a more diverse and inclusive society [22,23]. We aimed to build a safe and less stressful virtual environment for people with autism, where they can become familiar with the workplace environment and practice interacting with others without fear of failure. Moreover, with inclusive design in mind, we conducted an iterative design process to incorporate the voices of people with autism into the VR program.

We first conducted a preliminary study to draw design suggestions for the VR program [24]. To elicit end-user feedback, we created a 5-minute video prototype demonstrating the overarching concept and the use scenario of the WorkplaceVR program. Using the video prototype, we conducted semistructured interviews with 20 participants, including individuals with autism (employees, job seekers, etc), managers of companies where people with autism are currently working, psychiatrists, and professional counselors, to uncover the various needs of individuals with autism at workplaces. Through the interview results, we identified the following key design insights. First, the participants emphasized the importance of designing realistic VR scenarios to help users engage in the program by reflecting on their personal experiences or challenges at work. Second, the system should guide users to reflect on their emotions and behaviors. Finally, during the data reflection phase, participants with autism should be able to take control of expressing their thoughts and behaviors and build confidence.

With insights from the preliminary study, we developed our WorkplaceVR program (Figure 1). The system comprises 2 parts: the simulation VR scenario and the physiological data visualization interface. WorkplaceVR was developed using the Unity3D engine (Unity Technologies) and the SteamVR plug-in (Valve Inc) and runs on a Windows 10 (Microsoft) PC with an Intel Core i7, GeForce RTX 2070 graphics card, and 16 GB RAM. A head-mounted display (HMD), VIVE Pro Eye VR headset (HTC) [25], and Empatica E4 wristband (E4 band, Empatica Inc) [22] were worn by the users for viewing the virtual world and sensing physiological signals.

We conducted a pilot study with 4 neurotypical participants to assess the study protocol and identify any risks and challenges that users might experience in trying WorkplaceVR. The participants were introduced to the VR system, experienced WorkplaceVR using an HMD, and then asked to provide feedback on the overall experiences, including task difficulty and scenario length, visual components of the interface, and side effects of the VR experience if there were any (eg, motion sickness, headache, and visual disturbance). On the basis of participants’ suggestions, we added additional visual cues, such as a shining effect to objects (eg, the bill receipt) for users to quickly find and interact with in the 3D environment and an arrow user interaction icon to lead the eye to a specific direction or object to draw attention to specific objects or areas of interest when needed. Another problem raised by the users was regarding the test environment setup. To facilitate an environment where the users can feel safe and private, we placed a blackout curtain in front of the desk where they stood during the experiment.

**WorkplaceVR: System Design**

In this section, we elaborate on the features and design considerations of the VR system, WorkplaceVR, based on the self-efficacy theory by Bandura [19,20].

**Designing VR Contents for Mastery Experience**

VR can offer mastery experiences by simulating real-world scenarios that can engage users with autism to accomplish tasks related to workplace settings. To achieve this, our design...
approach focused on three key factors: (1) the inclusion of authentic work scenarios, (2) the integration of evidence-based social skills, and (3) the incorporation of multiple levels of difficulty within the scenarios.

1. Realistic work scenarios: we used café scenarios with work-related realistic visuals and auditory stimuli that provide a sense of surrounding in the immersive environment. For example, we designed café furniture and appliances (e.g., a cash machine and a coffee machine) and instrumental background music to create a relaxing café atmosphere. Specifically, we selected a situation in which the barista must inform a customer that the drink they ordered is ready from among the situations that are required when working in a café (e.g., greeting customers upon entry and taking customer drink orders).

2. Evidence-based social skills intervention: in specific café-based scenarios involving interactions between a barista and a customer, we incorporated evidence-based social skills (e.g., active listening, initiating conversations, and not interrupting when someone else is talking) derived from interventions supported by previous research [26,27]. In addition, we included context-specific skills (such as informing customers about available options, verifying order accuracy, and problem-solving in unforeseen circumstances) sourced from the café service training manual [28,29].

3. Levels of scenarios: for the users to build self-efficacy, it is important to help them reduce anxiety and fear of failure by designing attainable goals that gradually increase in difficulty [19,30-32]. Thus, we designed 4 scenarios with 2 levels of difficulty: basic and advanced. The basic level requires the users to explore and interact with a simulated 3D environment to practice basic conversational skills as baristas. Following this, the users are involved in starting a conversation with a customer avatar by offering a polite greeting, taking orders, and serving beverages such as coffee. The 2 scenarios at this level have the same structure but different order details (number and types of beverages). When the users perform these tasks successfully, they are moved on to the advanced-level scenarios. At the advanced level, the scenario adds complications that arise within the conversation with the customer avatar (e.g., a customer avatar claiming that their drink order is wrong).

**Physiological Data and Visualization Interface**

In the self-efficacy theory, Bandura [19,20] explains that people can shape their perceptions and beliefs about their capabilities by examining one’s physiological and emotional states. Therefore, one of our goals in designing the VR program was to help users better understand their emotions and behaviors by identifying their physiological states and behaviors in social situations [12,33]. We aimed to support people with autism to become more aware of their physiological and emotional states when they face stressful or anxiety-raising situations in the virtual work environment. Hence, we collected the users’ behavioral and physiological data during the VR intervention using the E4 band and HMD. We synchronized the time stamp information transmitted from E4 with the HMD time stamp information using the open-source framework Flask [34]. In addition, we presented a data visualization interface for users to understand and interpret their data along with their recorded performance videos, as illustrated in Figure 2. On the interface, we present (1) anxiety-related physiological measures (e.g., temperature, electrodermal activity [EDA], and heart rate [HR]; Figure 2B); (2) changes in voice volume (Figure 2C); and (3) detection of eye contact (Figure 2D). The following explains how we defined, processed, and visualized each data type: anxiety moments, voice volume, and eye contact.

1. Anxiety moments: anxiety moments were defined using the time stamp on data collected when the sensors detected radical changes in the signal because of the users’ physiological reactions to the stimuli. We calculated anxiety moments using an anomaly detector provided by the Microsoft Azure machine learning algorithm [35]. This involved 5 sensor types—HR, EDA, temperature, interbeat intervals, and blood volume pulse—provided by the E4 band. We presented the anxiety moments on the interface with a time stamp range in seconds; each time stamp is linked to the exact time of the performance video. Users can click on a time range to watch their performance video on the left, as shown in Figure 2A.

2. Voice volume: voice data were collected using the microphone of the HMD. We then presented the voice volume level as a graph with time stamps to help users recognize changes in their voice tone and volume (Figure 2C).

3. Eye contact: to help participants understand their eye movements during their moments of anxiety, we used a region of interest (ROI), which is a specific area within an image or video selected for analysis. In our study, we set the faces of the customer and manager avatars as ROIs. We used the box collider of Unity3D and measured whether a user gazes at ROIs through Tobii G2oM [36], which is a machine learning algorithm that can accurately ascertain on which objects a user focuses. If the participant looked anywhere in that location, it was labeled as seen (1); if they looked at another area, it was labeled not seen (0) in 1-second units. On the basis of the labeling results, we presented the seen labeled periods as a bar graph to help users understand how they make eye contact with people during VR scenarios, as presented in Figure 2D.
Verbal Persuasion Through User-Driven Insights

In this study, we focused on facilitating a data-driven reflection process to leverage verbal persuasion for users with autism to engage in the sense making of their physiological data. Our goal was to provide an opportunity for users with autism to navigate their physiological and behavioral data results, identify their own strengths and interests, and gain confidence in speaking about themselves. Thus, the participants were given an active role in interpreting their data and answering questions such as What insights have you gained from the data about your characteristics or strengths? and What are your goals, taking data into account?

Phase 2: Implementation of the VR Program With Participants With Autism

We recruited participants with autism aged >16 years who are either currently employed or unemployed but plan on job searching in the near future. Our inclusion criteria for participants with autism were people who (1) have been diagnosed with autism; (2) can verbally articulate their thoughts, feelings, and experiences; (3) have no difficulties wearing an HMD for VR (eg, sickness or headaches while in VR or vision impairment, such as anisometropia); and (4) understand the study procedure and agree to participate. We posted the study posters to autism-related web-based communities on social networking services and web-based bulletin boards of autism-related institutions (eg, developmental disability social welfare centers and employment agencies for people with disability). In addition, we placed the flyers on the offline bulletin board of a child and adolescent psychiatry outpatient clinic in a hospital and a private counseling center for individuals with autism to outline the participant demographics.

Ethical Considerations

To ensure ethical conduct, our study received approval from the Institutional Review Board of Seoul National University Bundang Hospital (institutional review board number: B-2202-736-302). As this study involved the collection of sensitive information (eg, physiological data) from participants, we informed the participants about the data collection process and obtained their informed consent before they participated in the study. The researchers explained the consent form to the participants with autism in easily understandable terms. We also clarified to the participants with autism that any personal information, VR data and sensor data collected during the research would be anonymized for analysis and securely discarded to protect their privacy. In addition, we offered a compensation of US $50 for the research participants.

Measures

Perceived Self-Efficacy for VR Social Skill Training Scale

We developed the Perceived Self-Efficacy for VR Social Skill Training Scale (PSES-VR), an 8-item questionnaire with a 5-point Likert scale, to evaluate whether our VR intervention changed people’s beliefs regarding the self-efficacy of practicing social skills at the workplace. We developed this scale by modifying the Perceived Self-Efficacy questionnaire based on the theory of perceived self-efficacy by Bandura [37] and referencing existing scales, including the Perceived Social Self-Efficacy Scale [37], Perceived Improvement Scale [38], and Self-Efficacy Scale for Social Workers [39]. The questionnaires were revised to address the target social skills in the VR scenario. The 6 items consisted of communication skills required in general conversation situations. The 2 items evaluate whether participants respond appropriately to the situation required in the VR program scenario. Higher total scores on the items indicate better self-efficacy related to the social skills of the participants. The participants were asked to complete the same PSES-VR survey before the user study session and after the VR experience. All items of the PSES-VR are attached in Multimedia Appendix 1.
**iGroup Presence Questionnaire**

We used the iGroup Presence Questionnaire (IPQ), a 14-item questionnaire with a 7-point Likert scale, to investigate how users perceive a sense of presence of our VR system. The IPQ scale consists of 4 components: a general sense of being there (1 question), the sense of spatial presence (5 questions), involvement (4 questions), and experienced realism (4 questions), measuring the level of perceived presence during the VR experience [40]. Higher scores on the 4 components, as well as the total scores, indicate a better sense of presence in the VR system as perceived by the participants. All items of the IPQ are attached in Multimedia Appendix 1.

**Study Procedure**

Our study procedure included three stages: (1) pre-experiment, (2) VR experiment, and (3) postexperiment reflection phase. Figure 3 presents a summary of the procedure.

**Before the Experiment**

To start, the researchers provided a brief overview of the study purpose and conducted a brief interview with participants, asking questions about their work-related experiences and previous experiences with VR. Next, the researchers introduced WorkplaceVR with explanations of the contents of the VR system and sensor devices that the participants should wear during the VR experiment. After that, the participants had a tutorial session to learn the 2 basic-level scenarios through video and scripted role-play.

**VR Experiment**

When the participants were ready for the VR experiment, they were asked to wear the E4 band, which is used for sensing the physiological signals, and to put on the VR headset. We informed the participants that they could stop at any time if they started experiencing motion sickness. After they wore the devices, we asked the participants to explore the café environment to adjust to the VR environment. Then, all participants performed the same 4 scenarios in the order of basic to advanced scenarios in VR for approximately 10 to 15 minutes.

**After the Experiment**

After the VR experience, we took a 5-minute break and conducted a data reflection on their VR performance with sensor data. In the data reflection process, participants were presented with various types of data related to their own anxiety moments, voice volume level, and eye contact while video recording their performance in VR using the interfaces and were asked to reflect on their performance based on these data. Here, we informed the participants that the data may not be accurate because of technical issues, so there can be errors or missing data in the interface. In the data reflection process, we asked the following to help participants reflect their data, such as “How was your performance in the VR scenarios?” “What were your strengths and challenges while performing the VR scenarios?” or “Did you find anything new about your behaviors while reviewing the data?” After the data reflection, we conducted a semistructured interview about the overall VR experience and data reflection. We asked participants how the VR content (eg, tasks and levels of scenarios, VR environment, and avatar design), data-driven reflection, and protocol (eg, watching a video and role-playing before the VR experiment) could be used to better understand their emotions and responses in potential workplace settings. We also asked questions to elicit feedback on the usefulness of data-driven reflection and its potential impact on their self-understanding of their behavior.

**Data Analysis**

For qualitative data analysis, the interview data were transcribed, coded, and analyzed based on open coding and thematic analysis [41] using ATLAS.ti (version 7; Scientific Software Development GmbH). A total of 3 researchers individually read the interview transcripts and generated open codes that were discussed among the research team to identify patterns and build themes around VR experience and data-driven reflection regarding self-efficacy. The coding procedure was iterative; it concluded once the researchers agreed that the themes were consistent and a distinct set of themes surfaced. Interviews and surveys were conducted and documented in Korean, and analyses were performed in the same language. This ensured that the original responses were analyzed with the utmost precision. For the quotations used in the article, an English-Korean bilingual-speaking researcher initially translated the responses into English and then revised them through consultation with a proofreading expert.

For quantitative analysis, we analyzed the scale results (IPQ and PSES-VR) of the participants with autism. As a case-control
group was not included in this study, it was difficult to confirm the statistical significance of the IPQ scale in identifying the presence of the VR system in participants with autism. For the IPQ scale, we calculated the mean and SD of the items corresponding to each of the 4 IPQ components and used it descriptively to analyze the VR presence of participants with autism. For the PSES-VR, we performed a paired 2-tailed t test to analyze whether there was a statistically significant change in perceived self-efficacy after using our VR system. For anxiety moments and eye contact sensor data, we conducted a 1-way ANOVA to analyze whether there was a statistically significant difference among the 4 scenarios in WorkplaceVR.

Results

Overview of the VR Experience

A total of individuals with autism participated in the study, including 2 women and 12 men, with an age range of 16 to 34 years. Table 1 shows the baseline characteristics of the participants.

There were no participants who reported difficulties such as motion sickness and headache during the VR experiments. Most (12/14, 86%) participants succeeded in completing the advanced-level scenarios without any support. In total, 2 participants (ND8 and ND13) completed the advanced-level scenario with minimal prompts from the researchers, such as guiding participants to find where the receipt was placed in the table.

Table 1. Demographic information about participants.

<table>
<thead>
<tr>
<th>Code</th>
<th>Sex</th>
<th>Age (y)</th>
<th>Work experience (work period)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND1</td>
<td>Male</td>
<td>23</td>
<td>Food service experience at a fast-food restaurant (1 y)</td>
</tr>
<tr>
<td>ND2</td>
<td>Male</td>
<td>32</td>
<td>Office worker (8 y)</td>
</tr>
<tr>
<td>ND3</td>
<td>Male</td>
<td>23</td>
<td>Undergraduate student or part-time job (assistant at a counseling center; 2 y)</td>
</tr>
<tr>
<td>ND4</td>
<td>Male</td>
<td>24</td>
<td>Part timer for a cleaning service and daily paid jobs (event staff; 4 y)</td>
</tr>
<tr>
<td>ND5</td>
<td>Female</td>
<td>22</td>
<td>Undergraduate student or daily paid jobs (PowerPoint presentation design; 1 mo)</td>
</tr>
<tr>
<td>ND6</td>
<td>Male</td>
<td>19</td>
<td>Undergraduate student and no working experience</td>
</tr>
<tr>
<td>ND7</td>
<td>Male</td>
<td>27</td>
<td>Undergraduate student or café barista (1 y)</td>
</tr>
<tr>
<td>ND8</td>
<td>Male</td>
<td>25</td>
<td>Remote worker (data entry in Excel) or designer (making web-based banners; 3 y)</td>
</tr>
<tr>
<td>ND9</td>
<td>Male</td>
<td>20</td>
<td>Undergraduate student and no working experience</td>
</tr>
<tr>
<td>ND10</td>
<td>Male</td>
<td>20</td>
<td>Undergraduate student or part-time work experience (warehouse loading truck job and restaurant server; 3 mo)</td>
</tr>
<tr>
<td>ND11</td>
<td>Male</td>
<td>20</td>
<td>Customer service agent (8 mo); staff at a central radio management service (4 mo)</td>
</tr>
<tr>
<td>ND12</td>
<td>Female</td>
<td>27</td>
<td>Part-time worker (pet care; 1 y)</td>
</tr>
<tr>
<td>ND13</td>
<td>Male</td>
<td>33</td>
<td>Cleaning and maintenance (2 y)</td>
</tr>
<tr>
<td>ND14</td>
<td>Male</td>
<td>21</td>
<td>Freelancer (6 mo)</td>
</tr>
</tbody>
</table>

Quantitative Assessment of the VR System Use

Our results showed a significant increase in the perceived self-efficacy of participants with autism ($P=.02$) before and after experiencing WorkplaceVR. In the IPQ result, the mean scores for “general presence” (or the “sense of being there”) and “experienced realism” were higher than the mean scores for the other 2 components (spatial presence and involvement). Table 2 presents the results of the questionnaires.

Table 2 presents the physiological sensor data of the participants in the 4 VR scenarios.

Our hypothesis was that participants with autism would experience more anxiety moments in the advanced scenarios. Consistent with the hypothesis, the participants had more anxiety moments in the advanced scenarios than in the basic scenarios ($P<.001$). This result shows that our physiological sensing data can effectively act as a proxy indicator for anxiety experienced by participants with autism. This is because individuals with autism tend to encounter elevated levels of anxiety in unexpected situations, such as the more advanced scenarios that we introduced.
**Table 2.** Results of the Self-Efficacy Scale and System Evaluation.

<table>
<thead>
<tr>
<th>Scales</th>
<th>Values, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSES-VR</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>21.86 (7.33)</td>
<td>.02&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Post</td>
<td>39.07 (7.52)</td>
<td></td>
</tr>
<tr>
<td><strong>IPQ</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td>N/A&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>General presence</td>
<td>4.07 (1.86)</td>
<td></td>
</tr>
<tr>
<td>Spatial presence</td>
<td>3.87 (0.59)</td>
<td></td>
</tr>
<tr>
<td>Involvement</td>
<td>3.71 (0.79)</td>
<td></td>
</tr>
<tr>
<td>Experienced realism</td>
<td>3.89 (1.34)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>PSES-VR: Perceived Self-Efficacy for VR Social Skill Training Scale.

<sup>b</sup>There was a significant increase in the perceived self-efficacy of participants with autism before and after their experience with WorkplaceVR.

<sup>c</sup>IPQ: iGroup Presence Questionnaire.

<sup>d</sup>N/A: not applicable.

**Table 3.** Physiological data of participants at each scenario.

<table>
<thead>
<tr>
<th>Sensor data</th>
<th>Basic scenarios</th>
<th>Advanced scenarios</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
<td>Second</td>
<td>First</td>
</tr>
<tr>
<td>Anxiety moments&lt;sup&gt;e&lt;/sup&gt; (number of points), mean (SD)</td>
<td>2.43 (3.56)</td>
<td>3.21 (3.36)</td>
<td>9.43 (4.96)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Anxiety moments are defined as the timestamps detected by the Azure algorithm’s anomaly detector when sensor signals (eg, electrodermal activity and heart rate) display fluctuations in response to users’ physiological reactions to stimuli.

<sup>b</sup>There were significantly more anxiety moments in the advanced scenarios than the basic scenarios.

**Qualitative Analysis of the VR System’s Impact on the Self-Efficacy**

In this study, qualitative data analysis revealed 3 major themes that correspond to 3 of the 4 constructs of the self-efficacy theory by Bandura: physiological responses, verbal persuasion considered as encouragements, and mastery experience [19]. Below, we report how these findings can shed light on the feasibility of using the WorkplaceVR system to promote self-efficacy.

**Impact of the VR Program on Self-Efficacy of Participants With Autism**

**Self-Awareness of Physiological and Emotional State Through Data Reflection**

In the data reflection process, we provided an interface for visualizing the sensor data (ie, anxiety moments, voice volume level, and eye contact) to help the participants better recognize their emotions and behaviors in relation to their VR experiences. All our participants reviewed whether the presented data accurately reflected their emotions and behaviors. While reviewing their physiological signals (eg, HR, IBI, and EDA) presented by the VR system, the participants explained what they were feeling, thinking, or doing. For example, ND6 noticed that his HR increased when he was experiencing tense situations in VR and reported that he was aware of the stressors and reactions, drawing from his prior experiences:

> I know that my heartbeat increases when I am in situations that make me anxious or nervous. I remember my heart beating so fast on the night of June 29th. There was also thunder that night [...] It was like the nervous feeling I get when I check my grades. [ND6]

Moreover, the participants recognized and described the patterns of when and why they felt anxious in certain situations. By watching the recorded video of their performance at “anxiety moments,” participants actively described when they were most stressed or anxious while carrying out the simulated workplace tasks (eg, when a customer avatar approaches or when a customer avatar makes a sudden request to the user). For instance, ND2 had 3 events marked as “anxiety moments,” all of which were related to tasks where the user was required to start a conversation with the customer avatar. After playing back the recorded videos for all 3 events, ND2 reported that the data well represented his characters and that he usually becomes anxious when he has to initiate conversations with people at work:

> Oh, I’m quite sensitive sometimes, especially before I start talking, because I feel like I have to say...
something important. I also tremble when I start talking or see someone for the first time. [ND2]

By reflecting on their physiological and emotional states and behaviors in advanced scenarios, it enabled participants to recollect their related experiences in the real world and talk about goals to tame anxiety. For example, after reflecting on anxious moments, ND6 discussed ways in which he can improve his comfort in similar situations in the future:

I don’t want to panic again when a problem arises. For example, I shouldn’t be too stressed about my exam results. Whether it be a good or bad result, what’s important is accepting the result and looking beyond it. [ND6]

ND4 realized that he almost always becomes exhausted after meeting other people:

When I come home after meeting someone, I get so tired and have a headache. I always wondered what the reason was. Now that I think of it, I think it was because I got so nervous and tense in social situations. [ND4]

In addition, some participants suggested that having more diverse data (eg, facial expressions, body movements, and gestures) could help them better reflect on and understand their behaviors:

It would be helpful to see how I make gestures and move my body because it is something that is hard to notice in real life. Having an observation camera might work. [ND7]

**Verbal Persuasion Through Self-Expression in Data Reflection**

We found that the data-driven reflection process of our system can help participants with autism better understand themselves and build positive self-beliefs about who they are and what they can achieve. This opportunity allowed participants with autism to engage in verbal persuasion experiences of self-efficacy [42,43], which means that they could speak about their own strengths and interests and gain confidence in expressing themselves:

I was worried that I am not usually good at making eye contact. But seeing the data, I am surprised to realize that I am, in fact, doing quite well on this. I’m feeling more confident about myself. [ND10]

In addition, some participants reported that they wanted to use behavioral and physiological data to explain the strengths and characteristics they recently discovered to others so they could express their thoughts or arguments more clearly. For example, ND12 reported that she feels misunderstood by people when they have the wrong impressions or ideas about her. She proposed ways to use the data to effectively communicate her opinions to decrease potential conflicts with others:

I often find myself in a state of persistent frustration when communicating with others. I believe that people do not truly understand or accept my thoughts or feelings. [...] However, with the physiological sensor data I currently possess, I am confident that I can convince others. [ND12]

**Mastery Experience Through Realistic VR Scenarios**

Our interview results suggested that participants with autism experienced a sense of mastery while using the VR program. For example, the participants with autism reported that the successful completion of realistic workplace interpersonal situations in a VR environment boosted their confidence in handling similar situations in the real world:

You have to taste the fruits of success in order to gain confidence and not be afraid of failure. I felt this was an important factor for me to go on to the advanced level because failure can have a huge damaging effect. [ND4]

This mastery experience was also supported by our VR system design that gradually increases the level of task difficulty from the basic-level scenarios to help users respond to unexpected situations with confidence:

Although it was new, it didn’t deviate too much from what I already experienced, so I could handle it. [ND14]

Consistent with the IPQ result, the participants also reported a high level of presence in WorkplaceVR. The participants explained that this was because our VR program provided high-fidelity simulation embodying realistic visual content (eg, “Face Mask Required” signs on the walls of the virtual café considering COVID-19 pandemic situations) and immersive scenarios where they could have a naturalistic social interaction with virtual customers and colleagues (eg, using gestures to communicate and giving receipts to customers). For example, notable observations indicating high engagement included participants attempting tasks such as making coffee or using the cash machine, even though these actions were not part of the assigned tasks (Figure 1). The lifelike experiences participants experienced during our VR program could have played a role in cultivating a sense of mastery.

**Discussion**

**Principal Findings**

**Overview**

This study demonstrated that the VR program, which enables individuals with autism to experience work-related social scenarios and reflect on their VR practice through physiological and behavioral data visualization, can significantly increase the individuals’ perceived self-efficacy in practicing social skills within a workplace context. The interview data further showed how the data reflection of VR practice can influence self-efficacy. The user-driven data review practice allowed individuals with autism to reflect on their physiological data, that is, by promoting self-awareness of their emotions, gaining insights into their real-world behaviors that they were unaware of, and self-advocating their characteristics to others based on their data. In particular, participants could understand when and why they feel anxious, enabling them to proactively devise strategies for self-comfort in anticipation of similar anxiety-provoking situations. Moreover, the increased self-awareness about the underlying causes of their anxiety and
related behaviors motivated them to communicate their experiences and advocate for their needs with others.

**Promoting Self-Efficacy by Promoting Self-Awareness About Physiological States**

We found that the physiological data reflection helped individuals better understand their emotional responses. This increased self-awareness that participants with autism obtained through our VR system motivated them to take their learnings from reflection in the real world to better explain and advocate for themselves to others. Therefore, in this section, we discuss how the self-reflection interface should be designed to present physiological and behavioral data in a way that encourages individuals to reflect on their experiences.

According to Bandura [19], recognizing and managing one’s own emotions and physiological states is essential in promoting self-efficacy, as this affects people’s decision-making process and performance. In line with the theory, our study found that behavioral and physiological sensor data (eg, eye contact, voice volume, EDA, and HR) could be used to support people with autism to become aware of their emotional states and, in turn, establish strategies to respond effectively to intense emotions. The participants became aware of their current affect state by mapping their physiological responses on the interface and sometimes wanted to examine data in depth to improve their self-understanding (eg, facial expression, standing posture, and hand gestures). Specifically, reflecting on the physiological data taken during the VR experience while watching the playback of the sessions allowed participants with autism to revisit how they felt and behaved in VR situations in specific moments that heightened their anxiety. This reflection reminded them of similar situations that induce anxiety in their daily lives, such as when they had to initiate conversations or when a conversational partner is approaching them. This increased self-awareness about their emotional responses—why they were anxious or nervous at specific times—further allowed participants with autism to make a resolution: how they might manage their emotional reactions in everyday lives.

Moreover, we found that the data reflection allowed participants with autism to gain insights about their real-world behaviors that they were unaware of (eg, “Why I was anxious when I had a conversation with coworkers?” or “Why I was always tired when I went to a place with a lot of people?”). Therefore, VR interventions for promoting self-efficacy could be designed to provide opportunities for users to investigate their emotional reactions through data reflection and to connect the insights to their real-world practices. To encourage reflection, Fleck and Fitzpatrick [44] suggest incorporating reflective questions into technology to prompt users to think about their behaviors and provide justifications or explanations for knowledge, actions, or events. Therefore, VR systems can present reflective questions such as “What about this scenario made you feel anxious?” to accompany data reflection and to provide scaffolding for individuals with autism to consider how their performance in the scenarios relates to their real-world experiences.

Finally, when reflecting on the physiological data marked during the VR simulations, the system can guide users to raise their emotional awareness by relating it to their real-world experiences. It is important for them to understand what factors trigger their emotional and physical responses, why these factors affect them in a certain way, and how they should respond to such emotions. Previous research suggests that careful observation of one’s behavior, either by themselves or by others, might be the most informative and applicable source of emotional self-awareness [45,46]. Similarly, in our study, participants with autism identified the situations where they felt anxious while reviewing the anxiety moments data and watching the recorded videos of their VR performance and described their feelings by recalling prior related experiences. Through this process, they planned more specific ways to respond to anxiety, which could be applicable to their real-world interpersonal situations.

**Toward Data-Driven Self-Advocacy**

Verbal persuasion, involving encouragement from others and self-advocacy practice, is an important source of self-efficacy because it can help individuals shape self-beliefs that they have the skills or knowledge to succeed in a given situation and have confidence in themselves [19,47]. Our study suggests that the data-driven reflection process could have a similar effect as verbal persuasion. Our participants with autism reported that they often received negative feedback about their behaviors from others but experienced validation of behaviors through our system, for example, when the data interface indicated that their voice volume was lower than they expected. These results show that data reflection creates an opportunity for individuals with autism to experience verbal persuasion through identifying what they can do well in VR scenarios and fostering internal motivation to apply what they learned about themselves in the real world.

Furthermore, the participants with autism wanted to advocate for their characteristics or strengths identified in the data reflection to others to resolve conflicts or difficulties in their interpersonal relationships (eg, family members’ negative comments on the behaviors of autistic individuals). This finding suggests that data reflection could help individuals with autism to advocate for themselves in their daily lives and workplace. To design systems that can effectively support self-advocacy, our findings suggest the importance of presenting data relevant to their daily lives and supporting them to use the data to reflect on their behaviors, build confidence, and foster self-advocacy.

In our study, reflective questions [48] enabled users with autism to take time to understand themselves and translate their thoughts and concerns into positive resolutions based on the data. The questions included the following: “[reviewing the anxiety moment data] Have you ever encountered a similar situation in real life? If so, why did you feel that way? How do you typically respond to stressful events that make you anxious?” Although the participants could not directly manipulate the data visualization interface, future studies are needed to uncover how the interface can be designed to engage users with autism to better reflect on their personal interests, skills, and experiences. This could be approached by visualizing data with metaphors familiar to individuals with autism and
customizing the user interface to reflect users’ priorities and topics of interest [49-51].

Limitations
Although our VR system could provide individuals with autism with opportunities to promote self-efficacy, there are several limitations. First, in our study, we only included participants with autism who are able to communicate and interact with others. This decision was made in our study because WorkplaceVR was designed to focus on a specific population with autism. However, future research should explore how VR interventions can also benefit participants with autism who have different communication abilities. To estimate anxiety moments, we used the algorithm offered by Microsoft Azure. Although our participants confirmed that the anxiety predictions were aligned with their subjective feelings (eg, anxiety and nervousness), future studies should investigate and apply more rigorous algorithms that can predict the anxiety levels of participants. In this study, we used the PSES-VR, a questionnaire written in Korean, and the IPQ, which was translated into Korean. However, neither of the 2 measures had been previously validated in the Korean population. Finally, we could not confirm whether participants’ experiences in the study would translate to their real-life situations through poststudy observations.

Conclusions
This study investigated how the VR system promotes the self-efficacy of individuals with autism for their success at work. For this, we presented WorkplaceVR, a VR system that allows users to experience work-related social situations and data reflection of the user’s behavioral and physiological data. Through the VR experiment and data reflection, we confirmed that the VR system significantly improved the perceived self-efficacy of participants with autism. Our study results revealed that the VR system provided participants with autism with an opportunity to have a mastery experience in VR scenarios, self-awareness of their emotional states, and self-advocacy of their strengths and characteristics through data reflection. By addressing the expectations and challenges in the VR system for people with autism, these results contribute not only supporting researchers who design the technology for autistic people but also helping individuals with autism have a successful work experience.

Acknowledgments
The authors used generative artificial intelligence, ChatGPT (OpenAI, version 3.5, 2023), to correct the grammatical errors in the manuscript.

Data Availability
The data sets generated during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Contents of the two scales for evaluating social skills related to the perceived self-efficacy of the participants and the perceived sense of presence of our virtual reality system: (1) Perceived Self-Efficacy for VR Social Skill Training Scale and (2) iGroup Presence Questionnaire.

References


25. VIVE Pro 2. VIVE. URL: [https://www.vive.com/us/product/vive-pro-eye/overview/] [accessed 2023-08-16]


29. Service and training manual. Smith Coffee and Cafe. 2019 Jan 30. URL: [https://static1.squarespace.com/static/5ad7d558f93fd4e494252090t/5c5d4de4f0d601a4d772292/1549062373251/SmithTraining.pdf] [accessed 2023-12-22]


36. Tobii G2oM. Tobii. URL: https://vr.tobii.com/sdk/solutions/tobii-g2om/ [accessed 2023-08-16]

Abbreviations

- **E4 band**: Empatica E4 wristband
- **EDA**: electrodermal activity
- **HMD**: head-mounted display
- **HR**: heart rate
- **IPQ**: iGroup Presence Questionnaire

https://formative.jmir.org/2024/1/e52157

JMIR Form Res 2024 | vol. 8 | e52157 | p.155

(page number not for citation purposes)
PSES-VR: Perceived Self-Efficacy for VR Social Skill Training Scale
ROI: region of interest
VR: virtual reality

Please cite as:
Promoting Self-Efficacy of Individuals With Autism in Practicing Social Skills in the Workplace Using Virtual Reality and Physiological Sensors: Mixed Methods Study
JMIR Form Res 2024;8:e52157
URL: https://formative.jmir.org/2024/1/e52157
doi:10.2196/52157
PMID:38206652

©Sung-In Kim, So-youn Jang, Taewan Kim, Bogoan Kim, Dayoung Jeong, Taehyung Noh, Mingon Jeong, Kaely Hall, Meelim Kim, Hee Jeong Yoo, Kyungsik Han, Hwajung Hong, Jennifer G Kim. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Mapping the Cardiometabolic Patient Experience and Self-Care Behaviors to Inform Design, Implementation, and Persistent Use of Digital Health Care Solutions: Mixed Methods Study

Jan Liska^1, MSc; Marie Mical^1, PharmD; Christophe Maillard^1, MA; Cécile Dessapt^1, CMPP, PhD; Europa Bendig^2, Dip Ing; Daniel Mai^2, MA, PhD; John D Piette^3, MS, PhD; Sabina De Geest^4,5, RN, PhD; Guillaume Fontaine^6,7,8,9, RN, PhD

^1Sanofi, Paris, France
^2STURM und DRANG, Hamburg, Germany
^3Department of Health Behavior Health Education, School of Public Health, University of Michigan, Michigan, MI, United States
^4Institute of Nursing Science, Department Public Health, University of Basel, Basel, Switzerland
^5Academic Center for Nursing and Midwifery, Department of Public Health and Primary Care, Katholieke Universiteit Leuven, Leuven, Belgium
^6Ingram School of Nursing, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada
^7Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Montreal, QC, Canada
^8Centre for Nursing Research, Jewish General Hospital – Centre intégré universitaire de santé et de services sociaux West-Central, Montreal, QC, Canada
^9Centre for Implementation Research, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, QC, Canada

Corresponding Author:
Jan Liska, MSc
Sanofi
46-48 Avenue de la Grande Armée
Paris, 75017
France
Phone: 33 141247000
Email: Jan.Liska@sanofi.com

Abstract

Background: Cardiometabolic conditions including acute coronary syndrome (ACS) and type 2 diabetes (T2D) require comprehensive care and patient engagement in self-care behaviors, and the drivers of those behaviors at the individual and health system level are still poorly understood.

Objective: We aim to gain insights into self-care behaviors of individuals with cardiometabolic conditions.

Methods: A convenience sample of 98 adult patients with ACS and T2D was recruited in the United States, Germany, and Taiwan to participate in a mixed methods study using ethnographic methods. All participants completed 7-day web-based diaries tracking their level of engagement, and 48 completed 90-minute web-based semistructured interviews between February 4, 2021, and March 27, 2021, focusing on themes including moments of engagement. Qualitative analysis identified factors influencing self-care practices and a Patient Mind States Model prototype.

Results: Patient reports indicate that many patients feel social pressure to adhere to treatment. Patients’ experience can be understood within 5 categories defined in terms of their degree of engagement and adherence (“ignoring,” “struggling,” “juggling,” “controlling,” and “reframing”).

Conclusions: For people living with ACS and T2D, the self-care journey is defined by patterns of patient experiences, which can identify areas that tailored digital health care interventions may play a meaningful role.

(JMIR Form Res 2024;8:e43683) doi:10.2196/43683

KEYWORDS
self-care; adherence; digital health; design; implementation; coronary; type 2 diabetes; care; patient engagement; behavior; interview; treatment; tool; digital tool; support
Introduction

The development of effective self-care behaviors is essential for patients living with chronic health conditions [1]. However, the definition of self-care behavior varies considerably between medical disciplines. From recent concept analyses, a broad term for self-care has been defined as “the ability to care for oneself through awareness, self-control, and self-reliance in order to achieve, maintain, or promote optimal health and well-being” [2]. Indeed, inadequate engagement in these self-care ability driving areas, suboptimal adherence to medication, and failure to enact healthy behaviors (eg, smoking cessation and physical activity) can compromise patients’ quality of life and health outcomes [1]. Self-care is particularly important for patients diagnosed with cardiometabolic conditions such as acute coronary syndrome (ACS) and type 2 diabetes (T2D) [3,4]. Despite this, rates of adherence to a care plan including but not limited to medication and lifestyle modifications remain suboptimal for this population [5-8], and a number of patient-specific factors (eg, depressive symptoms, attitudes toward management, and daily activities) that impact self-care and outcomes have been identified [9].

Despite efforts to establish the psychosocial factors underlying self-regulatory or self-management behaviors [10,11], the typical driver of such studies is to assess how these factors may impact adherence. Behavioral models have been considered within this research, such as the capability, opportunity, motivation, and behavioral model and intervention mapping; however, the volume of frameworks available and a fragmented, confusing taxonomy remain barriers to their effective use [12]. Equally, these frameworks are often generalized and applicable in a range of clinical scenarios and are not exclusive to diabetes or ACS.

This model design approach from a general perspective may explain in part why these models have had a limited impact in diabetes or cardiovascular care, where individual behavior plays a dominant role in patient outcomes [13]. Establishing a behavioral framework that is tailored for those with a cardiometabolic condition will form an important step toward understanding attitudes in clinical care and how these may innately fluctuate over time.

The determinants of effective self-care practices among people living with cardiometabolic diseases have been well-established and include self-efficacy, social support, cognitive skills, and positive attitudes [14-16]. Health care professionals (HCPs) and care systems play a central role in influencing self-care through these determinants as well as by optimizing health service delivery, addressing financial burdens, promoting engagement with technology, and encouraging community support initiatives [17,18]. As self-care determinants include both patient-level and system-level drivers, effective interventions must address both environmental factors and individual patient needs to achieve sustainable effects [19,20]. Designing these interventions requires a deep understanding of patients’ needs as well as their broader life context [21]. Patient activation–focused frameworks such as the Patient Activation Measure (PAM) or the Social, Psychological, Usage, Rational model have helped design interventions that improve medication adherence and lifestyle behaviors [22,23]. However, given ongoing changes in health systems’ use of digital patient support tools in the context of the COVID-19 pandemic, the ways in which tools such as the PAM and Social, Psychological, Usage, Rational model can best be used to support patient self-care needs further exploration. In particular, the COVID-19 pandemic has led to major shifts in clinical practice, including an increase in the use of digital health care approaches [24].

To better understand the ways in which health systems can support patients using digital support tools in the era of the pandemic, we conducted a survey of adult patients with ACS and T2D as well as their caregivers in the United States, Taiwan, and Germany. Our goal was to help establish a behavioral model based on the survey results, which would help inform how digital solutions may be utilized to improve adherence to therapy. This series of interviews and diaries, which drew on grounded theory and phenomenology, aimed to (1) provide deep human-centric insights into the behavioral dimension of the patient experience for people living with T2D or ACS and (2) identify the moments and motivational triggers in a patient’s life that have a strong impact on behavior and health outcomes.

Methods

Research Design

The survey consisted of 2 distinct phases, with all participants completing 7-day web-based, ethnographic diaries tracking their level of engagement in self-management and 48 participants completing web-based, semistructured interviews focused on themes including moments of engagement.

This survey used methods from the field of qualitative research, grounded theory [24] and phenomenology [25,26]. Grounded theory has been used to understand the processes through which patients manage new or chronic health problems [27], and as such, it has particular relevance to elucidating chronically ill patients’ experiences. Using grounded theory, data are collected and analyzed, and then a theory based on the resulting data was developed. The approach was designed to generate a theoretical explanation using both inductive and deductive approaches to a social phenomenon (ie, chronic disease self-care in this survey) from empirical data rather than a preconceived framework. Phenomenology is well suited to study the self-care of new or chronic health problems as it is based on the assumption that there is an essence to what people live with every day, and it aims to depict the basic structure of this experience [25]. As a qualitative research method that is particularly useful to study affective, emotional, and intense human experiences, phenomenology is a study of people’s conscious experience of their everyday life and social action [25].

Ethical Considerations

Informed consent was obtained from all individual participants included in this survey. The research is a qualitative behavioral research survey but not a clinical study or clinical survey. This research was conducted in accordance with the organization’s intended regulations for qualitative market research studies and no personal data or sensitive information were collected or presented in this publication. Information was recorded by the
in investigator in such a way that the identity of the human subjects cannot be directly or indirectly ascertained. This study was exempt from the institutional review board oversight in accordance with exemption guidelines listed in the 45 Code of Federal Regulations Part 46 and the Secretary’s Advisory Committee on Human Research Protections Recommendations on benign behavioral intervention of the Health and Human Services regulation in United States [28,29]; the 2022 European Pharmaceutical Market Research Association Code of Conduct for the market research conducted in Germany [28]; and Article 5 of the Human Subjects Research Act of the Ministry of Health and Welfare, Republic of China (Taiwan), and the “Exempt Review Categories for Human Research” announced by Department of Health, Taiwan [29,30].

**Sampling Methods**

Patients diagnosed with either ACS or T2D and cardiovascular comorbidities were identified and recruited in 3 countries: the United States, Germany, and Taiwan. These countries were selected based upon the Hofstede 6D model of cultural dimensions. The 3 countries were noted for their diversity of cultural dimensions that may influence and nurture patient behavior while holding similar maturity of health care systems.

We aimed to recruit equal proportions of men and women as well as patients with a variety in types of health insurance. Potential patients were members of an ongoing research panel who have agreed to be approached for research studies and have provided basic demographic and health status information. Patients with ACS and T2D were screened for eligibility using elements for ACS and T2D patient profiles, as well as survey-specific data regarding demographics, digital behavior, and personality profiles. As part of the screening process, patients were asked if they use any technology, either app or device, to manage their condition. All patients screened were interested and open to the idea of digital health care (ie, the use of digital tools to support tracking or managing measurements used to help manage their condition).

Screening was conducted by an external recruiting agency who preselected relevant candidates out of their patient panel. In total, 32 participants per country were selected out of approximately 48 recommended profiles per country from an external recruiting agency. An appropriate mix of people was selected by the research lead, striving for maximum variation in the sample among those who fulfill the sampling criteria.

Our goal was to recruit 16 patients with ACS from each country, including 8 who had undergone an invasive procedure (eg, percutaneous coronary intervention and coronary artery bypass graft). We also sought to recruit roughly equal numbers of patients with ACS who were aged 40-50 years and 50-70 years. All patients had to be eligible for enrollment in a cardiac rehabilitation program (whether this was digital or analog was recorded); approximately 60% (10/16) should have attended rehabilitation, while 40% (6/16) should not. A focus was placed on those with higher severity levels for their condition and those with progressed treatment regimens. Eligibility for enrollment in a cardiac rehabilitation program was determined as a proxy for ACS severity levels and an indicator of having undergone a medical procedure. In recruiting patients with ACS, our goal was to represent a diversity of family situations (single, married, and with or without children) and a range of experiences using self-care technology for managing cardiovascular conditions.

Finally, we sought to recruit 10-16 caregivers of patient with ACS to assist in setting up patient support with the diaries, 8 of whom would also serve as interview participants.

Our goal was to recruit 16 patients with T2D from each country, with equal numbers having less severe (eg, hypertension) and more severe (eg, heart arrhythmia and heart failure) comorbidities. These groups were each further equally split into those on oral antidiabetic drugs and those who had initiated basal insulin in the past 12 months (18-24 months if recruitment was difficult). All participants in the T2D sample were aged 40-70 years, with the group being split equally between those who currently use health care technology (an app or device) and those who are considered lapsed users of health technology. Within the screening process, patients with T2D were asked if they use any technology, either an app or device, to manage their disease. Lapsed users confirmed that they used technology to manage their condition but no longer do so.

Both the ACS and T2D samples were split into equal-sized groups of patients deemed to be fully adherent with medical advice and treatment and those considered either partly adherent or nonadherent. This was based on the assumption that the behavioral reasons for partial and nonadherence were similar. Adherence levels were determined based on patients’ self-reported assessment during the respondent recruitment process. Patients were asked a range of multiple-choice and open-ended questions, the results of which were used to quantitatively determine their adherence level. For instance, patients were asked to what extent they agree with the statement, “I am confident that I can follow through on medical recommendations my health care provider makes, such as changing my diet or regular exercise;” with which they could strongly agree, somewhat agree, somewhat disagree, or strongly disagree.

**Data Collection**

To increase internal validity of the survey through triangulation, 2 methods of data collection common in grounded theory and phenomenology research were used [26]. First, patients were asked to complete web-based diaries for 7 days, in which they recorded information including their daily experience of engagement with notable life events, health, and conditions (eg, photos and video) for acute versus chronic; a mind map of helpers (people, tools, and institutions); and observations of self-concept versus reality (projections and narratives around the self-care experience).

The diaries consisted of 5 chapters, with each one completed once within 7 days. Each chapter was completed in consecutive order: (1) “Me and My Body,” (2) “My Story,” (3) “My Day-to-Day Life with My Condition,” (4) “My Daily Health Regimen,” and (5) “My Helpful Tools and Resources.” Daily diary exercise was recorded based on open-ended questions with private responses recorded via video, audio, photographs, a map, or written statement.
Specifically, respondents were asked to answer a range of open-ended questions over the course of 5 days. They were structured in six thematic buckets: (1) personal background, (2) me and my body, (3) daily health regimen, (4) day-to-day life with my condition, (5) helpful tools and resources, and (6) my health story. In topic 2, we asked, for instance, “Please recall particularly pleasurable and positive moments during the day when you felt happy, in control, proud of yourself, content with regards to a) general state of mind, b) relating to health, and c) relating to your condition.”

For the interviews, participants were screened by an external recruiting agency, relying on their patient panels, and an appropriate mix of people was selected by the research lead, striving for maximum variation in the sample among those who fulfill the sampling criteria. In the ACS sample, patients and their caregivers took part in dyad interviews. In the T2D sample, interviews were carried out on a one-to-one basis between interviewers and patients. These interviews aimed to capture key moments of engagement to gain an understanding of levers of change or opportunities; elucidate the role or influence of the caregiver; and explore a prototype of the patient experience map produced from patient diary content. The “SturmundDrang” team of cultural researchers and anthropologists created the interview guide in an iterative approach. This guide provided information on field research preparations. Designed as a rough framework of topics and themes to explore, this document offered guidance for the ethnographic video interviews. The same discussion guide was used across all countries, providing open-ended questions and sufficient room for exploration in different cultural contexts. The respective researcher could modify the guides in each country, if deemed necessary.

Data Analysis
Data were analyzed using a constant comparative method [31], which involves comparing 1 segment of data in an interview with another segment of data to determine similarities and differences [26]. The diary data was treated as 1 hermeneutic unit that was then qualitatively analyzed in direct comparison with the interview data. The interview data formed a second hermeneutic unit. Diaries and interviews were conducted within 2 weeks, allowing the researchers to pursue a close comparative approach. These 2 hermeneutic units were then compiled in a comprehensive raw field note document for each market, outlining the total qualitative data set that could then be analyzed.

Analysis was guided by the principles of horizontalization, where all data are treated with equal weight, and phenomenological reduction, which is the process of continually returning to the essence of the experience to derive the inner structure (of disease self-care) in and of itself [26]. We grouped data together on similar dimensions of (1) patient modes of engagement and experience domains, (2) the ongoing process of gaining self-care expertise, and (3) patient mind states regarding disease self-care. Furthermore, we identified beliefs and biases, drivers of engagement and challenges to disease self-care.

In particular, interviews were analyzed to identify emerging themes through established social-scientific methods of data gathering, including writing semistructured fieldnotes [32], qualitative data analysis [33], and ethnographic insights building [34]. To establish key patient mind states, the data were clustered using spatial clustering, iterative loops, and narrative listening. Spatial clustering of codes and signals was used to group themes into a figure of overarching mind states to provide a framework for qualitative interpretation and a hypothesis. Following data collection, the resulting map was refined through iterative loops, whereby the data and theory were examined and re-examined by a team of 3 researchers; with each loop, the number of clusters and mind states was adapted and refined. During the final iteration loop, the mind states and their names were co-designed with the patients to ensure a patient-centric outcome. Narrative listening was used throughout, with patients describing and naming the chapters of their journey to allow for patient-led clustering and refining of the mind states [35].

Results
Sample Characteristics
In total, 98 patients with ACS or T2D and cardiovascular comorbidities were recruited, 32 from the United States and 33 from each of Germany and Taiwan. All 98 patients who were recruited completed the survey. All patients completed the 7-day ethnographic diaries (Germany or United States: February 4-11, 2021; Taiwan: February 17-24, 2021). Patients spent approximately 5 hours to complete the research questions. In total, 48 web-based interviews were conducted (Germany or United States: March 3-15, 2021; Taiwan: March 22-27, 2021), 24 with patients with T2D and 24 with patients with ACS and their caregivers. Respondents were selected for further interviews based on the depth of patient journey detail.

Survey Results
Overview
From the survey conducted through interviews and diaries, a number of theoretical themes were identified.

Insights Into the Behavioral Dimension of the Patient Experience
Interviews with patients living with T2D indicated that the disease continues to extract a heavy emotional toll. Participants commented: “The nurse made me feel like some pig, that all I do is sit around and eat and be obese”; “I used to eat gourmet food anywhere, anytime, and now all I have left is pay attention to controlling my diet”; “If I were to buy something to eat – I almost never do – then I would have to take everything apart and weight the ingredients individually; I can then calculated what units I need to inject afterwards”; and “I thought I’d just take a pill and everything will be ok, so I didn’t take it as seriously as I should.”

These patient insights highlighted how T2D is still a chronic condition that complicates every aspect of one’s life. Overarching themes from these interviews indicate that the condition is widely perceived to be a self-inflicted lifestyle disease; patients experience feelings of stigma, shame, self-blame, and a need to justify their lifestyle choices to acquaintances and HCPs. The general experience is underscored
by patient sacrifices and unsolicited social pressures to adhere to treatment regimens; generating a depressive, restrictive atmosphere rather than the positive outlook that could help patients pursue a better quality of life. The need for constant monitoring and tracking of the body increases adherence pressure and creates a heightened focus on hemoglobin A\textsubscript{1c} levels. Patients can also initially find diabetes easy to underestimate and ignore, adopting an acute rather than chronic mindset that makes it difficult to take ownership in the long term.

Acute events often are perceived by patients as inevitable consequences of living with chronic conditions. For example, 1 patient with ACS commented: “I learned that even without heart disease, unhealthy habits as far as eating and exercise can lead to stroke and heart failure. This really opened my eyes. At that point my life changed dramatically. I changed the way I shopped, I started eating healthier meals, and I began exercising regularly.” However, acute conditions may have a longer-term impact on social behaviors. ACS can lower social and career performance expectations; patients may feel forced out of the “rat race” by their condition or voluntarily take steps to reduce work-related burdens and pressures, as indicated by 1 patient who commented: “I see people around me going to work and realize how much my health prohibits me from working.”

Based on these moments and experiences, patients may feel that rehabilitation is a challenge and may not fully comprehend or appreciate the potential benefits. Once the advantages of rehabilitation are clearly communicated, the offering is compelling to patients if logistical barriers are no issue. Participants also described the gradual acquisition of self-care expertise as a journey, which can support them in working toward better health. They also see this journey as influenced by universal health-related experiences, where a patient undergoes a cognitive or emotional change regarding their health engagement and self-care. Although each patient journey is unique, we identified 3 modes of engagement that determine the engagement and behavioral patterns in self-care (Figure 1). Underlying these modes are clusters of experiences called “experience domains.” At different points in the patient journey, health-related experiences from any domain can affect the patient’s mode of engagement, in a process of fluid exchange and even overlap between domains.

Self-care can be seen as an ongoing process of gaining expertise on how to deeply incorporate management routines into patients’ day-to-day lives (Figure 2). Over the years of encountering the full breadth of health care–related experiences, patients work out the best ways to manage their conditions in the context of their individual lives. This process takes patients from awareness of the need for self-care, through acquiring the necessary practices and tools to learning how to use them successfully. A set of applicable self-care practices are developed, which become habits in the form of routines. Highly engaged patients eventually become informal self-care experts on their own body and health.

Figure 1. Patient modes of engagement and experience domains. HCP: health care professional.
Interviews and patient log data indicate that a patient’s willingness to engage in self-care and gain expertise may be shaped by the cultural context in which they reside. Some of the participants in the United States expressed a growing mistrust of medical expertise and institutions, with a lack of medical insurance also disrupting the continuity of care. For example, patients from the United States shared: “Never trust someone just because they have a medical degree.” And “I’m on Medicare and Medicaid, I’m disabled, cannot work and now I’m getting paid back for the things I used to say because the clinic won’t cover it.” Effective self-care may also present challenges in countries such as Germany, where many health care systems rely primarily on paper-based clinical records, owing to historical and ongoing data privacy concerns. This impacts the patient experience. A man with T2D from Germany shared: “I am fortunate that my two doctors are located in one facility and can therefore coordinate closely. All the information [about my treatment] is centrally stored and can be viewed at the facility at any time.” Pseudoscientific or alternative approaches to health care, such as homeopathic or organic products as well as spiritual practice, also remain prevalent in Germany and Taiwan; however, participants in Taiwan expressed a high level of trust in HCPs’ expertise and authority, with 1 participant commenting: “I’ll strictly follow the doctor’s advice as best as possible.”

These differing experiences across countries was a common theme during the survey. Although, patients in the United States expressed a mistrust toward institutions, those surveyed were enthusiastic about trying new digital health solutions. However, this attitude was not reflected in Germany, with privacy concerns limiting the uptake of digital solutions. Many patients remain cautious around such technologies. Further, I patient with T2D in Germany stated: “I don’t know what I would want [for a digital solution]...and what is possible privacy wise.” Both US and German attitudes toward digital health care stand in contrast with the opinions of those patients in Taiwan. There has been an increasing adoption of digital solutions among those with T2D in Taiwan, with many perceiving such innovations as convenient and readily integrable.

**Theorization of Patient Engagement: the Patient Mind States Model (PMM)**

Based on analyses of the survey through patient interviews, we identified 5 patient mind states according to degree of patient engagement, adherence, and the experience domains that drive engagement (namely, the condition-, context-, and self-driven experiences; Figure 3). These mindsets comprise what we have called the Patient Mind States Model (PMM), which articulates these 5 mind states regarding disease self-care. A mind state is defined as a patient’s mental and emotional attitude toward self-care; these states are not related to age, gender, sociodemographic criteria, or culture of persons living with cardiometabolic conditions. The mind state has a large influence on a patient’s receptivity toward support and their ability to develop more healthy behavior. Patients’ mind states are not constant; shifts in mind state can be driven by external forces such as seasonal cycles and life changes. Self-care engagement and maturity of patients with chronic conditions change with their mindset. The 5 mind states identified are the following:

- **Ignoring:** patients carry on with life as it was prior to their diagnosis; depending on their culture and the extent to which they are stuck in an acute mindset, they may believe that a drug can fix the problem. For example, a 68-year-old woman with T2D from Germany commented: “I don’t really feel sick, because I don’t notice anything. I don’t have a different life in terms of ‘before’ or ‘after’ the diagnosis. Actually, I ignore the disease.”

- **Struggling:** patients feel overwhelmed and anxious; in desperately trying to make sense of what is happening, they can be paralyzed and oscillate between desperation and overambition. A 59-year-old woman with ACS from Germany shared: “I sleep very badly, have fears about my health. I feel completely overwhelmed and don’t know how to get out of this dilemma.”

- **Juggling:** people want to focus on positive aspects of life, which are viewed as part of the healing process; they see themselves as trying to find a balance between their own wants and the demands of their condition. One 64-year-old man with ACS from Germany commented: “Life should still be fun, you have to continue to participate in it, even if you are sick...I do not want to miss on the taste of a beer, a glass of wine or a good meal.”

*Figure 2. The ongoing process of gaining self-care expertise.*
• Controlling: patients have high familiarity with the effects of food and exercise on their body; they constantly learn more about their condition and have a high use of tools to maintain control and promote a feeling of self-reliance. As noted by a 63-year-old man with T2D from the United States: “Being in touch with my body, it has an equal ‘voice’ in determining what is best for me. This helps me to maintain an aggressive mindset and to immediately determine any abnormalities I might be experiencing and take appropriate action.”

• Reframing: people have learned to control their disease and focus on achieving life goals; they use organic methods and vitamins or supplements combined with physical exercise, breathing techniques and meditation to reduce stress. A 61-year-old man with T2D indicated: “Just because someone has a Medical degree doesn’t mean they don’t make mistakes, listen to their advice, but also do some research for yourself. Never blindly follow anyone or anything but find your own ways.”

These mind states may share common traits through their drivers for engagement, as detailed in Figure 1. Through this survey, each mind state is not exclusively self-, context-, or condition-driven, with each having a predominant motivation.

**Figure 3.** The Patient Mind States Model (PMM)—this model contains patient mind states regarding disease self-care identified in patients with chronic cardiometabolic disease.

---

**Discussion**

**Principal Findings**

Each mind state in the PMM is associated with specific beliefs and challenges and is susceptible to particular drivers of engagement that can be used to propel effective self-care behavior (Multimedia Appendix 1). In the ignoring mind state, these drivers of engagement include a fear of mortality and sense of urgency, social pressure from loved ones, and relatable role models who defy the stigma. As for the “why me” frame of mind of the struggling mind state, affirming feedback from HCPs, realistic prognosis that shows possible points for medical intervention, and a sense of urgency that prioritizes the disease serve as drivers of engagement. In the juggling mind state, a mind state that focuses on the positive aspects of life as part of the healing process, clear images of cause and effect, gratification through joyful activities, and companionship with peer patients on their disease journey drive engagement. Moreover, people in the controlling mind state are driven to engage through curiosity for, and excitement about, innovation and what is novel; recognition of progress and being awarded for success; and having a sense of medical expertise. When patients are in the reframing mind state, drivers of engagement include an awareness that certain rules are malleable, deep trust in the individual’s own capabilities, and joy of helping peer sufferers. These emotional and social drivers of engagement across each mind state highlight the need for a holistic approach beyond the current physiological and intellectual drivers of engagement in health care, from gratification through device use to feelings of self-efficacy via immediate feedback and to encourage patient self-care behaviors.

Partitioning a person’s mind state into different stages is not a recent concept, with frameworks such as the transtheoretical model (TTM) examining different behavioral states that evolve over time [36]. The TTM consists of several stages of change and different behavioral processes that drive people to transition to a different state (eg, contemplation, determination, and action). However, the underlying assumption behind this model that decision-making is linear and unidirectional does not recognize how people may move back and forth between such...
states over their lifetime, further deviating from expected behavior change trajectory by exponentially changing social and technological context. The PMM described in this survey allows us to nuance further TTM by its nonlinear nature where patient mind states do not necessarily occur in sequential fashion and may even coexist.

It should also be noted that existing behavioral models are used as tools to support clinical decision-making. In diabetes care, the PAM has been leveraged to predict the potential course of outcomes and how underlying social factors may contribute to activation levels. Mean scores obtained from various PAM instruments, such as PAM-13, may offer a concise summary of a person’s knowledge, skills, and motivation. However, the assessment of drivers underlying activation levels and broader consideration of patient context and its evolution are often decoupled from a PAM assessment. In a recent study on patient activation in individuals with T2D, generic health status topics, distress, and social support were all assessed in questionnaires separate from the primary PAM survey [37]. The PMM may offer complementary perspective for assessment of the context for an individual’s mind state through 1 centralized survey and establish the patient modes of engagement and wider experiences, as detailed in Figure 1. Further studies will be required to elucidate how the PMM could form the basis of a behavioral tool in practice.

**Opportunities for Future-Focused Digital Solutions**

Digital health care has an important role in chronic patient journey. However, many digital patient support programs may suboptimally tailor and target their support based on these important differences, in part because of a lack of data integration across platforms. Guidance on the day-to-day implementation of digital solutions is lacking, so patients often achieve success through trial and error.

Given a patients’ potential frustration with digital tools, engendering a level of comfort and trust in the technology is an important step to effective self-care. Along with privacy concerns, evidence for digital solutions remains as key challenges in establishing digital health as a viable solution for patient self-care [38-40].

Each patient has a unique experience in self-care, with the previous discussion outlining how we can understand the changing mind state of these individuals and how digital solutions may proffer opportunities to improve self-care.

For chronic care digital interventions, our findings suggest that it is important to tailor support to a patient’s mind state, with personal drivers of engagement potentially leading to optimized patient outcomes, adherence, and self-care expertise. Guiding patients throughout their individual health journey to a life worth living is critical and should be based on the individual, attainable life goals, and intelligently balanced compromises that undergo constant revision in the ever-changing context.

These dynamic patient experience mapping refined with help of the PMM may form a more optimal basis in which to effectively integrate digital solutions that enable and support disease self-care, while considering more holistically the context of those living with cardiometabolic conditions. These insights also may warrant further studies in the field of patient adherence and sustainable behavioral change. It will be of interest to further investigate the underlying motivations behind a change in patient mind state and how digital health care may help move individuals from “struggling” to “controlling” mind states, for instance, and effectively ignite intrinsic patient motivation drivers.

These initial data may form a basis of future studies through the validation and refinement of the PMM and the relationship between patient mind states and chronic disease self-care. In particular, future research should clarify patient self-care behaviors and attitudes toward specific digital health care interventions as a critical part of digital intervention design and development processes and verify that the user experience of participants with these 2 conditions in these 3 countries is consistent with patients in other contexts. Equally, it will be critical to understand more deeply the levers of progressive self-care expertise acquisition and use by patients.

The limitations of this survey include its geographic profile (only 3 countries) and the limited sample size of the patient populations (owing to the specific inclusion criteria). Participants who completed the survey may not be representative of the general patient population, as the survey was conducted via the web and patients participated on a voluntary basis. This approach may have introduced a selection bias, such that only the most motivated or educated patients were included. The educational needs of a representative patient population may have therefore been underestimated. Further, the PMM has been generated based on the inputs from participants who were managing ACS and T2D and has not included people living with cardiovascular chronic conditions alone or type 1 diabetes. This may result in that the PMM may not fully embrace the whole spectrum of cardiometabolic patient profiles. The focus on the social aspect of survey respondents (role of caregivers in building patient’s self-care motivation) has not been sufficient to frame more distinctly in the PMM. However, this survey had reflected the real-life experiences of patients in different clinical and geographic settings.

**Conclusions**

Any single journey with an acute or chronic condition is consistently shaped by moments, as well as motivational triggers, which may impact a patient’s growing expertise in self-care. Through this behavioral science survey, a heuristically useful framework has been defined on the underlying nature of how patients engage with self-care, which requires further testing and adaptations. Patients will gradually gain expertise in self-care before acquiring more confidence in proactively using a range of practices and tools. However, this path may be shaped by the patient’s mind state, which will impact treatment adherence and their willingness to engage in self-care practices.

For digital health care solutions to be fully integrated into the patient care journey, it is important to understand how such tools should be tailored to a patient’s mind state and how these states may shift when digital solutions are adopted. It will also be important to understand that such solutions may adapt according to changes in patient mind states.
Acknowledgments
Medical writing support for this paper was provided by Matthew Gunther, a medical writer for Ashfield Health, a UDG Healthcare company. Writing support was funded by Sanofi.

Data Availability
Qualified researchers may request access to documents related to this survey. Further details on Sanofi’s data sharing criteria, eligible studies, and process for requesting access can be found online. This survey was funded by Sanofi.

Conflicts of Interest
SDG received consultancy fees from Novartis and Sanofi. CD, CM, JL, and MM are Sanofi employees and may hold shares or stock options in the company. The other authors have nothing to disclose.

Multimedia Appendix 1
Key features of different patient mind states.

References


Abbreviations

ACS: acute coronary syndrome
HCP: health care professional
PAM: Patient Activation Measure
PMM: Patient Mind States Model
T2D: type 2 diabetes
TTM: transtheoretical model
Linguistic Variables and Gender Differences Within a Messenger-Based Psychosocial Chat Counseling Service for Children and Adolescents: Cross-Sectional Study

Zeki Efe, MSc; Sabrina Baldofski, PhD; Elisabeth Kohls, PhD; Melanie Eckert, MA; Shadi Saee, MSc; Julia Thomas, MA; Richard Wundrack, MA; Christine Rummel-Kluge, MD

1Department of Psychiatry and Psychotherapy, Medical Faculty, Leipzig University, Leipzig, Germany
2Department of Psychiatry and Psychotherapy, University Leipzig Medical Center, Leipzig University, Leipzig, Germany
3Krisenchat gGmbH, Berlin, Germany
4Department of Psychology, Chair of Personality Psychology, Humboldt Universität zu Berlin, Berlin, Germany

Corresponding Author:
Christine Rummel-Kluge, MD
Department of Psychiatry and Psychotherapy
Medical Faculty
Leipzig University
Semmelweissstr 10
Haus 13
Leipzig, 04103
Germany
Phone: 49 341 9724464
Email: Christine.Rummel-Kluge@medizin.uni-leipzig.de

Abstract

Background: Text messaging is widely used by young people for communicating and seeking mental health support through chat-based helplines. However, written communication lacks nonverbal cues, and language usage is an important source of information about a person’s mental health state and is known to be a marker for psychopathology.

Objective: The aim of the study was to investigate language usage, and its gender differences and associations with the presence of psychiatric symptoms within a chat counseling service for adolescents and young adults.

Methods: For this study, the anonymized chat content of a German messenger–based psychosocial chat counseling service for children and adolescents ("krisenchat") between May 2020 and July 2021 was analyzed. In total, 661,131 messages from 6962 users were evaluated using Linguistic Inquiry and Word Count, considering the following linguistic variables: first-person singular and plural pronouns, negations, positive and negative emotion words, insight words, and causation words. Descriptive analyses were performed, and gender differences of those variables were evaluated. Finally, a binary logistic regression analysis examined the predictive value of linguistic variables on the presence of psychiatric symptoms.

Results: Across all analyzed chats, first-person singular pronouns were used most frequently (965,542/8,328,309, 11.6%), followed by positive emotion words (408,087/8,328,309, 4.9%), insight words (341,460/8,328,309, 4.1%), negations (316,475/8,328,309, 3.8%), negative emotion words (266,505/8,328,309, 3.2%), causation words (241,520/8,328,309, 2.9%), and first-person plural pronouns (499,698/8,328,309, 0.6%). Female users and users identifying as diverse used significantly more first-person singular pronouns and insight words than male users (both P<.001). Negations were significantly more used by female users than male users or users identifying as diverse (P=.007). Similar findings were noted for negative emotion words (P=.01). The regression model of predicting psychiatric symptoms by linguistic variables was significant and indicated that increased use of first-person singular pronouns (odds ratio [OR] 1.05), negations (OR 1.11), and negative emotion words (OR 1.15) was positively associated with the presence of psychiatric symptoms, whereas increased use of first-person plural pronouns (OR 0.39) and causation words (OR 0.90) was negatively associated with the presence of psychiatric symptoms. Suicidality, self-harm, and depression showed the most significant correlations with linguistic variables.

Conclusions: This study highlights the importance of examining linguistic features in chat counseling contexts. By integrating psycholinguistic findings into counseling practice, counselors may better understand users' psychological processes and provide more targeted support. For instance, certain linguistic features, such as high use of first-person singular pronouns, negations, or
negative emotion words, may indicate the presence of psychiatric symptoms, particularly among female users and users identifying as diverse. Further research is needed to provide an in-depth look into language processes within chat counseling services.

**KEYWORDS**

e-mental health; chat counseling; crisis; helpline; linguistic; language; Linguistic Inquiry and Word Count; LIWC; psychiatric symptoms

**Introduction**

Childhood and adolescence are known for their biological, social, and psychological changes as vulnerable periods, in which young people are at an increased risk for experiencing mental health problems. It is also known that an early age of onset of mental illness is a risk factor for poor mental health conditions in adulthood [1,2]. The use of mental health care services for adolescents and young adults can have a positive influence on their attitudes, beliefs, and behaviors, which are known to be important predictors of their later mental health [3]. A growing number of studies indicate that children, adolescents, and young adults use the internet to seek help for their mental health problems because the digital environment is familiar and easily accessible, offers anonymity, and accommodates their need for independence [4-8]. Nearly all young people aged 12 to 19 years in Germany (94%) own a smartphone [9].

With the increased use of smartphones, text messaging has become the primary communication tool for today’s youth [10]. Studies on text messaging usage with mobile phones have shown that adolescents experience text messaging as a quick, easy, convenient, playful, and inexpensive way of communication [11-14]. In line with this trend, a number of crisis helplines and similar services have begun to offer online support services such as chat or email counseling [15-18]. Studies have shown that adolescents prefer texting to talking when seeking help for mental health problems and find it easier to write than to express serious concerns verbally [18,19]. Recent studies support the acceptance, feasibility, and usability of online support services, especially among young people [7,20].

However, written language lacks nonverbal stimuli. Recent research has shown that facial expressions and prosody have an influence on the recognition of a speaker’s intention in face-to-face communication [21,22]. In fact, in comparison to face-to-face interactions, people report higher levels of miscommunication when texting. This might represent a barrier in messenger-based counseling and may make it difficult for young people to understand and interpret the intentions of online counselors [23,24]. This lack of nonverbal cues can also make it more difficult for crisis line counselors to establish and maintain a therapeutic relationship [25,26]. In some studies, counselors reported greater difficulty and a lower perceived ability to establish a therapeutic relationship in the digital environment compared to a face-to-face counseling or therapy setting [25,27]. In addition to nonverbal stimuli, language usage is an important source of information in the therapeutic context, and the way people use words conveys a great deal of information about themselves and their current situation [28].

Language reflects both conscious and unconscious thoughts and feelings [29]. Linguists distinguish 2 aspects in the study of language: the formal and the content features of language. The formal aspect concerns grammar, syntax, reaction speed, speech tempo, etc, whereas content features consider vocabulary and word choice [30]. The investigation of “lexical diversity” thus allows a better insight into the cognitive diversity of people [31,32].

Thus, in the 1970s, it was evidenced for the first time that language use can be a specific marker for psychopathology, especially depression [30]. It was found that individuals with depression use more first-person singular pronouns (ie, “I,” “my,” “me,” and “mine”) in both spoken and written language [33,34], supporting cognitive theories of depression [33], which indicate that depression is associated with an increased self-focus. Recent research has shown that increased use of certain words, for example, sad (eg, “crying,” “grief,” and “sad”) or sleep (eg, “asleep” and “bed”), correlates positively with higher levels of depressive symptoms [35]. Further studies also found gender differences in language usage. For example, it was found that women tend to use more language related to thoughts, emotions, senses, negations, and verbs in the present or past tense than men [36]. Furthermore, women were shown to be more likely to use first-person singular pronouns than men, which is consistent with the higher prevalence of depression in women [36,37]. Besides first-person singular pronouns and negative emotion words, causation words (eg, “because”) were also found to be used more by people having depression [34,38]. In line with this, research confirmed the Seligman theory for learned helplessness, which postulates that individuals at risk for depression attribute the cause of a negative event as being internal, global, and stable, by showing that young adults with negative attributional styles were more likely to develop clinically significant depression than those without such attributional styles [39]. There are also studies indicating negative attributional styles as predictive factors for developing depressive symptoms when experiencing negative life events [40-42]. Due to the trend of digitalization in mental health care, linguistic investigations have been conducted in the digital environment as well. In doing so, a positive association was found between Twitter posts indicating loneliness and mental health problems of the users [43]. In another study, which examined the language usage of users of loneliness forums, it was found that these users tended to use words associated with sadness or a desire for social contact, that is, their overall language leaned toward words with negative valence [44]. Regarding associations with psychopathology, the newest findings indicate that individuals with depressive symptoms used fewer complex syntactic constructions, such as adverbial phrases, perhaps because these require greater cognitive effort.
The population of young smartphone users and texters remains a vulnerable and underserved group in crisis counseling, which is why further research on outcomes and the effectiveness of specific communication and counseling strategies is needed [45,46]. To date, there have been no attempts to examine the chat content of crisis counseling services with regard to their linguistic structure. For this purpose, anonymized chat messages from a messenger-based psychosocial chat counseling service, krisenchat (German for “crisis chat”), were used to examine (1) which linguistic indicators and gender differences can be identified within the messages of chat users and (2) how these linguistic indicators are associated with the presence of psychiatric symptoms. Based on the existing literature, it was hypothesized that female users would be more likely to use first-person singular pronouns, negations, and insight words than male users and users identifying as diverse. Additionally, it was hypothesized that higher use of first-person singular pronouns, negations, negative emotion words, and words indicating causation would be associated with a higher likelihood of the presence of psychiatric symptoms among users.

Methods

Sampling and Data Collection

For the purpose of this study, anonymized chat data from all users receiving counseling between May 2020 and July 2021 were extracted from the krisenchat database. Data extraction and preparation were performed by authors affiliated with krisenchat (ME, SS, JT, and RW) so that chat content remained within the krisenchat database. The anonymized chat data included metadata on the chat (total number of messages and words sent by users during the whole counseling process, and number of sessions) and information about the user that counselors identified and noted during the counseling process (sociodemographic information, such as gender and age, and topics of users’ concerns). krisenchat counselors were volunteers and had a background in psychosocial studies. In addition, they underwent a structured 2-month training in chat-based counseling. Regarding gender, counselors had 3 options (male, female, and diverse) to mark in their documentation. They were encouraged to record the identified gender and not the biological sex of the users. The gender “diverse” included individuals identifying as nonbinary or diverse, or indicating to be unsure about their gender identity. For more information about the study design and the nature of krisenchat, we referred to the initial evaluation study of krisenchat [20]. Linguistic variables were determined using Linguistic Inquiry and Word Count (LIWC; see below for details).

The sample examined in this study was based on the previous evaluation of krisenchat, in which the sample consisted of those who completed a subsequent feedback survey after the counseling session [20]. Thus, out of a total of 11,031 users in the above-mentioned time period, 6962 (63.1%) completed a feedback survey. The chat messages of these 6962 users were analyzed. In total, 661,131 messages (mean 94.96, SD 259.46) from 26,614 chat sessions (mean 3.82, SD 6.24) with a total word count of 8,872,154 (mean 1274.37, SD 2954.57) were analyzed.

Psychiatric Symptoms

The presence of psychiatric symptoms was assessed during the counseling process and noted by krisenchat counselors. The identification of psychiatric symptoms was derived from the concerns reported by the users. Counselors distinguished between the presence of the following symptoms: depression, anxiety, suicidality, self-harm, addictive behavior, eating disorders, flashbacks, and obsessive-compulsive symptoms. Additionally, symptoms were summed up into a dichotomous variable “psychiatric symptoms” to indicate the presence or absence of psychiatric symptoms (0, “not present”; 1, “present”).

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 27.0 (IBM Corp.). A 2-tailed α value of .05 was applied to statistical testing. First, descriptive statistics were performed for sociodemographic variables and linguistic variables of the total sample. Additionally, Kruskal-Wallis H tests (because of nonnormality of the linguistic variables) were used to identify

Ethical Considerations

The Medical Faculty of the University of Leipzig approved this study on August 3, 2021 (372/21-ek). Users were informed about the data protection and privacy policy of krisenchat when they first contacted the counseling service. The chat counseling only began after confirming the policy with “Yes.” Participants in the study confirmed informed consent via an opt-in function before taking part in the feedback survey.

Measures

Linguistic Variables

LIWC is a software for dictionary-based quantitative text analysis [29]. LIWC performs an automated 1-word analysis based on a lexicon with more than 80 categories (ie, language variables, descriptors, linguistic dimensions, psychological dimensions, concerns, informal language, and punctuation) including a total of 18,711 words. In 2008, the German version of the lexicon was developed, and good equivalence was confirmed for the majority of LIWC categories [47]. The tool has been used in various studies on personality, social, and clinical psychological frameworks and for the analysis of therapeutic essays, everyday communication, or computer-based communication, and it can therefore be considered a reliable software program for quantitative text analysis [47-52]. LIWC counts the number of words within the lexicon over a whole chat and assigns them to categories. The output file includes all categories of the lexicon. All variables, except summary variables, are expressed as percentages of the total word count of a respective chat. Based on previous findings [34-37,44,53,54], the following linguistic variables were considered in this study: first-person singular (eg, “I,” “me,” and “mine”) and first-person plural pronouns (eg, “we,” “us,” and “our”), negations (eg, “no,” “not,” and “never”), positive emotion words (eg, “love,” “nice,” and “sweet”), negative emotion words (eg, “hurt,” “worried,” and “sad”), cognitive process words such as words related to insight (eg, “think” and “know”), and words related to causation (eg, “because” and “effect”).
gender differences in use, that is, metadata for the number of sessions, messages, and words of each user. Then, a 1-way multivariate analysis of variance (1-way MANOVA) was conducted to test for gender differences in language usage controlling for word count. Gender was considered as an independent variable, and all 7 linguistic variables (ie, first-person singular pronouns, first-person plural pronouns, negations, positive emotion words, negative emotion words, insight words, and causation words) were considered as dependent variables. Post-hoc univariate ANOVAs were conducted separately for every linguistic variable. Bonferroni correction was applied to account for multiple testing. Then, binary logistic regression analysis was conducted to examine the predictive effect of linguistic variables (first-person singular and plural pronouns, negations, positive and negative emotion words, insight words, and causation words), age, and gender (recoded into a set of dummy variables with “male” as the reference variable) on the presence of psychiatric symptoms.

The amount of explained variance as shown by Nagelkerke $R^2$ was interpreted as follows: $R^2 > 0.20$, “acceptable” or small effect size; $R^2 > 0.40$, “good” or average effect size; and $R^2 > 0.50$, “very good” or large effect size [55]. Additionally, Spearman correlation coefficients ($\rho$) were reported between linguistic variables. Finally, with the aim to examine the deeper relationship between linguistic variables and psychiatric symptoms, explorative Spearman correlations ($\rho$) between all 7 linguistic variables and all categories of psychiatric symptoms (suicidality, self-harm, depression, anxiety, eating disorder symptoms, flashbacks, obsessive-compulsive symptoms, and addictive behavior) were computed and interpreted as follows: $\rho = 0.10$, small effect size; $\rho = 0.30$, moderate effect size; and $\rho = 0.50$, large effect size [56].

### Descriptive Statistics of Linguistic Variables

In the total sample, the mean percentage of first-person singular pronouns among all words of a user during the whole counseling process was 11.59% (SD 2.46%), indicating that on average, more than one-tenth of all words written throughout all chat messages was a first-person singular pronoun (“I,” “me,” “my,” or “mine”). The next most used linguistic categories were positive emotion words (mean 4.85%, SD 1.70%) and insight words (mean 4.05%, SD 1.32%). The mean percentage of negations among all words of a user during the whole counseling process was 3.76% (SD 1.54%). Furthermore, negative emotion words were used with a mean percentage among all words of 3.23% (SD 1.27%). Causation words were used with a mean percentage among all words of 2.93% (SD 1.04%). Finally, first-person plural pronouns were least frequently used with a mean percentage among all words of 0.43% (SD 0.63%).

### Gender Differences in Linguistic Variables

Gender-specific descriptive statistics are displayed in Table 2. A 1-way MANOVA showed statistically significant differences in linguistic variables between genders ($F_{14, 11,932}=8.945$; $P<.001$; partial $\eta^2=0.01$; Wilk $\Lambda=0.979$). Post-hoc univariate ANOVAs were conducted separately for every linguistic variable. Separate ANOVAs and respective Bonferroni-corrected post-hoc tests showed that when controlling for word count, there were statistically significant differences in the use of first-person singular pronouns between genders ($F_{2,5972}=49.780$; $P<.001$; partial $\eta^2=0.05$) and sent significantly more messages ($z=-4.211$; $P<.001$; $r=0.05$) and sent significantly more messages ($z=-3.247$; $P<.001$; $r=0.04$) and words ($z=-5.349$; $P<.001$; $r=0.07$) than male users, whereas there were no significant differences between female users and users identifying as diverse (number of sessions: $P=0.13$; number of messages: $P=0.18$; number of words: $P=0.22$). Users identifying as diverse also attended a significantly higher number of sessions ($z=-3.441$; $P<.001$; $r=0.04$) and sent significantly more messages ($z=-2.972$; $P<.001$; $r=0.04$) and words ($z=-3.639$; $P<.001$; $r=0.05$) than male users.

### Results

#### Sociodemographic Characteristics

The average user was 17 years old (mean 16.55, SD 3.45 years; range 8-25 years), and most users were female (female: 4988/5978, 83.4%; male: 881/5978, 14.7%; diverse: 109/5978, 1.8%). A large number of all users (4841/6962, 69.5%) contacted the counseling service due to psychiatric symptoms. Further concerns identified were psychosocial distress (eg, school-related problems, family-related problems, bullying, etc; 2370/6962, 34.0%) or emotional distress (eg, grief, lovesickness, anger, and loneliness; 2101/6962, 30.2%) [20].

The users participated in an average of 3.82 (SD 6.24) counseling sessions and sent an average of 94.96 (SD 259.46; range 2-11,512) messages with an average of 1274.37 (SD 2954.57) words throughout the counseling process. Additional testing indicated that there were gender differences in the numbers of sessions ($\chi^2=22.849$; $P<.001$), messages ($\chi^2=14.863$; $P<.001$), and words ($\chi^2=33.036$; $P<.001$). The results are presented in Table 1. Subsequent post-hoc tests indicated that female users attended a significantly higher number of sessions ($z=-4.211$; $P<.001$; $r=0.05$) and sent significantly more messages ($z=-3.247$; $P<.001$; $r=0.04$) and words ($z=-5.349$; $P<.001$; $r=0.07$) than male users, whereas there were no significant differences between female users and users identifying as diverse (number of sessions: $P=0.13$; number of messages: $P=0.18$; number of words: $P=0.22$). Users identifying as diverse also attended a significantly higher number of sessions ($z=-3.441$; $P<.001$; $r=0.04$) and sent significantly more messages ($z=-2.972$; $P<.001$; $r=0.04$) and words ($z=-3.639$; $P<.001$; $r=0.05$) than male users.

### Table 1. Gender-specific differences in metadata (N=5978).

<table>
<thead>
<tr>
<th>Metadata variables</th>
<th>Male^a, mean (SD)</th>
<th>Female^a, mean (SD)</th>
<th>Diverse^a, mean (SD)</th>
<th>$\chi^2$ (df)^b</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session count</td>
<td>3.23 (4.17)^c</td>
<td>4.17 (6.94)^d</td>
<td>5.43 (7.56)^d</td>
<td>22.849 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Message count</td>
<td>79.05 (145.20)^c</td>
<td>103.74 (287.43)^d</td>
<td>189.14 (512.32)^d</td>
<td>14.863 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Word count</td>
<td>1068.80 (1965.98)^c</td>
<td>1392.55 (3279.22)^d</td>
<td>2166.18 (4653.08)^d</td>
<td>33.036 (2)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

^aReduced sample size owing to missing data on gender.

^bTest statistic for the Kruskal-Wallis $H$ test.

^cDifferent letters indicate significant differences between the groups, while same letters indicate no significant differences between the groups.

### References

[55] Spearman correlation coefficients ($\rho$) between all 7 linguistic variables and all categories of psychiatric symptoms (suicidality, self-harm, depression, anxiety, eating disorder symptoms, flashbacks, obsessive-compulsive symptoms, and addictive behavior) were computed and interpreted as follows: $\rho = 0.10$, small effect size; $\rho = 0.30$, moderate effect size; and $\rho = 0.50$, large effect size [56].

[56] Reduced sample size owing to missing data on gender.

[57] Test statistic for the Kruskal-Wallis $H$ test.

[58] Different letters indicate significant differences between the groups, while same letters indicate no significant differences between the groups.
Predicting Psychiatric Symptoms by Linguistic Variables

The binomial logistic regression model was statistically significant ($\chi^2_3=25.0; P=.002$), resulting in a small amount of explained variance, as shown by Nagelkerke $R^2=0.124$ (Table 3). Of the 10 variables entered into the regression model, all but 3 contributed significantly to the presence of psychiatric symptoms: first-person singular and plural pronouns, negations, negative emotion words, causation words (all $P<.001$), and female gender ($P=.005$), while positive emotion words ($P=.08$), insight words ($P=.90$), and diverse gender ($P=.57$) showed no significant effects. Using first-person plural pronouns was associated with a lower likelihood of reporting psychiatric symptoms (odds ratio [OR] 0.39), as did using more causation words (OR 0.90). In contrast, a higher use of first-person singular pronouns was associated with an increased likelihood of reporting psychiatric symptoms (OR 1.18), as did using more negative emotion words (OR 0.39), as did using more negations (OR 1.11) or negative emotion words (OR 1.15). Finally, being female (OR 1.18) or having a higher age (OR 1.04) was also associated with an increased likelihood of the presence of psychiatric symptoms.

Table 2. Gender-specific differences in language usage (N=5978).

<table>
<thead>
<tr>
<th>Linguistic variables</th>
<th>Male$^a$ (n=881), mean (SD)</th>
<th>Female$^a$ (n=4988), mean (SD)</th>
<th>Diverse$^a$ (n=109), mean (SD)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-person singular pronouns</td>
<td>10.81 (2.60)$^b$</td>
<td>11.71 (2.43)$^c$</td>
<td>11.79 (2.55)$^c$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>First-person plural pronouns</td>
<td>0.47 (0.84)</td>
<td>0.42 (0.59)</td>
<td>0.32 (0.43)</td>
<td>.05</td>
</tr>
<tr>
<td>Negations</td>
<td>3.62 (2.24)$^b$</td>
<td>3.76 (1.39)$^c,d$</td>
<td>3.94 (1.49)$^b,d$</td>
<td>.007</td>
</tr>
<tr>
<td>Positive emotions</td>
<td>4.87 (1.80)</td>
<td>4.86 (1.67)</td>
<td>4.68 (1.79)</td>
<td>.61</td>
</tr>
<tr>
<td>Negative emotions</td>
<td>3.13 (1.37)$^b$</td>
<td>3.24 (1.24)$^e$</td>
<td>3.02 (1.43)$^b,c$</td>
<td>.01</td>
</tr>
<tr>
<td>Insight words</td>
<td>3.81 (1.38)$^b$</td>
<td>4.07 (1.30)$^e$</td>
<td>4.20 (1.42)$^e$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Causation words</td>
<td>2.99 (1.12)</td>
<td>2.92 (1.02)</td>
<td>3.04 (0.92)</td>
<td>.09</td>
</tr>
</tbody>
</table>

$^a$Reduced sample size owing to missing data on gender.

$^{b,c,d}Different letters indicate significant differences between the groups, while same letters indicate no significant differences between the groups.

$\chi^2=25.0; P=.002$
Table 3. Prediction of psychiatric symptoms by language usage (N=6962).

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>P value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.04</td>
<td>0.01</td>
<td>22.04</td>
<td>&lt;.001</td>
<td>1.04 (1.02-1.06)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.16</td>
<td>0.06</td>
<td>7.94</td>
<td>.005</td>
<td>1.18 (1.05-1.32)</td>
</tr>
<tr>
<td>Diverse</td>
<td>0.12</td>
<td>0.21</td>
<td>0.33</td>
<td>.57</td>
<td>1.13 (0.75-1.70)</td>
</tr>
<tr>
<td>First-person singular pronouns</td>
<td>0.05</td>
<td>0.01</td>
<td>15.78</td>
<td>&lt;.001</td>
<td>1.05 (1.03-1.08)</td>
</tr>
<tr>
<td>First-person plural pronouns</td>
<td>−0.94</td>
<td>0.05</td>
<td>301.21</td>
<td>&lt;.001</td>
<td>0.39 (0.35-0.44)</td>
</tr>
<tr>
<td>Negations</td>
<td>0.11</td>
<td>0.02</td>
<td>32.14</td>
<td>&lt;.001</td>
<td>1.11 (1.07-1.16)</td>
</tr>
<tr>
<td>Positive emotions</td>
<td>−0.03</td>
<td>0.02</td>
<td>3.09</td>
<td>.08</td>
<td>0.97 (0.95-1.01)</td>
</tr>
<tr>
<td>Negative emotions</td>
<td>0.14</td>
<td>0.02</td>
<td>43.45</td>
<td>&lt;.001</td>
<td>1.15 (1.11-1.21)</td>
</tr>
<tr>
<td>Insight words</td>
<td>−0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>.90</td>
<td>1.00 (0.96-1.04)</td>
</tr>
<tr>
<td>Causation words</td>
<td>−0.11</td>
<td>0.03</td>
<td>13.80</td>
<td>&lt;.001</td>
<td>0.90 (0.85-0.94)</td>
</tr>
<tr>
<td>Constant</td>
<td>−0.87</td>
<td>0.24</td>
<td>13.07</td>
<td>&lt;.001</td>
<td>0.42 (N/A)</td>
</tr>
</tbody>
</table>

a: B: regression coefficient.  
b: Degrees of freedom were 1 for all Wald statistics.  
c: OR: odds ratio.  
d: N/A: not applicable.

The results of an additional correlation analysis between all linguistic variables are reported in Table 4. Among others, significant findings included a negative association between first-person singular pronouns and first-person plural pronouns (ρ=-0.24; P<.001). In line with this, first-person singular pronouns were positively correlated with negations (ρ=0.17; P<.001) and negative emotion words (ρ=0.15; P<.001).
### Table 4. Spearman correlation coefficients between linguistic variables (N=6962).

<table>
<thead>
<tr>
<th>Variable</th>
<th>First-person singular pronouns</th>
<th>First-person plural pronouns</th>
<th>Negations</th>
<th>Positive emotions</th>
<th>Negative emotions</th>
<th>Insight words</th>
<th>Causation words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First-person singular pronouns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>1</td>
<td>-0.24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.02</td>
</tr>
<tr>
<td>P value</td>
<td>_b</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.02</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>22</td>
</tr>
<tr>
<td><strong>First-person plural pronouns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>-0.24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>-0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.01</td>
<td>-0.10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.07&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>_b</td>
<td>&lt;.001</td>
<td>.34</td>
<td>&lt;.001</td>
<td>.01</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Negations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>-0.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.02</td>
<td>0.13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>_b</td>
<td>.045</td>
<td>.19</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Positive emotions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.01</td>
<td>-0.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>-0.06&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>.002</td>
<td>.34</td>
<td>.045</td>
<td>_b</td>
<td>&lt;.001</td>
<td>.023</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Negative emotions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.02</td>
<td>-0.06&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.07&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.19</td>
<td>&lt;.001</td>
<td>_b</td>
<td>.023</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Insight words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.01</td>
<td>&lt;.001</td>
<td>.023</td>
<td>.023</td>
<td>_b</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Causation words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>-0.02</td>
<td>-0.07&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.07&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td>P value</td>
<td>.22</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>_b</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Statistical significance.

<sup>b</sup>Not applicable.

### Associations Between Linguistic Variables and Psychiatric Symptoms

Finally, an exploratory correlation analysis indicated evidence for the relationship between linguistic variables and various psychiatric symptoms, which are displayed in Table 5. Suicidality, self-harm, depression, and anxiety showed the most significant correlations to linguistic variables. In particular, the use of first-person singular pronouns was positively associated with suicidality (ρ=0.11; P<.001) and self-harm (ρ=0.10; P<.001). The use of first-person plural pronouns was negatively associated with suicidality (ρ=−0.10; P<.001) and depression (ρ=−0.14; P<.001). The use of negations was positively associated with suicidality (ρ=0.18; P<.001) and self-harm (ρ=0.12; P<.001). Finally, the use of negative emotion words was positively associated with depression (ρ=0.15; P<.001) and anxiety (ρ=0.15; P<.001).
### Table 5. Spearman correlation coefficients between linguistic variables and psychiatric symptoms (N=6962).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Suicidality</th>
<th>Self-harm</th>
<th>Depression</th>
<th>Anxiety</th>
<th>Eating disorder symptoms</th>
<th>Flashbacks</th>
<th>Obsessive-compulsive symptoms</th>
<th>Addictive behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First-person singular pronouns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.11$^a$</td>
<td>0.10$^a$</td>
<td>0.08$^a$</td>
<td>−0.01</td>
<td>0.05$^a$</td>
<td>−0.02</td>
<td>0.00</td>
<td>−0.01</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.46</td>
<td>&lt;.001</td>
<td>.08</td>
<td>.75</td>
<td>.33</td>
</tr>
<tr>
<td><strong>First-person plural pronouns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>−0.10$^a$</td>
<td>−0.08$^a$</td>
<td>−0.14$^a$</td>
<td>−0.04$^a$</td>
<td>−0.04$^a$</td>
<td>−0.01</td>
<td>−0.03$^a$</td>
<td>−0.04$^a$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.26</td>
<td>.007</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>Negations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.18$^a$</td>
<td>0.12$^a$</td>
<td>−0.06$^a$</td>
<td>−0.06$^a$</td>
<td>0.01</td>
<td>−0.02</td>
<td>−0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.40</td>
<td>.07</td>
<td>.28</td>
<td>.80</td>
</tr>
<tr>
<td><strong>Positive emotions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>−0.05$^a$</td>
<td>−0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>−0.02</td>
<td>−0.02</td>
<td>−0.02</td>
<td>−0.03$^a$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.08</td>
<td>.41</td>
<td>.87</td>
<td>.06</td>
<td>.19</td>
<td>.09</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Negative emotions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.04$^a$</td>
<td>0.03$^a$</td>
<td>0.12$^a$</td>
<td>0.15$^a$</td>
<td>−0.03$^a$</td>
<td>0.02</td>
<td>0.02$^a$</td>
<td>0.00</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.002</td>
<td>.03</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.02</td>
<td>.18</td>
<td>.045</td>
<td>.82</td>
</tr>
<tr>
<td><strong>Insight words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.03$^a$</td>
<td>−0.00</td>
<td>0.08$^a$</td>
<td>−0.01</td>
<td>−0.01</td>
<td>−0.01</td>
<td>0.00$^a$</td>
<td>−0.04$^a$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.005</td>
<td>.73</td>
<td>&lt;.001</td>
<td>.65</td>
<td>.67</td>
<td>.61</td>
<td>.045</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Causation words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>−0.06$^a$</td>
<td>−0.03$^a$</td>
<td>−0.03$^a$</td>
<td>−0.02</td>
<td>−0.02</td>
<td>−0.04$^a$</td>
<td>0.01</td>
<td>−0.01</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.007</td>
<td>.03</td>
<td>.10</td>
<td>.13</td>
<td>&lt;.001</td>
<td>.63</td>
<td>.32</td>
</tr>
</tbody>
</table>

$^a$Statistical significance.

### Discussion

**Principal Findings and Comparison With Prior Work**

The findings of this study provide first-time valuable insights into the psycholinguistic characteristics of children, adolescents, and young adults seeking psychosocial support through a messenger-based crisis counseling service (krisenchat). Previous findings that examined psycholinguistic characteristics in association with mental health, which however focused on texts (eg, from social media or online therapies), could be identified in the chat context as well. Specifically, linguistic variables were found to be associated with the presence of psychiatric symptoms [43,44,57]. The use of first-person singular pronouns, negations, and negative emotion words increased the likelihood of the presence of psychiatric symptoms, while the use of first-person plural pronouns and causation words was associated with a lower likelihood of the presence of psychiatric symptoms. Female gender was also associated with an increased likelihood of the presence of psychiatric symptoms, which is consistent with the higher prevalence of psychiatric symptoms in women [58]. Gender differences were found, with female users exhibiting more frequent use of certain linguistic features. Previous linguistic analyses, especially in the digital context, such as social media platforms, have been performed with a focus on the presence of psychiatric symptoms or associations with psychiatric symptoms, in particular, depressive symptoms [59]. Linguistic analyses of social media have proven useful in predicting depression, anxiety, loneliness, personality disorders, or other mental health issues [43,57,60-62]. As there are no other comparative studies in this field, the focus of the below comparison of the present results with previous findings relies on correlates of linguistic variables with the presence of psychiatric symptoms, especially depression or anxiety.

**First-Person Pronouns**

Starting with the most frequently used linguistic variable among those examined, an increased use of first-person singular pronouns was associated with a higher likelihood of the presence of psychiatric symptoms. Additionally, it was determined that they were used more often by female users and users identifying as diverse than by male users. In contrast, the linguistic variable of first-person plural pronouns was found to be the least used and did not predict the presence of psychiatric symptoms.
Taking into account the most frequently mentioned concerns among the users of krisenchat (see [20]), involving psychiatric symptoms, such as depression or anxiety, and the higher prevalence of depression in female samples [63], these findings are consistent with previous findings on language usage. Thus, in line with cognitive theories of depression (eg, according to [33]) indicating that depression is associated with an increased self-focus, previous research showed that individuals with depression used more first-person singular pronouns (ie, “I,” “my,” “me,” and “mine”) in both spoken and written language [33,34]. Similarly, first-person plural pronouns (ie, “we,” “us,” and “our”) were used significantly less by depressed individuals, which may be attributed to social isolation or lack of social integration and social engagement in the context of depression [64-66]. However, the increased use of first-person singular pronouns may also be a marker for increased vulnerability to stress and negative emotionality and not directly for depression [67,68].

Negations and Causation Words

Negations were found to be significantly associated with the presence of psychiatric symptoms and were used more frequently by female users than male users or users identifying as diverse. Contradicting the hypothesis, the present results indicate that the use of causation words reduces the likelihood of the presence of psychiatric symptoms, while no significant differences were found between genders. In line with this finding, previous studies indicated that low use of causation words is associated with positive treatment outcomes in treatment for personality disorders [53]. It was found that the use of fewer cognitive words, such as causation words, was associated with a more coherent personal story [69]. A recent study underlined this finding by pointing out that patients having depression tend to use significantly more aligned sentences than bringing them into a logical chain compared with a healthy control group [34]. This cross-sectional study design does not allow to draw conclusions about the trend of the use of causation words throughout the chat counseling. The meaning of these divergent results deserves further specific longitudinal research on the development and change of language use across chat counseling.

Negative and Positive Emotion Words

Elevated use of negative emotion words was associated with an increase in the presence of psychiatric symptoms. Negative emotion words were found to differ between genders, that is, more negative emotion words were used by female users than by male users. This is in line with previous research indicating a significantly higher general use of emotion words by women than men, while men were found to use more anger words [70,71]. This is also in accordance with findings showing that more frequent use of negative emotion words, including anxiety, sadness, and anger words, was positively correlated with higher anxiety and depression levels [54].

Depressive Symptoms

Taken together, the results underline a higher likelihood of depression (or anxiety) in users using more self-focused language (first-person singular pronouns), and more negative and fewer positive (emotion) words. The response style theory of rumination in depression, which could also be proven for worry in generalized anxiety, explains that symptoms of repetitive self-focused negative thinking become habitual over time [72,73]. In terms of language usage, this theory suggests that people with high levels of depression or anxiety might communicate using more self-focused language (ie, more I-related pronouns), and more negative and fewer positive words, and that this tendency may become habitual and outside of conscious awareness. In line with this, linguistic analyses of text-based therapy found reductions in the use of first-person singular pronouns, even though language usage was not being focused on in the treatment [74]. Moreover, changes in the use of positive and negative emotion words and words indicating certainty (eg, “always” and “never”) could be identified during the treatment for depression. Researchers interpret these findings as changes in cognitive processes [75].

Suicidality

In relation to depression, it is also important to keep suicidality in mind because messenger-based chat counseling services are used in acute crises, such as suicidality [76]. In accordance with the present findings, previous studies found that suicidal behavior is associated with the use of more I-related pronouns [77]. Likewise, in accordance with the present findings, previous literary analyses indicated that suicidal poets also used fewer first-person plural pronouns than nonsuicidal poets [66]. In addition, these studies showed that the use of more absolutist language, that is, superlatives and intensifiers (eg, “absolutely,” “completely,” “all,” “none,” etc), was associated with suicidality [60,77,78].

The results of this study indicate that by considering language usage, differences in the user population can be discovered and may also be linked to psychopathology. Thus, language usage should be integrated into the counseling strategy. In the context of computer-based analyses, it was found that in addition to the standardized diagnostic tools used to confirm a psychiatric diagnosis, linguistic research showed that systematic analyses of clients’ language may be used to reliably classify them into diagnostic groups [34]. Additionally, computer-based methods were shown to distinguish persons with depression from other clinical subgroups [79]. Thus, linguistic or, in general, qualitative analyses of text messages seem to be advisable to examine chat-based counseling services in more depth. This is even more true than for social media platforms because text messages are less influenced by social desirability, facilitating more granular visibility into changes in linguistic patterns [80].

Implications for Counselors in Chat-Based Counseling

Multiple implications for psychosocial chat counseling have emerged from the findings. First, counselors may use psycholinguistic analyses as an additional tool for assessing the mental health state of users. By monitoring the usage of specific linguistic features, counselors can identify individuals at risk for psychiatric symptoms and tailor interventions accordingly. For example, in cases of elevated use of first-person singular pronouns, negations, and negative emotion words, particularly for female users and users who identify as diverse, a more in-depth exploration of psychiatric symptoms, especially
Acknowledgments

We acknowledge support from Leipzig University for open access publishing. The Federal Ministry of Health funded this project (ZMI1-2521FEP001).

Strengths and Limitations

To the best of our knowledge, this study is the first to examine language usage in a messenger-based crisis counseling service among youth and young adults. In addition to the strength of the large sample size, this study acts as a reference and comparison for further studies in this area, in part because of the use of the internationally recognized method LIWC. Nevertheless, some limitations have to be taken into account. Owing to the retrospective study design, the data rely on convenience sampling, which limits the generalizability to a more mixed-gender population. It should also be noted that the counseling service is offered in German-speaking countries, which is why cultural and linguistic differences in language usage must be taken into account, and generalizability is limited. An international comparison between counseling services in different countries, cultures, and languages would provide insights into similarities and differences. Assuming that people in crisis reach out to the chat counseling service, high use of emotion words seems somewhat expected. Since the nature of the study was cross-sectional and the words were counted across all chat messages (ie, in chronological order or within sessions), no data can be provided on the trend of the word count for either positive or negative emotion words, which is why no indications can be derived on whether the word count relates only to the beginning of the chat or also to the progression throughout the chat counseling. Longitudinal studies examining changes in language use across consecutive chat sessions may provide further insights into these associations. Consideration must also be given to the nested data structure, which cannot be clearly read owing to the format used to provide data by LIWC. Therefore, for future studies, in addition to the trend of the word count throughout the chat counseling, the consideration of levels (eg, within a message and during a session) is recommended. Therefore, a qualitative analysis of the chats may be beneficial for providing more in-depth insights into individual language usage as well as concerns. A qualitative approach would also ensure the quality of the classification of the presence of psychiatric symptoms. Likewise, no standardized measurement instruments were used, which in turn opens up further opportunities for future research, for example, implementing symptom-specific questionnaires to examine associations between (changes in) symptom severity and language usage.

Conclusion

This study underlines the options, possibilities, and chances of examining psycholinguistic characteristics in psychosocial online chat counseling services for children and adolescents. The identified associations between specific linguistic features and the presence of psychiatric symptoms provide valuable insights for the development of targeted interventions. By considering psycholinguistic findings in the counseling practice, counselors may enhance their understanding of the psychological processes of users and their interventions to offer a more targeted service for children and adolescents seeking help. Nevertheless, further research is needed to investigate the mechanisms underlying linguistic patterns and explore the effectiveness of linguistic-based interventions. At the same time, this would allow further research on the link between specific indicators and changes in specific psychopathology.
Authors’ Contributions
ZE, SB, and CRK designed the study. The data set was prepared by ME, SS, JT, and RW. ZE performed the statistical analysis. ZE and SB drafted the article. All authors approved the final manuscript.

Conflicts of Interest
ZE, SB, EK, ME, SS, JT, and RW confirm no conflicts of interest. ME, SS, JT, and RW are paid employees at krisenchat gGmbH. CRK received lecture honoraria from Recordati and Servier outside and independent of the submitted work.

References


44. Andy A. Studying How Individuals Who Express the Feeling of Loneliness in an Online Loneliness Forum Communicate in a Nonloneliness Forum: Observational Study. JMIR Form Res 2021 Jul 20;5(7):e28738 [FREE Full text] [doi: 10.2196/28738] [Medline: 34283026]


Abbreviations

LIWC: Linguistic Inquiry and Word Count
OR: odds ratio
Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Rosie, a Health Education Question-and-Answer Chatbot for New Mothers: Randomized Pilot Study

Quynh C Nguyen1, PhD; Elizabeth M Aparicio2, PhD; Michelle Jasczynski3, PhD; Amara Channell Doig2, PhD; Xiaohe Yue1, MSc; Heran Mane1, BSc; Neha Srikanth1, MSc; Francia Ximena Marin Gutierrez2, MSW; Nataly Delcid1; Xin He1, PhD; Jordan Boyd-Graber3, PhD

1Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, MD, United States
2Department of Behavioral and Community Health, University of Maryland School of Public Health, College Park, MD, United States
3Department of Computer Science, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD, United States

Corresponding Author:
Quynh C Nguyen, PhD
Department of Epidemiology and Biostatistics
University of Maryland School of Public Health
4254 Stadium Dr.
College Park, MD, 20742
United States
Phone: 1 301 405 6425
Email: qtnguyen@umd.edu

Abstract

Background: Stark disparities exist in maternal and child outcomes and there is a need to provide timely and accurate health information.

Objective: In this pilot study, we assessed the feasibility and acceptability of a health chatbot for new mothers of color.

Methods: Rosie, a question-and-answer chatbot, was developed as a mobile app and is available to answer questions about pregnancy, parenting, and child development. From January 9, 2023, to February 9, 2023, participants were recruited using social media posts and through engagement with community organizations. Inclusion criteria included being aged ≥14 years, being a woman of color, and either being currently pregnant or having given birth within the past 6 months. Participants were randomly assigned to the Rosie treatment group (15/29, 52% received the Rosie app) or control group (14/29, 48% received a children’s book each month) for 3 months. Those assigned to the treatment group could ask Rosie questions and receive an immediate response generated from Rosie’s knowledgebase. Upon detection of a possible health emergency, Rosie sends emergency resources and relevant hotline information. In addition, a study staff member, who is a clinical social worker, reaches out to the participant within 24 hours to follow up. Preintervention and postintervention tests were completed to qualitatively and quantitatively evaluate Rosie and describe changes across key health outcomes, including postpartum depression and the frequency of emergency room visits. These measurements were used to inform the clinical trial’s sample size calculations.

Results: Of 41 individuals who were screened and eligible, 31 (76%) enrolled and 29 (71%) were retained in the study. More than 87% (13/15) of Rosie treatment group members reported using Rosie daily (5/15, 33%) or weekly (8/15, 53%) across the 3-month study period. Most users reported that Rosie was easy to use (14/15, 93%) and provided responses quickly (13/15, 87%). The remaining issues identified included crashing of the app (8/15, 53%), and users were not satisfied with some of Rosie’s answers (12/15, 80%). Mothers in both the Rosie treatment group and control group experienced a decline in depression scores from pretest to posttest periods, but the decline was statistically significant only among treatment group mothers (P=.008). In addition, a low proportion of treatment group infants had emergency room visits (1/11, 9%) compared with control group members (3/13, 23%). Nonetheless, no between-group differences reached statistical significance at P<.05.

Conclusions: Rosie was found to be an acceptable, feasible, and appropriate intervention for ethnic and racial minority pregnant women and mothers of infants owing to the chatbot’s ability to provide a personalized, flexible tool to increase the timeliness and accessibility of high-quality health information to individuals during a period of elevated health risks for the mother and child.

Trial Registration: ClinicalTrials.gov NCT06053515; https://clinicaltrials.gov/study/NCT06053515
KEYWORDS

chatbot; health information; maternal and child health; health disparities; health equity; health informatics; preventive health care; postpartum care; patient education; newborn care; prenatal care; mobile phone

Introduction

Background

Maternal morbidity and mortality have remained persistent problems in the United States and disproportionately affect women and birthing people from racial and ethnic minoritized backgrounds owing to embedded racism and bias across the medical and public health systems [1-3]. More concerning, >80% of maternal deaths in 2019 were designated as preventable by the Centers for Disease Control and Prevention’s (CDC’s) maternal mortality review committees [4]. In particular, the perinatal period is associated with high risk of depression and anxiety among mothers and is a leading cause of maternal mortality in the United States [1,4,5]. Other leading causes include hemorrhage, cardiovascular and coronary conditions, and substance use disorders [4,5].

Timely and reliable health information may help to reduce the adverse outcomes during pregnancy and in the postpartum period [6]. Children of single parents, with low household income, of a minority group, or whose parents perceived them as being more susceptible are seen more frequently in the emergency department [7-10]. Health education interventions have been shown to reduce emergency department use among infant caregivers [11]. Seeking health information on websites is common among soon-to-be and new parents; however, the quality of information and sources found on the web about pregnancy, birth, parenting, and maternal health were rated by pregnant women and new parents as having varying quality, or the information found was not sufficiently specific to fully answer questions [12,13]. Currently, some popular programs for these vulnerable populations involve resource-intensive home visits, which face challenges in scaling to assist more mothers owing to staff and cost constraints, or nonpersonalized SMS text messages that may not directly address an individual’s questions [14-18].

Recognizing that facilitating maternal and child health equity in the United States will require intervention at all levels of the socioecological model to address the deficits in medical and public health research and practice, our research team selected an innovative approach. This approach consisted of developing a maternal and child health information chatbot that would be iteratively improved through a multiyear, community-engaged research process. The chatbot addresses some limitations of previous strategies by providing personalized health information based on the users’ needs, is readily available at any time, and can include participants nationwide. Rosie offers timely health information from verifiable websites such as children’s hospitals or the CDC to help parents navigate infant care and find clinically correct information to tackle health issues as they arise. In addition, Rosie offers reminders about preventive care visits for infants (eg, well-baby visits) that can also encourage greater continuity of care, which have been shown to reduce emergency department visits for infants [19].

The research team developed the chatbot, named Rosie, to be able to respond to user questions about parenting, pregnancy, and infant development with vetted, trustworthy web-based sources. Question-answering (QA) chatbots such as Rosie, unlike informational sites and frequently asked questions (FAQs) pages, support their users with personalized responses based on the user’s input. They provide users with the unique opportunity to enter their questions in their own words and receive responses to their questions. We built the corpus or index of maternal and child health information using the information derived from expert sources such as federal agencies (eg, the National Institute for Child and Human Development), hospitals (eg, Mayo Clinic), and professional medical organizations (eg, the American Academy of Pediatrics). These sources provided vital information regarding topics such as pregnancy, parenting, infant development, maternal health, and postpartum care.

Chatbots developed to support maternal mental health and parenting have been shown to be a promising intervention needing further evaluation [20-22]. A mixed methods review of literature led by Chua et al [20] suggests that maternal and child health information chatbots have high acceptance among pregnant women and new parents; however, the reviewed papers noted that both development teams and test users expressed preferences for refining the language used in the responses to be more humanlike and for the chatbots to be familiarized with informal, descriptive language to be more adept at generating answers for users who may be describing symptoms or may not know the medical term for the topic of interest. Recognizing these recommendations from the literature, our research team used a multimethod approach to receive substantive, high-quality feedback from participants when launching our pilot evaluation of our maternal and child health information chatbot, Rosie.

Study Objectives

We conducted an experimental pilot study to examine the feasibility, acceptability, and appropriateness of Rosie for our target audience. The pilot study also allowed us to test all the software and equipment, study protocols, and staff coordination to enable remedies before scaling to a full randomized controlled trial. Pilot data were also used to enhance the accuracy of Rosie’s responses by refining the existing models and fine-tuning the mechanisms and heuristics. The focus was on determining access to the resources and the capability of implementing the components and activities of the intervention as planned. Challenges in the provision of any component or the performance of any activity of the intervention were identified and potential solutions were determined. A sample size of 30 was chosen to be within the typical sample size range of a phase 1 clinical trial, according to the National Institutes...
of Health definitions, to conduct the study protocols and elicit participant feedback about the Rosie app with sufficient representation from our target group. Preintervention and postintervention tests were completed to qualitatively and quantitatively evaluate Rosie and describe any changes across key health outcomes including postpartum depression and frequency of emergency room visits to inform the full trial’s sample size calculations. Analyses of these data allowed us to present the preliminary findings, which should be interpreted as preliminary evidence, given that the pilot study was not powered to assess treatment-control differences, and this was not the main objective of the pilot study.

Methods

Development and Functionality of Rosie, the Chatbot

Rosie was customized to meet the needs of the target audience through continuous community feedback. Over the course of 3 years, our research team conducted community listening sessions, >20 community demonstrations of Rosie [23], and focus groups with pregnant women and new mothers of color. With this feedback, we customized Rosie to respond to health topics that mothers requested such as feeding tips, sleep advice, and information about rashes and fevers. To the Rosie app, we also added a set of the most popular questions that were asked by mothers as an FAQs page and provided a list of additional resources (eg, Supplemental Nutrition Assistance Program) benefits and emergency hotlines). Moreover, we added the requested video library that features things such as how to swaddle a baby, change a diaper, or perform cardiopulmonary resuscitation.

To build Rosie’s robust knowledge base, we collected, scraped, and extracted text from 60 sources, including websites of government agencies, hospitals, and professional medical organizations. A corpus of documents about maternal and infant health was built by scraping text from these vetted web domains using Trafilatura, a Python package and a web document processing tool called Scrapy that extracts text from HTML source code [24]. Each web document was then parsed into approximately 73,000 passages by applying a set of heuristics that retain sentence context. These passages were edited as necessary to serve as answers to the mothers’ questions and were used in a question generation model, probably asked questions (PAQs), to produce likely questions from users. The generated questions and their source passages were reviewed by annotators, who either edited both the question and the passage as necessary or discarded the pairs that were unhelpful, inaccurate, or incomprehensible. The answers augmented the existing knowledge base.

In addition, the research team supplemented the knowledge base by manually writing 350 question-and-answer pairs based on feedback from focus groups and community events with pregnant women and new mothers of color, who asked Rosie questions and identified topics of particular interest (eg, rashes and infant sleep). Only verified sources of health information were used in Rosie’s knowledge base. Sources with sponsored or commercial content were excluded.

Rosie’s underlying QA system uses an unsupervised, dense passage retrieval model. When users ask questions to Rosie, the retrieval model finds relevant content from the knowledge base that best answers the questions. Rosie also provides a source link in her responses, which can direct users to the websites from where the answer was extracted.

To better communicate with the users and understand their needs, we implemented an intent classification model using Rasa, a conversational artificial intelligence software. This classification model is used to categorize users’ text and respond accordingly. For example, it can identify greetings, thank you messages, and requests for information. It is also able to detect potential mental and physical health emergencies and send alerts via Slack, an instant messaging program, to our team members. Upon detection of a possible health emergency, Rosie sends emergency resources and relevant hotline information. In addition, a study staff member, who is a clinical social worker, reaches out to the participant within 24 hours to follow-up.

Rosie was built using Flutter, an open-source user interface software platform by Google, and designed to be compatible with both iPhone and Android devices. The Rosie mobile app has a log-in page with Google authentication, a chat window page that allows users to ask Rosie questions and rate the answers, an FAQs and resources page, and a medical disclaimer page reminding users that Rosie is an informational tool that does not replace professional medical advice and care (Figure 1).
After downloading and logging into the Rosie app, users are asked to enter their estimated due date or their infant’s date of birth. With the users’ permission, Rosie sends push notifications with daily health tips that are generated based on the due date or birth date, thus allowing for personalized advice based on the week-by-week progression of the user’s pregnancy or infant’s health. The past 7 days’ tips are also saved on an app page for users to refer to if needed. The app development team also created an app monitoring system that can detect server-related issues or app interruptions and notify team members via Slack for prompt troubleshooting and resolution. All conversations between users and Rosie were securely stored in Firebase Database.

Recruitment and Enrollment

This was a prospective randomized controlled pilot study involving a mobile app intervention, Rosie the chatbot. To clarify the methods, we followed the CONSORT (Consolidated Standards of Reporting Trials) checklist (Multimedia Appendix 1). From January 9, 2023, to February 9, 2023, participants were enrolled on a rolling basis until the target sample size was met (N=30) for the 3-month randomized pilot study. Participants were recruited using social media posts, including targeted advertisements, and through partnerships with community-based organizations. In addition, our research assistants contacted mothers who had completed an interest form at a previous Rosie community event or focus group.

Interested potential participants completed a screening questionnaire to determine eligibility. Inclusion criteria included being aged ≥14 years, being a woman of color, and either being currently pregnant or having a baby aged ≤ 6 months. Research assistants contacted each of the 248 potential participants to complete a brief video call to assess eligibility, explain study details, and obtain informed consent for participation. We identified fraudulent interest forms through video calls and review of interest survey meta-data, including IP addresses, to filter out potential participants who were falsely claiming to meet the inclusion criteria or were residing outside the United States.

Eligible participants were randomized into the control group (a monthly children’s book club) or the treatment group (Rosie, the chatbot) using a web-generated table with 15 slots for each study arm, for a total of 30 enrolled participants. After a participant assigned to the Rosie treatment group was unable to fully enroll owing to technical issues, we recruited an additional participant as a replacement. In addition, a mother in the control group experienced a stillbirth during the pilot study and did not complete the postintervention test. Thus, the final analytic sample was 52% (15/29) Rosie treatment group members and 48% (14/29) control group members (Figure 2). Among the 29 participants who were successfully recruited for the pilot study, 3 (10%) were recruited from partner organizations, 6 (21%) were recruited based on our interest forms at past Rosie events, and the remainder (n=20, 69%) were recruited using social media advertisements.
The control group participants were mailed a children’s board book once a month. The control group was modeled after similar programs across the United States that provide free monthly books to children [25]. Books were selected based on feedback from focus groups and community demonstrations with new parents or pregnant women of color, who expressed a desire to have books featuring diverse families.

At the initial enrollment meeting, our research team provided the Rosie treatment group participants instructions about how to install the Rosie app on their smartphones and provided a walkthrough of how to use the app and how to provide feedback about Rosie’s responses to their questions. We emailed each participant a weekly user engagement summary and sent reminder SMS text messages to encourage the use of the app.

### Data Collection, Outcome Measures, and Analysis

We assigned each participant an identification number to link pretest and posttest data and to track progress through the pilot study using clinical trial management software. Enrolled participants completed a pretest and posttest Qualtrics survey. Pretest surveys included questions about demographics (maternal age, race and ethnicity, education, household size, and health insurance), whether they were pregnant or parenting a young infant, and their due date or their baby’s birth date (as applicable). Pretest surveys also included the Patient Health Questionnaire–9 (PHQ-9) depression scale [26] and assessed the frequency of emergency room visits for infants.

Posttest surveys administered at the 3-month follow-up assessed pregnancy outcomes (birth weight and gestational age), emergency room visits for infants, PHQ-9 depression scale, and group-specific questions. Rosie treatment group members were asked how often they used Rosie and whether they experienced any of the following issues while using the app (eg, “application crashed,” “took too long to get a response,” “was difficult to use,” and “was not satisfied with the answer[s] to my question[s]”). In contrast, the control group members were asked to rate how much they agreed with the following statements on a 5-point Likert Scale ranging from 5 (strongly agree) to 1 (strongly disagree): the books I received were of good quality, the content of the books I received is a good match for my baby’s needs, the books were helpful to me during my pregnancy or parenting my infant, the books were enjoyable to me during my pregnancy or parenting my infant, participating with the book club was easy for me, and I would recommend the book club to other parents.

Both groups were also asked open-ended questions to obtain qualitative feedback. Rosie treatment group members were asked the following open-ended questions: (1) Besides answering your questions, what other features would you like to see on an application like Rosie? (2) Do you have any concerns about using Rosie? If so, please tell us about them; and (3) Do you have any additional feedback to help us build the best Rosie app possible? In an open-ended question, the control group members were asked to provide any additional feedback or ideas about their experience with the book club.

Descriptive statistics of the study sample were calculated based on group assignment. Pretest and 3-month posttest values were examined for postpartum depression and emergency room visits for infants. To examine the statistical significance of between-group differences at baseline, Fisher exact tests were used for categorical variables and Wilcoxon rank sum tests were used for continuous variables. For the within-group pretest and posttest comparisons, paired 2-tailed $t$ tests and McNemar tests were applied. Furthermore, 2-sample 2-tailed $t$ tests were used to compare the between-group differences in the pretest to posttest changes. Qualitative feedback was organized and
presented separately for the Rosie treatment group and book club control group members.

**Ethical Considerations**

The consent form was read aloud by research assistants to each participant and verbal informed consent was obtained. Written consent was obtained through the completion of an web-based form, and a copy of the consent form was sent via email to each enrolled participant. Participants were also encouraged to use the app and were told that if they were continuously enrolled for 3 months and actively engaged with the app by asking questions to the chatbot at least once a week, they would receive a gift card worth US $50. In addition, participants were told that if, at the end of the 3-month pilot study, they were among the top 20% of active users, defined by the number of unique questions sent to Rosie, they would also receive a tablet preloaded with children’s books. Participants were also given a gift card worth US $15, disbursed through a participant incentive distribution platform, Tango, upon completing the pretest and posttest Qualtrics surveys. The study was reviewed and approved (institutional review board study ID: 1556200) by the institutional review board of the University of Maryland, College Park, based on procedures for studies involving human participants.

**Results**

**Overview**

Baseline key demographic characteristics were not statistically significantly different between the Rosie treatment and control groups (Table 1). The mean age of mothers for both groups was 31.7 (SD 4.7) years. Approximately half of the mothers (9/15, 60%) were pregnant, and the other half (6/15, 40%) took care of young infants. Among those with infants, the mean age of infants was 4 months among Rosie treatment group members and 4.6 months among control group members. Most participants (9/15, 60%) were African American or Black, with the remainder being Asian, Hispanic or Latina, or multiracial (Table 1).

<table>
<thead>
<tr>
<th>Table 1. Total enrollment demographics*.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant characteristics</strong></td>
</tr>
<tr>
<td>Age (years), mean (IQR)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>African American or Black</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
</tr>
<tr>
<td>Multiracial</td>
</tr>
<tr>
<td>Currently pregnant, n (%)</td>
</tr>
<tr>
<td>9 (60)</td>
</tr>
<tr>
<td>Parenting infant, n (%)</td>
</tr>
<tr>
<td>6 (40)</td>
</tr>
<tr>
<td>Currently pregnant and parenting infant, n (%)</td>
</tr>
<tr>
<td>0 (0)</td>
</tr>
<tr>
<td>Infant age, months (Q1-Q3)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
</tr>
<tr>
<td>High school</td>
</tr>
<tr>
<td>Associate degree</td>
</tr>
<tr>
<td>Bachelor degree</td>
</tr>
<tr>
<td>Master degree</td>
</tr>
<tr>
<td>Professional degree</td>
</tr>
<tr>
<td>Average family size, mean (range)</td>
</tr>
</tbody>
</table>

*To examine the statistical significance of between-group differences at baseline, Fisher exact tests were used for categorical variables and Wilcoxon rank sum tests were used for continuous variables.

**Acceptability of Rosie**

More than 87% (13/15) of Rosie treatment group members reported using Rosie daily (5/15, 33%) or weekly (8/15, 53%) across the 3-month study period. Most users reported that Rosie was easy to use (14/15, 93%), and that they received a response from Rosie quickly (13/15, 87%). The remaining issues identified included crashing of the app during attempted use (8/15, 53%), and they were not satisfied with some of Rosie’s answers (12/15, 80%; Table 2).
Table 2. Acceptability statistics for Rosie treatment group (n=15).

<table>
<thead>
<tr>
<th>Questions and response options</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often did you use Rosie?</td>
<td></td>
</tr>
<tr>
<td>Monthly or less</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Multiple times a day</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Once daily</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Weekly</td>
<td>8 (53)</td>
</tr>
<tr>
<td>Did you experience any of the following issues while using the app?</td>
<td></td>
</tr>
<tr>
<td>Application crashed</td>
<td>8 (53)</td>
</tr>
<tr>
<td>It took too long to get a response</td>
<td>1 (7)</td>
</tr>
<tr>
<td>It was difficult to use</td>
<td>2 (13)</td>
</tr>
<tr>
<td>I was not satisfied with the answer(s) to my question(s)</td>
<td>12 (80)</td>
</tr>
</tbody>
</table>

Health Results: Quantitative
Pilot results suggested better health outcomes for the Rosie treatment group compared with the control group; however, between-group differences did not reach statistical significance. The estimated change in Rosie participants’ PHQ-9 mean depression scores from baseline to posttest period was −3.66 (SD 4.55) among Rosie treatment group participants compared with −2.77 (SD 4.92) among control group members (Table 3). This decline in depression scores between pretest and posttest period was only statistically significant for the Rosie treatment group (P=.008) and not the control group (P=.07). None of the participants from either group reported any emergency room visits for infants at baseline, but this percentage increased to 23% (3/13) for the control group members versus 9% (1/11) for the Rosie treatment group members (Table 3). Notably, 10 (67%) out of 15 mothers who were pregnant at baseline gave birth by the 3-month posttest period, and it could be a possible reason why emergency room visits for infants increased during the posttest period for both groups.

Table 3. Health and health behavior outcomesa.

<table>
<thead>
<tr>
<th>Maternal depression scaleb</th>
<th>Rosie (treatment group)</th>
<th>Book club (control group)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest period, mean (SD)</td>
<td>5.33 (4.43)</td>
<td>5.31 (3.33)</td>
<td>N/A</td>
</tr>
<tr>
<td>Posttest period, mean (SD)</td>
<td>1.67 (2.64)</td>
<td>2.54 (2.96)</td>
<td>N/A</td>
</tr>
<tr>
<td>Post-pre change, mean (SD)</td>
<td>−3.66 (4.55)</td>
<td>−2.77 (4.92)</td>
<td>.62</td>
</tr>
<tr>
<td>Values, n (%)</td>
<td>15 (100)</td>
<td>13 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Any emergency visit for infants</td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Pretest period, n (%)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Posttest period, n (%)</td>
<td>1 (9)</td>
<td>3 (23)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pre-post change, n (%)</td>
<td>+9.09</td>
<td>+23.08</td>
<td>.60</td>
</tr>
<tr>
<td>Values, n (%)</td>
<td>11</td>
<td>13</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aFor the within-group pretest and posttest comparisons, paired 2-tailed t tests and McNemar tests were applied. Paired 2-tailed t tests comparing pretest and posttest Patient Health Questionnaire depression scores were statistically different for the treatment group (P=.008). No other within-group comparisons were statistically significant at P<.05. Moreover, 2-sample 2-tailed t tests were used to compare the between-group differences in the pretest to posttest changes. P values assess pre- to postperiod changes for treatment versus control groups.
bThe sample size for emergency room visits was smaller because this outcome was assessed among mothers with infants (excludes currently pregnant mothers during the posttest period).
cN/A: not applicable.

Rosie Results: Qualitative Feedback
The Rosie participants provided considerable qualitative feedback about their experiences (Textbox 1). Participants expressed that they liked having a personal library to ask all their pregnancy and parenting questions, but improvements were needed in both the user experience and the content of responses. Participants commented that the quality of Rosie’s responses to pregnancy-related questions seemed to be low in accuracy compared with questions about infant caretaking. Participants noted that they observed improvements in the app’s functionality as the trial progressed and that the addition of an FAQs library and daily tips about baby’s development were helpful.
Textbox 1. Qualitative feedback about Rosie.

**Domain and feedback**

**Strengths**
- “I really enjoyed having a personal library to ask all the questions. Especially because after losing my baby, I didn’t receive a ton of targeted ads since my questions were limited to this app.’’
- “I liked it when the daily tips were added to the app. Before that, they were only notifications that I couldn’t go back to and often couldn’t fully read. I also liked when the previous day’s answers became available for viewing, so I didn’t have to screenshot or save links.”
- “I like everything else.”

**Points system**
- “Additionally, the point system didn’t fit with the original study. We were asked to use Rosie a minimum of once a week with a preference for more frequent use. But then the points were awarded for daily use. It was discouraging to not earn points, but with a newborn at home it was a struggle to do anything other than feed him daily.”

**Content and technical concerns**
- “Perhaps some answers were not very accurate for the age of my baby; they were for younger or older kids.”
- “Answers were not specific enough, lots of glitches with the app.”
- “The answers to the questions were often inaccurate.”
- “It does not answer me well.”
- “Not efficient. All of the answers take you to the same website. Using the app was a waste of time.”
- “My pregnancy-related questions were often answered inaccurately. The baby questions were mildly better, but if the chatbot is intended for both than it needs more training related to pregnancy symptoms and side effects.”
- “I had some trouble with the push notifications as well. At first, they were accurate for the number of weeks along I was in my pregnancy. Then they started to speed up, telling me that I was as many as two weeks ahead of my baby’s gestational age [e.g., it said I was 38 weeks when I was really 36]. I went in and reset it, using the same due date that I started with, and it continued to be incorrect. The app itself didn’t seem to have the same problem, just the notifications.”
- “Due to the high number of inaccurate responses, I was not motivated to continue using the app. I tried to stick with it, but to be honest this chatbot and the accompanying app have a long way to go before they’re ready for implementation.”
- “Finally, last week I had some issues with the app where it was giving me answers in a mix of text and source code formatting. Everything looked like a hyperlink but the links themselves did not work. I received an update notification, updated the app, and the problem persisted. It did fix itself after a couple of days though.”
- “At times when typing a question at the moment of submitting the keyboard would stay open and would not allow the user to hit submit. I had to exit out the app completely and reopen it and it would work again.”
- “Sometimes I had urgent questions, it could’t be use because it was under maintenance often.”
- “Hard to update.”
- “Rosie crashed a few times [like over multiple days when I tried asking a question].”

**Suggestions**
- “Sometimes Rosie gave me some answers that were not related to what I wanted to know, I understand sometimes terms can apply to two different things but perhaps Rosie can ask Do you mean this (1) or this (2), and then one chooses what is closer to the question one is asking. It happened to me a couple of times but I don’t remember the specific question.”
- “Answers were not specific enough, lots of glitches with the app.”
- “More accuracy with the answers would be great!”
- “Just implement user feedback.”
- “It would be nice to see some statistics [e.g., x% of kids do x by whatever age].”
- “I was pretty unlikely to visit the website Rosie referred me to. I think I would be more likely to view info right on the screen [even a click box with additional text or pics etc].”
- “Random suicide hotline warning was a little bit abrupt and unexpected.”
- “Answers were unrelated sometimes.”
- “Maybe something more personal with the week we are on if its more geared for pregnancy.”
- “I would say maybe [a] different app for moms and another version for expecting moms only due to the fact that some symptom questions were meant for a child rather than me who is pregnant. Overall the app is a great idea and good help/support for all. Thanks for the opportunity.”
• “When I ask questions, it didn’t answer my answers so I suggest to add more keywords for more accurate answers.”
• “Live chat.”
• “Maybe a bilingual app?”
• “A way to keep a record of my baby’s weight and height.”
• “Citations, more images, voice feature.”
• “Chance to chat with other moms and create a community of peer to peer questions and answer library. I really wanted to talk to other moms who had experienced loss early in their pregnancy too.”
• “It should have a lot of tips for healthy living.”
• “Sometimes I would forget what question I already asked because the previous questions would disappear. It could be nice to have an archive of questions.”
• “An actual calendar.”

As a result of the feedback, the development team has added more source websites to Rosie’s knowledge bank and has expanded the FAQs section of the app to include topics such as descriptions of the full schedule of well-baby visits and immunizations in the first 2 years of life. As this pilot study was designed to be part of a broad iterative process, negative, neutral, and positive feedback are all integral parts of refining the app’s functionality and expanding its knowledge base.

Book Club Results: Qualitative Feedback

Participants were extremely positive in their feedback about the book club, with participants rating the books as being of high quality and as a helpful tool for parenting (Textbox 2 and Table 4). All participants (14/14, 100%) agreed that they would recommend the book club to other new parents. Participants appreciated the “diversity and bilingual aspects of the books” and that their infants “really enjoyed the Global Babies book and loved to stare at the faces [presented in the books].” A participant offered a recommendation for an additional children’s book by an author whose works focus on social justice leaders in the United States to be offered in the book club.

Textbox 2. Qualitative feedback about the book club.

<table>
<thead>
<tr>
<th>Domain and feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths</td>
</tr>
<tr>
<td>“I so much love it.”</td>
</tr>
<tr>
<td>“I loved the diversity and bilingual aspects of the books.”</td>
</tr>
<tr>
<td>“Baby really enjoyed the global babies book and loved to stare at the faces.”</td>
</tr>
<tr>
<td>Suggestions</td>
</tr>
<tr>
<td>“The book[s] were good, but I think there are more popular/exciting book options for babies, especially books featuring babies of color. I am thinking of all the books by Jabari Asim for example. The last book was great though and the overall idea for a book club is fantastic. I loved knowing that new books were coming each month.”</td>
</tr>
<tr>
<td>“Oh, another thing: I got an automatic message from the book club quite frequently with the same message and it seemed redundant.”</td>
</tr>
</tbody>
</table>

Table 4. Acceptability statistics for the control group (n=13).

<table>
<thead>
<tr>
<th>Item</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate how much you agree with the following statements based your thoughts and experiences (1=strongly disagree; 5=strongly agree)</td>
<td></td>
</tr>
<tr>
<td>The books I received were of good quality</td>
<td>4.54 (0.14)</td>
</tr>
<tr>
<td>The content of the books I received is a good match for my baby’s needs</td>
<td>4.46 (0.18)</td>
</tr>
<tr>
<td>The books were helpful to me during my pregnancy or parenting my infant</td>
<td>4.38 (0.29)</td>
</tr>
<tr>
<td>The books were enjoyable to me during my pregnancy or parenting my infant</td>
<td>4.58 (0.19)</td>
</tr>
<tr>
<td>Participating with the book club was easy for me</td>
<td>4.77 (0.12)</td>
</tr>
<tr>
<td>I would recommend the book club to other parents</td>
<td>4.85 (0.10)</td>
</tr>
</tbody>
</table>
Health Results: Qualitative

A Rosie treatment group participant and a control group participant experienced pregnancy loss (a miscarriage and a stillbirth) during the study. An unexpected finding from one of the participants, who was assigned to the treatment group, was that using Rosie for maternal and child health information helped in shielding them from some emotional distress after their loss, with our participant stating the following:

I really enjoyed having a personal library to ask all the questions. Especially because after losing my baby, I didn’t receive a ton of targeted ads since my questions were limited to this app.

Recognizing that most of the currently available maternal and child health apps track user interactions for advertisers and feature advertisements, Rosie’s development as a no-cost, advertisement-free app may have additional benefits for mothers who value personal data privacy.

Discussion

Principal Findings

The pilot study demonstrates that Rosie is a feasible, acceptable, and appropriate intervention for pregnant women and new mothers of color. Rosie’s software was able to function with a given set of users and was generally able to generate responses to most of the asked questions. Our study found overall reduction in the PHQ-9 depression scale scores from baseline to the 3-month follow-up across both groups. However, the Rosie treatment group experienced a relatively large reduction from baseline to posttest follow-up. The reduction in maternal depression among both groups may correspond to known trajectories in maternal depression, which identified that depressive symptoms peak around birth and decrease as the infant ages [27,28]. Researchers have found variability in the timing and duration of perinatal and antenatal depression, but some studies have found that women who had depression symptoms during the antenatal period were likely to have more intense symptoms during pregnancy than during the postpartum period and that perinatal depression symptoms decreased over time [29,30]. With an expanded time frame of 12 months planned for intervention delivery during the full randomized controlled trial, the research team will be more able to precisely track trends in depressive symptoms and identify what, if any, congruence exists in our participants and the current literature about maternal depression and other mental health symptoms.

Although our small sample size and study design limit our ability to identify the causal pathways between Rosie and changes in depressive symptoms, our findings indicate that an association may exist between the use of the Rosie app and low maternal depression owing to increased parental confidence about their own health and infant caretaking through increased access to accurate health information. Rosie’s ability to provide rapid, accurate response with high-quality sources may also reduce the cognitive burden that pregnant women and new parents described in previous studies that emphasized that sorting through information, making comparisons, and determining the quality of the source of information were significant stressors. In addition, the Rosie app may also reduce maternal depression because it can help provide support to mothers who may not otherwise have access to many health-related supports and resources.

The low rates of emergency room use for infants in the Rosie treatment group compared with the control group aligns with previous study hypotheses that health information provided by Rosie can decrease acute health care use. Nonetheless, this could have occurred through multiple channels including potentially greater use of preventive health care services and Rosie assisting with the identification of relevant health information or clinical guidelines to support infant care.

The qualitative feedback the Rosie participants provided aligns with the conclusions obtained by Chua et al [20] during their review of maternal health chatbots that the first evaluations of these interventions often yield a need for improvement in the language models to understand the variety of ways in which users may ask questions about their pregnancy and child and to provide more precise and accurate responses to these questions.

New Rosie Features in Response to the Pilot Study

User experiences and feedback about the Rosie app has informed the further development of Rosie and continued precision of the QA model. For each of Rosie’s responses, users were able to click “thumbs up” or “thumbs down” to indicate their satisfaction or dissatisfaction with Rosie’s response to their question, and approximately 35% of questions received this additional level of feedback from users. The team analyzed this feedback and enhanced Rosie’s knowledge base by including topics that were not covered in previous iterations of the QA model and further refined the QA model based on the issues identified by the participants and the research team. We analyzed and discussed these interactions weekly with the goal of improving Rosie and initiating improvements in the user experience.

In addition, as a result of user feedback, we have expanded Rosie’s knowledge bank by >10 folds from 75,000 passages to >1.8 million passages extracted from 400,000 documents from verified health sources such as the CDC, National Institutes of Health, Mayo Clinic, and children’s hospitals. Rosie’s previous knowledge bank was restricted to only maternal and infant care questions, but Rosie users had requested information about topics such as managing chronic conditions, food safety and preparation, mental health, and self-care. The expanded corpus now enables mothers to ask any health-related question.

Strengths

This pilot study adds to the existing literature about chatbots broadly and their application in the context of maternal and child health. The team’s findings specific to reduction in maternal depression will help address one of the CDC-identified preventable causes of maternal death. In addition, low emergency room visits for infants suggest potential improvements in infant care and avoidance of some health crises. Our qualitative findings concur with those of previous studies, showing that improved precision in responses is needed [13,20]. Overall, participants found the chatbot as a helpful tool, and this intervention is delivered in a way that is easily...
accessible and usable. They also believe that it is an appropriate and acceptable approach for women of color who are pregnant or parenting an infant to receive reliable information. The feedback from our participants is invaluable in the preparation for scaling to a full randomized controlled trial. The use of a multimethod approach that obtained both quantitative and qualitative feedback resulted in a broad understanding of participants’ experiences and needs and addressed some gaps recognized in previous trials of chatbots designed for improving health knowledge.

Limitations

The recruitment of study participants was conducted primarily through web-based advertisements, potentially yielding a sample of participants who are overall more comfortable with using apps and their phones as their primary way of seeking health information than other women who are pregnant or parenting infants. Our sample, overall, was highly educated and the most (26/29, 90%) had health insurance, which may have reduced our ability to detect the experiences and health information needs of mothers without the same level of education or identify the needs of mothers whose lack of insurance may be associated with more variability in the use of emergency rooms. The small sample size of this pilot study resulted in low statistical power. Several health outcomes were found to be different in the posttest period when making between-group comparisons of the Rosie treatment and control groups, but differences did not reach statistical significance. In addition, detection of other between-group differences or predictive relationships between group assignment or demographic variables and outcomes of interest was limited by the small sample size. It was also not feasible to compare pregnant women and those parenting infants within groups at pretest and posttest periods to determine whether there were statistically significant differences owing to the small sample size. However, our approach helped to accomplish our goals for the pilot study and has facilitated a robust planning process for scaling to the full randomized controlled trial.

Conclusions

This pilot study showed that the prototype of the Rosie app is a feasible and usable innovation during pregnancy and postpartum period. This study provides valuable insight into using chatbots to help pregnant women and new mothers of color access reliable information the moment it is requested. Promising pilot results suggest that chatbots may reduce adverse health outcomes among ethnic and racial minoritized mothers; however, additional evaluation is warranted including a planned randomized clinical trial to evaluate the effects of Rosie on maternal and infant outcomes. If successful, chatbots such as Rosie can help address the existing health disparities in maternal and child health that have important intergenerational and downstream health consequences for the nation.

Acknowledgments

The study reported in this paper was supported by research grants from the National Institute on Minority Health and Health Disparities (QCN and EMA; grant number R01MD016037) and by the National Library of Medicine (QCN; grant number R01LM012849). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Authors do not have any conflicts of interests relevant to this study.

Data Availability

The deidentified versions of data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions

Study conceptualization, design, and implementation were completed by QCN and EMA. MJ and FXMG facilitated participant recruitment, enrollment, and data acquisition activities, with supervision from EMA. Development, build-out, and amendments of natural language process models and information corpus were completed by XY, HM, NPS, and JB-G. QCN, MJ, ACD, XY, HM, and XH analyzed, interpreted, and created the tables and diagrams based on the study’s quantitative data. MJ, ACD, and FXMG analyzed and interpreted the qualitative data. QCN, MJ, ACD, XY, HM, FXMG, and XH completed the analysis of current literature about chatbot interventions and implications of previous studies, provided critical analysis of integrated results, drafted and revised the iterations of the manuscript. QCN, EMA, MJ, ACD, XY, HM, NPS, FXMG, ND, XH, and JB-G edited and reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT (Consolidated Standards of Reporting Trials) checklist.

[PDF File (Adobe PDF File), 1262 KB - formative_v8i1e51361_app1.pdf]
References


Clinical Needs Assessment of a Machine Learning–Based Asthma Management Tool: User-Centered Design Approach

Lu Zheng¹, MS, PhD; Joshua W Ohde¹, PhD; Shauna M Overgaard¹, PhD; Tracey A Brereton¹, MS; Kristelle Jose¹, MS; Chung-II Wi²,³, MD; Kevin J Peterson¹, PhD; Young J Juhn²,³,⁴, MD, MPH

¹Center for Digital Health, Mayo Clinic, Rochester, MN, United States
²Precision Population Science Lab, Mayo Clinic, Rochester, MN, United States
³Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, United States
⁴Mayo Clinic Health System Research, Mayo Clinic, Rochester, MN, United States
*these authors contributed equally

Corresponding Author:
Lu Zheng, MS, PhD
Center for Digital Health
Mayo Clinic
200 1st Street South West
Rochester, MN
United States
Phone: 1 480 758 0664
Email: zheng.lu@mayo.edu

Abstract

Background: Personalized asthma management depends on a clinician’s ability to efficiently review patient’s data and make timely clinical decisions. Unfortunately, efficient and effective review of these data is impeded by the varied format, location, and workflow of data acquisition, storage, and processing in the electronic health record. While machine learning (ML) and clinical decision support tools are well-positioned as potential solutions, the translation of such frameworks requires that barriers to implementation be addressed in the formative research stages.

Objective: We aimed to use a structured user-centered design approach (double-diamond design framework) to (1) qualitatively explore clinicians’ experience with the current asthma management system, (2) identify user requirements to improve algorithm explainability and Asthma Guidance and Prediction System prototype, and (3) identify potential barriers to ML-based clinical decision support system use.

Methods: At the “discovery” phase, we first shadowed to understand the practice context. Then, semistructured interviews were conducted digitally with 14 clinicians who encountered pediatric asthma patients at 2 outpatient facilities. Participants were asked about their current difficulties in gathering information for patients with pediatric asthma, their expectations of ideal workflows and tools, and suggestions on user-centered interfaces and features. At the “define” phase, a synthesis analysis was conducted to converge key results from interviewees’ insights into themes, eventually forming critical “how might we” research questions to guide model development and implementation.

Results: We identified user requirements and potential barriers associated with three overarching themes: (1) usability and workflow aspects of the ML system, (2) user expectations and algorithm explainability, and (3) barriers to implementation in context. Even though the responsibilities and workflows vary among different roles, the core asthma-related information and functions they requested were highly cohesive, which allows for a shared information view of the tool. Clinicians hope to perceive the usability of the model with the ability to note patients’ high risks and take proactive actions to manage asthma efficiently and effectively. For optimal ML algorithm explainability, requirements included documentation to support the validity of algorithm development and output logic, and a request for increased transparency to build trust and validate how the algorithm arrived at the decision. Acceptability, adoption, and sustainability of the asthma management tool are implementation outcomes that are reliant on the proper design and training as suggested by participants.

Conclusions: As part of our comprehensive informatics-based process centered on clinical usability, we approach the problem using a theoretical framework grounded in user experience research leveraging semistructured interviews. Our focus on meeting
the needs of the practice with ML technology is emphasized by a user-centered approach to clinician engagement through upstream technology design.

*(JMIR Form Res 2024;8:e45391) doi: [10.2196/45391](https://doi.org/10.2196/45391)*

**KEYWORDS**

asthma; formative research; user-centered design; machine learning (ML); artificial intelligence (AI); qualitative; user needs.

**Introduction**

**Background**

Transparency, suitability, and adaptability are cited reasons for the chasm between advances in artificial intelligence (AI) and implementation in health systems [1]. Hindering implementation is a lack of transparency about the data used to make decisions and recommendations [2]. The conceptual suitability of, or aversion to, an algorithm in clinical use is practically governed by a clinician’s autonomous decision to engage with the tool [3]. The adaptability of the algorithm to local patient populations and unique workflows further increases the likelihood of adoption [4]. Logically, a proactive and systematic approach to addressing barriers to transparency, suitability, and adaptability may propel the wider implementation and adoption of AI in patient care [5].

Ultimately, the foundation of this approach is rooted in clinician engagement at the earliest stages of AI development [6]. Determining the user’s complex and diverse requirements for effective machine learning (ML)–based clinical decision support (CDS) tools requires a thorough understanding of the clinical utility of data sources and suitable designs to facilitate contact and response in appropriate settings [7]. This formative usability approach can be achieved through an empathetic and sustained relationship within a multidisciplinary team initiated by early-stage formative research and upstream technology design [8].

In a personalized medical practice aiming to optimize a clinician’s management of asthma, an efficient review of the condition’s characterizing features is critical [9]. Unfortunately, efficient and effective review of these data using electronic health records (EHRs) and timely clinical decisions are impeded by the varied format, location, and workflow of data acquisition, storage, and processing [10]. To support clinicians, we aim to develop an ML-based CDS tool that (1) predicts future risk of asthma exacerbation (AE; risk stratification and resource management), (2) provides this risk evaluation in the context of a summary of relevant information for asthma management (reduction of EHR review burden), and (3) offers options for actionable intervention.

As described in our published work, our AI evaluation plan uses a phased framework (Figure 1 [11]) to address technical performance, usability and workflow, and health impact, and iteratively follows our model documentation steps [12]. This paper describes how we conduct phase 0 and phase 1, highlighting user experience (UX) design and formative research through clinical user shadowing and interviews.

**Figure 1.** Phased research framework for evaluation of AI applied to the A-GPS project based on Park et al [11]. A-GPS: Asthma Guidance and Prediction System; AI: artificial intelligence; FDA: United States Food and Drug Administration; IRB: institutional review board; NASSS: Nonadoption, Abandonment, Scale-up, Spread, and Sustainability; UX: user experience.

<table>
<thead>
<tr>
<th>Phase 0: Discovery and intervention</th>
<th>Phase 1: Technical performance and safety</th>
<th>Phase 2: Efficacy and side effects</th>
<th>Phase 3: Therapeutic efficacy</th>
<th>Phase 4: Safety and effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>- User-centered design work and user requirements specification through stakeholder interviews</td>
<td>- IRIS and FDA submission</td>
<td>- Iterative interface design with stakeholder focus groups</td>
<td>- Prospective study</td>
<td>- Postdeployment surveillance and maintenance audit</td>
</tr>
<tr>
<td>- Evaluation of user requirements against NASSS framework</td>
<td>- Algorithm model explainability</td>
<td>- Preclinical model performance and validation assessment</td>
<td>- Clinical validation and model documentation</td>
<td>- Model maintenance</td>
</tr>
<tr>
<td>- Map model outcomes to future prospective study or clinical trial</td>
<td>- UX design research through stakeholder shadowing and interviews</td>
<td>- Model implementation evaluation and validation</td>
<td>- Corrective and preventative action model documentation</td>
<td>- Model maintenance</td>
</tr>
</tbody>
</table>

Model documentation framework

Prepare | Develop | Validate | Deploy | Maintain

---

https://formative.jmir.org/2024/1/e45391
Theoretical Framework

Literature detailing ML-based CDS tool translation indicates an ineffective balance of tool intelligence with explainability, creating gaps in translation, implementation, and accountability [13]. This suggests the need for early engagement with clinical users to mitigate present gaps by gaining a comprehensive and multidisciplinary understanding of present challenges, as well as identifying requirements for development and integration that prioritize both intelligence and practical usability. This formative research required a structured methodology to gather unstructured qualitative data and draw reliable conclusions, especially when diverse aspects and roles are involved. Therefore, UX research methodologies were adopted throughout phase 0 and 1, including the Double Diamond design framework and participatory design method that strategically engaged clinical users to derive unmet needs and identify user requirements [14,15]. The Double Diamond design framework (Figure 2 [16,17]), a graphical guide following the phases of the design process, is widely used to customize and standardize the progression of UX research by incorporating iterative loops and feedback opportunities to progress development [16,18].

Figure 2. Dan Nessler [16] developed this revamped version of the Double Diamond process based on the British Design Council’s Double Diamond [17] (reproduced from Nessler [16], with permission from Dan Nessler). This paper demonstrated how we went through each step in the first diamond to discover and define “user needs,” thus complete certain aspects on the Phases 0 and 1. HMW: How-Might-We.

Study Objectives

The A-GPS tool is an ML-based CDS tool accessible from “within” the EHR workflow. It aims to summarize all asthma-related context information extracted from the EHR on 1 screen page [9,22]. The tool will be embedded with a functional component of the AE risk model (AE risk model), which applies ML algorithms to predict a patient’s risk of exacerbation in 1 year [23]. This study’s objectives were to qualitatively explore clinicians’ experience with the current asthma management system, identify user requirements to improve algorithm explainability and A-GPS prototype, and identify potential barriers to ML-based CDS system use. Research questions were developed to probe the challenges and pain points of gathering asthma-related information within the current asthma management system, thorough evaluation of clinical team member workflows, user requirements for prototype optimization, algorithm explainability and display.
Methods

Participant Selection and Recruitment
We invited a group of clinicians representing the key roles in asthma management to collect user requirements and listen to their suggestions for future implementation. These roles include physician, nurse practitioner, nurse, and coordinator in primary care and asthma management specialty. Most of them are practicing in the Department of Pediatric and Adolescents Medicine and Family Medicine outpatient practices, where A-GPS will be implemented. A total of 14 participants were recruited by email using a convenience sampling approach and scheduled for a one-on-one, 30 to 60-minute virtual interview.

Data Collection
This formative usability research was directed toward an understanding of user requirements and to facilitate optimal workflow integration, estimate the potential impact of health care delivery factors, and work capacity constraints on achieved benefit. We aimed to collect different facets of qualitative data to identify all stakeholders, understand user needs, probe for optimal tool design to support clinical decision-making and routine workflow for each group in a comprehensive manner. To obtain the clinical context of how the tool will be used in practice, 1 researcher shadowed both sites and described the general patient flow. Next, we scheduled a 60-minute virtual interview with each recruited participant. An introductory statement provided background on the ML-based CDS tool prototype and explained the goal of the interview, developing rapport with interview participants. Each interview session was composed of 2 parts. Part 1 was a routine 30-minute semistructured interview. Interview guides were created for stakeholders, clinicians who were part of the A-GPS project or practice leadership, such as a division or practice chair, and users, defined as those with no stake in A-GPS but are practicing clinicians. Detailed interview guides were attached as supplementary material (Multimedia Appendix 1). Stakeholders were asked specific questions regarding their role as stakeholders in A-GPS and as potential users. Nonstakeholder participants (users) were questioned about their experience and needs as end users. Within part 2 of the interview session, participants were invited to demonstrate an EHR walkthrough on their working computer. During this time, we observed how the clinician routinely uses the system and defines the asthma-related information required to make a medical decision. Additionally, follow-up questions were asked to explore their cognitive process. While part 1 focused on clinicians’ reported problems and individual opinions, part 2 allowed us to observe the current problems and workflow objectively. Each type of data supplemented the other to achieve problem-probing and user needs consolidation.

Data Analysis
UX specialized translational informaticians engaged with practice components to evaluate usability and workflow to determine effectiveness, efficiency, satisfaction, ease of use, explainability, and usage, as described in the AI Evaluation Framework by Overgaard et al [23]. Interviews were transcribed, reviewed, and coded by team members LZ, JWO, KAJ, and TAB. Using a web-based collaboration tool MURAL (Tactivos, Inc) [24], transcripts were coded by identifying emergent themes and categorized into primary research questions asking what, how, where, and when. Other themes included challenges or pain points, barriers to adoption, novel ideas, new insights, and stakeholder considerations. Subthemes were presented as opportunities for change using an HMW question format [19]. Figure 3 provides a brief look into the synthesis and analysis work completed using the MURAL tool. As for the EHR walk-through, we used the data as a reference to make the list of the acquired information in the EHR system.

Figure 3. MURAL (Tactivos, Inc) is an online collaborative tool. Key quotes from participant transcripts were added to the board, using colors to identify participants. Similar or consistent responses across participants were grouped within circles, as shown in the image, to form subthemes. AE: asthma exacerbation; AI: artificial intelligence; ED: emergency department; ER: emergency room; HMW: How-Might-We.
Ethical Considerations
This study was conducted as a quality improvement initiative defined by the Mayo Clinic Institutional Review Board and Mayo Clinic Policy. According to the Code of Federal Regulations, 45 CFR 46.102, the project does not require IRB review. Patient and provider consent were not required, but all participants provided oral consent. To protect participant privacy, the participant’s name and email were used for recruitment purposes only and never linked to audio or transcribed data. Data were used to improve the delivery of health care services at Mayo Clinic.

Results

Participant Characteristics
A total of 14 clinicians were interviewed across 4 Mayo Clinic Health System sites in Minnesota, including Rochester, Red Wing, Albert Lea or Austin, and Kasson. Of those who participated, 7 (50%) were women, 11 (79%) were physicians (Doctor of Medicine or Advanced Practice Registered Nurse) and 3 (21%) were nurses. Their roles included asthma specialist, allergist, pulmonologist, pediatrician, family medicine physician, asthma care coordinator, and medical resident.

Targeted Patient Population and Clinician Users of A-GPS Tool
Clinicians identified opportunities to enhance asthma management for the diagnosed and at-risk population through the use of the A-GPS tool. According to clinicians, the tool would best serve pediatric patients with a diagnosis of asthma and should also aim to cover pediatric patients not officially diagnosed but at high risk of developing asthma, including those identified with the following conditions: symptoms of wheezing or coughing, albuterol or oral steroids use, frequently reported respiratory conditions of pneumonia, lung infections, wheezing, or coughing.

Some patients probably have asthma that we don’t detect, but that’s where I think this tool would be helpful because maybe even though they don’t have a diagnosis of asthma, they’ve had wheezing, or other things listed in their diagnosis and problem list. That would be helpful to avoid missing those people. [P6]

Some kids had been given a bronchodilator because often at 18 months, they present with like viral induced wheezes, and we find it improves with albuterol. So, we get a response to albuterol, and we know that these kids are potentially likely to get asthma, but we typically don’t make that diagnosis until after two. [P4]

Patients with pediatric asthma may be seen by multidisciplinary clinical roles including allergy specialists, pulmonologists, pediatricians, asthma care coordinators, rooming nurses, emergency department physicians, and primary care providers. When identifying proper clinician users of the tool, it was reported that any role that needs to provide asthma management care in practice would benefit from accessing and using A-GPS tool. To capture potential role-based variance in user requirements, the routine workflows and information needs were asked for each participant. It was found that asthma management is coordinated care by dynamic care teams, however, participants demonstrated a preference for a shared view of the tool to gain a shared understanding of patient cases. Even though the responsibilities and workflows vary among different roles, the core asthma-related information and functions they requested were highly cohesive, which allows for a shared information view of the tool.

Usability and Workflow Aspects of the ML System
In general, clinicians welcomed the integration of the AE risk model into the patient’s EHR. In practice, the prediction results are expected to help facilitate preventive actions to support better asthma management. To accomplish this, the AE risk prediction results cannot simply be in the EHR, it also needs to notify clinicians and prompt the care team to follow-up with patients in an expedited fashion.

If risk prediction results are added and approaching the threshold, and you will get a message letting you know that that is happening, that would be the best way to go. Because then you can prevent the next exacerbation, rather than waiting to see the patient the next time they come to the clinic, by then patient might have been through a couple of exacerbations. [P2]

If we are getting this risk score and especially if it were telling them that this is somebody that is at high risk of relapses and recurrences of episodes, then we can make that effort to reach out to those individuals. That should be flowing in my mind. That should be going to our care teams. [P14]

Despite the goal of being proactive, obtaining clinician’s attention to the right patient at the right time in an acceptable format is an issue. Notification methods were suggested by participants and opinions varied based on roles.

When you open the chart, it be helpful to have that notification sent via an in-basket message so that we’re aware and could follow up sooner. There might be cases where we’re aware that they’re high risk, but we can just delete it if we already have that plan for follow-up. [P9]

Best Practice Advisory (BPA) kind of prompts the provider that some action needs to be taken in these areas. And it might be a nice opportunity. Or one of the things we have is emergency action plans. Some similar way that incorporates into an action that needs to be taken or addressed for this patient. [P6]

In-basket message was mentioned by many clinicians as a common type of active alarm. However, it is necessary to balance effective information delivery and alert fatigue as clinicians, especially physicians, receive various alarms and notifications from multiple channels in their daily work.

I would just like the color coding in the records. I do not know if an in-basket message would be effective because we get a lot of them. If it was, I like the message was really clear and can quickly know what it is for. [P13]
Probably not an alarm for high-risk cases. I can imagine people getting annoyed at that, but if it came up in care gaps on the storyboard, like a reminder of something needing to be reviewed, that would be nice.

I hate to say in-basket messages because that just generates another inbox that the provider doesn’t have time to handle. I think having these folks show up as high-risk followed up by our care coordinators is the right way of handling this. It should be a trigger to get care coordinators or nurses to schedule a visit with the patient, which is more important than notifying the provider.

Additionally, the alarm or notification should reach clinicians with proper guidance for the next steps.

In BPA, we see alarms as this bright red thing with exclamation points. We’re going to want to act on that, but how do you act on it? Like, does it prompt then if you go into your plan, will it prompt something where you get some choices, like high-risk, you know. Whatever the risk score is, here are some options for you and you click those, and it goes into an order set, and you can order it and you’re done.

User Expectations and Algorithm Explainability

Expectations and Perceived Impact
Participants reported that the A-GPS tool is expected to have a positive impact on clinicians’ workflow and patient experiences. Clinicians anticipate a positive impact on usability and workflow by (1) streamlining the review of asthma information, (2) providing patients with “proactive” rather than “reactive” care, (3) empowering patients with a deeper and more personalized understanding of their condition, and (4) improving outcomes. Participants find the tool can be helpful in several situations, such as preparing for an upcoming patient visit, following up on a patient’s condition remotely, and changing or refilling medications based on changes in a patient’s condition. With well-organized asthma information presented at the appropriate time in the workflow, clinicians expect they can save time reviewing information and the care team will have a consistent understanding of patient cases.

Importantly, proactive and preventive care is anticipated with the AE risk model, allowing the prioritization of resources to patients of greatest need, and reduction of AE, emergency department visits, and hospitalizations. Clinicians hope to perceive good usability of the model with the ability to note patients’ high risks and take proactive actions to manage asthma efficiently and effectively. Ideally, with attention-grabbing model output visualization, both patient and caregivers would be more engaged in home-based care after seeing future risks. The potential to further drive higher quality outcomes was identified in the potential to monitor the relationship between patient adherence to medication, symptoms, and other contributing factors.

Algorithm Explainability

User requirements of the AE prediction algorithm output emphasized interpretability, logical justification, and validation as is shown in Textbox 1. Specifically, known definitions and levels of risk categories must be explainable, leading to efficient patient classification and resource allocation. Visual indications of severity, such as red, yellow, and green to define high, medium, and low-risk categories paired with a numerical indication were required. Supporting contextual information such as flagging primary features impacting risk prediction and providing a summary of additional asthma management variables were key requirements. Supporting information should be easily accessible and presented as hovering capabilities or links to relevant data (eg, patient history and baseline diagnostics). To assist with algorithm explainability and informing next steps, users required supplemental information on how the prediction score was calculated, bolstered by comparative diagnostics (eg, individual and population baseline values). Clinicians expressed concern regarding accuracy and reliability without significant validation of the model. Requirements included documentation to support the validity of algorithm development and output logic, and a request for increased transparency to build trust and validate how the algorithm arrived at the decision. For successful integration, users require that strategic education and phased implementation must be offered. Education and regular reports on the clear demonstration of value was the preferred strategy to gain an understanding of appropriate A-GPS use and limitations. Examples of stated learning preferences included hands-on training, such as workshops presented at monthly meetings, regular follow-up communication and showcasing of successful use cases, and video tutorials. Importantly, users require a clear demonstration of value to ease adoption, achieved by a phased implementation approach (multisite).
Barriers to Implementation in Context

### Accuracy and Reliability

Many participants expect the AI model given by A-GPS will be validated for accuracy and reliability. They also stressed the importance of making the model explainable and transparent to users. Clearly explaining why the model predicted a specific risk score will allow users to understand the logic of the model and its relevance to the patient’s current asthma situation. Without demonstrating validation or providing transparency clinicians will lack trust in the tool and likely not use it, limiting its clinical value.

A patient-specific concern was the potential for unnecessary anxiety and emotional burden on patients and their caregivers when told the AE risk model deems the child at “high risk.” The fear that an asthma event could occur based on prediction model may provoke unnecessary changes in the child’s daily activities, such as not sending their child to school or letting them play outside. [P7]

### Clinician and Patient Concerns

Clinicians recognized the potential benefits of A-GPS but voiced several ethical concerns regarding the AE risk model. As is shown in Textbox 2, one concern was the misunderstanding of AI’s role in clinical practice and that AI will override clinician autonomy to make clinical decisions. However, 1 participant asserted that the goal of AI is to provide complementary information and that the clinician would still make the final clinical decisions. A similar concern was the impact the risk prediction model would have on a clinician’s intuition. More specifically, when the AE risk model contradicts the clinician’s judgment, the possibility of legal or ethical issues may arise depending on what action the clinician takes.

**Textbox 1.** Examples of interviewees’ statements about explainability of prediction outcomes.

- I want to be able to see this risk score. When the patient is in front of me, I also want to be able to see a whole lot more information about that patient, preferably in an easy to find format that I don’t have to go digging in Epic for it, like I currently do. [P14]
- I would probably like something simpler, like not necessarily a percentage. And then I like, okay, it’s red, which means they’re at high risk. In the background, I could know what that means. And if you want more information, then you could click and find why it is high. [P13]
- I think high, medium, low would you know, would be sufficient. And if you would have some colour popping up or even color-coded too, like they are low risk in green, medium in yellow. If they’re high-risk and in red, that certainly will get your attention. I also want to know what is putting them at risk. Is it the severity of symptoms, their need for oral steroids, their hospitalization and ED visits? So that would certainly be helpful to know exactly where their risk area is. [P9]
- I would like it to give me a percentage score rather than a level. So, if I had a model, I would like to see a percentage within a certain period. Like within two years or within one year, there is a percentage chance that there’ll be an asthma exacerbation. Some people like the simplicity of a one to five. But if I’m making my decisions, I kind of want a little bit more detail on it. Also, if I click on it, I would like to be able to see where and what data points they’re using to make these predictions because sometimes the data in the Epic chart is incorrect. [P4]

**Textbox 2.** Examples of interviewees’ statements about ethical concerns.

- Machine learning introduces a new wrench in things. Because now you’re not giving me a necessarily a recommendation, but you’re giving me insight that might either raise my intuition or lower it. How do you handle having that prediction result legally and ethically and everything else? [P12]
- As a pulmonologist, I am trying to understand how other systematic diseases impact asthma. So, I am also checking tests of other body systems and evaluating by talking to patients. For the populations I am seeing with asthma, hopefully, at some point artificial intelligence could help us, but I just do not see it at this point. [P2]
- This prediction score is not meant to override. This is complimentary information for you. I know you do mental calculations, but this is a data-driven calculation that gives you other complementary information. If there’s a discrepancy, is there anything you are thinking low in emotion, say “hi, just to think about it on this page.” So then, you know, you don’t have to go to that page, just look through another page of the sectional summary. [P5]

### Discussion

**Principal Findings**

Principal results are discussed by identified themes. In each theme, we started with “How Might We” questions to inspire discussions on challenges and opportunities.

**Usability and Workflow Aspects of the ML System**

Challenges and opportunities of usability and workflow aspects of the AI system prompt key questions such as the following: How might we incorporate the A-GPS tool to support workflows of different roles? Clinicians are tasked with a workload that involves increasing patient volumes, more complex diseases, and an overwhelming EHR system. Further, 1 goal of A-GPS is to help alleviate the time clinicians spend in the EHR to find asthma-related information and supplement the clinical decision process required to minimize a patient’s risk of AE. Successfully
incorporating A-GPS into the current workflow of various clinical roles is arguably as important as the tool itself. Participants suggested placing the A-GPS tool in the same location within the EHR and having the same view, regardless of roles. This will allow easier navigation in the EHR within the care team’s current practice as clinicians, nurses, and care coordinators frequently view each other’s screens during patient care.

**User Expectations and Algorithm Explainability**

Challenges and opportunities of user expectations and algorithm explainability prompt key questions such as the following: (1) *How might we communicate A-GPS results in a way that is explainable to patients?* Although patients and caregivers were not interviewed, they will receive some level of information from the A-GPS tool communicated to them by the clinician or care team. How the outcome of the AE risk model is explained to patients will be important. Limiting unnecessary anxiety or misunderstandings while still effectively communicating the model results needs to be carefully addressed so families can make appropriate decisions that improve the patient’s outcomes. Properly educating clinicians on how to explain the AE risk model to families is a step that could be tied into the overall education plan for A-GPS. (2) *How might we remove barriers to adoption to increase clinician buy-in?* The adoption of any new tool or technology rarely goes as planned, it takes time to achieve buy-in from users. To increase clinicians’ buy-in for A-GPS a few barriers should be addressed. First, clinicians need to see that the A-GPS tool is validated, accurate, reliable, and effective at saving time in the EHR. Communicating this data using various educational modalities can increase the reach among clinicians. Second, some clinicians will wait to see the value A-GPS brings to their colleagues before using it themselves. These individuals may be reached by leveraging clinician champions who believe the A-GPS tool improves UX with the EHR and patient outcomes. Ultimately, the best method to facilitate the adoption of the A-GPS tool is to ensure its functionality meets the user’s needs and expectations. Clinicians are more likely to use A-GPS if a clear and concise ML-based CDS tool is created that contains only asthma-related information with easy access to more detailed notes and test results. Moreover, clinicians may use and act on the AE risk model if they trust it, understand what the risk score means for their patient, and understand how the model came to its conclusion for their patient.

**Barriers to Implementation in Context**

Challenges and opportunities related to implementation and system use in context prompt key questions such as the following: How might we address ethical issues brought on by a difference between the AE risk model and a clinician’s professional judgment? The contradiction between the AE risk model and a clinician’s clinical assessment may pose an ethical and even legal issue. In practice, clinicians may feel pressured to act upon the model’s output in fear of legal challenges even if they believe the patient’s risk of AE is different based on their professional judgment. Although this topic deserves further exploration, it is reasonable to assume in educating clinicians that their clinical judgment supersedes the result from the AE risk model as the model does not take into consideration the multitude of variables the clinician assesses. Moreover, the reason for the development of the AE risk model in this context was to provide supplemental information to improve the care of patients with asthma, not replace the expertise of clinicians. Addressing this upfront with potential users should be a component of A-GPS implementation.

Acceptability, adoption, and sustainability of A-GPS are implementation outcomes that are reliant on the proper design and training as suggested by participants. Without following the guidance of clinicians given in this study the success of A-GPS will be limited, resulting in decreased user satisfaction and clinical effectiveness. To overcome potential barriers to implementation success several priority areas should be met. First, the A-GPS tool needs to be easily accessible within the EHR, in a location that is obvious and consistent across all clinical roles, contains all asthma-related information on a single page, and is visually concise and intuitive. To increase acceptability, the AE risk model needs to be validated for accuracy and reliability and made transparent to users. Transparency is necessary to build trust among clinicians and trust facilitates acceptance and adoption. The AE risk model output needs to be easily interpretable, clearly defined, and intuitive to improve adoption and sustainability. While the risk category output and descriptive statement is essential, an organized display of supporting information popping up in proper format is generally desired as part of the output. Except for the quality of model output, the importance of quality and strategy of education and training cannot be ignored. In a paper by Gordon et al [25], Mayo Clinic took a standardized and efficient approach to provide education and training sessions when implementing a new EHR system. The results demonstrated higher acceptance and confidence among users. This could be a good example for an A-GPS project in terms of successful implementation.

**Strengths and Limitations**

Conducting interviews with potential users of a new clinical tool not only gains insights into their needs but also encourages buy-in as seeking their input before implementation demonstrates that the research team values user feedback. To our knowledge, this study facilitated buy-in and support among participants as several thanked this study’s team for their efforts to understand user needs. For a broader scope of AI projects, the innovative multi-background collaboration between translational informatics, data science teams and engaged clinical stakeholders guided by the design framework at an early stage provides well-defined user requirements and implementation plans and delivers evidence documentation for upcoming implementation and validation stages. The method can be applied to a wide variety of CDS. In this project, the UX research method was used to engage with clinical stakeholders and prospective end users to gain a comprehensive and multidisciplinary understanding of the role A-GPS is expected to play in pediatric asthma care. Shadowing and interviewing clinical stakeholders were a source of engagement that gathered user requirements from the perspective of potential users, with the objective of informing tool development and translation efforts [21]. Once A-GPS tool is functioning technically, its fit
into the clinical workflow must be evaluated. Moreover, education and documentation must be provided to explain the algorithm and its limitations to effectively translate between the perspectives of experts who created and supported the technology and the perspectives of experts who use the solution to patients. Evaluating the interpretation needs of clinicians, preferences for the display of model output (eg, percentage vs binary threshold), and feature contributions will be assessed based on the data obtained from UX research efforts. Concurrently, the team will also engage clinician stakeholders in the development of model documentation to support explainability [26]. The data obtained from UX research will assist the translational informatics and data science teams in identifying the level of explainability needed to inform and validate the design of A-GPS and supplementally enhance existing workflows [27]. Strategic efforts to promote explainability include applying a documentation framework grounded in scientific research addressing known challenges. This encompasses interdisciplinary best practice reporting requirements that follow phases of model development (prepare, develop, validate, deploy, and maintain) for knowledge continuity throughout the solution's life cycle [12].

A limitation of this study is the generalizability of design and implementation suggestions as the practice environment of participants has different features. Even though they are all within Mayo Clinic Health Systems, some are working on hospital campuses, while some are from community-based primary care, which led to various user needs. Another limitation is that the perspective of patients and their caregivers were not evaluated in this study. This was purposeful as the intended users of A-GPS are clinicians but the impact the AE risk model may have on families, as stated in the Ethical Considerations section, should be explored further.

**Further Research**

Future research will evaluate the sustainability and scalability of user requirements for enterprise, national, and international adoption of the ML-based CDS tool. Ethical considerations of AI interpretation, patient engagement, and clinician autonomy warrant further investigation. Our research team will conduct multiple studies as we approach the future stages of efficacy and side effects, therapeutic efficacy, and safety and effectiveness planned in our phased comprehensive AI evaluation framework (stage 1). There are more questions to answer in the future: How might we demonstrate to providers the accuracy and reliability of the AI output? How might we define required transparency for AI output? How might we provide an efficient educational module for users and show validation measures to support and explain AI output? How might we handle alarm fatigue, including situations where patients do not respond to providers’ intervention? How might we improve ML-model predicting the risk of AE when the provider’s proper and timely intervention may reduce the performance of the model (eg, positive predictive value)?

**Conclusions**

We aimed to anticipate barriers to the translation of our pediatric asthma management ML-based CDS tool by engaging clinicians in prototype development and optimization leveraging UX research methodologies. In efforts to bolster the transparency, suitability, and adaptability of our solution we qualitatively evaluated user requirements and potential barriers in 3 overarching themes: usability and workflow aspects of the ML system, user expectations and algorithm explainability, and barriers to implementation in context. We presented findings specific to our tool’s risk evaluation in the context of a summary of relevant information for asthma management. This work contributes to phases 0 and 1 of our comprehensive informatics-based AI evaluation frameworks developed by our multidisciplinary team of clinicians, data scientists, translational informaticians, and UX experts at Mayo Clinic [23]. The transparent evaluation and documentation of AI applications in health care enhances clinician and patient trust, supports sharing of AI between hospitals, and increases standards and shared responsibility across the continuum of care. The results of this development study further enhance the model documentation of A-GPS aimed to ensure rigorous evaluation, transparency, and knowledge continuity [12]. A sustainability and scalability evaluation of user requirements will strengthen the potential for national and international adoption of A-GPS.

**Acknowledgments**

We would like to deeply thank Dr Manuel Arteta, Dr Valeria Cristiani, Joy Fladager Muth, Dr Gregory Garrison, Dr Margret Gill, Dr Jason Greenwood, Darcy Hall, Dr Martha Hartz, Dr Julian Lynaugh, Dr Donnchadh O’Sullivan, Dr Sarah Scherger, Dr Eric Schnaith, and Dr David Soma for finding their time for interview and sharing their valuable insights into usability of A-GPS in clinical practice setting. This project was supported by the Center for Digital Health, Precision Population Science Lab, the AI Program of Department of Pediatric and Adolescent Medicine. We acknowledge no generative AI usage in the preparation of this paper.

**Data Availability**

The data sets generated and analyzed during this study are not publicly available due to participants’ privacy, but are available from the corresponding author on reasonable request.

**Conflicts of Interest**

None declared.
Multimedia Appendix 1
Interview question list.

[DOCX File , 17 KB - formative_v8i1e45391_app1.docx ]

References


24. Tactivos Inc. MURAL. URL: https://www.mural.co/ [accessed 2023-12-07]


Abbreviations

A-GPS: Asthma Guidance and Prediction System
AE: asthma exacerbation
AI: artificial intelligence
CDS: clinical decision support
EHR: electronic health record
HMW: How-Might-We
ML: machine learning
UX: user experience

Edited by A Mavragani; submitted 23.01.23; peer-reviewed by L Zhou, B Senst, R Marshall; comments to author 05.05.23; revised version received 25.08.23; accepted 20.11.23; published 15.01.24.

Please cite as:
Zheng L, Ohde JW, Overgaard SM, Brereton TA, Jose K, Wi CI, Peterson KJ, Juhn YJ
Clinical Needs Assessment of a Machine Learning–Based Asthma Management Tool: User-Centered Design Approach
JMIR Form Res 2024;8:e45391
URL: https://formative.jmir.org/2024/1/e45391
doi:10.2196/45391
PMID:38224482

©Lu Zheng, Joshua W Ohde, Shauna M Overgaard, Tracey A Brereton, Kristelle Jose, Chung-Il Wi, Kevin J Peterson, Young J Juhn. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Exploring Adolescents’ Attitudes Toward Mental Health Apps: Concurrent Mixed Methods Study

Helene Høgsdal1, MSc; Henriette Kyrrestad1, PhD; Marte Rye1, PhD; Sabine Kaiser1, PhD
Regional Centre for Child and Youth Mental Health and Child Welfare - North, Faculty of Health Sciences, Tromsø, Norway

Corresponding Author:
Helene Høgsdal, MSc
Regional Centre for Child and Youth Mental Health and Child Welfare - North
Faculty of Health Sciences
UiT The Arctic University of Norway
Campus Tromsø
Tromsø, 9019
Norway
Phone: 47 77646619
Email: helene.hogsdal@uit.no

Abstract

Background: Adolescence is a critical time in which many psychological disorders develop. Mental health promotion is important, especially during this period. In recent years, an increasing number of mobile apps geared toward mental health promotion and preventing mental illness have been developed specifically for adolescents, with the goal of strengthening their mental health and well-being.

Objective: This study aims to explore adolescents’ attitudes toward mental health apps, as well as the perceived usefulness of mental health apps.

Methods: In this mixed methods study, a total of 183 adolescents (mean age 15.62, SD 3.21 years) answered a cross-sectional questionnaire, with 10 questions (eg, “What do you think about mental health apps in general?”). To complement the quantitative findings, individual interviews were conducted with 9 adolescents, during which they could elaborate on their opinions about mental health apps.

Results: A total of 30% (56/183) of the adolescents in the quantitative study had used a mental health app. Over half of the respondents (77/126, 61.1%) reported that they would use a mental health app if they had a mental health problem as well as that they thought mental health apps were somewhat or very useful (114/183, 62.3%). Availability was the most frequently reported advantage of mental health apps (107/183, 58.8%). Possible associated costs of mental health apps were the most frequently mentioned barrier to their use (87/183, 47.5%). Findings from the interviews also pointed to the importance of the availability of mental health apps as well as their credibility and potential to provide adolescents with autonomy when seeking mental health advice and help.

Conclusions: Overall, the results indicate that adolescents have a positive attitude toward mental health apps. However, adolescents are also more or less unaware of such apps, which might be one reason why they are often not used. The findings of this study have important implications for future research on mental health apps and for developers of mental health apps that target young people. The insights gained from this study can inform the development of more effective mental health apps that better meet the needs and preferences of adolescents.

(JMIR Form Res 2024;8:e50222) doi:10.2196/50222

KEYWORDS
mental health applications; mental health; adolescents; adolescent; youth; mobile health; app; apps; application; applications; opinion; opinions; cross sectional; survey; surveys; questionnaire
Introduction

Overview
The World Health Organization defines well-being as follows: “Well-being is a positive state experienced by individuals and societies. Similar to health, it is a resource for daily life and is determined by social, economic and environmental conditions” [1]. In today’s society, reference is often made to the importance of working to improve young people’s mental health and well-being as adolescence is a critical, transformative, and challenging time [2]. Research indicates that 10%-20% of adolescents worldwide experience mental health problems [3,4], and in approximately half of all mental disorders, the onset of symptoms occurs before the age of 15 years [5,6]. This illustrates that mental illness in adolescence can have negative effects also on adulthood [6,7]. Today, the need for mental health care and tools for young people is high, and it is argued that provided mental health services for adolescents are not sufficient to keep up with the growing demand [8]. It is therefore important to work toward developing targeted measures and evidence-based interventions that can help young people cope with stress and different challenges [9,9]. Furthermore, it is important to provide supporting tools and resources that can promote healthy habits and behaviors that contribute to overall well-being in adolescence.

In recent years, it has been suggested that mobile devices and the internet represent ideal tools to deliver such interventions to young people [10], and an increasing number of developers and researchers have been following this suggestion [11-14]. In a systematic review, Grist et al [13] highlight the importance of digital interventions developed for young people being of good quality. Furthermore, they point out that the development of digital products, intended for young people, should be designed in collaboration with young people, in order to develop customized and high-quality products for the target group. Thus, a broad understanding of what young people think about specific digital interventions can be beneficial for developing effective products suitable for the target group.

Mental Health Mobile Apps
Mobile apps (ie, apps) are tools where users can receive digital health interventions or information, these health apps have the intention of improving the user’s overall health [12]. Along those same lines, mental health apps are designed to improve mental health and well-being among their users [14-16]. These apps can be designed to address specific mental health disorders [17,18], or they can be more general, focusing on promoting mental health and preventing mental distress through a variety of tools and resources such as emotional self-monitoring and coping strategies [19,20].

One important advantage of mental health apps is their availability [16]. Users can gain access to them anytime and anywhere with a smartphone or other mobile devices. This is particularly useful for individuals who may not have easy access to mental health services due to geographic barriers or for individuals who might benefit from receiving frequent reminders or immediate support [15,16]. Furthermore, mental health apps can reach larger segments of the population compared with in-person therapy, and, in some cases, they have shown promising cost-effectiveness [10,16].

Adolescents’ Engagement With Mental Health Apps
Young people spend a considerable amount of time on their smartphones and use them for entertainment, social purposes, and to access information on different topics, including mental health [21-23]. There are several mental health apps available for adolescents, but as Grist et al [13] point out, the quality of the assessment on these apps is scarce. Research on mental health apps for adolescents indicates that young people are generally satisfied with the access to and ease of use of mental health apps [13,24]. Kenny et al [25] highlight in their study that adolescents prefer mental health apps to be safe, engaging, and easily accessible. The use of mental health apps can also allow people to maintain anonymity when seeking mental health advice or guidance, which is appreciated by young people, who often have a high threshold for help-seeking due to a desire for autonomy and the negative attitudes or stigma surrounding mental health services [26-28].

Nevertheless, the level of user engagement in existing mental health apps is relatively low or moderate and repeated long-time use of mental health apps after downloading them is rare [13,29,30]. This may indicate that despite the fact that young people find mental health apps appealing, it is not sufficient to ensure sufficient use over a long span of time. Regardless of the reasons, it is possible that low user engagement could decrease the overall effectiveness of mental health apps. Research on the effectiveness of mental health apps on adolescent mental health outcomes has revealed mixed results. Some research has failed to illustrate any effect [13,31,32], and others have indicated that mental health apps are promising and have the potential to provide improvement in mental health outcomes [14,33,34]. Yet, there is broad agreement that more research on the effectiveness of mental health apps is needed to fully understand their capacity to improve mental health outcomes [13,32,33]. In order to understand more about the effect mental health apps can have on young people’s mental health, it is also beneficial to examine what young people think about mental health apps and how they should appear in order to be effective for the target group.

This Study
The aim of this study is to explore adolescents’ attitudes toward mental health apps, and their perceived usefulness, using both quantitative and qualitative data. While the quantitative study focuses more on the use and usefulness of these apps as well as on their perceived advantages and disadvantages, the qualitative study examines adolescents’ more general thoughts on mental health apps and how they may be of help to them. Together these data sets provide important insight for the development of mental health apps. The results can help to ensure more user-targeted products and more user engagement, which in turn can contribute to a greater effectiveness of mental health apps.
Methods

Overview
This study used a convergent, parallel QUAN + qual mixed methods design [35]. Quantitative and qualitative data were collected simultaneously, and the qualitative study was used to complement the results of the quantitative study. The quantitative data are based on a cross-sectional survey and the qualitative data are based on semistructured interviews. A total of 2 authors (HK and SK) analyzed the quantitative data, and the other 2 authors (HH and MR) analyzed the qualitative data. The results of the 2 data sets are integrated in the discussion.

Quantitative Study

Participants and Procedure
Adolescents were invited to participate in the study during several events that took place in autumn 2022, including an innovation camp in Mosjøen (a municipality in Northern Norway), “The research days” at UiT The Arctic University of Norway, and the World Mental Health Day event organized by the municipality of Tromsø. To participate, adolescents were asked to scan a QR code, which led to a digital questionnaire [36]. Posters and flyers with the QR code were also distributed at various youth clubs and adolescent health centers in Northern Norway. The questionnaire was anonymous, and participation was voluntary. The final quantitative sample (N=183) consisted of 118 (64.5%) girls, 59 (32.2%) boys, and 6 (3.3%) adolescents that did not specify their gender. Participants were aged between 13 and 19 (mean 15.75, SD 1.65) years; 170 (92.9%) participants were Norwegian and 13 (7.1%) participants reported other nationalities (2.2% were Serbian, 1.1% were Swedish, and 3.8% did not specify their nationality).

Questionnaire

Overview
The questionnaire consisted of 10 questions. Information was collected on demographic characteristics like gender (boy, girl, and other), nationality (Norwegian, Serbian, Swedish, or other), and age (11 to 20 years). There were 5 or 6 questions about mental health apps that are described in the following and an open-ended question where the adolescents could give comments about the topic or the questionnaire in general.

Use of Mental Health Apps
Adolescents were asked 1 or 2 questions developed by Grist et al [37]: “Do you use (or have you used) any apps to help you with mental health problems?” with response options “yes” or “no.” Adolescents who answered “no” were asked a follow-up question: “If you had a mental health problem and there were apps available to help, would you use them?” with response options “yes” or “no.”

Perceived Usefulness of Mental Health Apps
Adolescents were asked: “What do you think of mental health apps in general?” with responses given on a 5-point Likert scale, ranging from (1) “not useful at all” to (5) “very useful.”

Advantages of Mental Health Apps
Adolescents were asked: “What do you think some of the advantages are of using an app for your mental health?” Adolescents were presented with 8 statements developed by Grist et al [37], such as “It is more private” and “I don’t have to talk to someone face to face.” One additional statement, “I can be anonymous” was added to the list by the authors. Adolescents could tick up to 3 statements with which they agreed the most.

Disadvantages of Mental Health Apps
Adolescents were asked: “What do you think some of the barriers are to using an app for your mental health?” Adolescents were presented with 9 statements developed by Grist et al [37], such as “I don’t trust apps,” “I am afraid someone will see the app on my phone,” and “It might cost money.” Adolescents could tick up to 3 statements with which they agreed the most.

Information About Mental Health
To examine where adolescents looked for information about mental health, they were asked: “Which media do you use to find information about mental health?” There were 7 response options as follows: Google, social media (eg, TikTok and Facebook), mental health apps, television, web-based newspapers, podcast or radio, and others, where adolescents could specify a different medium in an open-ended textbox. Adolescents were asked to tick all the media they used with no limitations.

Data Analytical Strategy
Data were analyzed with SPSS (version 29; IBM Corp). Descriptive statistics were calculated and included means, SDs, and frequency distributions. A multiple linear regression analysis was calculated to predict adolescents perceived usefulness of mental health apps, with gender (1=boy, 2=girl), age (11-20 years), and previous use of mental health apps (1=yes, 2=no) as predictors. A significance level of less than .05 was applied.

Qualitative Study

Participants and Procedure
Adolescents who answered the questionnaire were invited to contact the researchers if they wanted to participate in an interview and share their thoughts and opinions about mental health apps. A total of 9 adolescents (all Norwegian) chose to be interviewed (Table 1); 7 elected to be interviewed remotely (via telephone, Microsoft Teams, or Zoom) and 2 elected to be interviewed in person. Before each interview, the adolescents were informed that they were expected to talk about mental health apps, which were defined as mobile apps “developed with the thought of helping people manage their own mental health and wellness” [16]. All interviews were audio recorded on an Apple iPad with the app Diktatfon [38], encrypted, and sent directly to Nettskjema [36] before they were transcribed. Two adolescents requested to be interviewed together, hence, 1 interview was conducted with 2 adolescents present (informant 3 and informant 4). These adolescents were considered as individual informants despite being from the same interview.
Table 1. Overview of informants and their experience with mental health apps.

<table>
<thead>
<tr>
<th>Informant</th>
<th>Gender</th>
<th>Age (years)</th>
<th>Experience with mental health apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>No15</td>
<td>Girl</td>
<td>15</td>
<td>No</td>
</tr>
<tr>
<td>No16</td>
<td>Boy</td>
<td>16</td>
<td>No</td>
</tr>
<tr>
<td>Yes15</td>
<td>Girl</td>
<td>14</td>
<td>Yes</td>
</tr>
<tr>
<td>No16</td>
<td>Boy</td>
<td>16</td>
<td>No</td>
</tr>
<tr>
<td>No14</td>
<td>Girl</td>
<td>17</td>
<td>No</td>
</tr>
<tr>
<td>No17</td>
<td>Girl</td>
<td>17</td>
<td>Yes</td>
</tr>
<tr>
<td>No16</td>
<td>Boy</td>
<td>16</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes14</td>
<td>Girl</td>
<td>14</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Interviews**

All interviews were conducted in Norwegian. The interviewer followed a semistructured interview guide, however, the adolescents were encouraged to talk freely without too much interruption from the interviewer. There were three key questions in the interview guide: (1) “What are your thoughts on mental health applications that are aimed at adolescents?” (2) “What do you think mental health applications can do with regard to adolescents’ knowledge about mental health?” (3) “How do you think mental health applications may be of help to young people?” The interviewer asked follow-up questions when it was natural to go more in-depth on the information the adolescent provided. The interviews varied in duration from 15 to 30 minutes.

**Analytic Strategies**

The qualitative data were analyzed using thematic analysis inspired by Braun and Clarke [39]. Our analysis was based on a constructionist epistemological position, assuming knowledge is socially constructed and developed through communication and interactions between people [40]. Audio recordings were transcribed by the first author, and transcripts were imported into NVivo [41] to support the organization and coding of the data. Transcripts were read carefully by the first and third authors to identify meaning and patterns in the data. Initial codes were created for the data material, which were then sorted into potential themes. The themes were reviewed and defined in order to identify the essence of each one and the relationship between them. All the codes, themes, and presented quotes are based on data from the Norwegian language. The translation was done using a contextualized hermeneutic approach to translation [42] where the data are presented as close to the original context as possible, hence the quotations in particular are directly translated from Norwegian to English.

**Ethical Considerations**

The questionnaire in the quantitative study was anonymous and no personally identifiable information was collected. To complete the questionnaire, adolescents had to scan the QR code of their own volition. Moreover, before they could access the questionnaire, the adolescents were presented with information about the study and they were told that they agreed to participate by answering the questions. The adolescents were also informed that they could stop answering the questionnaire at any time without consequences.

The qualitative study was evaluated and approved by the Norwegian Center for Research Data (reference 631424). Adolescents of 16 years or older could consent to take part in the interview themselves, while active parental consent was required from adolescents younger than 16 years of age. Consent was retrieved via a digital consent form in Nettskjema [36]. All interview informants received a cinema gift card with a value of NOK 150 (approximately US $15) as compensation.

**Results**

**Quantitative Study**

Of the 183 adolescents who completed the web-based questionnaire, 56 (30.6%) adolescents had used a mental health app. Among adolescents who had not used a mental health app (126/183, 68.9%), approximately half (77/126, 61.6%) said they would use one if they had a mental health problem and there were apps available to help, while 48 (38.4%) adolescents said they would not. When asked “What do you think about mental health apps in general?” 25 (13.7%) adolescents answered that they were not or not very useful, 44 (24.0%) adolescents answered “neither”, and 114 (62.3%) adolescents found them somewhat or very useful (mean 3.6, SD 1.1). The regression analysis to predict the perceived usefulness of mental health apps, identified only gender as a significant predictor ($\beta$ = .35; $P < .001$) indicating that girls perceived mental health apps as more useful than boys.

The most frequently reported advantages of using a mental health app were as follows: “It will always be there when I need it” (107/183, 58.8%), “I don’t have to talk to someone face to face” (83/183, 45.4%), “It is more private” (75/183; 41.0%), and “I can be anonymous” (72/183, 39.3%; Table 2). The most frequently reported barriers were as follows: “It might cost money” (87/183, 47.5%), “I don’t know whether the information in them is accurate or true” (82/183, 44.8%), and “I am afraid someone will see the app on my phone” (78/183, 42.6%; Table 3). The most frequently cited media that adolescents used to find information about mental health were by far “Google” (151/183, 82.5%) and “Social media (for example Facebook, Instagram, TikTok)” (90/183, 49.2%; Table 4).
Table 2. Advantages of using mental health apps (N=183).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is more private</td>
<td>75 (41.0)</td>
</tr>
<tr>
<td>I don’t have to talk to someone face to face</td>
<td>83 (45.4)</td>
</tr>
<tr>
<td>It will always be there when I need it</td>
<td>107 (58.8)</td>
</tr>
<tr>
<td>I don’t have to wait to get information</td>
<td>45 (24.6)</td>
</tr>
<tr>
<td>I can get support and information whenever I need it</td>
<td>51 (27.9)</td>
</tr>
<tr>
<td>I don’t have to write things like my mood down on paper</td>
<td>13 (7.1)</td>
</tr>
<tr>
<td>It is personal to me</td>
<td>12 (6.6)</td>
</tr>
<tr>
<td>I can be anonymous</td>
<td>72 (39.3)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (3.3)</td>
</tr>
</tbody>
</table>

Table 3. Barriers to using mental health apps (N=183).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t trust apps</td>
<td>49 (26.8)</td>
</tr>
<tr>
<td>I don’t know whether the information in them is accurate or true</td>
<td>82 (44.8)</td>
</tr>
<tr>
<td>I would prefer to speak to someone face to face</td>
<td>46 (25.1)</td>
</tr>
<tr>
<td>I don’t think apps can help me</td>
<td>36 (19.7)</td>
</tr>
<tr>
<td>I am afraid someone will see the app on my phone</td>
<td>78 (42.6)</td>
</tr>
<tr>
<td>It might cost money</td>
<td>87 (47.5)</td>
</tr>
<tr>
<td>Other</td>
<td>17 (9.3)</td>
</tr>
</tbody>
</table>

Table 4. Most often used media to find information about mental health (N=183).

<table>
<thead>
<tr>
<th>Media</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>151 (82.5)</td>
</tr>
<tr>
<td>Social media (eg, Facebook, Instagram, TikTok)</td>
<td>90 (49.2)</td>
</tr>
<tr>
<td>Mental health mobile apps</td>
<td>23 (12.6)</td>
</tr>
<tr>
<td>Television</td>
<td>28 (15.3)</td>
</tr>
<tr>
<td>Web-based newspapers</td>
<td>8 (4.4)</td>
</tr>
<tr>
<td>Podcast or radio</td>
<td>26 (14.2)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (6.6)</td>
</tr>
</tbody>
</table>

Qualitative Study

**Theme 1: Accessibility—The Significance of Approachable Mental Health Apps**

A central theme that arose from the interviews, was the experienced and perceived accessibility of mental health apps. The adolescents highlighted that mental health apps were a readily accessible alternative to use in order to find information about health care as well as to get help. Further, they pointed out that mental health apps could be a more accessible helping tool than person-to-person offers. In addition, the adolescents talked about accessibility in terms of knowing about mental health apps’ existence, how appealing they are, and potential costs related to using them.

Many of the young people pointed out that they had neither heard of nor used mental health apps.

---

I don't really know about any health-promoting apps that young people use [informant 6]

Some adolescents reflected on reasons why they had not heard of them. For instance 1 adolescent stated that her lack of knowledge of mental health apps may be attributable to marketing campaigns that do not target young people directly or correctly.

I don't think, for example, that the makers of such apps have managed to reach out to young people [...]. I don't know if I'm perhaps in the wrong target group or something, but at least it hasn't really caught on yet, so to speak [informant 1]

The adolescents also pointed out that the information in the mental health apps must be easy to understand to be appealing to them. The adolescents stated that information found on internet might be challenging to comprehend, due to for example
the use of specialized terminology. Therefore, they highlighted that it is important that the information provided in mental health apps is specific to the topic that they want to learn more about.

I think that can be good, if you have a good app where the information is simply explained and that you understand it quite easily, then I think it can be used a lot [informant 1]

An app should be a place where people can read about exactly what they need and not so much medical stuff [informant 1]

The importance of free mental health apps was frequently mentioned in all interviews; adolescents stressed that apps should be equally accessible to everyone regardless of their economic situation. Further, some adolescents stated that charging a fee would weaken the credibility of such apps and call into question the developers’ intentions.

I’ve been on an app where it was obvious that they were going to make money off it. And then the information and everything seemed a bit wrong. Suddenly, I had to pay for a membership in order to continue to use the app; then they lost me [informant 3]

Theme 2: Trustworthiness—The Significance of Credible Mental Health Apps

Most adolescents stated that it is important that app developers are credible and that the information that is provided in apps comes from reliable sources.

It is important that young people find a reliable source. That it’s not just random journalists, or someone who is just trying to get clicks, but that it comes from someone who actually knows something about it. So that you know you are getting the right information [informant 2]

Some adolescents also assumed that it would be easier to judge whether the information on a mental health app is correct, as compared with determining the accuracy of the information on the internet. One adolescent stated that apps developed by established, credible sources would be easier to trust than those created by unknown sources.

I think it would be very easy to just take in everything that’s written there [in a mental health app]. Even the bad stuff. That is why I think that such an app should somehow be under Ung.no [a public information channel for young people in Norway] or something, some reliable sources [informant 5]

Theme 3: Autonomy—Mental Health Apps Can Help Adolescents Help Themselves

Several adolescents stated that self-determination was important to them, so they would appreciate an app, as it would allow them to manage their own mental health and difficulties without interference from adults.

In a way, you want to try to help yourself before it goes so far that you are dependent on others to help you [informant 1]

An app should contain information about what we can do ourselves to get better and where we can ask for help. Sometimes you may just need to get guidelines on how to do it yourself [informant 8]

Further, several adolescents pointed out that an app can be used to ask for help if one does not want to contact in-person services. I think it’s good that there are apps like that, that you don’t have to go and talk to someone, but that you can have an app on your phone that you have all the time [informant 9]

Most adolescents stated that the apps should include a way to communicate with someone.

If it [an app] can be made so that you can communicate with someone, then perhaps the threshold [of asking for help] will be even lower [informant 1]

Mm, it would have been good if the app contained something like that, a way to talk to a person online [informant 2]

Adolescents were also interested in interacting within the app in ways that did not involve another person. Some suggested that the app could contain some predefined questions that could be used to give the user a more personalized experience, or that the app could refer them to the right professional.

Several adolescents also pointed out that contacting in-person services can trigger undesirable actions, which might make it easier to trust an app.

And you are probably afraid that it will somehow..., you want to deal with it, but at the same time you are afraid that, for example, a nurse or someone will tell someone else and that there will be a lot of actions at school, and then everyone will know about it [informant 1]

Finally, a frequently mentioned topic in the interviews was the importance of being anonymous in the apps. Adolescents cited the value they placed on anonymity both if they needed information and if they needed actual help with their mental health, as this could make it easier for the adolescents to seek help.

It is important that you can remain anonymous, because not everyone likes to talk about such things if you are recognizable [informant 6]

Discussion

Principal Findings

In this study, we examined adolescents’ attitudes and general thoughts toward mental health apps using a concurrent, mixed methods approach. This research is an important contribution to understanding what young people think about mental health-promoting mobile apps, as well as their thoughts on what they should include and how they should appear.

Overall, the adolescents expressed positive attitudes toward mental health apps, even though few had experience with using them. A large proportion perceived such apps as useful tools.
that can help them cope with normal stresses of life, which is important in order to promote well-being and prevent mental health problems [43,44]. In the quantitative section of the study, availability was the most chosen benefit of using mental health apps. This is not surprising as adolescents in Norway spend a lot of time on their phones and on social media [21]. Privacy, as well as the possibility of remaining anonymous, was the subsequent most chosen advantage. These results are in accordance with those reported in previous literature [16,25,37]. Further, in the interviews, adolescents expressed that mental health apps can be of help at any time of the day, also when in-person services are not available. However, the adolescents expressed, that for mental health apps to be accessible, they need to also be easily understandable and have no costs attached.

It is well-known that adolescents tend to be reluctant to seek help, despite the importance of obtaining help and support for improved well-being [45-47]. In this study, adolescents expressed that mental health apps can lower adolescents’ threshold for asking for help, by always being available and by offering anonymity in the help-seeking process. Several interviewees also highlighted the empowering potential of mental health apps. The adolescents valued how an app could provide them with the ability to help themselves without interference from others. Meeting young people’s need for autonomy is important to ensure that they ask for help when they need it [28]. The findings from this study concur with previous findings, which suggested that adolescents value the autonomy that mental health apps provide [48].

Further, the adolescents emphasized that mental health apps should provide an anonymous way for them to talk about their problems, with either a web-based assistant or a real person. In addition, some adolescents reported fear of someone else seeing a mental health app installed on their phone as a deterrent to using such apps. Anonymity is often mentioned as an important reason why adolescents use web-based services when searching for help or advice regarding mental health [16,49]. Our findings support that adolescents view anonymity as an advantage of using mental health apps, which agrees with previous findings [16,25].

Another important aspect adolescents highlighted was the difficulty they might have in judging whether the information contained in a mental health app was accurate. Most adolescents reported that they used the internet (ie, Google and social media) to search for mental health-related information. The adolescents stated that mental health apps should be developed by credible organizations or individuals that adolescents already trust. Adolescents also stated that, the information presented in the app should be clear and concise in order to be engaging for them. Previous research have also shown the importance of content and appearance to ensure engaging mental health apps [25,50].

Although adolescents generally have a positive attitude toward the increased development and use of mental health apps, few of the adolescents in our study sample were familiar with such apps. These results are in line with previous research which indicates that adolescents are not highly engaged in mental health apps [13,29]. Indeed, some adolescents believed that they had not heard of such apps due to poor marketing strategies. They stated that, if the apps were meant to target adolescents, then they should be marketed to young people directly. However, there are currently strict rules regarding advertising to children and adolescents in Norway [51], which may pose challenges in devising effective marketing strategies to reach adolescents.

Strengths and Limitations
By including both quantitative and qualitative data, this study provides valuable knowledge and insight into adolescents’ perspectives and opinions on mental health apps. However, there are several limitations that must be taken into consideration. In the quantitative study, the sample size was relatively small and not representative. One should therefore be careful about drawing general conclusions from the findings. However, the results can be an important contribution to the field of research on what adolescents think about mental health apps. Moreover, the interviews were of short duration, which may have influenced the depth of the answers. However, the adolescents were encouraged to speak freely, and they were not interrupted until they had concluded. Follow-up questions were asked when they brought up topics that the interviewer found interesting or considered meaningful for the adolescent to elaborate on. Furthermore, because the qualitative study was intended to complement the quantitative study, we perceive the length and depth of the interviews as sufficient.

One interview was conducted with 2 adolescents together, based on the participants’ own desires. Since they stated it would increase their sense of security during the interviews, the request was accepted. However, we cannot exclude the possibility that the presence of another adolescent influenced the answers the adolescents provided.

Conclusions
Mental health apps can be a useful resource for adolescents, and several apps geared toward adolescents have been developed. Our findings show that there is a lack of knowledge about the existence of mental health apps among adolescents. A large proportion of adolescents expressed that they would use these apps if they knew they were available. This study shows that apps directed toward adolescents should be easily accessible, free of charge, and provide easily understandable information. Adolescents also emphasized the significance that apps should be developed by credible sources or institutions, offer a choice between human and web-based support, and enable users to remain anonymous while seeking help. Future development of mental health apps should take these considerations into account.

https://formative.jmir.org/2024/1/e50222

JMIR Form Res 2024 | vol. 8 | e50222 | p.213
(page number not for citation purposes)
Acknowledgments

The authors would like to thank the adolescents who participated in this study, both those who answered the web-based questionnaire and those who were interviewed. Further, the authors would like to thank Eirik Fjukstad and Annelene Moberg for helping with the recruitment and promotion of the study to adolescents in youth clubs and adolescent health centers in Tromsø. The publication charges for this paper have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors provided substantial contributions to this study. HH and MR analyzed the qualitative data and HK and SK analyzed the quantitative data. HH wrote the first draft of the manuscript. HK, MR, SK read, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

References

19. Herrman H, Jan
41. NVivo. QSR International. 2022. URL: https://www.qsrinternational.com/nvivo
37. Grist R, Cliffe B, Denne M, Croker A, Stallard P. An online survey of young adolescent girls' use of the internet and
36. Schoonenboom J, Johnson RB. How to construct a
35. Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: a systematic
34. Karasouli E, Adams A. Assessing the evidence for e-resources for mental health self-management: a systematic literature
30. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone
29. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
28. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people:
24. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
23. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place
18. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
17. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people:
15. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people:
13. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people:
12. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people:
10. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
9. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
8. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
7. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
6. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
5. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
4. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
3. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
2. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking
1. Wilson CJ, Rickwood DJ, Bushnell IA, Caputi P, Thomas SJ. The effects of need for autonomy and preference for seeking


©Helene Høgsdal, Henriette Kyrrestad, Marte Rye, Sabine Kaiser. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
The Impact of Social Media Use on Mental Health and Family Functioning Within Web-Based Communities in Saudi Arabia: Ethnographic Correlational Study

Bdour Alwuqaysi¹, MA; Alfie Abdul-Rahman¹, PhD; Rita Borgo¹, PhD
King's College London, London, United Kingdom

Corresponding Author:
Bdour Alwuqaysi, MA
King's College London
155 Wandsworth Road
Apt 3004, Sky Gardens
London, SW8 2FZ
United Kingdom
Phone: 44 07470334344
Email: bdour.alwuqaysi@hotmail.com

Abstract

Background: In recent years, increasing numbers of parents, activists, and decision-makers have raised concerns about the potential adverse effects of social media use on both mental health and family functioning. Although some studies have indicated associations between social media use and negative mental health outcomes, others have found no evidence of mental health harm.

Objective: This correlation study investigated the interplay between social media use, mental health, and family functioning. Analyzing data from 314 users, this study explores diverse mental health outcomes. The study places particular emphasis on the Saudi Arabian sample, providing valuable insights into the cultural context and shedding light on the specific dynamics of social media’s impact on mental well-being and family dynamics in this demographic context.

Methods: We collected data through a subsection of an anonymous web-based survey titled “The Effect of COVID-19 on Social Media Usage, Mental Health, and Family Functioning.” The survey was distributed through diverse web-based platforms in Saudi Arabia, emphasizing the Saudi sample. The participants indicated their social media accounts and estimated their daily use. Mental health was assessed using the General Health Questionnaire and family functioning was evaluated using the Family Assessment Device Questionnaire. In addition, 6 mental health conditions (anxiety, self-esteem, depression, body dysmorphia, social media addiction, and eating disorders) were self-reported by participants.

Results: The study demonstrates a pattern of frequent social media use, with a significant portion dedicating 3-5 hours daily for web-based activities, and most of the sample accessed platforms multiple times a day. Despite concerns about social media addiction and perceived unhealthiness, participants cited staying connected with friends and family as their primary motivation for social media use. WhatsApp was perceived as the most positively impactful, whereas TikTok was considered the most negative for our Saudi sample. YouTube, Instagram, and Snapchat users reported poorer mental health compared with nonusers of these platforms. Mental health effects encompassed anxiety and addiction, with age and gender emerging as significant factors. Associations between social media use and family functioning were evident, with higher social media quartiles correlating with a greater likelihood of mental health and unhealthy family functioning. Logistic regression identified age and gender as factors linked to affected mental health, particularly noting that female participants aged 25-34 years were found to be more susceptible to affected mental health. In addition, multivariable analysis identified age and social media use quartiles as factors associated with poor family functioning.

Conclusions: This study examined how social media affects mental health and family functioning in Saudi Arabia. These findings underscore the need for culturally tailored interventions to address these challenges, considering diverse demographic needs. Recognizing these nuances can guide the development of interventions to promote digital well-being, acknowledging the importance of familial connections in Saudi society.

(JMIR Form Res 2024;8:e44923) doi:10.2196/44923
KEYWORDS

social media use; mental health; family functioning

Introduction

Individual’s lives worldwide are now mostly impacted by social media [1]. By enabling individuals to retain offline connections and provide a more welcoming setting for emotional self-disclosure and help-seeking, social media might have a positive influence on their lives [2]. However, there have been rising worries about the possible detrimental impact of social media on individual mental health among certain scholars, governmental organizations, and the public [3]. Prior research has related social media use to multiple mental health conditions including body dissatisfaction, eating disorders, depressive symptoms, and social anxiety [4]. Facebook users, for instance, report higher body dissatisfaction than nonusers, according to a study involving male and female adults [5]. Another example of female adults revealed that although there was no difference between Instagram users and nonusers in terms of body dissatisfaction, Instagram users reported greater body monitoring than nonusers [6,7]. However, each of these works studied these conditions separately, and the combination of these conditions and how they are impacted by social media use has been overlooked. Given this, it becomes evident that there is a pressing need to conduct more in-depth investigations into the specific factors related to social media use that contribute to the development or exacerbation of these mental health concerns.

Family functioning pertains to how well family members fulfill their responsibilities and navigate life challenges [8]. It involves their interactions, growth, and responses to external influences [9]. Research highlights the need to investigate the potential risks posed by mobile device use to family functioning [10,11]. This suggests that families with flexibility and limited mobile device use may experience better cohesion and functioning, emphasizing the necessity for a more in-depth examination of the relationship between family functioning and social media use. Excessive social media use may lead to decreased emotional well-being, which negatively affects relationships [12]. For example, recent Indoneseian research connects social media addiction to mild depression in university students, emphasizing the importance of bracing mental health by promoting family relationships and religiosity while addressing social media overuse [13]. Another study of mental health outpatients suggested a potential association between family functioning and psychological distress [14]. Although research has explored the link between social media and factors such as social connectedness, friendship quality, emotional well-being, and interpersonal relationships [15,16], there is a research gap regarding the impact of social media use on family functioning. Thus, it is imperative to conduct comprehensive research on family functioning and mental health, while analyzing their correlation with social media use.

In this paper, we looked at the relationships between social media use, mental health, and family functioning. Specifically, we examined various mental health conditions such as anxiety, self-esteem, depression, body dysmorphia, addiction to social media, and eating disorders. We selected Saudi Arabia as our sample for this study because of the increasing rates of internet penetration and social media use (90%) [17]. This high prevalence provides a valuable opportunity to examine the potential impact of social media on mental health and family functioning in this specific cultural context. Family is a fundamental unit in Saudi Arabian culture, and understanding the dynamics altered by social media is essential to comprehending broader societal changes. Investigating this aspect will not only enrich the existing literature, but also provide valuable insights into the evolving role of technology in shaping familial relationships and dynamics within the Saudi Arabian context. Thus, Saudi Arabia offers a unique setting characterized by a blend of traditional values, rapid modernization, and an evolving digital landscape. These factors create an intriguing environment in which to investigate the effects of social media on mental health and family dynamics. Two studies on social media use in Saudi Arabia’s web-based community have been conducted, one highlighting the mental health problems associated with social media use [3] and the influence of social media on food consumption among individuals [18]. To our knowledge, there is no research on how social media use may affect family functioning in Saudi Arabia. Thus, our study aimed to provide insights tailored to this cultural context. We aimed to investigate the relationship between social media use and mental health outcomes, with a particular focus on age and gender differences. Gender plays a significant role in this context, and research indicates that women are more likely than men to experience these mental health issues and they also tend to use social media platforms more frequently than their male counterparts [19,20].

Methods

Procedures

The data collected in this study were gathered through a subsection of an anonymous web-based survey titled “The Effect of COVID-19 on Social Media Usage, Mental Health, and Family Functioning.” The survey was designed based on a review of previous studies and surveys on pandemic effects [21-24]. Recruitment for this survey used diverse web-based social media platforms such as Facebook and Twitter. The survey was distributed to participants in Saudi Arabia with a clear emphasis on targeting a Saudi sample. We obtained informed consent from all participants before they completed the survey. The survey was designed to be anonymous and voluntary. Participants were given the option to answer the survey in English or Arabic. The survey targeted social media users who were at least 18 years old, and included sections on demographics, COVID-19, social media use, and well-being. Refer to Multimedia Appendix 1 for more information on the survey.

Participants were asked to choose which of the following prominent social networking sites they had an account on: Instagram, Facebook, Twitter, Snapchat, Google+, Vine, Tumblr, Pinterest, YouTube, and others. Participants were also...
provided with the option to specify “Other” platforms or “I do not have any social media accounts.” Participants were asked to estimate how much time they would spend on each platform during a typical day of browsing. No time (0), <5 minutes (1), 5-15 minutes (2), 15-30 minutes (3), 30 minutes -1 hour (4), 1-2 hours (5), 4-6 hours (7), 6-8 hours (8), 8-10 hours (9), or >10 hours (10). Participants were asked to respond to a variety of questions regarding the activities they engaged in on social media in general (not on specific platforms). Only participants who indicated that they had at least one social media account were presented with questions on social media activities. The web-based survey used in this study, designed based on a review of previous studies and surveys on the pandemic, can serve as a model for future research on the effects of social media use on mental health and family functioning.

Measures

Mental Health Status

We used the General Health Questionnaire-12 (GHQ-12), a well-established 12-item self-report assessment tool, to assess the mental health status of the survey participants. The Likert Scale was used to score all 12 questions in the GHQ-12, generating 3 distinct statistical indicators: typical, suggestive evidence of distress, and severe problems with psychological distress. This tool is widely used to assess psychological distress and mental well-being. The validity and reliability of this scale have been demonstrated [25-27].

Family Functioning Status

We used the Family Assessment Device Questionnaire (FAD) to estimate participants’ family functioning status. FAD is a self-reported scale specifically designed to provide insights into the overall dynamics and functionality within a family. It assesses family relationships and identifies areas of potential dysfunction by adhering to the McMaster Model of Family Functioning. Within the scope of this study, participants were presented with the general functioning scale of the FAD, comprising 12 questions and yielding 4 distinct statistical outputs: healthy, almost healthy, almost unhealthy, and unhealthy. FAD was chosen for this study because it best suited the study objectives and demonstrated its validity and reliability [28-30].

Mental Health Conditions

We present 6 mental health conditions—anxiety, self-esteem, depression, body dysmorphia, addiction to social media, and eating disorders—collected through direct survey questions where participants self-reported whether social media affected those conditions. The inclusion of these conditions in our study is justified based on their prevalence, established links to social media use, public health importance, diverse impacts on mental health, and practical implications for interventions and policies [31-34]. Notably, these issues are dominant, with a high incidence of social media addiction among individuals experiencing these mental health challenges [35].

Pilot Testing

To assess the initial survey, 4 participants participated in a pilot test. User feedback was collected to identify potential problems. The study had improved readability and validity because of this iterative process.

Statistical Analysis

Descriptive statistics were reported as numbers and percentages for categorical variables. The mean and SD are reported for the numerical values. A score was calculated based on the frequency of social media access and the average time spent on social media. The participants were classified based on their scores into 4 quartiles (Q1-Q4). Participants in the first quartile had the lowest social media use, whereas those in the fourth quartile had the highest social media use. To calculate the mental health score for the GHQ-12, we summed the assigned values (0-3) for each response, with higher scores indicating a greater likelihood of mental health issues. For family functioning using the FAD, we assigned values (1-4) to each response, with higher scores indicating a greater likelihood of unhealthy family functioning. To assess self-reported mental health conditions, including anxiety, self-esteem, and depression, we analyzed respondent’s answers to identify correlations with their social media scores. A chi-square test was performed to determine the association between social media use, mental health, and family functioning scores. Logistic regression was performed to identify the factors associated with mental health and family functioning. SPSS 28 (IBM Corp) was used for the analysis, and statistical significance set at \( P<.05 \) is considered statistically significant.

Recruitment

A total of 314 social media users who participated in this study were surveyed between the periods of 2021 to 2023 across 2 rounds to validate the results. Furthermore, 74.5% \((n=234)\) were female, and 24.2% \((n=76)\) were male. Most participants \((n=293, 93.3\%)\) were from Saudi Arabia, whereas the rest \((n=21, 6.7\%)\) represented other nationalities because of the nature of web-based sampling. The Saudi sample serves as an interesting case study for investigating the impact of social media on mental health and family functioning. First, Saudi Arabia is a highly conservative society that is undergoing rapid modernization, with social media playing a significant role in this transformation. According to recent statistics, 29.10 million social media users in Saudi Arabia access it through their mobile devices [17]. Second, there is a lack of research on the effects of social media on mental health and family functioning in Saudi Arabia. Finally, given that Saudi Arabia is a highly collectivist society, family dynamics play a significant role in shaping individual behaviors and attitudes, making it an ideal context to explore the interplay between social media use and family functioning. The largest group of participants was aged 35-44 years old \((n=80, 25.5\%)\), followed by 55-64 years \((n=75, 23.9\%)\). Regarding psychological and medical conditions, most respondents did not report having any psychological \((n=271, 86.3\%)\) or medical condition \((n=216, 68.8\%)\). Regarding the respondents’ educational background, the highest reported level of education was a bachelor’s degree \((n=138, 43.9\%)\), followed closely by a doctorate \((n=81, 25.8\%)\). A small proportion of respondents reported having a master’s degree \((n=60, 19.1\%)\), whereas a minority reported having a high school degree \((n=26, 8.3\%)\). Only a small percentage of the respondents reported having no formal education \((n=9, 2.9\%)\). Regarding participants’
race, most respondents identified as Arab (n=284, 90.4%), whereas a minority identified as non-Arab (n=30, 9.6%). Table S1 in Multimedia Appendix 2 provide more details on the sample demographics.

**Ethical Considerations**

The Effect of COVID-19 on Social Media Usage, Mental Health, and Family Functioning survey was ethically approved by King’s College London ethics committee LRS-19/20-19717.

**Results**

The purpose of this report was to summarize the results of a survey on the perceived impact of social media platforms on mental health.

**Social Media Use**

The data show that most respondents accessed social media frequently, with (n=115, 36.6%) reporting that they go on social media every couple of hours. When asked about the amount of time they spent on social media per day, the most common response was 3-5 hours (n=86, 27.4%), followed by 1-2 hours (n=58, 18.5%). When asked about the duration of their social media sessions, most respondents reported spending approximately 15 minutes or less logged in (n=129, 41.1%). A significant majority of respondents (n=238, 75.5%) reported that they did not feel it was healthy to spend much time on the internet. Most respondents (n=267, 85.03%) reported accessing social media in the evening, whereas 25.48% (n=80) reported accessing it at midnight. When asked about their addiction to social media, 41.1% (n=129) of the respondents reported feeling addicted. Table 1 provides more details on the social media use of the participants. The most common reason for using social media was to keep in touch with friends and family (n=243, 77.39%). Other reasons included inspiration (n=160, 50.96%), browsing or wasting time (n=135, 42.99%), and entertainment (n=130, 41.4%). Around 29.94% (n=94) of the participants reported using social media for work or business purposes, and only 5.1% (n=16) reported using social media for dating or romantic purposes. Table 2 elaborates the reasons for using social media.
Table 1. Social media use among participants.

<table>
<thead>
<tr>
<th>How often currently do you go on social media?</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost never or rarely</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Just about every month</td>
<td>3 (1.0)</td>
</tr>
<tr>
<td>Every couple of days</td>
<td>10 (3.2)</td>
</tr>
<tr>
<td>Just about every day</td>
<td>83 (26.4)</td>
</tr>
<tr>
<td>Every couple of hours</td>
<td>115 (36.6)</td>
</tr>
<tr>
<td>Just about every hour</td>
<td>55 (17.5)</td>
</tr>
<tr>
<td>Every couple of minutes</td>
<td>47 (15.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>On average how much time do you think you spend on social media per day? (h)</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1</td>
<td>25 (8.0)</td>
</tr>
<tr>
<td>1-2</td>
<td>58 (18.5)</td>
</tr>
<tr>
<td>2-3</td>
<td>63 (20.1)</td>
</tr>
<tr>
<td>3-5</td>
<td>86 (27.4)</td>
</tr>
<tr>
<td>5-7</td>
<td>47 (15.0)</td>
</tr>
<tr>
<td>&gt;7</td>
<td>35 (11.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Every time you log in to social media, on average how long do you spend logged in?</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>About 15 min or less</td>
<td>129 (41.1)</td>
</tr>
<tr>
<td>About 30 min</td>
<td>83 (26.4)</td>
</tr>
<tr>
<td>About an hour</td>
<td>65 (20.7)</td>
</tr>
<tr>
<td>More than an hour</td>
<td>37 (11.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Do you feel it is healthy to spend that much time online?</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>237 (75.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>77 (24.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When do you currently access social media? (multiple answers allowed)</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning (from 5 AM to 11:59 AM)</td>
<td>149 (47.45)</td>
</tr>
<tr>
<td>Afternoon (from 12 PM to 6 PM)</td>
<td>125 (39.81)</td>
</tr>
<tr>
<td>Evening (from 6 PM to 11:59 PM)</td>
<td>267 (85.03)</td>
</tr>
<tr>
<td>Midnight (exactly 12 AM to 4:59 AM)</td>
<td>80 (25.48)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Do you consider yourself addicted to social media?</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>185 (58.9)</td>
</tr>
<tr>
<td>Yes</td>
<td>129 (41.1)</td>
</tr>
</tbody>
</table>
Table 2. Reasons for using social media.

<table>
<thead>
<tr>
<th>What do you use social media for? (multiple answers allowed)</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keeping in touch with friends and family</td>
<td>243 (77.39)</td>
</tr>
<tr>
<td>Event planning</td>
<td>49 (15.61)</td>
</tr>
<tr>
<td>Buying and selling</td>
<td>50 (15.92)</td>
</tr>
<tr>
<td>Inspiration</td>
<td>160 (50.96)</td>
</tr>
<tr>
<td>News about COVID-19 (coronavirus) pandemic</td>
<td>125 (39.81)</td>
</tr>
<tr>
<td>To make new friends</td>
<td>8 (2.55)</td>
</tr>
<tr>
<td>To find employment</td>
<td>13 (4.14)</td>
</tr>
<tr>
<td>To browse or time waste</td>
<td>135 (42.99)</td>
</tr>
<tr>
<td>To raise awareness</td>
<td>62 (19.75)</td>
</tr>
<tr>
<td>To provide support to others</td>
<td>58 (18.47)</td>
</tr>
<tr>
<td>To share your posts</td>
<td>66 (21.02)</td>
</tr>
<tr>
<td>To work</td>
<td>76 (24.20)</td>
</tr>
<tr>
<td>None of the above</td>
<td>3 (0.96)</td>
</tr>
</tbody>
</table>

Participants Attitudes

When asked if social media distracted them when they needed to be productive, 33.8% (n=106) of the respondents reported that it did. In contrast, 66.2% (n=208) of the respondents reported that social media does not distract them when they need to be productive. The data also revealed that a significant majority of respondents (n=238, 75.5%) did not care about how many people like or view their posts or pictures, whereas 24.5% (n=77) of respondents reported that they do care. When asked about cyberbullying on social media, 12.1% (n=38) of the respondents reported that they had been cyberbullied in some way, whereas 87.9% (n=276) reported that they had not. In terms of how social media affects self-esteem, only 30.9% (n=97) of the respondents reported thinking negatively about their body when seeing pictures of a person who has the body type they desire. Similarly, only 14.3% (n=45) of respondents reported feeling depressed when seeing posts about intriguing events in other people’s lives. When asked if they accept friend requests or followers from people they do not know to appear more popular, 84.7% (n=266) of the respondents reported that they did not, whereas 15.3% (n=48) reported that they did.

Social Media Impact via Self-Reporting

Participants were asked to rate each platform on a scale of 0 to 5, with 0 indicating that they did not use the platform and 5 indicating that it was the most positive social media platform in their opinion. After those who reported not using the platform, an average score was calculated based on each participant’s opinion, with a lower score indicating a negative impact and a higher score indicating a positive impact.

The platform perceived as having the most positive effect was WhatsApp (4.08), followed by Telegram (3.86), and Pinterest (3.85). The lowest score, indicating the most negative outcome, was observed for TikTok (1.98), followed by Snapchat (3.02). Table S2 in Multimedia Appendix 2 presents a comprehensive analysis of the perceived impact of each social media platform, as reported by the participants. The effects of individual platforms are shown in Figure 1. In addition, Figure 2 provides an overview of the mean scores indicating the perceived impact of social media platforms. Regarding self-reported mental health conditions, the most common effects were anxiety (n=217, 69.09%) and social media addiction (n=206, 65.45%). Other reported effects included depression (n=108, 34.55%), self-esteem (n=97, 30.91%), body dysmorphia (n=69, 21.82%), and eating disorders (n=46, 14.55%) as illustrated in Table 3. However, 25.45% (n=80) of the respondents who reported an impact stated that social media did not affect them. Finally, when asked about the emotions they experienced when using social networking sites, the most common responses were inspiration (n=159, 50.64%), motivation (n=83, 26.43%), and happiness (n=78, 24.84%). Other reported emotions included the fear of missing out (n=31, 9.87%), boosted self-esteem (n=64, 20.38%), jealousy (n=11, 3.5%), and rejection (n=9, 2.87%). Only 2.55% (n=8) of the respondents reported experiencing lower self-esteem when using networking sites. Table S3 in Multimedia Appendix 2 elaborates on social media use habits and their impact on participants.

Regarding the impact of social media on relationships with family members, 76.8% (n=241) of respondents reported that social media did not affect their relationships, whereas 23.2% (n=73) of respondents reported that it did. Among those who reported an impact, 15.3% (n=48) reported a positive effect, whereas 15.9% (n=50) reported a negative effect. When asked if they had a web-based persona, 91.1% (n=286) of the respondents reported that they did not, whereas only 8.9% (n=28) reported that they did.
Figure 1. The perceived effect of each social media platform as indicated by the participants.

Figure 2. Mean scores for the perceived effect of social media platforms.
Table 3. Social media impact.

<table>
<thead>
<tr>
<th>What impact has social media had on your mental health?</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>38 (69.09)</td>
</tr>
<tr>
<td>Self-Esteem</td>
<td>17 (30.91)</td>
</tr>
<tr>
<td>Depression</td>
<td>19 (34.55)</td>
</tr>
<tr>
<td>Body dysmorphia</td>
<td>12 (21.82)</td>
</tr>
<tr>
<td>Addiction to social media</td>
<td>36 (65.45)</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>8 (14.55)</td>
</tr>
<tr>
<td>It has not affected me</td>
<td>14 (25.45)</td>
</tr>
<tr>
<td>None of the above</td>
<td>22 (40.00)</td>
</tr>
</tbody>
</table>

Social Media Impact via Statistical Tools

Regarding mental health, most respondents (n=274, 87.3%) had a score that indicated mental health within the normal range, whereas (n=40, 12.7%) had a score that indicated affected mental health. For family functioning, most respondents (n=269, 85.7%) had a score that indicated healthy or almost healthy family functioning, whereas (n=45, 14.3%) had a score that indicated unhealthy or almost unhealthy family functioning. Table 4 shows the mental health and family functioning of the participants.

Statistical analysis using the chi-square test showed a statistically significant association between social media use and mental health ($P<.001$). Participants in the higher social media quartiles had a higher percentage of affected mental health (26.7% for the fourth quartile and 14.6% for the third quartile) as compared with participants in lower quartiles of social media use (9.1% in the first quartile and 4.8% in the second quartile). Table 5 shows the association between social media use and mental health. Statistical analysis using the chi-square test showed a statistically significant association between social media use and family functioning ($P<.001$). Participants in the higher social media quartiles had a higher percentage of unhealthy or almost unhealthy family functioning (30% for the fourth quartile and 14.6% for the third quartile) as compared with participants in the lower quartiles of social media use (9.1% for the first quartile and 8.3% for the second quartile). Table 6 illustrates the association between social media use and family functioning. Logistic regression was performed to identify the factors associated with mental health. Age and sex showed statistically significant results in the multivariate analysis. Female participants were more likely to have affected mental health as compared with male participants (odds ratio [OR] 4.69, 95% CI 1.42-15.49; $P=.01$).

For age participants who were between 25 and 34 years were more likely to have affected mental health as compared with participants who were 18 to 24 years (OR 6.10, 95% CI 1.42-26.15; $P=.02$). Table 7 illustrates the factors associated with affected mental health.

Logistic regression was applied to identify the factors associated with unhealthy or unhealthy family functioning. The age and social media use quartiles showed statistically significant differences. For gender, female participants were more likely to have unhealthy or almost unhealthy family functioning as compared with male participants (OR 3.32, 95% CI 1.17-9.46; $P=.02$). Regarding social media use quartiles, participants in the fourth quartile were more likely to have unhealthy or almost unhealthy family functioning as compared with participants in the first quartile (OR 4.22, 95% CI 1.45-12.31; $P=.008$). Table 8 illustrates the factors associated with unhealthy or almost unhealthy family functioning.

Table 4. Mental health and family function of participants.

<table>
<thead>
<tr>
<th>Mental health</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected</td>
<td>40 (12.7)</td>
</tr>
<tr>
<td>Within normal range</td>
<td>274 (87.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family functioning</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy or almost healthy</td>
<td>269 (85.7)</td>
</tr>
<tr>
<td>Unhealthy or almost unhealthy</td>
<td>45 (14.3)</td>
</tr>
</tbody>
</table>
Table 5. Association between social media use and mental health.

<table>
<thead>
<tr>
<th>Social media use quartiles</th>
<th>Mental health Within normal range, n (%)</th>
<th>Affected, n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>80 (90.9)</td>
<td>8 (9.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q2</td>
<td>80 (95.2)</td>
<td>4 (4.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q3</td>
<td>70 (85.4)</td>
<td>12 (14.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q4</td>
<td>44 (73.3)</td>
<td>16 (26.7)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 6. Association between social media use and family functioning.

<table>
<thead>
<tr>
<th>Social media use quartiles</th>
<th>Family functioning Healthy or almost healthy, n (%)</th>
<th>Unhealthy or almost unhealthy, n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>80 (90.9)</td>
<td>8 (9.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q2</td>
<td>77 (91.7)</td>
<td>7 (8.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q3</td>
<td>70 (85.4)</td>
<td>12 (14.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q4</td>
<td>42 (70)</td>
<td>18 (30.0)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Table 7. Factors associated with affected mental health.

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th></th>
<th>Multivariable</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Odds ratio (95% CI)</td>
<td>$P$ value</td>
<td>Odds ratio (95% CI)</td>
</tr>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1 (N/A</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
</tr>
<tr>
<td>25-34</td>
<td>4.77 (1.29-17.68)</td>
<td>.02</td>
<td>6.10 (1.42-26.15)</td>
<td>.02</td>
</tr>
<tr>
<td>35-44</td>
<td>1.65 (0.43-6.32)</td>
<td>.47</td>
<td>1.74 (0.42-7.27)</td>
<td>.45</td>
</tr>
<tr>
<td>45-54</td>
<td>0.97 (0.22-4.36)</td>
<td>.97</td>
<td>1.31 (0.26-6.66)</td>
<td>.75</td>
</tr>
<tr>
<td>55-64</td>
<td>0.43 (0.08-2.25)</td>
<td>.32</td>
<td>0.39 (0.06-2.44)</td>
<td>.32</td>
</tr>
<tr>
<td>65-74</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
</tr>
<tr>
<td>Female</td>
<td>3.27 (1.13-9.52)</td>
<td>.03</td>
<td>4.69 (1.42-15.49)</td>
<td>.01</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
</tr>
<tr>
<td><strong>Residency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Saudi Arabia</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
</tr>
<tr>
<td>Outside Saudi Arabia</td>
<td>0.71 (0.16-3.15)</td>
<td>.65</td>
<td>0.96 (0.16-5.68)</td>
<td>.96</td>
</tr>
<tr>
<td><strong>Level of education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>0.67 (0.23-1.99)</td>
<td>.47</td>
<td>0.34 (0.10-1.23)</td>
<td>.10</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>0.55 (0.16-1.94)</td>
<td>.36</td>
<td>0.21 (0.04-1.00)</td>
<td>.05</td>
</tr>
<tr>
<td>Doctorate degree</td>
<td>0.34 (0.09-1.21)</td>
<td>.09</td>
<td>0.25 (0.06-1.15)</td>
<td>.08</td>
</tr>
<tr>
<td>None of the above</td>
<td>2.10 (0.39-11.43)</td>
<td>.39</td>
<td>1.86 (0.21-16.73)</td>
<td>.58</td>
</tr>
<tr>
<td><strong>Social media use quartiles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
</tr>
<tr>
<td>Q2</td>
<td>0.28 (0.11-0.69)</td>
<td>.006</td>
<td>0.41 (0.11-1.60)</td>
<td>.20</td>
</tr>
<tr>
<td>Q3</td>
<td>0.14 (0.04-0.44)</td>
<td>.001</td>
<td>1.47 (0.49-4.47)</td>
<td>.49</td>
</tr>
<tr>
<td>Q4</td>
<td>0.47 (0.20-1.09)</td>
<td>.08</td>
<td>2.36 (0.78-7.16)</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Do you consider yourself addicted to social media?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
</tr>
<tr>
<td>Yes</td>
<td>0.59 (0.30-1.15)</td>
<td>.12</td>
<td>0.83 (0.36-1.94)</td>
<td>.67</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Table 8. Factors associated with unhealthy or almost unhealthy family functioning.

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th></th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Odds ratio (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Odds ratio (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td></td>
<td>1 (N/A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td></td>
<td>1.24 (0.42-3.69)</td>
<td>.69</td>
<td></td>
<td>1.43 (0.40-5.09)</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td></td>
<td>0.82 (0.28-2.41)</td>
<td>.72</td>
<td></td>
<td>0.83 (0.25-2.77)</td>
<td>.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td></td>
<td>0.86 (0.28-2.66)</td>
<td>.79</td>
<td></td>
<td>1.08 (0.29-4.07)</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td></td>
<td>0.41 (0.12-1.37)</td>
<td>.15</td>
<td></td>
<td>0.39 (0.09-1.62)</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td></td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td></td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>2.84 (1.08-7.49)</td>
<td>.04</td>
<td></td>
<td>3.32 (1.17-9.46)</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to say</td>
<td></td>
<td>4.73 (0.41-54.20)</td>
<td>.21</td>
<td></td>
<td>3.89 (0.29-52.68)</td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Saudi Arabia</td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outside Saudi Arabia</td>
<td></td>
<td>0.28 (0.04-2.16)</td>
<td>.22</td>
<td></td>
<td>0.46 (0.05-4.04)</td>
<td>.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td></td>
<td>0.75 (0.26-2.22)</td>
<td>.61</td>
<td></td>
<td>0.51 (0.15-1.68)</td>
<td>.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree</td>
<td></td>
<td>0.30 (0.07-1.23)</td>
<td>.09</td>
<td></td>
<td>0.27 (0.06-1.26)</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctorate degree</td>
<td></td>
<td>0.73 (0.23-2.31)</td>
<td>.59</td>
<td></td>
<td>0.89 (0.24-3.27)</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None of the above</td>
<td></td>
<td>2.10 (0.39-11.43)</td>
<td>.39</td>
<td></td>
<td>3.36 (0.45-24.99)</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social media use quartiles</td>
<td></td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Q1</td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td>0.91 (0.31-2.63)</td>
<td>.86</td>
<td></td>
<td>0.94 (0.30-2.93)</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td>1.71 (0.66-4.43)</td>
<td>.27</td>
<td></td>
<td>1.89 (0.65-5.46)</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td>4.29 (1.72-10.68)</td>
<td>.002</td>
<td></td>
<td>4.22 (1.45-12.31)</td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you consider yourself addicted to social media?</td>
<td></td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td>1.00 (N/A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>1.45 (0.77-2.73)</td>
<td>.25</td>
<td></td>
<td>0.72 (0.33-1.58)</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Discussion

Principal Findings

This study’s insights, grounded in a sample from Saudi Arabia, provide a culturally specific lens at the intersection of social media use and mental health. The prevalence of frequent social media access and reported durations align with the global trend [36], highlighting the pervasive nature of these platforms in the Saudi context. However, it is crucial to interpret these findings within the cultural framework of Saudi Arabia, where familial and social ties are of significant importance [37]. The primary motivation for social media use, namely staying connected with friends and family, resonates strongly with the cultural emphasis on community bonds in Saudi society. This underscores the integral role that social media plays in facilitating and maintaining relationships, which is a culturally significant function. The findings of our study suggest that social media is a popular and frequently used technology in Saudi Arabia, with a significant proportion of users expressing concern about their use habits. Unlike previous research, which analyzed social media platforms and their effect on mental health [38-41], our study applied an in-depth investigation across platforms to evaluate each platform’s impact on mental health. We first tested the effect of time spent on each platform by participants and found no particular association between time spent on various platforms and mental health. Despite small negative correlations between time spent on YouTube, Instagram, and Snapchat and body satisfaction and a small positive correlation between time spent on YouTube and depressive symptoms. Our findings...
imply that future research might benefit from changing attention from time spent generally perusing platforms to participant’s attitudes when engaging with social media platforms. The platform-specific analysis, with WhatsApp emerging as the most positively perceived and TikTok as the most negatively perceived, was influenced by cultural preferences and content norms in the Saudi context. Understanding these variations is essential for tailoring interventions and guidelines to align with the cultural values and sensitivities of the Saudi population. Participants’ attitudes toward social media platforms may have affected their mental health. For instance, these platforms may be a hotspot for frequent and unjustified comparisons of appearances, which might be harmful to mental health. Performing more appearance comparisons with others and thinking that others are more attractive than you are on social media were both independent predictors of lower body satisfaction, more eating disorders, and higher levels of eating disorders. Our results support previous research on teenagers and adults, emphasizing the significance of appearance comparisons as a potential mechanism through which social media use may be detrimental to mental health [42,43].

Our results further imply that, despite their emphasis on physical appearance, these comparisons may have a detrimental effect on issues that are not just related to beauty, such as body satisfaction and eating disorders, but also on general mental health (such as depressive symptoms and anxiety). In our results, compared with men, women frequently paid greater attention to and regarded their beauty as a measure of their self-worth. Therefore, women may be less satisfied with their appearance and more depressed than men are. This is because women may engage in more frequent appearance comparisons on social media [44,45]. The identified effects on mental health, particularly anxiety and addiction, have cultural implications. Given the societal importance placed on mental well-being in Saudi Arabia, these findings underscore the need for targeted mental health awareness and support initiatives within a cultural context.

Logistic regression findings indicate that age and gender are factors associated with affected mental health and unhealthy family functioning. This is in line with a study that emphasized the significance of demographic factors when studying mental health in a Saudi sample [46]. According to our logistic regression findings, age and gender significantly influenced mental health and family functioning in the Saudi context. Female participants had a higher likelihood of experiencing mental health issues (OR 4.69, 95% CI 1.42-15.49; P=.01), emphasizing the need for gender-specific support. In addition, participants aged 25-34 years were more likely to face mental health challenges than those aged 18-24 years (OR 6.10, 95% CI 1.42-26.15; P=.02), suggesting the importance of age-targeted interventions. In terms of family functioning, female participants were more likely to report unhealthy dynamics (OR 3.32, 95% CI 1.17-9.46; P=.02), whereas older individuals in higher social media use quartiles were more likely to experience such challenges (OR 4.22, 95% CI 1.45-12.31; P=.008). Recognizing these age and gender dynamics is vital for tailoring mental health and family support strategies in Saudi Arabia. The statistical tools revealed associations between social media use and mental health, as well as family functioning, emphasizing the need for culturally informed strategies to address potential challenges.

The statistical analysis revealed significant associations between social media use, mental health, and family functioning within the Saudi Arabian context, underscoring the importance of culturally informed strategies. Higher social media quartiles exhibited a notable correlation with a greater likelihood of affecting mental health and unhealthy family functioning. These findings emphasize the nuanced interplay between web-based activities and individual well-being as well as the broader impact on familial relationships. Considering these associations, it is crucial to develop interventions and support mechanisms that are culturally sensitive and tailored to the unique sociocultural dynamics of Saudi Arabia. Recognizing the intricate relationship between social media use and mental health outcomes, along with its implications for family functioning, is the key to fostering digital well-being in this cultural context.

**Limitation and Implications**

It is important to consider the following limitations when interpreting the results of our study. More in-depth longitudinal studies are needed to explore the association between social media use and mental health over time. The sample’s specificity of the sample to Saudi Arabia’s demographic and cultural context may restrict the generalizability of the results to more diverse populations. To enhance the external validity, future research should aim for a broader and more representative sample that encompasses a range of cultural, socioeconomic, and demographic backgrounds. The study’s implications of this study are multifaceted and have significant relevance for the development of targeted interventions and public health initiatives in Saudi Arabia. First, the identified associations between social media use and mental health outcomes underscore the need for specific culturally sensitive interventions. Tailored mental health programs can address the distinct challenges faced by different demographic groups, such as female participants and individuals aged 25 to 34 years, who were found to be more susceptible to affected mental health. These interventions could include educational campaigns, support groups, and digital resources tailored to the cultural nuances of the Saudi context. Moreover, the observed link between social media use and family functioning emphasizes the interconnected nature of web-based behavior and familial relationships. Culturally informed strategies should not only address individual well-being, but also promote healthier family dynamics in the digital age. Public health campaigns can play a pivotal role in raising awareness of the potential impact of social media on family relationships and fostering open discussions within families and communities about responsible digital practices. This study not only contributes to the global discourse on social media, mental health, and family functioning but also offers nuanced insights specific to Saudi Arabia. Recognizing and understanding these cultural nuances are paramount for developing effective policies, educational programs, and support systems that promote positive mental health outcomes tailored to the sociocultural landscape of Saudi Arabia.
Conclusions
This study investigated the perceived impact of social media platforms on mental health and family functioning in a Saudi Arabian sample. The findings reveal important insights with implications for public health initiatives and targeted interventions. This study highlighted the observable association between social media use, mental health, and family functioning. Notably, age and gender have emerged as significant factors influencing mental health and unhealthy family functioning. This underscores the necessity for culturally sensitive strategies to address these identified challenges and tailor interventions to the specific needs of different demographic groups. Recognizing the nuanced associations observed in this study can inform the development of interventions that promote digital well-being, considering the crucial role of familial ties in the societal framework of Saudi Arabia.

Data Availability
The data sets generated during and analyzed during this study are available from the author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Effect of COVID-19 on social media use, mental health, and family functioning survey.
[DOCX File, 3525 KB - formative_v811e44923_app1.docx ]

Multimedia Appendix 2
Statistical analysis tables.
[DOCX File, 33 KB - formative_v811e44923_app2.docx ]

References
12. Christensen SP. Social media use and its impact on relationships and emotions. Brigham Young University. 2018 Jun 1. URL: https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=7927&context=etd [accessed 2023-12-19]


23. Covid-19 social study - UK. University College London. URL: https://redcap.idhs.ucl.ac.uk/surveys/?s=TTXKND8JMK [accessed 2023-12-19]


Abbreviations

- FAD: Family Assessment Device Questionnaire
- GHQ-12: General Health Questionnaire-12
- OR: odds ratio

©Bdour Alwuqaysi, Alfie Abdul-Rahman, Rita Borgo. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Enabling Personalization for Digital Cognitive Stimulation to Support Communication With People With Dementia: Pilot Intervention Study as a Prelude to AI Development

Nick Hird1, PhD; Tohmi Osaki2, PhD; Samik Ghosh3, PhD; Sucheendra K Palaniappan3, PhD; Kiyoshi Maeda2, MD, PhD

1Aikomi Ltd Co, Yokohama, Kanagawa, Japan
2Faculty of Rehabilitation, Kobe Gakuin University, Kobe, Japan
3Data Science and Engineering, SBX Corporation, Tokyo, Japan

Corresponding Author:
Nick Hird, PhD
Aikomi Ltd Co
Yokohama Blue Avenue 12th Floor
4-4-2 Minatomirai
Yokohama, Kanagawa, 220-0012
Japan
Phone: 81 70 4538 2854
Email: nick.hird@aikomi.co.jp

Abstract

Background: Maintaining good communication and engagement between people with dementia and their caregivers is a major challenge in dementia care. Cognitive stimulation is a psychosocial intervention that supports communication and engagement, and several digital applications for cognitive stimulation have been developed. Personalization is an important factor for obtaining sustainable benefits, but the time and effort required to personalize and optimize applications often makes them difficult for routine use by nonspecialist caregivers and families. Although artificial intelligence (AI) has great potential to support automation of the personalization process, its use is largely unexplored because of the lack of suitable data from which to develop and train machine learning models.

Objective: This pilot study aims to evaluate a digital application called Aikomi in Japanese care homes for its potential to (1) create and deliver personalized cognitive stimulation programs to promote communication and engagement between people with dementia and usual care staff and (2) capture meaningful personalized data suitable for the development of AI systems.

Methods: A modular technology platform was developed and used to create personalized programs for 15 people with dementia living in 4 residential care facilities in Japan with the cooperation of a family member or care staff. A single intervention with the program was conducted with the person with dementia together with a care staff member, and for some participants, smell stimulation was provided using selected smell sticks in conjunction with the digital program. All sessions were recorded using a video camera, and the combined personalized data obtained by the platform were analyzed.

Results: Most people with dementia (10/15, 67%) showed high levels of engagement (>40 on Engagement of a Person with Dementia Scale), and there were no incidences of negative reactions toward the programs. Care staff reported that some participants showed extended concentration and spontaneous communication while using Aikomi, which was not their usual behavior. Smell stimulation promoted engagement for some participants even when they were unable to identify the smell. No changes in well-being were observed following the intervention according to the Mental Function Impairment Scale. The level of response to each type of content in the stimulation program varied greatly according to the person with dementia, and personalized data captured by the Aikomi platform enabled understanding of correlations between stimulation content and responses for each participant.

Conclusions: This study suggests that the Aikomi digital application is acceptable for use by persons with dementia and care staff and may have the potential to promote communication and engagement. The platform captures personalized data, which can provide suitable input for machine learning. Further investigation of Aikomi will be conducted to develop AI systems and create personalized digital cognitive stimulation applications that can be easily used by nonspecialist caregivers.

(JMIR Form Res 2024;8:e51732) doi:10.2196/51732

https://formative.jmir.org/2024/1/e51732
KEYWORDS
dementia; digital technology; communication; engagement; cognitive stimulation; artificial intelligence; AI

Introduction

Background

The lack of effective drugs for dementia [1] means that, for the foreseeable future, high-quality care remains the best option to maintain quality of life (QOL) for persons with dementia. However, cognitive decline and behavior changes associated with dementia increase the complexity and difficulty of caregiving, often making it challenging for families and care staff [2]. In particular, responsive behaviors, known medically as behavioral and psychological symptoms of dementia (BPSD), are a range of neuropsychiatric disturbances that affect most persons with dementia and can greatly disrupt caregiving, causing both poor QOL for the person with dementia and physical and mental stress for their caregivers [3]. Communication and engagement between people with dementia and their caregivers lie at the heart of good-quality caregiving [4,5], which is usually provided in dyadic or triadic structures [6] formed by the person with dementia and professional care staff and family caregivers. Communication plays a key role in the successful functioning and quality of care relationships in these care structures as well as greatly influencing the well-being and QOL of everyone involved [7]. Unfortunately, the progression of dementia can significantly impair the communication process for both persons with dementia and their caregivers and lead to inadequate or 1-sided interactions. Overcoming communication issues requires skill, patience, and sensitivity on the part of caregivers, which further adds to the difficulties and stress of caregiving [8]. In addition, the lack of time and resources available for caregivers can deprioritize communication as an activity in itself and result in it being conducted while performing other care activities, which may not be sufficiently personal or meaningful to maintain the psychological well-being of the person with dementia [9].

Communication and engagement are also integral to person-centered care [10], which has been shown to support QOL and help manage responsive behaviors and is widely accepted as best practice in dementia care. Communication difficulties between persons with dementia and their caregivers can prevent adequate expression and understanding of needs, desires, and intentions for both [11], presenting a major barrier to implementing person-centered care [12] including in Japan [13]. Communication and engagement are also important for implementing most psychosocial interventions [14], which are widely used to support caregiving activities to maintain well-being and QOL and manage BPSD. Many psychosocial interventions are based on different types of cognitive stimulation activities, such as cognitive training [15], reality orientation [16], reminiscence therapy [17], multisensory stimulation [18], and music therapy [19]. Several psychosocial interventions have shown promising clinical evidence, in some cases comparable with drug therapies, especially when implemented at the individual level. One of the most well-validated psychosocial interventions is cognitive stimulation therapy (CST) [20], which is based on person-centered care and consists of systematic protocols that combine reminiscence therapy and reality orientation designed to promote enjoyable and meaningful activities for persons with dementia. Clinical studies in groups have demonstrated improved cognition and QOL for persons with dementia [21] and improved caregiver relationships when conducted at the individual level (individual CST; iCST) [22]. Culturally adapted CST protocols have been developed for >20 countries, including CST-J for Japan [23], although its adoption in Japanese care settings remains limited, especially at the individual level.

The lack of trained care staff and practical difficulties associated with the regular and consistent implementation of psychosocial interventions have generated considerable interest in the use of digital technologies [24,25], a trend that was accelerated by the COVID-19 pandemic [26]. Recently, several applications have become available to promote personalized communication and engagement between people with dementia and their caregivers [27], and personalization has been recognized as an important factor in obtaining sustainable benefits [28]. Digital storytelling is a promising approach based on the well-established life story book concept in reminiscence therapy, which uses a person’s own and other relevant content to create fully individualized interventions [29]. In digital storytelling, the physical materials commonly used to facilitate life story book interventions, such as photographs, books, and memorable objects, are replaced with digital media, such as images, videos, and audio. In addition to engagement, digital storytelling aims to help the person tell their own story and has also been used outside dementia in other areas of mental health. Feasibility studies with digital storytelling applications have shown improvements in memory, QOL, and depression [30] as well as additional benefits from the use of digital media. Similar results have been observed in both Western and Asian contexts [31], indicating the wide potential of this approach. However, one of the difficulties of digital storytelling is that preparation requires digital skills, time, and effort [32], which presents a significant adoption barrier for many caregivers and families. The need to create personalized content is avoided through a digital iCST application [33] that uses a pool of precreated generic stimulation activities, including quizzes and games designed to promote engagement, that can be used according to the interests and preferences of the users. A feasibility study showed promising results [34]; however, the lack of personalized content and content diversity was identified as an issue for maintaining engagement and long-term use. Given the importance of personalization, reducing the time and effort required for tailoring applications is a key concern to facilitate broader adoption by nonexpert care staff and families. Data generated by digital technologies can be used to develop artificial intelligence (AI) systems to support caregiving for dementia [35], which can aid the personalization process. However, the development of personalization applications is still at an early stage, and the lack of high-quality personalized data sets related to communication and engagement for people with dementia hinders progress in this area. To address this, a prototype
application called Aikomi was developed to support communication with people with dementia while also capturing high-context personalized data, including behaviors, that would provide a suitable source for developing machine learning models to automate the personalization process. Such AI systems could reduce the personalization burden of digital cognitive stimulation applications and enable their use by nonspecialist caregivers to support communication and engagement with people with dementia in real-world settings.

**Aims and Objectives**

The aim of this pilot study was to evaluate the Aikomi application in Japanese residential care homes for its potential to (1) create and deliver personalized cognitive stimulation programs and promote communication and engagement between people with dementia and their usual care staff and (2) capture meaningful personalized data suitable for the development of AI systems.

**Methods**

**Technology Development**

The Aikomi application was designed by a multidisciplinary team with expertise in occupational therapy (TO); clinical dementia care and psychiatry (KM); digital health (NH); engineering, data science, and machine learning (SG and SKP); the design process also received input from frontline care staff in Japan. The design goal was to create an application that could be routinely used by care staff and family members to promote communication with people with dementia living in both residential care and community settings and would not require expert knowledge, training, or long preparation times. The technology goals were to (1) construct the technology platform using a modular architecture with stand-alone applications for each key function to allow for “plug-and-play” integration of software and hardware applications and (2) incorporate an open connectivity platform to allow for convenient data flow between application modules and facilitate the development of machine learning applications. The connectivity platform used was Garuda (Garuda Alliance) [36], which is a community-driven open connectivity platform previously developed by one of the authors (SG). The modular design provides flexibility to develop fit-for-purpose applications as well as “future-proof” the platform to incorporate other technologies and data sources to enhance optimization and scaling of the personalization process.

A 6-step workflow for using the application was designed through informal discussions with professional care staff and families (Figure 1). The first step is to conduct an interview with families or care staff to obtain information about the person with dementia, including their life history, hobbies, interests, preferences, and abilities, which are used to create a standardized personal profile. On the basis of this personal profile, relevant digital media content is selected or created and, if available, the person’s private content obtained from the family is digitized (as necessary) and uploaded to the system. Next, relevant content is compiled into audiovisual stimulation programs, which are displayed as the intervention. Finally, the behavioral responses of the person with dementia during the intervention are recorded and analyzed to adapt and optimize the content and stimulation programs for the next intervention. To minimize the burden of the personalization process on families and care staff, their participation is limited to the initial interview and the intervention itself. The remaining personalization processes were conducted manually by the research team for this pilot study and will be automated in the future.

The software for all components (modules) was developed as an independent web application whose integration was facilitated by the Garuda connectivity platform. The content management system (CMS) is a repository for generic digital media content, such as images, videos, and audio files. It allows users to upload, create, edit, and store content that can be used by all persons. A separate private CMS module performs the same function as the CMS for the private content of the person, but its use is restricted to this person only. The simulator module is a function to create stimulation elements (called STIMs) from the content in the CMS and private CMS. A STIM consists of a short audiovisual sequence created from combinations of image, audio, and video data and is the building block used to create stimulation programs. The user has complete flexibility to create and select STIMs and compile them in any sequence order to generate the stimulation program. The personal CMS and simulator modules operate at the level of each individual and are not accessible to other users and caregivers. The simulator module also includes a function to create and edit a standardized personal profile of the person. The home and control modules are both used to conduct the intervention. The home module displays the stimulation programs viewed by the person with dementia, and the control module is used by the caregiver to select and control the display of the stimulation program on the home module during the intervention. The control module has functions to pause the program to talk or go back to a previous STIM or forward to a new STIM depending on the response of the person with dementia. A web meeting function was integrated into the home and control modules to allow for the remote use of the Aikomi application, but it was not used in this pilot study, in which only in-person sessions were conducted. The design of the functions and user interfaces for the home and control modules was conducted in collaboration with care staff primarily for use with tablets, but they can also be used on a PC. An example of the home and control modules during use is shown in Figure 2.

The response dashboard module enables the storage, review, and data analysis of video recordings of the person with dementia during the intervention. It was intended for video recordings to be made using the camera on the home tablet; however, this function was not fully operational at the time of the pilot study. Hence, a remote camera was used to record the interventions, and the data were subsequently uploaded to the response dashboard. Each module was a separate web application integrated with the Garuda connectivity platform. The tablets used in the pilot study were installed with SIM cards to avoid the need to use the local network at the residential care home, which can sometimes be unreliable.
Figure 1. Aikomi workflow and system modules.

**Aikomi workflow**

- interview with family or caregiver
- create personal profile for user

**Aikomi system modules**

- **Content management system**
  - upload, create, edit and store generic digital media contents

- **Private content management system**
  - upload, create, edit, and store a user’s own digital media contents

- **Simulator**
  - create, edit, store, and schedule stimulation elements and programs

- **Home**
  - display stimulation programs to person with dementia

- **Control**
  - caregiver operations to select and play stimulation programs

- **Response dashboard**
  - record behavioral responses of person with dementia to program

Figure 2. Home and control module display.
Pilot Study

Study Design and Participants

The study was conducted using Aikomi as a single intervention lasting approximately 30 minutes with 15 people with dementia living in 4 urban residential care homes in Japan and their usual care staff. The participants were nominated by the care managers at each facility based on the following selection criteria: (1) a diagnosis of dementia, (2) displaying negative BPSD such as anxiety or apathy, (3) no hearing or visual difficulties that would prevent using a tablet, and (4) agreement from their family to participate in the study. Participants with a diagnosis of frontotemporal dementia or who had other mental conditions were excluded from this study. All the care staff members who participated in this study were qualified professionals at the residential facilities.

Procedure

The protocol and timeline for the pilot study are shown in Figure 3.

Participant Selection

Care managers at participating care homes explained the details of Aikomi and the pilot study to the families of candidate persons with dementia using materials provided by the research team. After the explanation, written consent was obtained from the families who agreed to participate in the study.

Program Preparation

The research team conducted in-person or telephone interviews with family members to obtain information about the person’s life story, interests, and preferences as well as other topics that the family thought could be meaningful. In addition, the families were asked to provide any suitable family photos if they had them. In one case, a family member was not available for interview, and it was instead conducted with the care manager at the facility. The obtained information was used to create a profile for each person, and relevant digital content was selected and used to create STIMs that corresponded to items in the profile, such as hometown, childhood, family, work, life and cultural events, hobbies, sports, travel, pets, and music. The duration of each STIM ranged from 30 seconds to 3 minutes, and each stimulation program was created by compiling 10 to 20 STIMs in a sequential order expected to be easy for the person with dementia to follow. For this pilot study, all non–family-derived content was obtained by the research team from publicly available sources, and the programs were prepared by the research team within approximately 2 weeks following the family interview.

Intervention

The intervention was conducted as a 1:1 session with the person with dementia and the care staff member seated next to each other at a table in a quiet area of the care home. The home tablet was placed at a comfortable viewing distance for the person with dementia and the care staff member. The control tablet was operated by a research team member seated on the opposite side of the table. A camera was placed at an appropriate position to record the behavioral responses of the person with dementia and care staff member during the intervention. As the application was to be operated by the research team, the care staff were only given a brief overview of the device, and a rehearsal was conducted before the session to familiarize them with the intervention conditions. In addition, to allow the person with dementia to lead responses to the stimulation program as much as possible, the care staff were requested to adopt a passive role during the intervention and respond appropriately according to the person’s behavior and mood, although prompting according to their own judgment was permitted. The stimulation program was started as soon as the participant was seated and appeared comfortable, and apart from initial greetings, the research team member avoided direct interaction with the person with dementia during the session unless actively addressed by them. The intervention began with a few general reminiscence STIMs to act as a “warm up” and accustom the person with dementia to the device, after which STIMs related to personalized topics were played. The selection of STIMs was adjusted according to the mood and responses of the participant, and if a participant showed good engagement with a particular theme, the STIM was repeated or a related STIM was shown. In cases in which
the participant showed little engagement, the STIM was stopped and a STIM of a different topic was shown. The target time for each intervention was 30 minutes but was shortened if the participant appeared tired or uninterested, and the STIM was changed immediately if the participant appeared uncomfortable. The intervention was concluded by showing STIMs such as nature or landscapes as a “cool down” period, and if the participant was agreeable, smell stimulation was conducted using 5 synthetic smells that were embedded on paper strips (such as those used for sampling perfumes) supplied by a commercial smell product company [37]. The decision to use the smell sticks was made after asking each person with dementia and the care staff member after completion of the main stimulation program. The 5 smells used for the evaluation were chocolate, miso (fermented soy paste cooking ingredient), grass, earth, and soap. A smell stick was provided to both the participant and the caregiver without revealing the identity of the smell, and at the same time, a STIM related to the smell was displayed on the home module, for example, showing a bar of a famous brand of chocolate for the chocolate smell. Following the first smell, the remaining smell sticks were similarly used individually according to the participants’ interest in continuing.

**Evaluation Scales**

**Baseline**

BPSD assessment of the participants was conducted during the month before the intervention by the appropriate care staff using the Neuropsychiatric Inventory–Nursing Home version (NPI-NH) [38]. Cognitive ability was recorded using the most recent cognitive assessment conducted at the care home using one of the following cognitive scales: Mini-Mental State Examination (MMSE) [39], revised Hasegawa Dementia Scale (HDS-R) [40], or Nishimura Mental State Scale for the Elderly (NM) [41]. In total, 3 different cognitive assessments were used because of the different types of cognitive tests routinely used at each participating care home.

**Intervention**

Engagement during the intervention was measured using a Japanese translation of the Engagement of a Person with Dementia Scale (EPWDS) [42] prepared by the research team. The EPWDS is a 10-item assessment measuring positive and negative engagement in 5 domains—affect, visual, verbal, behavioral, and social—and was administered soon after the intervention by the attending care staff member, who was also asked to provide their comments on the intervention. Psychological well-being was measured using the 6 motivational and emotional dysfunction items of the Mental Function Impairment Scale (MENFIS) [43]. The MENFIS assessment was conducted at the care home by a care staff member on 3 consecutive days spanning the intervention: the day before the intervention (day 1), on the day of the intervention a few hours after it was conducted (day 2), and the day after the intervention (day 3). The care staff member was also asked to provide their written comments for each assessment. All 3 MENFIS assessments were conducted by the same care staff member, who was not always the same care staff member who was present at the intervention. As the study was intended as a preliminary pilot study, the EPWDS and MENFIS scores were only used as a guide for the qualitative assessment of the intervention.

**Ethical Considerations**

Ethics approval was obtained from the Kobe Gakuin University Human Research Ethics Committee (sourin 18-14). Written approval to conduct the study was obtained from the directors of each of the participating care facilities as well as the families of all the persons with dementia participating in the study.

**Results**

**Overview**

The characteristics of the 15 participants enrolled in the study are shown in Table 1. A total of 5 (33%) people did not meet all the selection criteria: 2 (13%) were “suspected” to have Alzheimer disease but had not received a formal diagnosis of dementia, and 3 (20%) did not show BPSD as defined by the NPI-NH. However, it was decided to include all 5 in the pilot study as the primary purpose was to evaluate Aikomi for its acceptability and communication effects on persons with dementia, and there was agreement to proceed from the care staff and families. The people who did show BPSD were reported to have a wide range of symptoms.
Table 1. Baseline characteristics of persons with dementia in the pilot study.

<table>
<thead>
<tr>
<th>Person</th>
<th>Care home</th>
<th>Sex</th>
<th>Age (y)</th>
<th>Dementia type</th>
<th>Cognitive scores</th>
<th>BPSD</th>
<th>Care level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>Male</td>
<td>93</td>
<td>ADd</td>
<td>8</td>
<td>17</td>
<td>ND</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Female</td>
<td>89</td>
<td>AD</td>
<td>7</td>
<td>11</td>
<td>ND</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>2</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>ND</td>
<td>21</td>
<td>ND</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>Male</td>
<td>79</td>
<td>AD</td>
<td>ND</td>
<td>24</td>
<td>ND</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>Male</td>
<td>76</td>
<td>AD</td>
<td>ND</td>
<td>8</td>
<td>ND</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>Female</td>
<td>87</td>
<td>AD</td>
<td>ND</td>
<td>17</td>
<td>ND</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>Female</td>
<td>84</td>
<td>AD</td>
<td>ND</td>
<td>16</td>
<td>ND</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>Male</td>
<td>79</td>
<td>AD</td>
<td>ND</td>
<td>20</td>
<td>ND</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>Female</td>
<td>83</td>
<td>AD</td>
<td>ND</td>
<td>9</td>
<td>ND</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>ND</td>
<td>15</td>
<td>ND</td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td>Female</td>
<td>90</td>
<td>AD</td>
<td>6</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>0</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>14</td>
<td>D</td>
<td>Female</td>
<td>77</td>
<td>NDj</td>
<td>ND</td>
<td>ND</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>D</td>
<td>Female</td>
<td>90</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>25</td>
</tr>
</tbody>
</table>

a: group home; B: residential nursing home; C and D: older adult rehabilitation facility.
b: Reported dementia diagnosis.
c: BPSD: behavioral and psychological symptoms of dementia.
d: Japanese long-term care insurance rating scale [44].
e: HDS-R: revised Hasegawa Dementia Scale [40].
f: MMSE: Mini-Mental State Examination [39].
g: NM: Nishimura Mental State Scale for the Elderly [41].
h: NPI-NH: Neuropsychiatric Inventory–Nursing Home version [38].
i: AD: Alzheimer disease.
j: ND: not determined.

Engagement

The responses of the persons with dementia during the intervention are shown in Table 2 and Multimedia Appendix 1. The duration of the intervention ranged from 15 to 38 minutes, most persons (13/15, 87%) showed strong positive responses to at least 1 of the STIMs, and none showed discomfort toward Aikomi or requested the session to be stopped. In total, 67% (10/15) of the participants had an EPWDS score of >40 (out of a maximum of 50), indicating both a high incidence of positive engagement and a low incidence of negative engagement. The remaining 33% (5/15) of the participants showed an EPWDS score of >30, which was due to low positive engagement scores (lack of engagement with the stimulation program) rather than the high incidence of negative responses such as anger, anxiety, or discomfort, which were not observed for any participant. Interestingly, person 14, who showed few positive responses during the intervention itself, spoke to thank the research team after the session was completed, which the care staff member said was highly unusual behavior for them.

The types of STIM topics and the responses they generated were analyzed from the video recordings and are shown in Table 3. The most common STIM topic to generate good engagement was family photos, which prompted self-initiated talk and the identification of persons they recognized. However, some participants (6/15, 40%) struggled to recognize their family members and even themselves. One person (person 3) showed no response, and person 4 responded most strongly to photographs of herself in early adult life rather than of her family. Music was also a popular STIM topic, prompting several participants to initiate singing and clapping to both traditional Japanese children’s songs (doyo) and Japanese popular music. The STIM of popular singers known to be liked by the persons with dementia often generated verbal dialogue with the care staff (persons 4, 6, 8, and 13). Japanese traditional arts, such as the tea ceremony, dance, and calligraphy, generated good responses from most women who had performed them (persons 2, 4, 7, 8, 11, and 13), and sports themes (baseball, sumo, and boxing) were popular for all the men (persons 1, 5, 6, and 9). Work-related themes induced mixed responses; however, STIMs related to actions performed during their work (eg, using a
Japanese typewriter, counting money, or preparing fish) generated responses from 33% (5/15) of the participants (persons 2, 6, 7, 8, and 15). Wartime navy service STIMs generated strong positive engagement from person 1 and prompted a detailed recollection of his experiences; this is described in more detail in case study 1.

Table 2. Engagement during the session with the Aikomi device.

<table>
<thead>
<tr>
<th>Person</th>
<th>EPWDS score</th>
<th>Caregiver written comment provided after the session (translation from Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>He showed very spontaneous reaction to family and the warship.</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>She concentrated for volleyball and knitting but could not remember family faces or names. She has severe memory loss.</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>She quickly responded to music, but the response is similar to what the care staff can obtain using tablets.</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>She was able to concentrate due to the music, and was very focused for one singer. Usually, her concentration doesn’t last for 5 minutes, it was very unusual for her to maintain concentration for 30 minutes.</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>He turned to look at the tablet as soon as the images appeared and focused on talking about them, including in great detail about the movies. This is behavior not usually observed by the care staff.</td>
</tr>
<tr>
<td>6</td>
<td>34</td>
<td>Despite being a shy person, he was able to sing in front of everyone. He usually can’t recognize things, but was able to for some pictures. He was anxious because he couldn’t understand many things.</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>She did hand gestures to songs and Japanese dance and hula dance. She did not say much because she is a naturally reserved person, but she showed concentration and seemed excited.</td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>She looked at the tablet and talked continuously but not related to the themes shown. She showed good responses to music and smell.</td>
</tr>
<tr>
<td>9</td>
<td>46</td>
<td>He talked about his mother and explained to us in detail about his old hobbies. He was anxious because he was able to detect the smells.</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>She showed good expression and mood. Usually she doesn’t continue laughing, and I think it was due to the continuous stimulation.</td>
</tr>
<tr>
<td>11</td>
<td>45</td>
<td>She clapped her hands and sang to the music, which is the same response as she regularly shows with karaoke.</td>
</tr>
<tr>
<td>12</td>
<td>32</td>
<td>She had pneumonia before the test, which reduced her will, and she only responded weakly to themes.</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
<td>She is usually not a person who can show good concentration but was calm and concentrated during the test and could recall her memories.</td>
</tr>
<tr>
<td>14</td>
<td>40</td>
<td>She showed most interest in the old photos. At the end of the test, she smiled and said “thank you”.</td>
</tr>
<tr>
<td>15</td>
<td>47</td>
<td>She sang along to the music and looked nostalgic when watching the old photos.</td>
</tr>
</tbody>
</table>

\(^a\)EPWDS: Engagement of a Person with Dementia Scale. Range of scores is 10-50, higher scores indicate higher level of positive engagement, lower scores indicate higher levels of disengagement or negative engagement [42].

\(^b\)Translation by the research team.
Table 3. Engagement by stimulation element (STIM) topic.

<table>
<thead>
<tr>
<th>Person</th>
<th>Self-initiated talk</th>
<th>Prompted talk</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Family, navy service</td>
<td>Hometown, baseball, photography</td>
<td>Viewed war service STIM several times, adding new anecdotes each time. Responded strongly to family photos.</td>
</tr>
<tr>
<td>2</td>
<td>Volleyball, Japanese typewriter, kimono</td>
<td>Family, music, cooking, childhood</td>
<td>Husband of participant 1 but did not respond to family photos. Was good at volleyball. Used typewriter at work.</td>
</tr>
<tr>
<td>3</td>
<td>None</td>
<td>Music</td>
<td>Only responded to music (sang). Did not react to personal topics.</td>
</tr>
<tr>
<td>5</td>
<td>Archery, 100-km walk, food, pigeon racing, gardening, childhood</td>
<td>Family, school, life, pet, hot spring bath, hometown</td>
<td>Talked in detail about participating in 100-km walk. Said that gardening STIM showed incorrect way to grow orchids. Talked about difficult times during childhood.</td>
</tr>
<tr>
<td>6</td>
<td>Boat race, Japanese singer, sumo, food</td>
<td>Baseball, cooking fish, family, music, cultural event</td>
<td>Nervous at first but calmed down when the program started. Talked about working at a fish restaurant. Did not recognize many family members. Sang to music.</td>
</tr>
<tr>
<td>7</td>
<td>Family, Hawaii, and Japanese arts</td>
<td>Bank, counting money</td>
<td>Talked about cousin who lived in Hawaii. Remembered how to count money. Looked closely at calligraphy and flower arranging.</td>
</tr>
<tr>
<td>8</td>
<td>Yakitori (grilled chicken), dog, Japanese poetry</td>
<td>Japanese singer, family, music, hometown</td>
<td>Talked continuously but not related to STIM. Stopped talking to look at yakitori and Japanese poetry. Sang to music.</td>
</tr>
<tr>
<td>9</td>
<td>Family, movies, boxing, baseball</td>
<td>Detective novel, dog</td>
<td>Talked about family life and children and in great detail about films, naming actors and directors. Identified boxers and baseball players and talked in detail.</td>
</tr>
<tr>
<td>10</td>
<td>Watercolor painting, hill walking, athletics</td>
<td>Family, knitting, television drama, food</td>
<td>Said that she wanted to try watercolor painting and knitting. Named some mountains she had climbed. Said she was fast at running at school.</td>
</tr>
<tr>
<td>11</td>
<td>Family, hometown, music, tea ceremony</td>
<td>Schools, school sports, music, cooking</td>
<td>Could identify more family members on repeated viewing. Repeated name of hometown several times. Watched tea ceremony closely for several minutes. Sang to music.</td>
</tr>
<tr>
<td>12</td>
<td>None</td>
<td>None</td>
<td>Did not look at the screen for most of the session. Showed no reaction to any STIM.</td>
</tr>
<tr>
<td>13</td>
<td>Hometown, Japanese singer, koto (Japanese musical instrument), childhood</td>
<td>Kobe, childhood songs</td>
<td>Talked about growing up in Tokyo and sad life of Japanese singer. Watched koto playing closely for several minutes.</td>
</tr>
<tr>
<td>14</td>
<td>None</td>
<td>None</td>
<td>Looked continuously at the tablet but did not show any reaction to any STIM</td>
</tr>
<tr>
<td>15</td>
<td>Family, travel, childhood, hotel work</td>
<td>Paper making, hot spring bath, music, son’s work</td>
<td>Talked about family and climbing Mount Fuji with her son. Recalled working at a hotel and her son’s company.</td>
</tr>
</tbody>
</table>

aTalk initiated by the person with dementia without prompting.  
bPerson responded to prompting by the care staff member.  
cOn the basis of a video review by the research team.

Smell
A total of 60% (9/15) of the participants tried the smell stick stimulation in conjunction with paired audiovisual stimulation, and 13% (2/15) of the participants tried all 5 smell sticks. The responses are shown in Table 4. In most cases, the smell sticks led to pleasurable responses from both the participant and the care staff member, with more laughter than was observed during the audiovisual programs. Almost all the people with dementia (8/9, 89%) had difficulty explicitly identifying the smell; however, majority (6/9, 67%) were able to notice that the smells were different, and recognition of the smell identity did not appear to affect the engagement with the caregivers. Although it was not the intention, some participants perceived the smell program as a test, which may have caused some confusion (person 15) and prompted one person to say that she had lost her sense of smell (person 5). This shows that care must be exercised when using smell stimulation to reduce the risk of causing anxiety to users who may no longer have a sense of smell or be stressed by the inability to correctly identify the smells.
Table 4. Response of participants to smell sticks combined with audiovisual stimulation.

<table>
<thead>
<tr>
<th>Person</th>
<th>Smell stick used</th>
<th>Responsea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
<td>Second</td>
</tr>
<tr>
<td>1</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td></td>
<td>Smelled the stick and, when prompted, looked at the picture. Did not identify either smell verbally but said the name of the chocolate brand and soap brand shown in the pictures in each case. Appeared to enjoy the experience.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Chocolate</td>
<td>Miso</td>
</tr>
<tr>
<td></td>
<td>Smelled the stick and said that she could not smell anything for chocolate but could smell something for miso.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td></td>
<td>Said that she could not smell either smell and, unprompted, said that she had lost her sense of smell.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td></td>
<td>Said that she liked the smell and talked about chocolate. Recognized the second smell as soap and said she liked it. Noticed that the fourth smell was different but could not identify it. Said that the fourth smell was a food smell but she did not know what it was. Talked and laughed during the smell session.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td></td>
<td>Noticed that the first stick had a smell and then said that it smelled delicious and talked about the picture. Did not notice the second smell and said that his sense of smell was poor. Talked and laughed during the smell session.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td></td>
<td>Very actively smelled the first stick and laughed but did not identify the smell. Noticed that the second smell was different. For the third, fourth, and fifth smells, the participant actively smelled the sticks but said that they smelled the same. Talked and laughed during the session.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td></td>
<td>Actively smelled the first stick and asked what it was. Said “not bad” to the second and third stick and noticed that all smells were different but could not recognize them when looking at picture prompts. Talked and laughed during the session.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Chocolate</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Smelled the stick but said that it had no smell.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td></td>
<td>Smelled the first stick but said that she could not smell anything. Appeared reluctant to try the second smell.</td>
<td></td>
</tr>
</tbody>
</table>

aOn the basis of a video review by the research team.
bNot used.

Well-Being and Behavior Changes

The results of the MENFIS well-being assessment of the persons with dementia conducted by the care staff over the consecutive 3-day intervention period are shown in Table 5. In total, 13% (2/15) of the participants showed improvement (reduction in MENFIS score) on the third day compared with the first, and 7% (1/15) of the participants (person 13) showed a reduction only on the second day (after the intervention). All other participants showed similar or identical MENFIS scores for all 3 days except for person 12, who showed worsening over the 3 days, which was attributed to her catching a cold during the period of the intervention. The data for persons 14 and 15 were disregarded as the care staff member mistakenly conducted the second-day assessment before the Aikomi session instead of after. In many cases, the participants had no recollection of the intervention the following day. Little inference can be drawn from the MENFIS data except that there appeared to be no adverse effects on the well-being of the people with dementia after using the Aikomi application.
Table 5: Well-being and behavior changes over the course of the intervention.

<table>
<thead>
<tr>
<th>Person</th>
<th>MENFIS score</th>
<th>Care staff written comments (translation from Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>16</td>
</tr>
</tbody>
</table>

aMENFIS: Mental Function Impairment Scale (6 noncognitive items). Day 1: evaluation the day before the session; day 2: evaluation on the day of the session after it was conducted; day 3: evaluation the day after the session.

bTranslation by the research team.

cBPSD: behavioral and psychological symptoms of dementia.

dDLB: dementia with Lewy bodies.
eRecorded before the session was conducted.

Selected Case Studies

Person 1

Although not formally experiencing BPSD according to the NPI-NH, the care staff member said that person 1 usually displayed apathetic behavior, and the staff member was very surprised by the amount of self-initiated dialogue in response to the family and wartime navy service STIM, which was shown several times during the test. The STIM included a photograph of the ship he served on as well as a Japanese navy song. The person became conversational and could recall the names of former navy colleagues as well as events in detail, providing new information and anecdotes with every repetition of the STIM. The care staff used this information to continue to manually provide stimulation using photographs on the following day, which is likely to have contributed to the continuing improvement trend observed in the MENFIS score from the second to the third day. In contrast, person 2, who was his spouse and lived at the same facility but showed more advanced dementia symptoms, did not respond strongly to the same family STIM used for person 1. However, at the request of the care staff member, an additional session with the Aikomi application was conducted with both persons 1 and 2 watching the family STIM together. In this case, person 1 provided prompts to person 2, who showed a much stronger response than when she viewed the STIM with the care staff member.

Person 10

The caregiver reported that the person had shown agitated behavior in the period before the intervention but was able to come into the room and quickly settled when the session started. She responded to several of the STIMs and actively talked about several of the themes presented. As she was an excellent athlete during her school days, the program included a STIM video of a modern high school girls’ 100-meter race, in which she showed great interest, including making a humorous comment about the lack of clothes worn by current athletes. In addition, she showed quiet concentration on the STIM related to her hobbies of watercolor painting and hill walking and told the care staff member that she wanted to do both again and recalled the names of several mountains that she had climbed. During the smell intervention, she laughed and engaged with the care staff member and research team and appeared to enjoy the experience even though she did not recognize any of the smells. The care staff member said that her behavior during the intervention was not usual and that she was not agitated when she returned to her living area. However, this was not reflected in the MENFIS.
scores, which were assessed by a different care staff member and showed no changes over the 3-day period.

**Person 13**

During the intervention, she showed good responses to STIMs about her hometown (Tokyo), her hobby of playing the *koto* (musical instrument), and a popular Japanese singer. She momentarily cried on 2 occasions when viewing her hometown and the Japanese singer, but her sadness did not persist and was followed by self-initiated talk to the care staff member about her wartime childhood and the Japanese singer’s unhappy life story. The care staff member reported that she usually could not show good concentration, and they were surprised that she was calm and could concentrate during the intervention and recall her memories. In the evening after the test, the care staff member reported that she said that she felt good and that other residents noticed that she was not as agitated as usual; however, on the day after the test, her agitated behavior returned to normal, and this pattern was reflected in the MENDIS scores.

**Discussion**

**Principal Findings**

Maintaining communication and engagement for persons with dementia is integral to caregiving, implementing person-centered care, and managing BPSD. It is also vital for facilitating the positive aspects of care that foster good QOL and sustainable caregiving relationships [45]. Over the last decade, several digital applications have become available to support communication and engagement for people with dementia [27], and to the authors’ knowledge, this pilot study is the first investigation of a digital application for personalized multi-sense stimulation conducted in Japan. All participants in this study were able to accept viewing the programs on the tablet, and these preliminary findings are broadly consistent with those of previous studies that demonstrate the importance of tailoring cognitive stimulation content to the individual profile and abilities of each person to obtain good engagement [28]. As reported for other applications, themes based on personal photographs or music often generated positive responses from participants, although for many people in the study, these themes did not generate the strongest responses during the intervention (Table 4). Instead, a wider range of themes specifically associated with the person’s lived experiences, such as wartime service (person 1), and interests such as gardening (person 5) and Western movies (person 9) were often the ones that resulted in prolonged and in-depth participant-initiated communication with care staff. This is in agreement with digital storytelling studies that have demonstrated the need to use a diverse range of relevant topics to adequately personalize interventions to obtain good engagement [29,30]. In addition, the flexible use of stimulation themes that provide options for caregivers to select and adapt the content of the stimulation program according to the mood and responses of the user. The usefulness of this function was illustrated in person 1, who showed an unexpectedly strong reaction to the war-themed pictures triggering extended dialogue with the caregiver that was facilitated by several repetitions of the same STIM. In contrast, when person 13 started to make negative comments while viewing a STIM, it was possible to quickly change to a new STIM about her hobby, after which she started to make positive comments. Furthermore, the ability of the Aikomi application to easily modify and add new content enables the convenient creation of new personalized stimulation programs each time the Aikomi device is used. This is important for longer-term use to minimize or avoid the repeated use of the same content, which may lead to reduced interest from users and caregivers and has been found to be an issue for applications restricted to a fixed pool of content [33]. In addition, the flexible use of personally targeted activity content has been reported to promote curiosity and encourage self-directed learning in people with dementia [48] and suggests that customizable digital cognitive stimulation applications such as Aikomi may be able to expand the range and depth of self-expression of people with dementia.

A secondary objective of this study was to explore the potential benefits of using paired smell and audiovisual stimulation to promote improved engagement compared with audiovisual stimulation alone. Currently, the use of smell stimulation in dementia care is limited and mainly focused on aromatherapy approaches to improve mood and reduce responsive behaviors using natural oils. However, other studies have shown that smell may play a role in triggering autobiographical and implicit memory [49], and it was thought that synthetic smells associated with daily life experiences could encourage not only the person with dementia but also the care staff member to share their own experiences, as well as being enjoyable. The engagement during paired smell and digital stimulation appeared to involve more smiling and laughter for both the participant and care staff.
member compared with responses to the audiovisual program alone. Interestingly, this was the case even when the person with dementia did not identify the smell and suggests that the combined use of digital audiovisual stimuli with smell may be more effective than smell alone [50].

From its inception, the Aikomi platform has aimed to develop AI capabilities to minimize the time and expertise required by non-specialist care staff and families to use personalized digital applications. This is a critical issue for adoption and sustained use in dementia care, where caregivers are often older and not “digital natives” and have limited time or support to learn how to use and personalize applications [51]. Currently, most digital applications support personalization using two types of approaches: (1) supported collaboration with families to create bespoke interventions or (2) provision of predesigned content for on-demand selection during application use [28]. The bespoke approach often focuses on personal reminiscence and identity-reinforcing applications such as digital storytelling [29,30] and memory books [52,53], but preparation requires extensive family involvement over weeks or months, which may be challenging to sustain. Conversely, on-demand selection approaches, usually providing content curated by expert research teams, allow for immediate use and scalability but may fall short of generating sufficient interest and maintaining long-term use and be more suitable for social interaction [54] and activity-based applications such as iCST [33], music [55], and games [56]. The data obtained by digital technologies open up new opportunities to use machine learning to develop automation that can overcome these personalization barriers as well as optimize and adapt interventions. However, the lack of available high-quality and personalized data sources for people with dementia has limited progress in AI development for dementia care [35], especially for applications to support communication and engagement. To address this, Aikomi’s modular architecture facilitates the seamless capture of user data across key function domains: personal profile, content tagging, sequence ordering, and response analysis. This approach can combine the precision of the bespoke approach with the convenience of using the prepreserved on-demand content. In this pilot study, data from the interviews and the provided content were used with the prepreserved generic content to create bespoke programs in approximately 2 weeks, a reduction from the 6 to 8 weeks reported for digital storytelling [28]. In addition, it was possible to evaluate the performance of personalization using high-context data such as personal profiles, content and sequence attributes, and behavioral responses captured by the Aikomi platform. A related approach was reported by another personalized digital application called Scrapbook, which demonstrates the importance of obtaining multiple personal context–related data inputs to enable analysis [57]. Although AI system development was not the focus of this pilot study and personalization was conducted manually by the research team, this study demonstrates the potential of using the Aikomi platform as a tool to generate personalized data for AI development, and a preliminary investigation was conducted to create machine learning models to automate the personalization process, which is reported elsewhere [58]. In addition, chatbot technology, pioneered by a reminiscence intervention called ReminX [59], demonstrates an alternative trajectory for AI-driven personalization in dementia care.

Although no effects on BPSD were expected from this single-intervention study, the transient behavior changes reported by care staff for some persons (persons 1, 10, and 13) after using the Aikomi application suggest that there may be potential for investigating Aikomi to affect more lasting behavior changes related to BPSD. Although there is currently only a weak clinical evidence base to support the use of cognitive stimulation to manage BPSD, the use of personalized digital interventions embedded with data capture functions may offer the potential to not only create more effective therapeutics but also generate personalized monitoring data that can provide more robust clinical evidence. In the last few years, digital therapeutics (DTX) has emerged as a new category of regulatory-approved medical products that are distinct from drugs and medical devices [60,61]. To date, no DTX interventions for dementia have received regulatory approval, but several applications have been developed for supportive care [62]. DTX for dementia was pioneered by ReminX [59], which was designated as a breakthrough medical device by the Food and Drug Administration (FDA), and more recently by CST Assistant [63], which is a clinical evidence–supported CST-based game application that is now commercially available in Europe. Given the continuing challenges in developing effective and affordable drug therapies for dementia and BPSD, digital interventions for personalized cognitive stimulation such as Aikomi may have potential for clinical development as DTX and offer non-drug options for the management of BPSD and improvement in QOL. In addition to the primary objective of addressing therapeutic goals, the importance of preserving care relationships and creating opportunities for positive aspects of care for caregivers is also gaining increasing attention in dementia care [64,65]. With the growing shortage of professional caregivers, which is particularly acute in Japan, digital technologies that can promote meaningful engagement and improve QOL for both people with dementia and caregivers may become important tools to foster greater participation by family and informal caregivers in caring for their loved ones.

**Limitations**

As this study was the first evaluation of the Aikomi application conducted in a care setting and limited to a single intervention to confirm its safety and acceptability for people with dementia, all inferences are preliminary and need to be confirmed via further multiple-intervention studies. Furthermore, although the aim was for Aikomi to be used independently by the care staff, the research team was required to be present during the interventions for this pilot study, which was a potential source of bias. This was necessary as the Aikomi platform was still a prototype, and the care staff were not familiar with operating digital technologies and required support to use Aikomi and overcome any technical difficulties. In fact, no application-related technical problems were encountered during the intervention, and connectivity issues were largely avoided by using tablets with SIM cards and addressed before the intervention. However, the provision of appropriate training and technical support for care staff to set up, use, and maintain the Aikomi application was not investigated in this pilot study.

https://formative.jmir.org/2024/1/e51732

**JMIR FORMATIVE RESEARCH**

**Hird et al**
Another limitation was that most staff members were unfamiliar with conducting EPWDS and MENFIS evaluations, and the evaluations of some participants were conducted by multiple care staff members because of shift changes, which may have led to some inconsistencies in the results. Furthermore, the EPWDS is not yet available in Japanese, and an unvalidated translation prepared by the research team was used. These limitations need to be addressed in future studies.

Future Research
The next step is to conduct multiple-intervention studies to investigate the effects of longer-term use of the Aikomi application when the program content of each intervention is adapted based on the behavioral responses of the person with dementia. The data derived from these studies will become the basis for developing machine learning models to create algorithms that can optimize personalization and increase the convenience for the care staff to use the system. To generate evidence to support the use of the Aikomi application in dementia care, it is necessary to conduct clinical trials to evaluate its effects on QOL and BPSD, from which its potential for further development as a digital therapeutic can be assessed. Separately, use by family caregivers in their own homes will be investigated to obtain feedback and data that will guide the development of the Aikomi application for community use. In addition, further work is needed to increase the pool of diverse and culturally relevant content that has been curated for people with dementia to reduce the lead time required to prepare stimulation programs. In this pilot study, all non–family-derived content was selected or created by the research team from publicly available sources, and copyright issues were not considered because of the noncommercial nature of this research. However, ensuring copyright compliance for the use of all digital media content, for both content owners and content providers, is a significant issue that must be addressed before the commercial deployment of personalized cognitive stimulation approaches. Finally, to enable more convenient use of smell stimuli with cognitive stimulation, automated smell delivery devices such as diffusers should be investigated for integration with the Aikomi platform.

Conclusions
This pilot study demonstrated that the Aikomi application was able to create personalized cognitive stimulation programs that were acceptable for use in Japanese care homes and may have the potential to promote communication between people with dementia and their care staff. The use of smell stimuli paired with audiovisual stimulation was found to promote enjoyable interactions for many users. In addition, the Aikomi platform captured several types of personalized data, including the behavioral responses during the intervention, which enabled a detailed analysis of the stimulation content preferences of each person with dementia. These results indicate that the Aikomi application has the potential to be used as a tool to provide personalized cognitive stimulation and also generate high-context data suitable for the future development of AI systems to automate the personalization process. Further research will be conducted to develop the Aikomi application as a communication tool that can be easily used by nonexpert care staff and families in residential and community care settings to enhance care relationships and positive aspects of care and aid the therapeutic management of BPSD.

Acknowledgments
The authors would like to thank all the people with dementia and their families and participating care staff for their support and cooperation in participating in this study. The help and advice of the care managers at the participating care homes is gratefully appreciated. Ken Inoue (chief executive officer, Promotool Corporation) is thanked for his kind advice and supply of the smell sticks.

Data Availability
The data sets generated and analyzed in this study are not publicly available as the data contain information that could compromise the privacy of the research participants. The data are available from the corresponding author upon reasonable request.

Authors' Contributions
NH, SG, and SKP collaboratively developed the Aikomi technology platform. KM and TO provided guidance throughout the development as well as advice and supervision of the pilot study in Japanese care homes. NH supported the conduct of the pilot study.

Conflicts of Interest
NH is a cofounder and chief scientific officer of Aikomi. NH and the SBX Corporation are shareholders of Aikomi.

Multimedia Appendix 1
Full details of engagement during the session with the Aikomi device.

References

https://formative.jmir.org/2024/1/e51732

XSL•FO RenderX


https://formative.jmir.org/2024/1/e51732

JMIR Form Res 2024 | vol. 8 | e51732 | p.246

(page number not for citation purposes)


34. Promotool. URL: https://www.promotool.jp [accessed 2023-12-19]


37. Rai HK, Schneider J, Orrell M. An individual cognitive stimulation therapy app for people with dementia: development and usability study of thinkability. JIMIR Aging 2020 Nov 16;3(2):e17105 [FREE Full text] [doi: 10.2196/17105] [Medline: 33196451]

38. Rai HK, Schneider J, Orrell M. An individual cognitive stimulation therapy app for people with dementia and their carers: protocol for a feasibility randomized controlled trial. JIMIR Res Protoc 2021 Apr 08;10(4):e24628 [FREE Full text] [doi: 10.2196/24628] [Medline: 33800508]


Abbreviations

AI: artificial intelligence
BPSD: behavioral and psychological symptoms of dementia
CMS: content management system
CST: cognitive stimulation therapy
DTX: digital therapeutics
EPWDS: Engagement of a Person with Dementia Scale
Designing and Developing a Mobile App for Management and Treatment of Gestational Diabetes in Nepal: User-Centered Design Study

Aarthi Shanmugavel¹, MPH; Prabin Raj Shakya², MS; Archana Shrestha³⁴⁵, PhD; Jyoti Nepal⁴, MPH; Abha Shrestha⁶, MD; Jean-Francois Daneault⁷, PhD; Shristi Rawal⁸, PhD

¹Department of Health Informatics, School of Health Professions, Rutgers, The State University of New Jersey, Piscataway, NJ, United States
²Biomedical Knowledge Engineering Lab, Department of Dentistry, Seoul National University, Seoul, Democratic People's Republic of Korea
³Institute for Implementation Science and Health, Kathmandu, Nepal
⁴Department of Public Health, Kathmandu University School of Medical Sciences, Dhulikhel, Nepal
⁵Department of Chronic Disease and Epidemiology, Center of Methods for Implementation and Prevention Science, Yale School of Public Health, New Haven, CT, United States
⁶Department of Obstetrics and Gynecology, Dhulikhel Hospital, Dhulikhel, Nepal
⁷Department of Rehabilitation and Movement Sciences, School of Health Professions, Rutgers University, Newark, NJ, United States
⁸Department of Clinical and Preventive Nutrition Sciences, School of Health Professions, Rutgers, The State University of New Jersey, Newark, NJ, United States

Corresponding Author:
Shristi Rawal, PhD
Department of Clinical and Preventive Nutrition Sciences
School of Health Professions
Rutgers, The State University of New Jersey
Stanley S Bergen, Jr Building, Suite 923
65 Bergen Street
Newark, NJ, 07107
United States
Phone: 1 973 972 2710
Fax: 1 973 972 7403
Email: shristi.rawal@rutgers.edu

Abstract

Background: Mobile apps can aid with the management of gestational diabetes mellitus (GDM) by providing patient education, reinforcing regular blood glucose monitoring and diet/lifestyle modification, and facilitating clinical and social support.

Objective: This study aimed to describe our process of designing and developing a culturally tailored app, Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital (GDM-DH), to support GDM management among Nepalese patients by applying a user-centered design approach.

Methods: A multidisciplinary team of experts, as well as health care providers and patients in Dhulikhel Hospital (Dhulikhel, Nepal), contributed to the development of the GDM-DH app. After finalizing the app’s content and features, we created the app’s wireframe, which illustrated the app’s proposed interface, navigation sequences, and features and function. Feedback was solicited on the wireframe via key informant interviews with health care providers (n=5) and a focus group and in-depth interviews with patients with GDM (n=12). Incorporating their input, we built a minimum viable product, which was then user-tested with 18 patients with GDM and further refined to obtain the final version of the GDM-DH app.

Results: Participants in the focus group and interviews unanimously concurred on the utility and relevance of the proposed mobile app for patients with GDM, offering additional insight into essential modifications and additions to the app’s features and content (eg, inclusion of example meal plans and exercise videos). The mean age of patients in the usability testing (n=18) was 28.8 (SD 3.3) years, with a mean gestational age of 27.2 (SD 3.0) weeks. The mean usability score across the 10 tasks was 3.50 (SD 0.55; maximum score=5 for “very easy”); task completion rates ranged from 55.6% (n=10) to 94.4% (n=17). Findings from the usability testing were reviewed to further optimize the GDM-DH app (eg, improving data visualization). Consistent with social cognitive theory, the final version of the GDM-DH app supports GDM self-management by providing health education...
and allowing patients to record and self-monitor blood glucose, blood pressure, carbohydrate intake, physical activity, and gestational weight gain. The app uses innovative features to minimize the self-monitoring burden, as well as automatic feedback and data visualization. The app also includes a social network “follow” feature to add friends and family and give them permission to view logged data and a progress summary. Health care providers can use the web-based admin portal of the GDM-DH app to enter/review glucose levels and other clinical measures, track patient progress, and guide treatment and counseling accordingly.

**Conclusions:** To the best of our knowledge, this is the first mobile health platform for GDM developed for a low-income country and the first one containing a social support feature. A pilot clinical trial is currently underway to explore the clinical utility of the GDM-DH app.

**KEYWORDS**
mHealth; mobile health; gestational diabetes; telehealth; usability testing; LMICs; low- and middle-income countries; user-centric design; social cognitive theory; South Asians; maternal health; diabetes; diabetes mellitus; daily glucose monitoring; hospital; medical institution; health center; clinical utility; Nepal; low income; clinical trial; focus group; interview; health care provider; medical practitioner; mobile app; application; digital health; app; apps; health education; web based; self-monitoring; glucose; physical activity; intervention

**Introduction**

Gestational diabetes mellitus (GDM), defined as hyperglycemia with onset during pregnancy, is a major public health issue worldwide. South Asians, who represent approximately one-fourth of the world’s population, are at a disproportionately higher risk of GDM [1-3], and the prevalence of GDM is increasing rapidly in South Asian countries, including Nepal [3-5]. Geographically situated between the two epicenters of the global diabetes epidemic, India and China [6], Nepal has a reported GDM prevalence ranging from 6.6% to 28% [7-9]. These estimates are alarming as GDM is associated with serious adverse perinatal outcomes and unfavorable long-term cardiometabolic consequences in both women and their children [10-14]. Although the short-term health and economic burden of GDM is substantial [15,16], its long-term implications are even more concerning, particularly among South Asian populations who are known to develop cardiometabolic complications at a relatively lower BMI than people with European ancestry [17-19]. Among women with GDM, those with South Asian ethnicity are also known to have a significantly higher risk of developing type 2 diabetes (T2D) compared to other ethnic groups [3,20]. Scalable and cost-effective solutions are thus needed to address the growing burden of GDM and its sequelae, particularly in low-resource South Asian countries, such as Nepal.

Successful GDM management relies on patient adherence to a complex care regimen, including dietary modification, adequate physical activity, weekly-to-biweekly antenatal follow-ups, and regular blood glucose monitoring and logging. Providing dietary and physical activity recommendations is a critical part of GDM management, but in resource-limited settings, such as Nepal, time for diet/lifestyle counseling often competes with other components of care. Face-to-face counseling for diet/lifestyle is also fraught with low participation rates and high attrition, as it is resource intensive for the health care providers and poses time and travel barriers for the patients [21,22]. Additionally, in many countries like Nepal, GDM counseling is only performed once, which is not conducive to facilitating a meaningful behavior change in diet and lifestyle. Mobile health (mHealth) technology provides new opportunities to circumvent these challenges [23] and support the treatment and management of GDM in low-resource settings. Mobile apps can aid in the management of GDM by providing patient education, reinforcing regular glucose monitoring and diet/lifestyle modification, and allowing health care providers to communicate and exchange health information with patients [24]. Mobile technology may offer cost-effective strategies to improve outcomes in patients with GDM by augmenting clinical care and empowering patients with GDM to self-manage their condition, yet this approach has not been tested previously in any low-income country [24].

App-based lifestyle interventions for GDM management are not common, especially in low-income countries, such as Nepal, where its prevalence is rapidly increasing [23,25,26]. To address this gap, our goal was to develop a mobile app that supports self-management and treatment among women with GDM in Nepal. In addition to making the app culturally tailored, our priority was also to design an app that matches the user needs and technological sophistication of the target users. Thus, taking a user-centered design approach [27], we developed the Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital (GDM-DH) app in collaboration with our target users, patients with GDM and providers, in Dhulikhel Hospital, a flagship university hospital in Dhulikhel, Nepal. Consistent with Bandura’s social cognitive theory (SCT) framework [28], the GDM-DH app supports GDM management by providing patient education, reinforcing regular blood glucose/carbohydrate monitoring, increasing self-efficacy for diet/lifestyle modification, and facilitating clinical and social support. Here, we aim to describe the app development process and features of the GDM-DH app.

**Methods**

**Overall Study Design**

The study was conducted at Dhulikhel Hospital, a community-based tertiary-level university hospital of Kathmandu University (Nepal). We took a user-centered design approach to develop a culturally tailored mobile app (GDM-DH)
for management of patients with GDM at the hospital. Figure 1 outlines the steps in GDM-DH app development. In the qualitative/requirement-gathering phase, patients with GDM were recruited for a focus group and structured interviews to show them the app prototype and obtain their feedback on its features and functions. Key informant interviews (KIIs) were conducted with clinicians and patients’ spouses. Incorporating and revising the app prototype based on user input, we built a minimum viable product (MVP), after which additional patients with GDM were recruited for usability testing including the think-aloud protocol [29]. The final GDM-DH app was developed following an iterative process of product design and user testing.

Figure 1. Schematic representation of the user-centered approach for GDM-DH app development among target users (women with gestational diabetes) in Dhulikhel Hospital, Nepal. GDM: gestational diabetes mellitus; GDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital; KII: key informant interview.

Recruitment Procedure and Inclusion Criteria
Located in a periurban setting, about 20 km from the capital city of Kathmandu in Dhulikhel, Nepal, Dhulikhel Hospital has a catchment population of 1.9 million people and delivers approximately 3000 babies annually. All pregnant women receiving antenatal care at the Obstetric Outpatient Department at Dhulikhel Hospital undergo routine screening for GDM at 24-28 weeks of gestation. Inclusion criteria were pregnant women who (1) received antenatal care at Dhulikhel Hospital, (2) received a GDM diagnosis (within the preceding year), (3) owned a smart phone, and (4) could understand and read Nepali. Patients with a confirmed GDM diagnosis were recruited into the study with the help of a senior obstetrician-gynecologist (OB-GYN; coinvestigator in the study) and other staff in the OB-GYN department at Dhulikhel Hospital. A convenience sampling strategy was used to recruit participants meeting the aforementioned inclusion criteria for usability testing (n=18) and qualitative user research (n=19); for the latter, participants were recruited until data saturation was achieved.

Ethical Considerations
The study protocol was approved by the Rutgers Newark Health Sciences Institutional Review Board (Pro2019001883) and the Ethical Review Board of the Nepal Research Health Council (NHRC; registration number 735/2019). Signed written informed consent was obtained from all participants by the research assistant at Dhulikhel Hospital. To ensure participant confidentiality, all documents including participant identifiers, such as the master list and consent forms, are stored separately in a locked cabinet and in a secure password-controlled Health Insurance Portability and Accountability Act (HIPAA)-compliant BOX folder. Only select research staff have access to the documents and folders containing participant identifiers and data. The participants (women with GDM and their spouses) received a mobile recharge card worth Nepalese rupees (NRs) 500 (US $3.77) to compensate for their time for the interview/focus group or usability testing.

App Development Stages

Prototype Development
A multidisciplinary team including experts in GDM, mHealth, and behavior and implementation sciences, as well as health care providers and patients at Dhulikhel Hospital, contributed to the development of the GDM-DH app. Content modules and features to be included in the app prototype were selected based on a literature review, theory-based behavioral strategies, discussions with subject matter experts, and international recommendations and guidelines (including the Package of Essential Noncommunicable [PEN] disease interventions for primary health care in low-resource settings) [30]. A series of meetings and a full-day workshop were conducted with the research team to select and finalize the features of the app prototype. During the meetings, app features were selected based on expected user needs, alignment with theory-based constructs for behavior change, and the logistical and economic feasibility of incorporating these features in the app. The meetings and deliberations spanned over several weeks until differences were resolved and consensus was reached.

Qualitative User Research
After finalizing the app’s content and features, a focus group and structured interviews were conducted to explore the perceived barriers to and facilitators of GDM management and to seek feedback on the GDM app prototype. A total of 12 women with a GDM diagnosis (either current or in the preceding 1 year) were recruited from Dhulikhel Hospital, 4 (33.3%) of whom participated in a focus group and the remaining 8 (66.7%) in structured interviews. KIIs were also conducted with health care providers (n=5) and spouses of patients with GDM (n=2). All interviews were audio-taped and transcribed verbatim.
The focus group and interviews explored the in-depth understanding of the target users’ views and opinions about GDM and its management, including knowledge and treatment gaps, perceived self-efficacy and barriers to GDM management, strategies to increase adherence to dietary/lifestyle management of GDM, and related social, cultural, and environmental factors. At the end of the focus group/interviews, the participants were given a demonstration of the app’s wireframe prototype, which is a schematic illustration that shows the app’s proposed interface, navigation sequences, and features and function. Feedback was then solicited from patients/providers/spouses on (1) the app dashboard, layout, and navigation; (2) usefulness of app features; (3) data entry burden; (4) usefulness of educational modules covered (5) clarity of graphs and data visualizations; and (6) additional features and content.

**Usability Testing**

Incorporating and revising the app prototype based on user input, we built an MVP, the simplest-possible version of the GDM-DH app, which retained the key features and functionalities of the app. The MVP was user-tested with 18 patients with GDM via the think-aloud protocol [29]. Individual 1-on-1 usability testing sessions were conducted in a private space and overseen by 2 facilitators; 1 facilitator led the session, while a designated notetaker recorded patients’ verbalizations. Usability testing consisted of a 2-step think-aloud protocol [32], in which the patients were asked to verbalize their thoughts as they navigated and completed various specified tasks (eg, profile setup, diet entry, weight visualization review, open video lesson) on the app. Patients were also asked to rate the difficulty of completing each task on a 5-point scale ranging from very easy to difficult. Additionally, they were asked to provide feedback on how the features and functions of the app could be improved upon [33]. Content analysis [34] was used to analyze and summarize the notes and verbalizations from the think-aloud protocol.

**Final GDM-DH App Development**

The facilitators’ notes and observations from the usability testing were compiled and scanned for indicators of usability problems experienced by the patients, such as annoyance, doubt, confusion, and slow/incomplete task completion. Based on the findings of the usability testing and recommendations provided by patients with GDM, a list of recommended modifications was compiled and discussed with the app developer (Ayata Inc, Kathmandu, Nepal). The results were used by the app developer to address the key usability barriers and patients’ preferences/feedback and develop a final version of the GDM-DH app for testing in a pilot clinical trial.

**Results**

**Prototype and Content Development**

Based on discussions with subject matter experts over a series of conference calls and workshops, we decided that the GDM-DH app suite would have 2 components: a mobile app for patients and a web-based portal for health care providers and researchers. The features and functionalities of the GDM-DH app were guided by Bandura’s SCT [28], which was selected as it has been widely applied in the dietary/lifestyle management of chronic health conditions [35] and is shown to be a suitable framework for promoting healthy behaviors among pregnant women, including those with GDM [36,37]. We focused our intervention modules on the SCT constructs of self-efficacy (confidence in one’s ability to take action and overcome perceived barriers to a behavior change), self-regulation/self-control (ability to understand and manage feelings, behaviors, and actions to achieve goals), behavioral capabilities (knowledge and skills needed to perform a given behavior), reinforcements (responses to a person’s behavior that increase or decrease the likelihood of occurrences), and outcome expectations/expectancies (anticipated outcomes of a behavior and values a person places on the probable outcomes of a behavior) [38]. Behavior change techniques (BCTs) [39] for the GDM-DH intervention content were selected based on the published literature [40] that maps the BCTs with the SCT constructs for behavior change (eg, the BCT of information about health consequences aligns with the SCT construct of outcome expectations). The SCT constructs observational learning (acquiring a new behavior by watching someone else performing it and observing their outcomes) and environment (physically external factors that can influence a behavior) were not targeted, as it was not feasible to achieve them at this time using the mobile app.

Using the SCT framework for behavior change, we decided that the content and features included in the GDM-DH app would support self-management of GDM by (1) providing health education, (2) helping patients identify and set target health goals (for diet, physical activity, and glucose levels), (3) enhancing their self-efficacy to meet target goals, and (4) facilitating desired support from family members. In SCT, self-monitoring of behavior is the first and most important step in self-regulating appropriate behavior changes [28]. Self-monitoring is also known to be a powerful behavior change strategy for changing diet and physical activity [41]. Hence, we decided that the core features of the GDM-DH app would facilitate the users to record and self-monitor their blood glucose, blood pressure, carbohydrate intake, physical activity, and gestational weight gain (GWG). The app would use automatic feedback and data visualization to aid in self-monitoring, as well as innovative technological features to minimize the self-monitoring burden, including visual aids for estimating carbohydrate portion sizes and smartphone GPS and accelerometer sensors for obtaining physical activity data [42]. Health care providers would be able to use the web-based portal of the GDM-DH app to enter/review blood glucose readings, track patient progress, and accordingly guide treatment and counseling [42].

In many South Asian countries, women are not the sole health decision makers, with mothers-in-law and husbands having a strong influence on their health decisions during pregnancy [43-45]. Additionally, family members are closely involved in a pregnant woman’s food selection and preparation, thus influencing her dietary behaviors. Hence, the team decided that the GDM-DH app would be designed to garner social support...
from family members by allowing the patient to add friends and family members to the app, providing them with access to track and follow the patient’s progress toward stated goals.

**Educational Modules**

The educational content for the app was adapted from international recommendations and guidelines, including PEN disease interventions for primary health care in low-resource settings [30]. The topics covered information included on GDM and associated risk factors, short- and long-term health consequences of GDM for mother and child, clinical and lifestyle management of GDM, dietary and physical activity recommendations for GDM, the role of social support in GDM management, and the proper use of insulin and a glucometer. The educational materials for the app included text- and image-based materials written at less than an eighth-grade reading level and brief 5- to 10-minute videos narrated by the health care providers in Dhulikhel Hospital.

**Qualitative User Research**

Qualitative findings from the focus group and interviews, which have been described in detail previously [46], provided insight into the app content and features, as well as design elements that needed to be added or modified. Briefly, we identified several facilitators of GDM management, including at the individual level (eg, concern for the baby’s health), family level (eg, spousal accompaniment to hospital visits, emotional support), and health system level (eg, universal GDM screening, team approach to management). Notable barriers included inadequate time for diet/lifestyle counseling during hospital visits, an abrupt change in the diet/lifestyle from pre- to post-GDM diagnosis, misconceptions around diet and physical activity, and social/cultural barriers, including food-centered traditions and festivities and a lack of decision-making power in the household. The majority of patients with GDM and their spouses indicated that they lacked sufficient information to manage GDM and were frustrated by frequent hospital visits.

They gave me a diet chart where it was written what to eat in the morning, daytime, and evening, but I didn’t receive any kind of other documents to read about my disease. [Patient with GDM on lack of information about GDM]

They think that like hypertension or thyroid medication, if they start to take an insulin injection, they have to use it throughout their life, and they think it is required for serious conditions only…you know there are a lot of myths. [Provider on common misconceptions]

There are traditions…you have to bless her and give her dai chiura; she has to eat many things. In that time, I think she ate curd and more sweet food; that is why earlier I think when I saw her last blood report, it was under control, but now again, it is all high. I think it is because she ate more during that time…these festivals…earlier too, her sisters came to bless, and she ate more curd and sweets, and again, she started eating rice instead of roti. So, I think that is why her blood sugar level is high again. [Spouse of a patient with GDM on social/cultural barriers to blood glucose management]

All participants agreed that the proposed mobile app and features would be useful and relevant to women with GDM. They believed it would help overcome existing barriers by empowering pregnant women with information and tools to manage GDM and track their progress.

If we had seen this app before, I think we would have been able to control blood sugar levels, and we would have been able to plan. I think after seeing this, it would have helped, it would be useful. [Spouse of a patient with GDM on the GDM-DH mobile app]

Just knowing that this app on its own records and tells you about your physical activity makes us alert…we will know how much more activity we need to do…it makes it easier. [Patient with GDM on the GDM-DH mobile app]

However, both patients with GDM and health professionals requested more content with respect to medical management and diet/lifestyle modification for GDM. Based on findings from the qualitative study, we changed some app features and design elements (eg, data visualization), in addition to modifying the educational materials and other resources to further tailor the GDM-DH app culturally. For example, the educational modules were revised to address specific cultural and social challenges faced by our patients (eg, food-centered festivals, long-held dietary/cultural practices surrounding pregnancy), and appropriate strategies were provided to problem-solve around these barriers. Based on our target users’ suggestions, we added example meal plans with locally available and culturally staple foods, video demonstrations of safe and culturally relevant exercises during pregnancy (eg, yoga, mild hiking, walking), and revised visual aids for carbohydrate estimation to include standardized pictures of staple Nepalese foods with common portion sizes shown in locally used utensils, such as plates, bowls, and cups.

**Usability Testing**

In total, 18 newly diagnosed patients with GDM participated in the usability testing [29] with the MVP. The mean age of patients in the usability testing was 28.8 (SD 3.3) years. All patients were married, and slightly more than half (n=10, 55.6%) were homemakers. The mean gestational age was 27.2 (SD 3.0) weeks, and the average number of years of schooling was 13.3 (SD 2.8) years.

Results from the think-aloud protocol are described in Table 1. The mean usability score across the 10 tasks was 3.50 (SD 0.55; maximum score=5 for very easy). The task completion rates ranged from 55.6% (n=10) to 94.4% (n=17) across the 10 tasks, with the lowest completion rate for the task requiring the patients to look up their next scheduled appointment on the app. All patients except 1 (5.6%) were able to successfully complete tasks requiring them to enter their weight and systolic and diastolic pressure into the app.

https://formative.jmir.org/2024/1/e50823

JMIR Form Res 2024 | vol. 8 | e50823 | p.254

(page number not for citation purposes)
<table>
<thead>
<tr>
<th>Usability testing task</th>
<th>Successful completion, n (%)</th>
<th>Very difficult, n (%)</th>
<th>Difficult, n (%)</th>
<th>Normal, n (%)</th>
<th>Easy, n (%)</th>
<th>Very easy, n (%)</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter fasting blood glucose levels.</td>
<td>16 (88.9)</td>
<td>2 (11.1)</td>
<td>0</td>
<td>5 (27.8)</td>
<td>11 (61.1)</td>
<td>2 (11.1)</td>
<td>3.8 (0.6)</td>
</tr>
<tr>
<td>Enter postprandial blood glucose levels and view the glucose chart.</td>
<td>16 (88.9)</td>
<td>2 (11.1)</td>
<td>0</td>
<td>5 (27.8)</td>
<td>13 (71.2)</td>
<td>0</td>
<td>3.7 (0.5)</td>
</tr>
<tr>
<td>Enter the systolic blood pressure level.</td>
<td>17 (94.4)</td>
<td>1 (5.6)</td>
<td>0</td>
<td>6 (33.3)</td>
<td>10 (55.6)</td>
<td>2 (11.1)</td>
<td>3.8 (0.6)</td>
</tr>
<tr>
<td>Enter the diastolic blood pressure level and view the blood pressure chart.</td>
<td>17 (94.4)</td>
<td>1 (5.6)</td>
<td>0</td>
<td>5 (27.8)</td>
<td>13 (71.2)</td>
<td>0</td>
<td>3.7 (0.5)</td>
</tr>
<tr>
<td>Open a video on GDM nutrition.</td>
<td>16 (88.9)</td>
<td>2 (11.1)</td>
<td>0</td>
<td>3 (16.7)</td>
<td>12 (66.7)</td>
<td>1 (5.6)</td>
<td>3.7 (0.8)</td>
</tr>
<tr>
<td>Add a friend or family member in the app.</td>
<td>12 (66.7)</td>
<td>6 (33.3)</td>
<td>1 (5.6)</td>
<td>4 (22.2)</td>
<td>5 (27.8)</td>
<td>7 (38.9)</td>
<td>3.2 (1.0)</td>
</tr>
<tr>
<td>Enter weight.</td>
<td>17 (94.4)</td>
<td>1 (5.6)</td>
<td>0</td>
<td>4 (22.2)</td>
<td>12 (66.7)</td>
<td>2 (11.1)</td>
<td>3.9 (0.6)</td>
</tr>
<tr>
<td>Find out the daily step count from today.</td>
<td>12 (66.7)</td>
<td>6 (33.3)</td>
<td>1 (5.6)</td>
<td>3 (16.7)</td>
<td>6 (33.3)</td>
<td>7 (38.9)</td>
<td>3.2 (1.0)</td>
</tr>
<tr>
<td>Figure out how many carbohydrates were consumed at breakfast today.</td>
<td>12 (66.7)</td>
<td>6 (33.3)</td>
<td>0</td>
<td>4 (22.2)</td>
<td>6 (33.3)</td>
<td>8 (44.4)</td>
<td>3.22 (0.8)</td>
</tr>
<tr>
<td>Find out when the next appointment is.</td>
<td>10 (55.6)</td>
<td>8 (44.4)</td>
<td>2 (11.1)</td>
<td>5 (27.8)</td>
<td>5 (27.8)</td>
<td>6 (33.3)</td>
<td>2.8 (1.0)</td>
</tr>
</tbody>
</table>

aGDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital.
bGDM: gestational diabetes mellitus.
cPatients were asked to rate the difficulty of completing each task on a 5-point scale ranging from 5 for “very easy” to 1 for “very difficult.” The mean (SD) score represents an average score for the corresponding task across all patients.

**Modifications to the GDM-DH App After Usability Testing**

Findings from the focus group, interviews, and usability testing were reviewed to identify recurring themes of feedback with respect to the GDM-DH app’s content, usability, navigation, and functionalities. These findings allowed us to gain insight into participants’ thought processes with regard to app use, places where they encounter difficulties, and ways to improve the app’s usability. As shown in Textbox 1, these themes served as valuable insights that were used by the app developer to address the key usability barriers and participants’ preferences/feedback and develop a final version of the GDM-DH app for testing in a pilot clinical trial. For example, considering that nearly half of the usability testing participants struggled in identifying their upcoming antenatal appointments, we decided that the upcoming appointments would be shown in a list, in addition to the calendar. Additionally, we included a comprehensive video tutorial on how to navigate the GDM-DH app and use its features. In the web-based portal, in addition to the features decided on by the research team and software development company, we incorporated new features requested by the providers, including a patient finder tool, data export customization, and additional analytics in the dashboard, such as the average number of antenatal visits per patient and the percentage of patients under medical therapy for GDM.
Textbox 1. Modifications requested and incorporated in the Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital (GDM-DH) app based on user research and usability testing among target users (women with gestational diabetes mellitus [GDM]) in Dhulikhel Hospital, Nepal.

### Educational resources
- Use simple and clear language in educational resources, avoiding jargon and using terms understandable to the target users.
- Provide clear guidance on healthy food selections and appropriate portion sizes, including example meal plans with locally available foods.
- Clarify common misconceptions around diet and physical activity during pregnancy.
- Address specific cultural and social challenges faced by target users with respect to diet/lifestyle modification and provide appropriate strategies to problem-solve around those barriers.
- Add information about the signs and symptoms of hypoglycemia, along with practical strategies for effective management.
- Clarify criteria for when insulin or medication is indicated for GDM.
- Add examples of physical activity and exercise videos that are appropriate for pregnant women.
- Include information about contraindications for physical activity during pregnancy, including warning signs to stop exercising.
- Add an educational module on how family members can provide support to women with GDM.

### App interface and features
- Use bigger font sizes; enlarge and make the images clearer.
- Revise the visual aids to include standardized pictures of common food portion sizes in locally used utensils, such as plates, bowls, and cups.
- Add reminders to input blood glucose, weight, diet, and blood pressure data.
- Use pop-ups to confirm data input and avoid double entries.
- Use bar graphs instead of line graphs to present blood glucose, weight, and blood pressure visualizations.
- Make glucose, weight, diet, and blood pressure data visible both as a list and in graphical format.
- Show upcoming appointments in a list, in addition to the calendar.
- Add the hospital hotline number under Help and Support.
- Include a video tutorial on how to navigate the GDM-DH app and use its features.

### Web-based portal
- Include a patient finder feature to search for a patient quickly.
- Add a function to customize data export based on specific parameters and layout.
- Add dashboard analytics based on patient data (e.g., average number of antenatal visits per patient, percentage of patients under medical therapy for GDM)

The decision to retain or disregard the requested modifications was based on their potential influence on behavioral and clinical outcomes, budget feasibility, the implementation time frame, and their potential impact on the app’s scalability in the future. For example, meal plans were added because they could be easily incorporated with minimal time and cost but would have maximal health gains. However, our target users also suggested that we add a feature to connect and chat with other app users with GDM, as well as a platform to communicate directly with health care providers via the app; these features were not added due to logistical, technical, and funding constraints but may be considered in the future.

### Final GDM-DH App
Based on the SCT framework, the final version of our culturally tailored GDM-DH app supports GDM management by providing patient education, reinforcing regular blood glucose/carbohydrate/weight monitoring, increasing self-efficacy for diet/lifestyle modification, and facilitating clinical and social support. A detailed description and justification of the app features as well as their alignment with the SCT constructs are provided in Table 2.
### Table 2. Description and justification of the GDM-DH\(^a\) app features designed for self-management and treatment of Nepalese women with GDM\(^b\).

<table>
<thead>
<tr>
<th>Feature and description</th>
<th>Rationale</th>
<th>SCT(^c) constructs</th>
</tr>
</thead>
</table>
| **Educational modules** | The level of knowledge about GDM is significantly associated with self-management efficacy and glycemic control [47-49]. It also facilitates better information retention as patients can go through the lessons at their own pace and revisit them at their convenience. | - Self-efficacy  
- Behavior capabilities  
- Outcome expectancies |
| Educational modules consist of text- and image-based materials and brief videos covering various health and nutrition topics related to GDM and its management. | - Self-regulation  
- Self-efficacy |
| **Blood glucose monitoring** | Self-monitoring of glucose levels is associated with an increase in self-efficacy and better glycemic control [50]. Data visualizations increases patient awareness and helps health care providers with timely and informed clinical decision-making. | - Self-regulation  
- Self-efficacy |
| Blood glucose levels can be logged in for fasting and postprandial levels 3 times daily. These data are displayed in color-coded graphs (red and green bars for above- and within-target ranges, respectively). | - Self-regulation  
- Behavior capabilities |
| **Carbohydrate monitoring** | It builds the user’s self-efficacy for understanding and changing their carbohydrate intake patterns. | - Self-regulation  
- Behavior capabilities |
| The app incorporates standardized pictures of local Nepalese foods with common portion sizes to help the user estimate and track calories and carbohydrate (grams) in their meals. | - Self-regulation  
- Behavior capabilities |
| **Blood pressure monitoring** | Data visualization increases patient awareness and helps health care providers with timely and informed clinical decision-making. | - Self-regulation  
- Self-efficacy |
| Users can log(track their systolic and diastolic blood pressure. These data are displayed in color-coded graphs (red and green bars for above- and within-target ranges, respectively). | - Self-regulation  
- Self-efficacy |
| **GWG\(^d\)** | Weight monitoring builds the user’s self-efficacy for understanding and managing their GWG. | - Self-regulation  
- Self-efficacy |
| Based on weights entered by the user, the app creates a graph comparing the user’s weekly GWG rate with the recommended guidelines for optimal GWG, depending on the pre-pregnancy BMI. | - Self-regulation  
- Self-efficacy |
| **Physical activity** | It builds the user’s self-efficacy for understanding and changing their physical activity patterns. | - Self-regulation  
- Self-efficacy |
| The app integrates with the Google-Fit app to pull and graph physical activity data, including the step count. | - Self-regulation  
- Self-efficacy |
| **Appointment reminder** | The reminder system enables patient adherence to the antenatal care regimen [51]. | - Reinforcement |
| The app has an in-built calendar, which the users can use to record and view upcoming antenatal appointments. | - Reinforcement |
| **Social network** | The app helps a user garner social support from friends/family and offers a source of accountability, motivation, and shared experience. | - Reinforcement |
| Via a social network “follow” feature, the patient is able to list 1 or more friends/family members as contacts in the app and give them the permission to view their logged data or progress summary. | - Reinforcement |
| **Web-based portal** | It streamlines the providers’ workflow, as they can quickly look at patient data visualizations to understand patient behaviors and progress and accordingly guide their treatment and counseling. | - Reinforcement |
| Health care providers can use the web-based admin portal to register a new patient, as well as enter, update, or review clinical and other patient-related information (glucose/blood pressure/weight/diet, clinical history/notes). | - Reinforcement |

\(^a\)GDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital.  
\(^b\)GDM: gestational diabetes mellitus.  
\(^c\)SCT: social cognitive theory.  
\(^d\)GWG: gestational weight gain.
Mobile App
The mobile app, which is patient facing, includes 6 feature icons on its home page: (1) Blood Glucose, (2) Food Intake, (3) Blood Pressure, (4) Weight, (5) Physical activity, and (6) Appointment (Figure 2). Using these features, the app allows patients to record and self-monitor their blood glucose, blood pressure, carbohydrate intake, physical activity, and GWG and track them over time. The app has a goal-setting feature and uses innovative technological features to minimize the self-monitoring burden, such as visual aids for carbohydrate estimation and integration with the Google-Fit app to automatically log physical activity data. Based on the data entered, the app provides automatic feedback about blood glucose, blood pressure, and GWG via a feedback engine that compares the user data to existing guidelines and recommendations. The app also generates visual displays summarizing their blood glucose, blood pressure, diet, physical activity, and weight patterns, allowing the user to easily monitor their alignment and progress toward target goals. In addition to the self-monitoring features, multimedia video- and text-based modules are included in the app as educational resources. The GDM-DH app also includes a social network “follow” feature, allowing the user to list 1 or more friends/family members as contacts in the app and give them the permission to view their logged data or progress summary. The app has an in-built calendar, which the users can use to record and view upcoming antenatal appointments.
Figure 2. Key app features of the GDM-DH app designed for self-management and treatment of Nepalese women with GDM. GDM: gestational diabetes mellitus; GDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital.

Web-Based Portal

The web-based portal can be securely accessed by researchers and health care providers from any device that supports modern web browsers. The web-based portal has features for patient management, data capture and review, and data dashboard/visualization. Health care providers can register a new patient; enter, update, or review clinical and other patient-related information (patient vitals, measurements, clinical notes, medications, etc); and schedule or make changes in appointments. Using Apache Kafka, the web-based portal syncs with the mobile app and allows providers to access data and graphs summarizing the patient’s diet, physical activity, weight, blood pressure, and blood glucose patterns. This streamlines the providers’ workflow and allows them to easily track patient progress and accordingly guide their treatment and counseling.
Researchers and admin users can use the web-based portal to add new users, add/update the app modules/images/visualization, and audit changes made by users.

Data Security
Data from the mobile app are stored in a HIPAA-compliant, secure server hosted by Amazon Web Services. MongoDB is used to implement the database service, which is a free, open source, no–Structured Query Language (no-SQL) database program tailored to support big data. It is encrypted and access-controlled using tokens to ensure it cannot be accessed outside the app. Apache Kafka is the core of the streaming service to ensure reliability, high availability, and scalability. All communications are transmitted using the Secure Sockets Layer (SSL) standard protocol, and data are encrypted at rest to ensure security. App access is controlled using a secure username-password combination.

Discussion
Principal Findings
Self-management of GDM is vital for controlling blood glucose levels and minimizing complications for both mother and baby [10,52]. In this paper, we described the design and development of GDM-DH, a culturally tailored GDM management app targeted for use among patients and health care providers in Nepal. Following the SCT framework [53], the GDM-DH app assists in self-management of GDM by enhancing the patients’ knowledge of and self-efficacy in adhering to blood glucose monitoring and recommended diet and physical activity regimens. With respect to the health care providers, the app’s web-based portal offers easy data visualization to track patient progress and treatment response, facilitating informed clinical decision-making at the point of care.

The growing literature highlights the importance of culturally tailoring health interventions, that is, adapting the intervention content and instructions according to the target users’ culture, diet, language, religion, customs, and beliefs [54-56]. Several studies [56-59], including systematic reviews [60], have found that culturally tailored programs and interventions are effective at improving disease knowledge, behavioral outcomes (eg, physical activity), access to care, and clinical outcomes, including glycated hemoglobin (HbA1C) levels in patients with diabetes. Our GDM-DH app development incorporated a user-centered design approach that actively involved end users and used ethnographic and human-computer interaction methodologies to better understand and meet their needs, the user-centered design approach is especially paramount to developing culturally tailored mobile interventions, ensuring app engagement, and promoting digital health equity in low-income countries, such as Nepal [54]. Although mobile technology has been widely applied and proven efficacious for self-management of diabetes outside of pregnancy [29,32,41,61], mobile app–based lifestyle interventions for GDM are just emerging, even in high-income countries [62-68]. To date, there are only a few published randomized controlled trials that have evaluated mobile app–based solutions for GDM management [63-68]. Two recent reviews [69,70] on app-based interventions for GDM concluded that most existing studies were of moderate quality and were underpowered to detect effects on perinatal outcomes but, overall, indicated improved glycemic control in the mobile intervention groups compared to standard care alone. However, most existing app-based interventions for GDM management focus on remote blood glucose monitoring, with manual feedback from health care providers [24,69,70], which can be resource intensive and burdensome for both providers and participants, thus limiting the potential for widespread dissemination and impact, particularly in low-resource settings, such as Nepal. Additionally, despite evidence showing that lifestyle and T2D interventions based on behavior change theory are more effective [71-73], we found only 2 studies [67,68] that incorporated relevant theories in their mobile intervention for GDM. Furthermore, only 2 studies [65,67] involved input from target users during app design and development, which is critical for ensuring the effectiveness and acceptability of evidence-based interventions [74].

Strengths
Our GDM-DH app overcomes existing limitations and represents an advance over previous mobile interventions for GDM as it provides a comprehensive solution for GDM management without the need for additional work from health care professionals and incorporates user-centered principles and theory-based BCTs to meet the specific needs and technological literacy of our target users. To the best of our knowledge, the GDM-DH app is also the first to contain a social support component by including a social network feature. Although the educational content and custom food tracker (and visual aids for calorie/carbohydrate estimation) in our GDM-DH app were specifically designed for our target population of Nepalese women, they can be easily adapted and scaled to other contexts and populations by applying similar user-designed principles.

Limitations
Several limitations of the GDM-DH app and its development process are worth mentioning. First, the number of participants in our usability study was limited. Additionally, during usability testing, we may have observed the best-case scenario for comfort and confidence in using the app, leading us to overestimate the true usability and technological proficiency. Second, we did not use a structured framework, such as the Delphi method [75], to organize and structure the discussions to guide our GDM-DH app and intervention development, which will make it difficult for others to replicate our study procedures. Nonetheless, we used many of the elements of the Delphi framework, including an iterative approach, the use of experts, and group-based responses. Similarly, although we did not use a structured framework to guide the modification and optimization of the GDM-DH app and intervention content, our prespecified criteria, including feasibility, scalability, and affordability, align with existing frameworks designed to optimize and evaluate an intervention prior to implementation (eg, multiphase optimization strategy, or MOST [76], and Acceptability, Practicability, Effectiveness, Affordability, Spill-Over Effects, and Equity, or APEASE [77], criteria). Due to resource limitations, we were unable to address all the needs and...
suggestions provided by our target users. For instance, the current version of the app has limited social support, but future versions could incorporate features such as a discussion forum to foster a stronger network and support system among users. The manual entry of blood glucose and blood pressure levels is also a limitation. Since the app was designed to address the cultural barriers and technological literacy of a specific population, generalizability is a potential issue, and adapting the app to other settings and contexts would require a similar level of user research and testing among the target populations. Large-scale cluster-randomized clinical trials at multiple urban and rural sites in Nepal are needed to establish the effectiveness and generalizability of the GDM-DH app to women with GDM across the country.

Future Research

In the future, we plan to further optimize the GDM-DH app by including Bluetooth-enabled data entry and advanced smartphone functionalities, such as multimedia push notifications, and gamification features, which have been shown to increase retention and improve engagement with mHealth interventions [78-80]. Push notifications enable on-the-go delivery of intervention content, providing the necessary trigger and reinforcement when the specific intervention is most needed or is most convenient for the user. Gamification (application of game elements) features provide entertainment and intrinsic/extrinsic motivation (eg, point scores, badges, levels, a leaderboard) to promote sustained engagement with the app [81-85].

Conclusion

The GDM-DH app targets specific needs identified by our target population in our pilot research and has unique features, including a social support feature, visual aids for carbohydrate estimation, and a comprehensive support system without imposing an additional provider burden. A proof-of-concept pilot clinical trial (NCT04198857) to study the feasibility, acceptability, and preliminary efficacy of the GDM-DH app is currently underway.

Acknowledgments

This project was supported by the National Institutes of Health/Fogarty International Center (grant #NIH 1R21TW011377-01).

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions

Study concept and design were managed by SR; acquisition, analysis, and interpretation of data, AS (Aarthi Shanmugavel), PS, AS (Archana Shrestha), JN, AS (Abha Shrestha), J-FD, and SR; drafting of the manuscript and literature review, AS (Aarthi Shanmugavel) and SR; critical revision of the manuscript for important intellectual content, AS (Aarthi Shanmugavel), PS, AS (Archana Shrestha), JN, AS (Abha Shrestha), J-FD, and SR; and statistical analysis, JN and SR. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References


43. Hmone MP, Dibley MJ, Li M, Alam A. A formative study to inform mHealth based randomized controlled trial intervention on pregnancy outcome of pregnant women with gestational diabetes mellitus (GDM) and recommendations for developing a mobile app for GDM. 2022 Presented at: NUTRITION 2022; June 14–16, 2022; Online. [doi: 10.1093/cdn/nza061_007]


Abbreviations

**BCT:** behavior change technique  
**GDM:** gestational diabetes mellitus  
**GDM-DH:** Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital  
**GWG:** gestational weight gain  
**HIPAA:** Health Insurance Portability and Accountability Act  
**KII:** key informant interview  
**mHealth:** mobile health  
**MVP:** minimum viable product  
**PEN:** Package of Essential Noncommunicable  
**SCT:** social cognitive theory  
**T2D:** type 2 diabetes
Preoperative Virtual Reality to Expose Patients With Breast Cancer to the Operating Room Environment: Feasibility and Pilot Case Series Study

Jordana L Sommer1,2, MA; Kristin Reynolds2,3, PhD; Pamela Hebbard4,5, MD; Michael S D Smith6, MSc; Natalie Mota3,7, PhD; W Alan C Mutch1, MD; Jessica Maples-Keller8, PhD; Leslie Roos2, PhD; Renée El-Gabalawy1,2,3,5,7, PhD

1Department of Anesthesiology, Perioperative and Pain Medicine, University of Manitoba, Winnipeg, MB, Canada
2Department of Psychology, University of Manitoba, Winnipeg, MB, Canada
3Department of Psychiatry, University of Manitoba, Winnipeg, MB, Canada
4Department of Surgery, University of Manitoba, Winnipeg, MB, Canada
5CancerCare Manitoba, Winnipeg, MB, Canada
6National Research Council Canada, Winnipeg, MB, Canada
7Department of Clinical Health Psychology, University of Manitoba, Winnipeg, MB, Canada
8Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, United States

Corresponding Author:
Jordana L. Sommer, MA
Department of Anesthesiology, Perioperative and Pain Medicine
University of Manitoba
AE207 Harry Medovy House
671 William Avenue
Winnipeg, MB, R3E 0Z2
Canada
Phone: 1 204 787 4713
Email: sommerj@myumanitoba.ca

Abstract

Background: Clinically elevated preoperative distress and anxiety are common among patients undergoing cancer surgery. Preoperative interventions have been developed to mitigate this distress and anxiety but are inconsistent in efficacy and feasibility for broad implementation.

Objective: This preliminary pilot study aims to assess the feasibility and utility of a newly developed virtual reality (VR) intervention to expose patients awaiting breast cancer surgery to the operating room environment and a simulation of anesthetic induction.

Methods: Patients undergoing breast cancer surgery (N=7) were assigned to the VR intervention or control (treatment as usual) group and completed self-report measures of distress and anxiety before surgery, on the day of surgery, and after surgery (5 and 30 d postoperatively). Those in the intervention group trialed the VR simulation 1 to 2 weeks preoperatively and provided qualitative and quantitative feedback. We assessed the feasibility of recruitment capability and study design and evaluated participants’ impressions of the intervention using self-report rating scales and open-ended questions. We also descriptively examined distress and anxiety levels throughout the duration of the study.

Results: Recruitment occurred between December 2021 and December 2022 and progressed slowly (rate: 1 participant/7 wk on average; some hesitancy because of stress and being overwhelmed). All participants who consented to participate completed the entire study. All participants were female and aged 56 (SD 10.56) years on average. In total, 57% (4/7) of the participants were assigned to the intervention group. On average, intervention participants spent 12 minutes engaged in the VR simulation. In general, the intervention was rated favorably (eg, clear information, enjoyable, and attractive presentation; mean % agreement 95.00-96.25, SD 4.79-10.00) and as helpful (mean % agreement 87.50, SD 25.00). Participants described the intervention as realistic (eg, “It was realistic to my past surgical experiences”), impacting their degree of preparedness and expectations for surgery (eg,
“The sounds and sights and procedures give you a test run; they prepare you for the actual day”), and having a calming or relaxing effect (eg. “You feel more relaxed for the surgery”).

Conclusions: This preoperative VR intervention demonstrated preliminary feasibility among a sample of patients undergoing breast cancer surgery. Results and participant feedback will inform modifications to the VR intervention and the study design of a large-scale randomized controlled trial to examine the efficacy of this intervention.

Trial Registration: ClinicalTrials.gov NCT04544618; https://clinicaltrials.gov/study/NCT04544618

(JMIR Form Res 2024;8:e46367) doi:10.2196/46367

KEYWORDS

virtual reality; preoperative anxiety and distress; breast cancer surgery; anesthesia; feasibility; pilot

Introduction

Preoperative Distress

Preoperative distress, a situational emotional reaction (eg, fear, worry, and helplessness), is common among surgical patients [1,2]. Extant research has identified several adverse health-related perioperative outcomes of preoperative distress in both the presence and absence of a mental disorder [3-7]. In particular, patients undergoing cancer surgery experience clinically meaningful elevated rates of preoperative distress, ranging from 23% to 77% in recent research [8-10]. Preoperative distress is also associated with various adverse health-related outcomes for patients undergoing cancer surgery specifically, including increased postoperative pain, nausea, discomfort, and fatigue, among others [11-15]. In recognition of its detrimental impact, the National Comprehensive Cancer Network (NCCN) designated distress as the “6th vital sign” [16].

Preoperative Psychological Interventions

Receiving a cancer diagnosis is a significant and life-altering event, often intensified by the necessity for major surgical intervention and an uncertain health trajectory. In considering the adverse health implications of psychological distress, several preoperative interventions (eg, education, relaxation training, and stress management) have been developed that seek to improve psychological and physical functioning before surgery by establishing realistic expectations of the surgical process and helping patients cope with surgery-related uncertainty and distress [17-24]. However, the literature reveals conflicting evidence regarding the efficacy of many such interventions [17-23]. Importantly, the interventions that are supported by evidence require delivery by licensed health care providers [21,22] and often require multiple sessions, rendering them impractical for large-scale implementation, particularly within the constraints of a publicly funded health care system.

Virtual Reality Interventions

Virtual reality (VR) interventions have shown considerable promise in reducing psychological distress in nonsurgical contexts [25-30]. Research in this area has examined the effectiveness of VR exposure therapy for the treatment of anxiety and posttraumatic stress disorder [25,27,28,31-33], where the user virtually and systematically confronts feared content to overcome anxiety. Patients often prefer using VR for exposure therapy over traditional in vivo exposure [34,35], and it may also be more straightforward to administer. This innovative technology has also been gaining popularity in broader medical contexts and has shown promising results in pain management [36-40] and cognitive and physical rehabilitation among various medical populations [37,41-43]. In contrast to therapist-guided VR exposure used in mental health settings, which may be used as a component of one-on-one psychotherapy over a duration of months, VR interventions developed for use in medical settings do not typically require a specialized health care professional to administer and can often achieve desirable outcomes following a shorter duration of use [37,44-47].

Preparatory Interventions for Stress Exposure

In a preoperative context, VR could be used to psychologically intervene before patients develop clinically elevated distress and are affected by the adverse downstream effects of distress (while also targeting any existing distress about surgery). This is similar to stress inoculation training, a form of cognitive behavioral intervention, aimed at psychologically preparing individuals for future exposure to a stressful environment through preliminary exposure to elements of that environment [48]. This form of intervention has been adapted using VR technology [48-52], and preliminary evidence supports reductions in predeployment distress for military personnel using such interventions to prepare for combat [50,53,54]. In fact, similar methods have been applied to psychologically prepare patients before surgery, including operating room (OR) tours before surgery [55], given that the OR environment is noted as distressing for many surgical patients [3,56,57]. Although this intervention was associated with reductions in preoperative distress [55], it has limited feasibility for broad administration because of the infrequent availability of ORs for such purposes and the limited resources and personnel to implement this intervention.

Preoperative Applications of VR

The use of VR to expose patients to the OR environment and preoperative process resolves some of these limitations. A few studies have implemented such interventions to target preoperative distress and other perioperative outcomes, largely among pediatric patients (all but one of the identified studies) undergoing variable types of surgeries (eg, general, neurological, and plastics or ear, nose, and throat) [58-64]. Small-scale meta-analyses examining this literature support the initial efficacy of such interventions in reducing preoperative distress [65-67], although some studies have used VR distraction interventions (eg, using games or relaxation) as opposed to...
exposing users to the OR environment. Importantly, this research is in its infancy, and only a few studies exist in this area to date, supporting the need for further exploration.

Gaps in the Literature

Although the preoperative VR interventions described in the preceding section demonstrate preliminary efficacy in mitigating preoperative distress and potential for broad implementation within the constraints of our health care system (ie, relatively low cost, do not require specialized professional training to administer, can be used repeatedly in different settings, translated into multiple languages, and adapted across surgery types), studies examining these interventions are not without limitations. First, most studies in this area have focused on samples of pediatric patients undergoing surgery; further research is needed to establish the efficacy of such interventions among adult patient samples. Second, no identified studies to date have evaluated a preoperative VR intervention using patients scheduled to undergo an oncological procedure, a population with elevated levels of preoperative distress [8-10]. Third, existing preoperative VR interventions have limited immersion capabilities (eg, lack of user embodiment [ie, the ability to visualize and manipulate virtual representations of the user’s body] and use of prerecorded virtual videos as opposed to a fully immersive virtual environment), which may weaken their impact on mitigating distress through reduced realism. Fourth, these studies lacked follow-up data beyond the acute postoperative phase (eg, <1 wk after discharge), which is needed to understand the long-term impact on postoperative recovery. Finally, many of these studies did not gather user feedback on the intervention, which is vital to help maximize the potential impacts of these interventions.

This Study

In light of these identified gaps, this study aims to assess the feasibility of, and preliminarily pilot (in case series format), a novel VR OR simulation targeting preoperative distress and anxiety among a sample of patients undergoing breast cancer surgery. Specifically, regarding feasibility, the aims are to assess recruitment capability and identify resulting sample characteristics, understand participants’ impressions of the study design and intervention, and evaluate data collection procedures and outcome measures. Finally, this study will also pilot-test the preliminary impact of the intervention on perioperative distress and anxiety in a case series format. The results of this study will inform modifications to the VR simulation and the design of a large-scale randomized controlled trial (RCT) to evaluate the efficacy of this intervention.

Methods

Overview

This study used a single-blind randomized design to assess the feasibility of and pilot the VR simulation to expose patients undergoing breast cancer surgery to the OR and preoperative process. This study represents an in-depth preliminary analysis of a larger pilot study (ClinicalTrials.gov; NCT04544618). Participants were assigned to the VR intervention group or the treatment-as-usual (control) group at the time of recruitment. Randomization was stratified according to the type of breast cancer surgery (with vs without reconstruction) and whether neoadjuvant chemotherapy was received to enable equal proportions of participants with these characteristics across study groups; research demonstrates differences in distress levels according to these factors [68]. All participants completed self-report measures 1 to 2 weeks before surgery (ie, baseline; VR group participants tested the intervention at this time), on the day of surgery, 5 days after surgery, and 30 days after surgery. Notably, the initial planned design included a third study arm (ie, non–surgery-related VR; Nature Treks), which was ultimately dropped because of recruitment challenges. Ethical amendments were approved supporting this change (and others noted in the Recruitment Capability and Sample Characteristics section), and the trial registry has been updated accordingly.

Participants

We recruited a convenience sample of adult patients undergoing breast cancer surgery by describing the study at patients’ surgical oncology appointments and preoperative information sessions and circulating study posters. Interested patients provided their contact information to enable a telephone discussion with the research coordinator and eligibility screening (those viewing the poster contacted the coordinator directly). Participants were eligible if they were being scheduled to undergo breast cancer surgery under general anesthesia at the Health Sciences Centre (a tertiary care hospital in Winnipeg, Canada) and could speak and read English. Those unable to provide informed consent or unable to participate in a VR intervention (eg, owing to visual or auditory impairment) were excluded. Our initial target was to recruit 15 participants per group, with a study aim to evaluate recruitment capability.

Procedure

Participants randomized to the VR group trialed the intervention 1 to 2 weeks before surgery (baseline). Those in the control group received no additional intervention beyond their standard medical appointments and optional preoperative information sessions (offered to all patients). All participants completed self-report measures at baseline (those in the VR group received additional measures to assess intervention feedback). On the day of surgery, preoperative distress and anxiety were reassessed while the participants were in the preoperative holding area and again in the OR before anesthetic induction. At 5 days and 30 days after the operation, all participants were readministered a subset of the baseline measures, and those in the VR group provided additional intervention feedback at the 5-day postoperative assessment. The participants in the VR group completed baseline measures in person (at the time of the intervention), and all participants completed the day-of-surgery measures in person. All other measures were completed online through the web-based survey platform, Qualtrics (Qualtrics International Inc).

Intervention

A VR development team at the National Research Council of Canada, in collaboration with coauthors (RE and JLS), developed the VR OR simulation for use in this research (a
technical paper describing the simulation more in depth is in progress). The simulation development stages included creating an initial prototype based on the observation of surgeries and consultation with medical personnel, developing an anesthetic induction script based on example scripts provided by several anesthesiologists, integrating input from an anesthesiologist (WACM) on the initial prototype, and refining the platform through multiple iterations of feedback from coauthors. The VR simulation begins with the participant sitting on an examination table (reflected as the OR table in the simulation), wearing the VR head-mounted display, and holding the controllers (enabling user embodiment and visualization and manipulation of virtual arms). The participant is instructed to imagine it is their day of surgery, including how they might be feeling that day. The participant then spends at least 5 minutes exploring the virtual OR, which includes relevant machinery and equipment, personnel, sounds, and details such as a mammogram displayed on a computer screen. This free exploration is followed by a scripted portion, where the participant is instructed to lie supine on the virtual OR table and is taken through a mock anesthetic induction process led by the virtual anesthesiologist and nurse; the patient is prompted to answer questions similar to those they will be asked on the day of surgery (eg, name, date of birth, type of surgery, and allergies) and is virtually prepared with monitoring devices by the nurse (eg, blood pressure cuff, a pulse oximeter, and electrocardiogram stickers and electrodes). The simulation ends after the virtual oxygen mask is placed on the patient’s mouth and the screen darkens (refer to Multimedia Appendix 1). We used the Oculus Rift S VR system (Meta Platforms) with a cable connection to a laptop computer for the intervention administration.

Ethical Considerations
This study was approved by the University of Manitoba Health Research Ethics Board (#HS23957). All participants provided written informed consent before participation. No participant-identifying information was included with the study data. Each participant was assigned a study identification number, which was used to collate participant data over the study duration. All participants were provided with a CAD $25 (US $18.94) gift card after completing the study, and the cost of parking for those attending an intervention appointment was reimbursed.

Measures

Preoperative Distress
A total of 4 self-report measures assessed preoperative distress, including 2 preoperative-specific scales (Preoperative Intrusive Thoughts Inventory [PITI] and Amsterdam Preoperative Anxiety Information Scale [APAIS]) and 2 nonspecific visual analog scales (NCCN Distress Thermometer and adapted “Anxiety Thermometer”). The PITI and APAIS were only administered at baseline and on the day of surgery, whereas the Distress and Anxiety Thermometers were assessed at all 4 time points (and in the OR before induction). At the 5-day postoperative follow-up, the participants were asked which measure best captured their experience of distress or anxiety, which will inform the selection of the primary outcome measure for the upcoming RCT. At 5-day postoperative follow-up, they were also prompted to retrospectively rate their level of distress/anxiety from 0 (no distress/anxiety) to 10 (extreme distress/anxiety) corresponding to 8 different “events” ranging from prediagnosis (average level of distress/anxiety before receiving a cancer diagnosis) until the 5-day follow-up.

The PITI is a validated and reliable 20-item self-report measure of preoperative anxiety [69]. Items (eg, “I worry that I won’t wake up”) are rated on a 4-point scale, ranging from 0 (not at all) to 3 (most of the time). Summing the items yields a total score ranging from 0 to 60, where a score ≥15 indicates clinically significant preoperative anxiety [69]. The APAIS is a validated and reliable 6-item measure of preoperative anxiety [70]. Items (eg, “I am worried about the procedure”) are rated on a 5-point scale, ranging from 1 (not at all) to 5 (extremely). A total score is calculated by summing all items, ranging from 6 to 30, where a score ≥10 indicates clinically elevated preoperative anxiety [70]. The NCCN Distress Thermometer is a visual analog scale that resembles a thermometer, with a scale ranging from 0 (no distress) to 10 (extreme distress) [71]. This has been validated among several oncology samples [72,73]. Distress is rated using a “past-week” timeframe (modified to present time in the OR), and a cut-off score of 4 indicates clinically elevated distress [73]. Because of the common interchangeable use of the terms distress and anxiety within the perioperative and oncology literature and the lack of clear differentiation between these terms, we adapted the Distress Thermometer to create an Anxiety Thermometer.

VR Impressions and Feedback
Participants provided self-reported feedback on the VR simulation at 2 different time points. Feedback measures were developed in accordance with previous research [74]. After trialing the intervention, the participants were provided with a list of statements about their experience using the simulation (eg, “I found the VR intervention was helpful”), which they rated from 0% (completely disagree) to 100% (completely agree). The participants are also asked whether they experienced any motion sickness during the intervention (0 [none] to 3 [severe]), followed by open-ended questions prompting intervention impressions (eg, what they liked or disliked) and whether they found the intervention worthwhile. Finally, the participants were asked about additional elements they wished to be included in the intervention and were offered multiple response options for selection (eg, being wheeled into the OR).

At the 5-day postoperative follow-up, the VR participants provided additional intervention feedback (eg, overall impressions). The participants are then asked if or how they think the intervention impacted their surgery or recovery, whether they disliked anything about it, and if they have any other suggestions for improvements. The participants are prompted to elaborate on their responses to these questions.

Presence
The iGroup Presence Questionnaire [75] assessed the presence associated with the VR intervention at baseline, defined as the sense of being in the virtual environment. This valid and reliable (Cronbach α= .87) self-report measure is comprised of 14 items (eg, “I had a sense of acting in the virtual space, rather than

https://formative.jmir.org/2024/1/e46367
operating something from outside”), which are rated on a 7-point scale (−3 [fully disagree] to +3 [fully agree]). Items are summed to create 3 subscale scores (spatial presence, involvement, and realism), where higher scores indicate increased presence in the virtual environment.

**Sample Characteristics**

Participants self-reported their sociodemographic characteristics and health history at baseline, including age (assessed continuously), sex (female or male), marital status (single, married or common law, divorced or separated, or widowed), highest level of education (high school or less or some college or higher), stage of breast cancer, type of breast cancer surgery (eg, lumpectomy or single or double mastectomy with or without reconstruction), whether they are receiving chemotherapy before surgery, history of prior surgeries, mental health service use since receiving their cancer diagnosis, and history of receiving a mental health diagnosis. Various other self-report measures were administered throughout the study (eg, assessing depression, coping, and quality of life) to determine their feasibility for inclusion in the upcoming RCT (by calculating the proportion of missing data).

**Analytic Strategy**

Descriptive statistics assessed the consent rate, recruitment speed, attrition rate, and sample characteristics. We calculated the participation rate among the intervention and control groups, and we assessed quantitative and qualitative intervention feedback descriptively. We then calculated the proportion of missing data, and descriptive statistics determined which measure of preoperative distress or anxiety was rated most favorably. Finally, we presented participants’ levels of distress and anxiety across the perioperative period (baseline to 30-day follow-up) descriptively in a case series format.

**Results**

**Feasibility Aims**

**Recruitment Capability and Sample Characteristics**

Recruitment was initiated on December 1, 2021. Between initiation and December 1, 2022, a total of 14 prospective participants were identified (n=5, 36% were identified in the final 2 months of recruitment). Of these 14 individuals, 12 (86%) contacted the study coordinator directly, 1 (7%) had their information provided by a health professional, and 1 (7%) expressed interest while attending a preoperative information session (in-person sessions were suspended until November 2022). Of these 14 individuals, 7 (50%) consented and participated, 4 (29%) were ineligible (eg, required to isolate until surgery or already had surgery), and 3 (21%) withdrew after providing verbal consent but before providing written consent (reasons: n=1, 33% too many appointments and unable to focus on anything else; n=1, 33% unwilling to come in person to try the VR; and n=1, 33% overwhelmed with family responsibilities; n=2, 67% had been randomized to the initial third arm before dropping out). The approximate recruitment speed for those who consented was 1 participant every 7 weeks, on average. In total, 57% (4/7) of the participants were assigned to the intervention group. Of those who provided written informed consent, 100% (7/7) completed the study. Because of ongoing recruitment challenges, the study target population was broadened 5 months after the initiation of recruitment to include any patients undergoing cancer surgery, as opposed to patients undergoing breast cancer surgery only. To date, no patients undergoing non–breast cancer surgery have expressed an interest in participating.

The participants were aged 56.43 (SD 10.56) years on average, and all were female. The participants were most commonly married (3/7, 43%), and the majority (5/7, 71%) had some college education or higher. The breast cancer stage of patients was most commonly uncertain or unknown (4/7, 57%). The most common surgical procedure was lumpectomy (4/7, 57%), and 43% (3/7) of the participants were planning to undergo reconstructive surgery. Most participants (6/7, 86%) had not received chemotherapy before their surgery, and most participants (6/7, 86%) had ≥1 prior surgeries. A total of 57% (4/7) of the participants reported receiving a mental health diagnosis in their lifetime (depression, anxiety, or substance use disorder), and 29% (2/7) of the participants indicated that they sought professional mental health support after receiving their cancer diagnosis. Most participants (4/7, 57% to 6/7, 86%) had clinically elevated preoperative distress or anxiety at baseline (Table 1).
Table 1. Sample characteristics for patients undergoing breast cancer surgery participating in the feasibility and pilot study to evaluate preoperative virtual reality (n=7).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>56.43 (10.56)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>7 (100)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Married or common law</td>
<td>3 (43)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Some college or higher</td>
<td>5 (71)</td>
</tr>
<tr>
<td>Stage of breast cancer, n (%)</td>
<td></td>
</tr>
<tr>
<td>Uncertain (0-1, 1-2, or other unknown)</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Stage 1</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Type of surgery, n (%)</td>
<td></td>
</tr>
<tr>
<td>Lumpectomy</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Single mastectomy without reconstruction</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Single mastectomy with immediate reconstruction</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Double mastectomy with immediate reconstruction</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Undergoing reconstruction, n (%)</td>
<td>3 (43)</td>
</tr>
<tr>
<td>Neoadjuvant chemotherapy, n (%)</td>
<td>1 (14)</td>
</tr>
<tr>
<td>History of prior surgery, n (%)</td>
<td>6 (86)</td>
</tr>
<tr>
<td>Sought professional mental health support since cancer diagnosis, n (%)</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Lifetime mental health diagnosis, n (%)</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Clinically significant preoperative distress or anxiety at baseline, n (%)</td>
<td></td>
</tr>
<tr>
<td>PITI&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4 (57)</td>
</tr>
<tr>
<td>APAIS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6 (86)</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>5 (71)</td>
</tr>
<tr>
<td>Intervention group, n (%)</td>
<td>4 (57)</td>
</tr>
</tbody>
</table>

---

<sup>a</sup>PITI: Preoperative Intrusive Thoughts Inventory.

<sup>b</sup>APAIS: Amsterdam Preoperative Anxiety Information Scale.

**Participant Impressions of the Study Design and Intervention**

All participants assigned to the control group completed the entire study. All participants assigned to the intervention group, who provided written informed consent, tested the intervention within 2 weeks before their surgery and completed the study. The participants in the intervention group spent 12 minutes engaged in the simulation, on average, and reported variable levels of presence while trialing the VR simulation (spatial presence: mean 8.75, involvement: mean 0.75, and realism: mean −2.50; refer to Table 2 for maximum ranges); the participants reported having a sense of being physically present in the virtual environment, with only partial attention devoted to the virtual environment, and moderate ratings of realism.
Table 2. Quantitative intervention impressions at baseline for patients undergoing breast cancer surgery assigned to the intervention group for the feasibility and pilot study evaluating preoperative virtual reality (VR; n=4).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>VR duration (min), mean (SD)</td>
<td>11.64 (1.08)</td>
</tr>
<tr>
<td>Presence: spatial presence subscale (maximum range: −15 to 15), mean (SD)</td>
<td>8.75 (3.30)</td>
</tr>
<tr>
<td>Presence: involvement subscale (maximum range: −12 to 12), mean (SD)</td>
<td>0.75 (1.26)</td>
</tr>
<tr>
<td>Presence: realism subscale (maximum range: −15 to 15), mean (SD)</td>
<td>−2.50 (3.70)</td>
</tr>
<tr>
<td>The way information was presented was clear and understandable (0%-100%), mean (SD)</td>
<td>95.00 (10.00)</td>
</tr>
<tr>
<td>I enjoyed my session with the VR program (0%-100%), mean (SD)</td>
<td>96.25 (4.79)</td>
</tr>
<tr>
<td>I could understand and act on the information provided by the program (0%-100%), mean (SD)</td>
<td>93.75 (7.50)</td>
</tr>
<tr>
<td>The program had a very attractive presentation (0%-100%), mean (SD)</td>
<td>95.00 (5.77)</td>
</tr>
<tr>
<td>I had to look for assistance when I used this program (0%-100%), mean (SD)</td>
<td>42.50 (43.49)</td>
</tr>
<tr>
<td>The VR program froze or stopped unexpectedly (0%-100%), mean (SD)</td>
<td>5.00 (10.00)</td>
</tr>
<tr>
<td>I found the VR intervention was helpful (0%-100%), mean (SD)</td>
<td>87.50 (25.00)</td>
</tr>
<tr>
<td>The VR intervention eased my anxiety/concerns about the OR&lt;sup&gt;a&lt;/sup&gt; (0%-100%), mean (SD)</td>
<td>55.00 (47.96)</td>
</tr>
<tr>
<td>The VR intervention eased my anxiety/concerns about the anesthesia (0%-100%), mean (SD)</td>
<td>60.00 (45.46)</td>
</tr>
<tr>
<td>The VR intervention eased my anxiety/concerns about my surgery (0%-100%), mean (SD)</td>
<td>46.25 (38.16)</td>
</tr>
<tr>
<td>The VR intervention worsened my anxiety/concerns about the OR (0%-100%), mean (SD)</td>
<td>37.50 (47.87)</td>
</tr>
<tr>
<td>The VR intervention worsened my anxiety/concerns about the anesthesia (0%-100%), mean (SD)</td>
<td>25.00 (28.87)</td>
</tr>
<tr>
<td>The VR intervention worsened my anxiety/concerns about my surgery (0%-100%), mean (SD)</td>
<td>30.00 (51.96)</td>
</tr>
<tr>
<td>Experienced motion sickness, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Participating in the VR intervention was worthwhile considering time commitment, n (%)</td>
<td>4 (100)</td>
</tr>
</tbody>
</table>

**Other elements you would have liked to be included, n (%)**

- Being wheeled into the OR 1 (25)
- Try on equipment (eg, oxygen mask) while engaged in the simulation 0 (0)
- Learn about the various machines I saw in the OR 1 (25)
- Ask the virtual anesthesiologist or nurse questions about my surgery 1 (25)
- None of the above 2 (50)
- Other (“real time pulse/heart rate”) 1 (25)

<sup>a</sup>OR: operating room.

There were minor technical difficulties for all 4 participants during the simulation (eg, difficulty finding the correct position lying down when prompted by the VR nurse), and the program needed to be restarted midway for 2 of the participants. In general, the participants found that the intervention presented information clearly, was enjoyable, easy to understand, and had an attractive presentation (mean% agreement range: 93.75-96.25, SD range: 4.79-10.00). It was also generally rated as helpful (mean% agreement 87.50, SD 25.00), and all participants considered participating in the VR intervention worthwhile considering the time commitment. The participants gave mixed ratings at baseline regarding the impact of the intervention on anxiety and concerns about the OR, anesthesia, and surgery. Given a list of suggestions for elements to be added to the intervention, a single participant selected each of the following: (1) being wheeled into the OR, (2) learn about the machines I saw in the OR, (3) ask the virtual anesthesiologist or nurse questions about surgery, and (4) other: “real time pulse/heart rate.”

Regarding open-ended feedback, multiple participants commented on the realism of the intervention, the impact of the intervention on expectation formation regarding surgery, and the calming or relaxing effect of the intervention (Textbox 1).
Qualitative intervention impressions for patients undergoing breast cancer surgery assigned to the intervention group for the feasibility and pilot study evaluating preoperative virtual reality (VR; n=4).

**Open-ended feedback (at baseline)**

- What did you like about the VR intervention
  - “It is very realistic”
  - “It was realistic to my past surgical experiences, it was interactive and could play a bit with it”
  - “What to expect”
  - “Just getting the feel of an OR” [operating room]

- What did you dislike about the VR intervention
  - “Nothing”
  - “Scary”
  - “The program calibrated my body position a few times and had to be reset which is why I was more present in the real world than in the VR world”
  - “Seemed like I was waiting for an hour until it told me to lie down”

- If you found it helpful: in what ways was the VR intervention helpful
  - “The sounds and sights and procedures give you a test run- prepares you for the actual day”
  - “Yes”
  - “Was helpful in that it reminded me of all the noises, lights, and people necessary in an OR”
  - “You feel more relaxed for the surgery”

- Explain why it was or was not worthwhile
  - “Gave me information and made me think of my feelings, made me feel better”
  - “I like to help with research and I’m curious about VR and mental health initiatives”
  - “Knowing what to expect”
  - “Think I can relax a bit now when it’s time for me to have my surgery”

**Assessed 5 d after surgery**

- Overall impression of the VR intervention
  - “It was very good, very real to life. I liked it”
  - “Head set didn’t work well”
  - “I had past surgery and it was familiar from memory and with current surgery experience”
  - “I thought it was a good way to help calm some of my fears”

- Elements from the OR that were missing from the VR which would have been helpful to include
  - “No I think they covered everything”
  - “If when they are putting stickers on etc. you would maybe lightly touch the spot”
  - “Not that I remember. I wasn’t paying much attention to what everyone was doing or the equipment”
  - “More condensed room, just focus on the 2 people in your face”

- Images or experiences from the VR intervention that stuck with you following the intervention
  - “No”
  - “The mask at the end”
  - “The nurse moved in on my too quickly and startled me because she was so close so suddenly”
  - “The lights”

- Components of in-hospital experiences on day of surgery that would have been helpful to include in VR
“The waiting in the surgical admitting area. Sitting for a long time in a chair in the gown with IV pick in”
“I didn’t get to wake up in the VR but it may be cool to wake up. You aren’t alone when they wake you up in case that unknown freaks people out”
“The actual experience happened a lot quicker than the virtual experience. Speed up the simulation”

How, if at all, do you think the VR simulation impacted your surgery or recovery
“It makes a person more relaxed in the operating room”
“Once they had me in the surgery room it was very fast”
“If I didn’t have past experience then it would have helped me a lot but I was already familiar”
“I believe it assisted me in that I was able to see the inside of an OR”

Was there anything you disliked about the VR intervention (if yes, please describe)
“No”
“Sadly the program had to be reset a bunch of times because...the orientation was off. It brought me out of it”
“It was way too long just sitting there and waiting for something to happen”

Suggestions regarding how we can improve the VR simulation
“No it was very informative”
“Have the room smaller and things not so far away. People need to be closer to you”

Data Collection Procedures and Outcomes Measures
Across all the time points, only 0.7% of the data were missing. Most participants (5/7, 71%) reported that, of the different measures assessing anxiety and distress, the PITI best captured their experiences.

Pilot Aim
Overview
Table 3 outlines the sample characteristics, and Table 4 outlines the perioperative levels of distress and anxiety of the participants in the control and intervention groups.

Table 3. Participant characteristics.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Group</th>
<th>Age (y)</th>
<th>Current surgery</th>
<th>Prior surgery</th>
<th>Mental health history</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Control</td>
<td>60s</td>
<td>Lumpectomy</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>P2</td>
<td>Control</td>
<td>60s</td>
<td>Lumpectomy</td>
<td>Single mastectomy (&gt; 10 y ago)</td>
<td>Depression</td>
</tr>
<tr>
<td>P3</td>
<td>Control</td>
<td>40s</td>
<td>Double mastectomy with immediate reconstruction</td>
<td>Broken arm and appendectomy</td>
<td>None</td>
</tr>
<tr>
<td>P4</td>
<td>Intervention</td>
<td>60s</td>
<td>Lumpectomy</td>
<td>“Replacements” and “abnormal cell removals”</td>
<td>Mental health leave (no diagnosis)</td>
</tr>
<tr>
<td>P5</td>
<td>Intervention</td>
<td>40s</td>
<td>Single mastectomy with immediate reconstruction</td>
<td>Thyroid surgery &gt; 5 y ago</td>
<td>None</td>
</tr>
<tr>
<td>P6</td>
<td>Intervention</td>
<td>50s</td>
<td>Lumpectomy</td>
<td>Lumpectomy, fibroids removed, hysterectomy, cervix and ovaries removed, and deviated septum repair</td>
<td>Depression and anxiety</td>
</tr>
<tr>
<td>P7</td>
<td>Intervention</td>
<td>50s</td>
<td>Single mastectomy without reconstruction</td>
<td>Arm and cesarean section</td>
<td>Depression and substance use disorder</td>
</tr>
</tbody>
</table>

aAge range.
Table 4. Perioperative distress or anxiety for patients ongoing breast cancer surgery participating in the feasibility and pilot study to evaluate preoperative virtual reality.

<table>
<thead>
<tr>
<th>Participant ID, group, and measure</th>
<th>Perioperative distress or anxiety&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Clinically elevated&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Preoperative</td>
</tr>
<tr>
<td>P1: control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI&lt;sup&gt;d&lt;/sup&gt; total</td>
<td>13.00</td>
<td>7.00</td>
</tr>
<tr>
<td>APAIS&lt;sup&gt;d&lt;/sup&gt; total</td>
<td>10.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>P2: control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>26.00</td>
<td>32.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>13.00</td>
<td>18.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>3.00</td>
<td>6.50</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>3.00</td>
<td>6.50</td>
</tr>
<tr>
<td>P3: control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>42.00</td>
<td>37.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>19.00</td>
<td>17.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>8.00</td>
<td>9.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>9.00</td>
<td>9.00</td>
</tr>
<tr>
<td>P4: intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>10.00</td>
<td>12.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>11.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>6.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>6.00</td>
<td>3.00</td>
</tr>
<tr>
<td>P5: intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>47.00</td>
<td>40.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>21.00</td>
<td>24.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>10.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>10.00</td>
<td>9.00</td>
</tr>
<tr>
<td>P6: intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>11.00</td>
<td>10.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>9.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>3.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>4.00</td>
<td>3.00</td>
</tr>
<tr>
<td>P7: intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>43.00</td>
<td>49.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>21.00</td>
<td>22.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>7.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>7.00</td>
<td>9.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values represent total scores on each measure at each time point.

<sup>b</sup>Values represent the number of times a score is clinically elevated across the total number of measurements.

<sup>c</sup>OR: operating room.

<sup>d</sup>PITI: Preoperative Intrusive Thoughts Inventory.

<sup>e</sup>N/A: not applicable; PITI and Amsterdam Preoperative Anxiety Information Scale are specific to the preoperative period and were not administered.
in the OR or during the postoperative phase.
\(^{f}\) APAIS: Amsterdam Preoperative Anxiety Information Scale.
\(^{g}\) Missing data because of surgery scheduling change.

**Retrospective Reports of Distress or Anxiety**

As shown in Figure 1, among the control group, ratings of distress/anxiety remained stable (P1) or increased (P2 and P3) between baseline (within 2-wk preoperatively) and being in the OR on the day of surgery. Among the intervention group, ratings decreased between baseline (within 2-wk preoperatively; when VR was administered) and being in the OR for 50% (2/4; P4 and P5) of the participants.

**Figure 1.** Retrospective reports of distress/anxiety among the control group and intervention group for patients undergoing breast cancer surgery participating in the feasibility and pilot study to evaluate preoperative virtual reality (VR). Blue guidelines outline the period between when the intervention group trialed the VR and participants’ day of surgery.

**Discussion**

**Overview**

To our knowledge, this is the first study to date to examine the feasibility of, and preliminarily pilot, a novel preoperative VR intervention exposing patients undergoing breast cancer surgery to the OR and preoperative process. Overall, despite some recruitment challenges, the intervention was generally rated favorably and described, on average, as 87.5% (SD 25.00%) helpful by participants. The results of this study will inform modifications made to the VR intervention and the study design of an upcoming RCT evaluating this intervention.

The newly developed VR intervention exposed patients undergoing breast cancer surgery to the OR environment (including machinery, sounds, personnel, and other medical items [eg, surgical tools and mammogram]) and preoperative process (from being seated on the OR table until completion of anesthetic induction). The simulation was developed to mimic the real-life OR and preoperative experience based on a large tertiary care hospital in Winnipeg, Manitoba. Compared with other recently developed preoperative VR interventions [58-64], this was designed to be more immersive through the integration of user embodiment (including visualization of one’s virtual body and real-time manipulation of virtual arms) and is one of the few interventions designed for adult use and the only such intervention developed and tested in Canada.

Although some technical difficulties arose during the intervention (eg, simulation needing to be restarted and slight delay in progression because of imprecise positioning of participant arm or body), likely detracting from immersion, the participants described the intervention as realistic and commented on its impact on feeling more prepared or knowing what to expect for surgery and feeling more relaxed or calm about their upcoming surgery. The participants also rated the intervention favorably in terms of enjoyment, clarity of information, attractiveness, and helpfulness. Although the sample size of this study limits our ability to establish trends regarding the impact of the intervention on distress and anxiety, the participants rated 46% to 60% agreement (SD range 38.16%-47.96%), on average, that the intervention eased their anxiety, and for half of the intervention group participants (2/4, 50%), retrospective ratings of distress/anxiety declined between trialing the intervention and being in the OR. Notably, the participants also rated 25% to 38% (SD range 28.87%-51.96%) agreement, on average, that the intervention worsened their anxiety (immediately postintervention), although they did not indicate this when providing feedback postoperatively. This may suggest the activation of the “fear structure” within the simulation, which is noted as an important component of anxiety-based exposure interventions [76].
Although preliminary data support the feasibility of the VR intervention, we encountered challenges regarding recruitment for the study. This may have been impacted by various factors including changes to surgical scheduling during the COVID-19 pandemic (noted in recent research on patients with cancer [77]), prospective participants' reported feelings of being overwhelmed and stressed by their own health or other commitments, and a strained health care system resulting in reduced resources to support recruitment (including canceling in-person preoperative information sessions for 10 months during the recruitment period, where recruitment was planned to take place). As noted, recruitment began improving over the final 2 months of the recruitment period, wherein 80% (4/5) of the individuals who expressed interest in the study provided consent to participate. Although speculative, this may suggest an impact of the changing centrality of the pandemic on recruitment capability. Interestingly, most participants (6/7, 86%) had a history of prior surgeries, which could have resulted in an increased willingness to participate. It may be worthwhile to consider modifications to our recruitment poster (eg, including the rationale for the intervention) to entice participation from those who have not undergone prior surgery. The study design elements, including data collection, intervention engagement, and participant retention, appear feasible based on the current data.

**Strengths and Limitations**

Despite the strengths of this study, including the novel preoperative VR intervention integrating user embodiment, evaluation of the feasibility of this intervention in a population with elevated estimates of clinically significant distress [8-10], collection of qualitative and quantitative intervention feedback, and inclusion of 2 iterations of postoperative follow-up data (5 and 30 days postoperatively; to be evaluated in an upcoming larger study), this study is not without limitations. First, recruitment challenges limited our sample size for this initial study; however, these challenges provided important information regarding the feasibility of implementing a larger study in the future. Second, there were a few technical difficulties encountered when administering the VR intervention, detracting from user immersion. Finally, although not directly investigated, distress in this population (and assessed using nonspecific measures) is likely to be influenced by many factors in addition to surgery. This particular intervention may not be very beneficial or impactful for those with primarily non-surgery-related distress.

**Implications**

Importantly, these limitations, along with the data collected as part of this study, provide important insights to inform modifications to the intervention and study design before the implementation of a large-scale RCT to evaluate the efficacy of this intervention. Regarding recruitment, we will consider ways to target enhancing the involvement of health care professionals in spreading awareness of the study to potentially eligible patients while continuing to attempt recruitment at the newly reinstated in-person preoperative information sessions. In addition, we will consider including additional information about the intervention (and thus removing participant blinding) as part of the recruitment process. Changes to consider for the VR simulation include modifying requirements for the user’s body positioning to avoid unnecessary interruptions and potentially adding elements that participants noted would have been helpful (eg, the opportunity to learn about OR machines and ask questions to the virtual anesthetist or nurse). On the basis of participant feedback, this intervention has the potential to reduce levels of preoperative distress/anxiety by helping participants form more realistic expectations of the day of surgery before their operation (thus potentially reducing their perception of threat associated with the preoperative experience and enhancing their perceived ability to cope with this stressor). In line with recommendations based on other VR exposure-based interventions [78], having repeated exposure to the simulation may enhance the potential impact on mitigating distress/anxiety. Thus, it may be beneficial to assess the utility of providing participants with a 2D “screen-capture” video recording of their VR trial to watch on their own device multiple times in between trialing the intervention and their surgery. This may be an important avenue for future research evaluating this intervention.

Overall, this study established the initial feasibility of a novel preoperative VR intervention to expose patients undergoing breast cancer surgery to the OR and anesthetic induction process. These results will inform the study design of an upcoming large RCT to further examine this intervention. Participant feedback supports the utility and acceptability of this intervention and will inform future adaptations to the simulation. If demonstrated as efficacious in upcoming research, this intervention has the potential to be adapted across multiple surgery types and implemented on a broad scale to help mitigate preoperative distress.

**Acknowledgments**

The authors acknowledge Gabrielle Logan and Bronwen Grocott for their contributions to this research and Catherine Proulx and Vincent Gagnon Shaigetz for their contributions to the development of the virtual reality software. This study received funding from the Canadian Institutes of Health Research Canada Graduate Scholarship-Doctoral Award, University of Manitoba Graduate Fellowship, University of Manitoba Sheu L Lee Family Scholarship in Oncology Research, Women’s Health Research Foundation of Canada Graduate Scholarship, National Register for Health Service Psychologists Doctoral Award (Sommer), University of Manitoba Start-Up Funds, and Tri-Agency New Frontier’s in Research Grant (El-Gabalawy).

**Data Availability**

Data access is restricted to protect the confidentiality of participants and in light of ongoing research to expand the data set.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Virtual reality simulation screen captures.
[PNG File, 1590 KB - formative_v81le46367_app1.png]

References


**Abbreviations**

APAIS: Amsterdam Preoperative Anxiety Information Scale  
NCCN: National Comprehensive Cancer Network  
OR: operating room  
PITI: Preoperative Intrusive Thoughts Inventory  
RCT: randomized controlled trial  
VR: virtual reality

---

**Edited by A Mavragani; submitted 08.02.23; peer-reviewed by W Trapp; comments to author 27.10.23; revised version received 16.11.23; accepted 04.12.23; published 17.01.24.**

Please cite as:


Preoperative Virtual Reality to Expose Patients With Breast Cancer to the Operating Room Environment: Feasibility and Pilot Case Series Study

JMIR Form Res 2024;8:e46367

URL: https://formative.jmir.org/2024/1/e46367
doi:10.2196/46367
PMID:38231570
Educational Video Intervention to Improve Health Misinformation Identification on WhatsApp Among Saudi Arabian Population: Pre-Post Intervention Study

Ebtihal Alsaad1,2*, DDS, MSc; Sharifah AlDossary1,2*, MSc, PhD
1Department of Health Informatics, College of Public Health and Health Informatics, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
2King Abdullah International Medical Research Center, Riyadh, Saudi Arabia
*all authors contributed equally

Corresponding Author:
Ebtihal Alsaad, DDS, MSc
Department of Health Informatics
College of Public Health and Health Informatics
King Saud bin Abdulaziz University for Health Sciences
Al Haras Al Watani St
Prince Mutib Ibn Abdullah Ibn Abdulaziz Rd, Ar Rimayah, Riyadh
Riyadh, PO Box 3660/ 11481
Saudi Arabia
Phone: 966 114299999
Email: ebtihalalsaad@gmail.com

Abstract

Background: Health misinformation can adversely affect individuals’ quality of life and increase the risk of mortality. People often fail to assess the content of messages before sharing them on the internet, increasing the spread of misinformation. The problem is exacerbated by the growing variety of digital information environments, especially social media, which presents as an effective platform for spreading misinformation due to its rapid information-sharing capabilities. Educational interventions have been developed to help consumers verify the validity of digital health information. However, tools designed to detect health misinformation on social media content have not been validated. Given the increased use of social media platforms, particularly WhatsApp, it is crucial to develop tools to help consumers assess the credibility of messages and detect misinformation.

Objective: The main objective of this study is to develop and assess an educational tool aimed at educating consumers about detecting health misinformation on WhatsApp. The secondary objective is to assess the association between demographic factors and knowledge levels.

Methods: The study used a single-arm, pre-post intervention design to evaluate the effectiveness of an educational video in improving participants’ ability to detect health-related misinformation in WhatsApp messages. In the first phase, an educational video intervention was developed and validated. In the second phase, participants were invited to complete a web-based survey that consisted of pre-evaluation questions, followed by the educational video intervention. Subsequently, they were asked to answer the same questions as the postevaluation questions.

Results: The web-based survey received 485 responses. The completion rate was 99.6% (n=483). Statistically significant associations existed between knowledge level and age, gender, employment, and region of residence (P<.05). The video intervention did elicit a statistically significant change in the participants’ abilities to identify misinformation in WhatsApp messages (z=–6.887; P<.001). Viewing the video was associated with increased knowledge about the following concepts: checking the “forwarded” label (P<.001), looking for spelling and grammatical errors (P<.001), analyzing the facts (P=.03), checking links (P=.002, P=.001), and assessing the photos and videos (P<.001). There was a statistically significant difference in knowledge level before and after the intervention (P<.001).

Conclusions: This study developed and evaluated the effectiveness of an educational video intervention to improve health misinformation identification on WhatsApp among the Saudi Arabian population. The results indicate that educational videos can be valuable tools for improving participants’ abilities to identify misinformation. The outcomes of this research can contribute to our understanding of what constitutes an effective tool for enhancing health misinformation awareness. Such interventions
may be particularly useful in combating misinformation among Arabic-speaking populations on WhatsApp, which may ultimately improve eHealth literacy. Limiting the prevalence and impact of misinformation allows people to make better-informed health decisions.

*JMIR Form Res 2024;8:e50211*  doi:10.2196/50211

**KEYWORDS**

misinformation; education; WhatsApp; intervention; pre-postintervention design; health literacy; educational; video; videos; consumer; consumers; patient education; survey; surveys; web-based information; health information; reliability; accuracy; reliable; social media

**Introduction**

**Background**

Researching health problems and learning about health via the internet has become a prevalent practice [1]. The level of credibility of this health-related information and the way it is used by patients, caregivers, and other health consumers have garnered the attention of health care providers and authorities [2]. There are many inaccurate sources of information on the internet, and this can lead to users becoming misinformed. According to Chou et al [3], health misinformation is defined as any health-related factual claim that is false according to recent scientific evidence. Misinformation about health can adversely affect quality of life and even increase one’s mortality risk [1].

When the COVID-19 pandemic first started, the amount of information related to this new global pandemic increased at an unprecedented rate. The volume of information, as well as the rate at which new information appeared, increased rapidly [4]. Global pandemics such as COVID-19 are likely to lead to the increased spread of misinformation as people explore massive amounts of information about the disease and its health implications. The term “infodemic” is used to describe the current media environment, which is characterized by an overflow of both true and false information. During the pandemic, individuals generally look for accurate, unbiased information, but these sources may be hidden among misinformation spread through the infodemic [5].

Due to its capacity to rapidly disseminate information, social media can serve as a platform for the propagation of misinformation. The abundance of available information can lead to the predominance of misinformation, thus negatively affecting cognitive, logical, and decision-making capacities. WhatsApp, Twitter, and Facebook are the most commonly used social media platforms for spreading false information. Since the beginning of the COVID-19 pandemic, internet use has expanded worldwide, which has resulted in the proliferation of incorrect information via social media [6].

Saudi Arabia, with a population of over 35 million, is the second largest Arab country [7]. The COVID-19 pandemic has had a significant impact on internet usage in Saudi Arabia, with a reported increase of 91.2% (n=28,775,889) in 2020. This represents a rise of 2.6 percentage points compared to the previous year [8]. Alshareef and Alotiby [6] used a web-based survey to investigate the most widely used social media platforms in Saudi Arabia, the proportions of Saudi Arabians who used these platforms to share information, and these users’ perceptions of the medical information shared on these platforms. According to their survey results, WhatsApp was used by 52.4% (n=144) of health care workers and 51.3% (n=500) of non–health care workers to circulate information. The findings of their study concluded that WhatsApp is the most commonly used social network among Saudi Arabians. COVID-19–related information is, therefore, more likely to be shared on this application [6].

Another study by Alasmari et al [9] found that social media platforms, with their capacity to quickly disseminate information, comprised the primary source of falsehoods spread in the community. Based on an examination of the social media platforms, the study revealed that WhatsApp users accounted for approximately 46% (n=41) of rumor sources on the internet in Saudi Arabia.

Additionally, research by Tan et al [10] examined daily WhatsApp use for receiving, forwarding, or discussing COVID-19–related content a in 1-week period. The results indicate that almost every respondent participated in conversations about COVID-19. However, users were more likely to share or receive forwarded messages than to engage in active, original conversations about COVID-19. A high volume of forwarded messages was observed; this is concerning because the developers of WhatsApp have linked forwarded messages with misinformation.

People rarely assess the content of messages before sharing them on social media platforms, and they frequently fail to verify whether the messages are accurate. Educating consumers about identifying misinformation and dealing with the infodemic is essential. The false information epidemic compromises public health as misinformation spreads throughout social media. It is critical to increase awareness about the nature of social media and how to use it effectively. Personal responsibility is the first and most crucial step in safeguarding our community from the harmful phenomena of misinformation [11].

To effectively access health-related information on the internet, consumers must be able to assess the quality of the information that they find. This is a crucial aspect of eHealth literacy. It remains difficult for digital health consumers to determine the quality of the information placed in front of them. The problem becomes more complex as the digital information environment becomes more complicated and heterogeneous, especially with the rise of social media, where anyone can spread information about health and where low-quality and misleading information...
spreads rapidly. Interventions are urgently needed to address this public health problem [12].

Several interventions have been developed to assist consumers in verifying the validity of digital health information [12]. A systematic review by Cusack et al [13] examined studies on educational interventions that aimed to improve knowledge of essential concepts, enabling health interventions to be evaluated for their impacts. According to the study, educational interventions, at least in the short term, can increase people’s knowledge and skills in evaluating health claims.

For the detection of health misinformation, interventions have been established based on instruments that allow anyone, including those with no prior medical background, to differentiate fact from fiction. However, these tools were designed for lengthy texts (such as text found on websites) and have not yet been validated for detecting health misinformation in social media content [14]. The majority of the tools developed were used to assess the quality of websites that provided health information. Considering the increased use of social media platforms—primarily WhatsApp—in Saudi Arabia for sharing health information, it is essential to develop tools that help consumers assess the credibility of messages and detect misinformation.

The first World Health Organization Infodemiology conference for managing the infodemic suggested evidence-based analysis and interventions to reduce the harmful effects of health misinformation during acute health events. Among the recommendations was the development of interventions that address factors that impact trust and resilience to misinformation at the individual, community, cultural, and societal levels [15].

**Theoretical Background**

A low level of health literacy has been recognized as one of the factors contributing to the infodemic. Other contributing factors include the widespread use of social media, quick publication processes, and preprint services. Rumor-spreading behavior also plays a role in the infodemic as do anxiety, distress, and fear [16]. According to a systematic review by Diviani et al [17], health literacy is essential when evaluating digital health information. Individuals’ abilities to find, evaluate, and use health information empowers them to actively deal with the misinformation they encounter on social media. In order to prevent people from automatically accepting health rumors as facts, health literacy must be improved [18].

In this research, the educational intervention concept was guided by the inoculation theory and the message interpretation process theory. According to the inoculation theory, previous experience helps individuals combat future attacks [19]. For example, literacy interventions may help audiences resist harmful media messages by providing them with the knowledge and skills necessary to reject them [20]. Based on the message interpretation process theory, exposure to message interventions influences subsequent decision-making when dealing with harmful information [21]. Both theories identify the role of an intervention or prior messages in influencing the cause of action [22].

The choice of the intervention media was guided by the cognitive theory of multimedia learning, which is built from the cognitive load theory and states that working memory contains 2 channels for acquiring and processing information: an auditory or verbal channel and a visual or pictorial channel. Although each channel has a limited capacity, the 2 can be used together to integrate new information more easily. Working memory can function at its best when both channels are used. However, 1 or both channels can become overloaded by a heavy cognitive load. It is possible to improve learning through the use of multimedia learning materials that manage the cognitive load across both channels. Furthermore, the cognitive theory of multimedia learning states that any learning should involve cognitive processing to be meaningful. Cognitive processing requires a learner to pay attention to the material presented, organize it mentally, and integrate it into prior knowledge [23].

**Objectives**

The main objective of this study is to develop and assess an educational tool aimed at educating consumers about detecting health misinformation on WhatsApp. The secondary objective is to assess the association between demographic factors and knowledge levels.

**Methods**

**Study Design**

The study used a single-arm, pre-post intervention design to evaluate the effectiveness of educational video in improving participants’ ability to detect health-related misinformation in WhatsApp messages. The study’s first phase was developing and validating an educational video intervention. In the second phase, participants were invited to complete a web-based survey that contained pre-evaluation questions, the intervention, and post-evaluation questions.

**Participants**

A web-based survey was distributed among the general Saudi population from November 24 to December 25, 2022. The survey was disseminated through social media networks (WhatsApp, Instagram, Twitter, Facebook, and Telegram), and the data were collected using Google Forms.

It has been estimated that 82% (n=29.50 million) of Saudis use social networks daily, with varying usage rates among different platforms. Among these platforms, WhatsApp is the most widely used social network with 87.4% (n=30.67 million) of internet users in Saudi Arabia, followed by Instagram (n=27.40 million, 78.1%), Twitter (n=25.23 million, 71.9%), Facebook (n=22.25 million, 63.4%), and Telegram (n=20.88 million, 59.5%) [24]. In order to target a wide range of the population, the web-based survey was disseminated across all of these social media platforms.

The study population consisted of social media users in the general population of the Kingdom of Saudi Arabia. The inclusion criteria were (1) having the ability to complete an anonymous survey questionnaire on the internet, (2) being at least 18 years of age, and (3) understanding Arabic.
The sample size was calculated using the Raosoft sample size calculator, based on the total population of Saudi Arabia (n=35,013,414), with a 95% CI [7,25]. This calculation yielded a minimum sample size of 385 using absolute error or precision of 0.05. This sample size is sufficient to detect a difference between pre and postscore with an effect size of 0.15 (small effect size) using a power of 80% and \( \alpha \) of .05.

The study population was targeted using a convenience sampling technique with no predetermined sampling frame. Convenience sampling is a nonprobability method in which individuals are sampled simply because they are “convenient” data sources [26].

Several specific methods were used for the recruitment process. First, as there are social accounts run by the public to share news and announcements related to each region in Saudi Arabia, the survey was distributed to these public social networking groups on various social media platforms. Second, the researchers approached social media influencers on different platforms to spread the survey to more participants. Third, the researchers asked all their social media contacts to consider completing the survey and sharing it with their contacts on social networks.

The web-based survey had 4 sections. In the first section, the participants were asked to provide demographic information. In the second section, they were given a set of pretest questions asking them to identify whether a WhatsApp message contained correct or false information. In the third section, the participants were shown an educational video. After finishing the video, the participants moved to the last section, which contained the same set of questions as the pretest.

There were no records of participant identity, and confidentiality was ensured. Upon completion of the survey, a message of thanks appeared. No incentives were offered for completing the survey.

**Intervention (Educational Video)**

**Educational Video Design**

This study used a short video intervention. The content of the educational video was developed based on three sources: (1) the recommendations on WhatsApp’s official website regarding how to prevent the spread of misinformation; (2) the World Health Organization’s (WHO’s) advice on how to navigate the infodemic and identify misinformation; and (3) the CRAAP test, a tool for evaluating the quality of a social media source by assessing its currency, relevance, authority, accuracy, and purpose [27-29]. The educational tool introduced 6 concepts that could be used to assess and identify misinformation in WhatsApp messages. These concepts included checking the “forwarded” label, looking for spelling and grammatical errors, reading beyond the headline, analyzing the facts, checking links, and assessing photos and videos (Table 1).

<table>
<thead>
<tr>
<th>Concept</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check the “forwarded” label</td>
<td>WhatsApp [27]</td>
</tr>
<tr>
<td>Look for spelling and grammatical errors</td>
<td>WHO(^a), WhatsApp, and CRAAP Test [27-29]</td>
</tr>
<tr>
<td>Read beyond the headline</td>
<td>WHO [28]</td>
</tr>
<tr>
<td>Analyze the facts</td>
<td>WHO, WhatsApp, and CRAAP Test [27-29]</td>
</tr>
<tr>
<td>Check links</td>
<td>CRAAP Test [29]</td>
</tr>
<tr>
<td>Assess the photos and videos</td>
<td>WHO [28]</td>
</tr>
</tbody>
</table>

\(^a\)WHO: World Health Organization.

The design of the educational video was based on literature guidelines for the design of health education messages [30]. Following Hugo recommendation, the construction of the educational material included the consideration of communication principles and sociocultural factors, including the literacy levels and language preferences of the audience, to design appropriate messages. When designing the audiovisual content, simplicity (text and visual composition) and the audience’s emotional involvement were considered [30]. The educational video was developed in classical Arabic to make it accessible to a wider audience. Multimedia Appendix 1 shows the developed educational video, while screenshots of the educational video are shown in Figure 1.
Educational Video Validation Process

Yusoff [31] recommended that when validating the content of a tool, a minimum of 6 (but no more than 10) experts should be involved in the assessment process. The validation of the educational material included assessments by 7 experts (Multimedia Appendix 2). For this study, it was essential that the specialists were well-versed in the Arabic language (spoken and written).

To identify the broad range of expertise needed, roles were categorized into different fields. Four roles were identified: (1) health informatics experts, (2) health education specialists, (3) infodemic managers, and (4) public health experts.

Each category was populated with a representative through purposeful sampling. A panel of experts was formed with at least 1 representative from each role; email was used to contact the representatives and provide information about the study. The validation forms and the educational video were delivered to the specialists via email.

The validation form was created by combining 2 tools from which items relevant to the study were selected. Questions Q2 through Q11 were adopted from the Educational Content Validation Instrument in Health developed by Leite et al [32], and questions Q1 and Q12 through Q17 were adopted from an audiovisual content evaluation instrument constructed by Rosa et al [33].

The final evaluation form contained 17 questions covering three areas: objectives, structure and presentation, and audiovisuals (Multimedia Appendix 3). The “objectives” section focused on purposes and goals, whereas the “structure and presentation” section emphasized organization, structure, strategy, sufficiency, and consistency. As for the “audiovisual” area, the emphasis was on the technological aspect. A score of 0 indicated disagreement, 1 indicated partial agreement, and 2 indicated strong agreement with the value of the items [34].

Educational Video Validation Result

A content validity index was used to analyze the results. Content validity indexes can be computed in 2 ways. One type of validity is item-level content validity indexes (I-CVIs), which consider the content validity of individual items. The other type is scale-level content validity indexes, which involve a scale’s overall content validity [35]. For the scale-level content validity,
calculations were conducted using the scale-level content validity index averaging method (S-CVI/Ave) as recommended by Polit and Beck [36].

The calculations were carried out manually. The items ranked “disagree” were scored as 0, whereas the items ranked “partially agree” and “strongly agree” were scored as 1 [34].

To obtain excellent content validity, the content of educational videos must have items with I-CVIs above 0.78 for (6 to 10 experts) and an S-CVI/Ave of 0.90 or higher [36]. When the I-CVI is below 0.78 and the S-CVI/Ave is below 0.90, the content modification should be considered for that particular educational video area.

As shown in Table 2, all items had I-CVIs greater than 0.78 (78%), indicating agreement between the experts’ answers. In terms of scale evaluation, all the 3 areas (objectives, structure and presentation, and audiovisual) had S-CVI/Aves above 0.90 (90%). In the objectives area, item 5 (“stimulates interest in the theme”) had the lowest specialist agreement score (6 out of 7 or an I-CVI of 0.86). The overall S-CVI/Ave for the objectives area was 0.97.

Table 2. A content validity index calculation using 7 expert ratings to validate the video’s content.

<table>
<thead>
<tr>
<th>Item</th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
<th>Expert 6</th>
<th>Expert 7</th>
<th>I-CVIa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.81</td>
</tr>
<tr>
<td>Q3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td>Q7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td>Q8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td>Q9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td>Q10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td>Q14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Q17</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.86</td>
</tr>
<tr>
<td>Proportion relevance</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Proportion relevance</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
</tr>
</tbody>
</table>

aI-CVI: item-level content validity index.
bThe S-CVI/Ave and average proportion of items judged as relevance across the 7 experts for objectives is 0.97.
cThe S-CVI/Ave and average proportion of items judged as relevance across the 7 experts for structure and presentation is 1.
dThe S-CVI/Ave and average proportion of items judged as relevance across the 7 experts for audiovisual is 0.94.

The structure and presentation area had the highest level of agreement (100% S-CVI/Ave). In the audiovisual area, 2 items (“the illustrations are expressive and sufficient” and “the characters/images are appropriate for the target audience”) had scores of 6 out of 7 or I-CVIs of 0.86. The overall S-CVI/Ave for the audiovisual area was 0.94.
Data Collection Tool

Development of the Survey

The study aimed to develop and assess an interventional tool to educate the Saudi population on how to identify health misinformation in WhatsApp messages. The developed Arabic survey contained 4 sections: demographic data, pretest questions, educational intervention, and posttest questions.

The first section (sociodemographic information) included 7 questions about each participant’s background information, including gender, age, educational level, employment status, region of residence, city of residence, and nationality. The second section included 8 questions that assessed each participant’s ability to identify misinformation based on the WhatsApp messages evaluation concepts mentioned in Table 1. Multimedia Appendix 4 shows the complete list of questions that were assigned to the concepts. The third section included the educational tool, which discussed 6 concepts that could be used to identify misinformation in WhatsApp messages. The fourth section included the same 8 questions as the second domain to measure the effectiveness of the educational tool.

The assessment questions were based on real examples representing the different domains of WhatsApp messages. The 8 pre and posttest questions included 5 messages with misinformation (based on messages circulated during the pandemic) and 3 with correct information (obtained from the Saudi Ministry of Health’s official website) [37]. The evaluation concepts, selection of the messages, and their relevancy were assessed by 2 authors (EA and SA). EA collected the messages and placed them under each domain, and SA assessed the relevancy; any uncertainty was resolved by consensus. Finally, a summary of the study’s goal was included in the survey, as was a statement assuring the respondents’ confidentiality.

The highest possible score for the survey questions was 8 points (correct answers were scored as 1 point; incorrect answers and responses of “I do not know” were scored as 0 points). We used a modified Bloom cutoff value of 75% to categorize participants’ knowledge. As a result, we considered participants with scores ≥75% to have high knowledge. Participants with scores below 75% were considered to have low knowledge. This cutoff value is based on previous publications [38,39].

Pilot Survey

The web-based survey was pilot tested, and a total of 31 participants responded. During the first piloting stage, the survey was sent to 15 participants, 3 of whom commented that the instructions needed clarification. Subsequently, an instruction section containing a description of the other survey sections was added at the beginning, and the survey was distributed again to 16 participants. In the second stage, no further comments were received; all participants indicated that the survey was clear. The internal consistency of the final survey was measured using Cronbach α. The scale had a Cronbach α of .847, demonstrating good internal reliability. The final version of the survey can be found in Multimedia Appendix 5.

Ethical Considerations

The study was approved by King Abdullah International Medical Research Center (reference RYD-22-419812-107000). The survey included a summary of the study’s purpose and a statement that, by completing the survey, the respondents agreed to participate in this research. The confidentiality of the study participants was ensured by not collecting identifiable data, encrypting files, and requiring a password to open or modify files.

Statistical Analysis

The demographic characteristics of the participants were reported using descriptive statistics, such as frequencies and percentages. Categorical variables were analyzed using chi-square tests to determine the associations between the demographic variables and the knowledge levels of the participants. A modified Bloom cutoff was used to categorize the knowledge levels. The normality of the variables was analyzed using the Kolmogorov-Smirnov test and the Shapiro-Wilk test. The median scores from before and after the educational video were compared using the Wilcoxon signed-rank test. The McNemar test for categorical data was used to compare the answers and differences in knowledge levels before and after the educational video intervention. The analyses were performed using SPSS Statistics (version 29.0; IBM). A P-value of .05 or less was considered statistically significant.

Results

Characteristics of Study Participants

The web-based survey received 485 responses, and 2 did not agree to participate, giving a 99.6% (n=483) completion rate. In total, 483 responses were analyzed. The socioeconomic characteristics of the participants are shown in Table 3. Most of the study respondents (n=457, 94.6%) were Saudis. More than half of the participants (n=300, 62.1%) were female, and more than half were in the age range of either 18-24 or 25-34 years (n=130, 26.9% and n=173, 35.8%, respectively). More than half of the sample (n=275, 56.9%) had bachelor degrees. With regard to the employment status, 45.5% (n=220) of the respondents were employed. The highest number of participants came from the eastern region (n=181, 37.5%), followed by the central (n=132, 27.3%) and western (n=83, 17.2%) regions.
Table 3. Participant demographics of the full sample (N=483) who participated in a web-based survey about the ability to identify health misinformation on WhatsApp messages between November and December 2022.

<table>
<thead>
<tr>
<th>Nationality</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saudi</td>
<td>457 (94.6)</td>
</tr>
<tr>
<td>Non-Saudi</td>
<td>26 (5.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-24</td>
<td>130 (26.9)</td>
</tr>
<tr>
<td>25-34</td>
<td>173 (35.8)</td>
</tr>
<tr>
<td>35-44</td>
<td>99 (20.5)</td>
</tr>
<tr>
<td>45-54</td>
<td>47 (9.7)</td>
</tr>
<tr>
<td>55 or older</td>
<td>34 (7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>183 (37.9)</td>
</tr>
<tr>
<td>Female</td>
<td>300 (62.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Educational level</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school and below</td>
<td>94 (19.4)</td>
</tr>
<tr>
<td>Diploma</td>
<td>39 (8.1)</td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>275 (56.9)</td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>75 (15.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employment status</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>114 (23.6)</td>
</tr>
<tr>
<td>Employed</td>
<td>220 (45.5)</td>
</tr>
<tr>
<td>Not employed</td>
<td>122 (25.3)</td>
</tr>
<tr>
<td>Retired</td>
<td>27 (5.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region of residence</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>181 (37.5)</td>
</tr>
<tr>
<td>Central</td>
<td>132 (27.3)</td>
</tr>
<tr>
<td>West</td>
<td>83 (17.2)</td>
</tr>
<tr>
<td>North</td>
<td>48 (9.9)</td>
</tr>
<tr>
<td>South</td>
<td>39 (8.1)</td>
</tr>
</tbody>
</table>

Despite some variations, the sample matched the age and sex distribution of the Saudi Arabian population. Similarities in regional distribution between our sample and the populations of certain regions were also evident in the sample, with 27.3% (n=132) from the central region and 9.9% (n=48) from the northern region aligning with the national census in Saudi Arabia (n=5,365,700, 28.5% and n=1,877,108, 9.9%, respectively). In the eastern region, our sample showed a higher representation at 37.5% (n=181) compared with 15.7% (n=2,949,854) reported in the census data [40]. Our sample also showed a difference in educational level, with 56.9% (n=275) of participants holding bachelor degrees, in contrast to the national statistic of 23% (n=2,812,477) [41].

Association Between Knowledge Level and Demographic Variables

The highest possible score for the survey questions was 8 points (correct answers were scored as 1 point; incorrect answers and responses of “I do not know” were scored as 0 points). We used a modified Bloom cutoff value of 75% (6 points) to categorize the participants’ knowledge. A knowledge score of ≥6 indicated a high level of knowledge, while a score of <6 indicated a low level of knowledge.

The associations between knowledge about identifying misinformation in WhatsApp messages and demographic variables were assessed using chi-square tests (Table 4). There were statistically significant associations between knowledge level and age, sex, employment, and region of residence (P<.05).
Table 4. Chi-square tests to examine the association between knowledge about identifying misinformation in WhatsApp messages before the intervention and demographic variables (N=483).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Low knowledge, n (%)</th>
<th>High knowledge, n (%)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nationality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saudi</td>
<td>136 (29.8)</td>
<td>321 (70.2)</td>
<td>0.529 (1)</td>
<td>.47</td>
</tr>
<tr>
<td>Non-Saudi</td>
<td>6 (23.1)</td>
<td>20 (76.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>22 (16.9)</td>
<td>108 (83.1)</td>
<td>43.030 (4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>25-34</td>
<td>37 (21.4)</td>
<td>136 (78.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>41 (41.4)</td>
<td>58 (58.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>22 (46.8)</td>
<td>25 (53.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 and above</td>
<td>20 (58.8)</td>
<td>14 (41.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>64 (35)</td>
<td>119 (65)</td>
<td>4.409 (1)</td>
<td>.04</td>
</tr>
<tr>
<td>Female</td>
<td>78 (26)</td>
<td>222 (74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school and below</td>
<td>35 (37.2)</td>
<td>59 (62.8)</td>
<td>7.289 (3)</td>
<td>.06</td>
</tr>
<tr>
<td>Diploma</td>
<td>16 (41)</td>
<td>23 (59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>71 (25.8)</td>
<td>204 (74.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>20 (26.7)</td>
<td>55 (73.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>19 (16.7)</td>
<td>95 (83.3)</td>
<td>21.378 (3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Employed</td>
<td>71 (32.3)</td>
<td>149 (67.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>36 (29.5)</td>
<td>86 (70.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>16 (59.3)</td>
<td>11 (40.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region of residence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>61 (33.7)</td>
<td>120 (66.3)</td>
<td>35.330 (4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Central</td>
<td>25 (18.9)</td>
<td>107 (81.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>13 (15.7)</td>
<td>70 (84.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>27 (56.3)</td>
<td>21 (43.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>16 (41)</td>
<td>23 (59)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effectiveness of the Intervention
The Kolmogorov-Smirnov test and the Shapiro-Wilk test indicated that the knowledge scores before and after the educational video were not normally distributed (P<.001). Since the distribution was not symmetric, it is negatively skewed, and there are a few outliers on the left side contributing to the skewness; nonparametric tests were used to assess statistical significance.

The median scores were assessed both before (median = 7, IQR = 5-8) and after (median = 8, IQR = 6-8) the video intervention, using the Wilcoxon signed-rank test. This comparison revealed that the video intervention did elicit a statistically significant change in the participants’ abilities to identify misinformation in WhatsApp messages (z=-6.887; P<.001).

Knowledge Questions
The proportions of correct answers per individual test question before and after the video intervention were compared using the McNemar test. Significant differences in the participants’ pre- and postintervention knowledge about identifying misinformation were found for specific questions (Table 5). Viewing the video was associated with increased knowledge about the following concepts: checking the “forwarded” label (P<.001), looking for spelling and grammatical errors (P<.001), analyzing the facts (P=.03), checking links (P=.002, P=.001), and assessing the photos and videos (P<.001).
Table 5. Participants’ answers before and after viewing the educational video (N=483).

<table>
<thead>
<tr>
<th>Domains and answer</th>
<th>Preintervention, n (%)</th>
<th>Postintervention, n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check the “Forwarded” label</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>387 (80.1)</td>
<td>423 (87.6)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>96 (19.9)</td>
<td>60 (12.4)</td>
<td></td>
</tr>
<tr>
<td>Look for spelling and grammatical errors</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>357 (73.9)</td>
<td>402 (83.2)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>126 (26.1)</td>
<td>81 (16.8)</td>
<td></td>
</tr>
<tr>
<td>Read beyond the headline</td>
<td></td>
<td></td>
<td>.061</td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>416 (86.1)</td>
<td>430 (89.0)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>67 (13.9)</td>
<td>53 (11)</td>
<td></td>
</tr>
<tr>
<td>Analyze the facts</td>
<td></td>
<td></td>
<td>.028</td>
</tr>
<tr>
<td>Q5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>418 (86.5)</td>
<td>435 (90.1)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>65 (13.5)</td>
<td>48 (9.9)</td>
<td></td>
</tr>
<tr>
<td>Check links</td>
<td></td>
<td></td>
<td>.54</td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>387 (80.1)</td>
<td>402 (83.2)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>96 (19.9)</td>
<td>81 (16.8)</td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>Correct</td>
<td>326 (67.5)</td>
<td>354 (73.3)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>157 (32.5)</td>
<td>129 (26.7)</td>
<td></td>
</tr>
<tr>
<td>Q8</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Correct</td>
<td>377 (78.1)</td>
<td>403 (83.4)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>106 (21.9)</td>
<td>80 (16.6)</td>
<td></td>
</tr>
<tr>
<td>Assess the photos and videos</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>333 (68.9)</td>
<td>368 (76.2)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>150 (31.1)</td>
<td>115 (23.8)</td>
<td></td>
</tr>
</tbody>
</table>

Improvement in Knowledge Level

The health misinformation education intervention involved 483 participants. Pretest results showed that 70.6% (n=341) of participants had high knowledge (score ≥6), while 29.4% (n=142) had low knowledge (score>6). After the posttest, 10.6% (n=51) of the sample had improved to high knowledge and 3.3% (n=16) had lower scores, indicating 77.8% (n=376) had a score of 6 or above. McNemar test determined that there was a statistically significant difference in knowledge level before and after the intervention (P<.001; Table 6).

Table 6. McNamar test to compare knowledge level before and after the intervention (n=483; P<.001).

<table>
<thead>
<tr>
<th>Before, n (%)</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>91 (18.9)</td>
<td>51 (10.5)</td>
</tr>
<tr>
<td>High</td>
<td>16 (3.3)</td>
<td>325 (67.3)</td>
</tr>
</tbody>
</table>
Discussion

Overview

This study aimed to design and evaluate the effectiveness of an educational video to improve the abilities of participants in Saudi Arabia to identify health misinformation within the WhatsApp app. The study used a single-arm, pre-post intervention design and was conducted on the web. The effectiveness of the intervention was assessed. Furthermore, the participants’ knowledge levels about identifying misinformation were assessed before and after the intervention as were the associations between the participants’ characteristics and their knowledge.

The proliferation of health-related misinformation on social media has raised public health concerns in many countries [42]. Chen et al [43] found that people with limited health literacy were more likely to trust health-related information found on social media and blogs. Thus, improving the public’s ability to evaluate health information may be necessary. In Saudi Arabia, WhatsApp is the most popular social network, which is used by 87.4% (n=30.67 million) of internet users [6,24]. This platform has been identified as one on which misinformation may be easily spread [44]; therefore, it was the platform on which this study focused.

Principal Results

The participants’ ability to identify misinformation in WhatsApp messages significantly improved following the educational intervention (P<.001). This result supports the finding of a systematic review by Cusack et al [13], which showed that, at least in the short term, educational interventions could improve knowledge and skills. This finding is also in line with the message interpretation process theory and the inoculation theory, in which interventions and prior messages are identified as factors that effectively protect against the harm caused by misinformation [19,21].

Additionally, the findings of this study suggest that literacy interventions combined with visual multimedia may improve misinformation detection. Apuke et al [45] found that participants who received visual multimedia education had better knowledge of literacy concepts than those who were educated without visual multimedia. Thus, as previously mentioned in this study, multimedia enhances memory, as stated in the cognitive theory of multimedia learning. The receiver’s exposure to the various message components makes it easier to integrate new information [23].

In this study, the items that assessed the following concepts (checking the “forwarded” label, looking for spelling and grammatical errors, analyzing the facts, assessing the photos and videos, and checking links) were significantly associated with improvements in the participants’ knowledge (P<.05 for all). However, the item related to the concept of reading beyond the headline was not significantly associated with improvement (P>.05).

It was noticed that WhatsApp messages with misinformation are characterized by requests to forward the message to many people. Further, many forwarded messages include fake information sources such as links or names and are vague about timelines, authors, and origins. Consequently, most forwarded messages are found to contain misinformation [46].

By recognizing the most prominent characteristics of health misinformation, users can improve their abilities to identify it on social media. Some studies have proposed criteria such as accuracy, authority, objectivity, and currency, but it is challenging for laypeople to evaluate these indicators. The role of such criteria is quite limited for general users, who, by definition, are not professionals [47]. Li et al [47] proposed a feature scheme and incorporated semantic, grammatical, and peripheral features of messages in evaluating their credibility. Their developed feature scheme allowed users to improve both their abilities to recognize health misinformation and their levels of digital health literacy.

This study found a significant association between knowledge level and age, sex, employment, and region of residence (P<.05 for all). Bapaye and Bapaye [48] noted that those engaged in elementary occupations and those older than 65 years of age were most likely to get false information from WhatsApp in a developing country. Workers in the health care industry were not immune from the impact of false information and were found to be just as susceptible as those in other professions.

The pretest results showed that 70.6% (n=341) of participants had good levels of knowledge about identifying misinformation on WhatsApp. There may have been factors contributing to an increase in knowledge, such as public awareness campaigns and government efforts. During the pandemic, the Saudi Arabian Ministry of Health conducted a comprehensive media campaign that included television, websites, and social media. Taking advantage of social media platforms, the Ministry of Health also engages with the public and the media. In addition to these early initiatives, efforts have been made to combat rumors and misinformation and engage the public in prevention and control measures [49].

Higher health literacy levels are associated with more favorable perceptions of health information. However, health literacy varies depending on the situation, and thus even those with high levels of health literacy may need help occasionally. For instance, those unfamiliar with medical language may find it challenging to distinguish between materials that provide accurate information and those with inaccurate information. Health care professionals and organizations must evaluate the population’s level of health literacy in order to ensure that people have access to adequate information when it matters most. Strategies like awareness-raising campaigns, community engagement, educational interventions, and training programs should be implemented when needed [50]. In accordance with the first Infodemiology Conference of the World Health Organization, public health authorities must create, evaluate, implement, and adapt tools and strategies for managing infodemics in acute public health crises in a manner that is suitable for their countries and situations [15]. This study’s findings may provide insight to public health authorities about developing an appropriate intervention for the population.
Strengths and Limitations
The strength of this study is that it involved developing and validating an educational video in the Arabic language to identify misinformation on WhatsApp. Few Arabic educational materials exist to combat misinformation. The study’s limitations include some sampling bias due to the use of convenience sampling, which is a nonprobability sampling technique. While this technique may limit the generalizability of the results, it was appropriate for our study because it is more cost-effective, faster, and more direct than other sampling techniques.

Conclusions
Health misinformation is an issue threatening public health because it dominates social media. Training people on the characteristics and practical applications of social media is urgently necessary. This study developed and evaluated the effectiveness of an educational video intervention to improve health misinformation identification on WhatsApp among the Saudi Arabian population. The results indicate that educational videos can be valuable tools for improving participants’ abilities to identify misinformation. The outcomes of this research can contribute to our understanding of effective tools for enhancing health misinformation awareness. These interventions can be particularly useful in combating misinformation in Arabic-speaking populations on WhatsApp, which may ultimately improve eHealth literacy. Limiting the prevalence and impact of misinformation allows people to make better-informed health care decisions. Our findings may also be helpful for health care professionals and organizations deciding on interventions suitable for providing access to adequate information to certain populations when needed.

Acknowledgments
The authors appreciate the support provided by the Office of Research in KSAU-HS, especially Mr Ahmed Aldakhil, for his valuable biostatistics consultations. The authors gratefully acknowledge King Abdullah International Medical Research Center for funding the publication of this paper.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
The study was designed and planned by EA and SA. EA took on the task of developing the educational video and survey. Data collection, results interpretation, and manuscript writing were all carried out by EA, SA reviewed the work. EA and SA discussed the findings, contributed to the final manuscript, and shaped the research.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The developed educational video (Health Misinformation Identification on WhatsApp).
[MP4 File (MP4 Video), 11898 KB - formative_v8i1e50211_app1.mp4 ]

Multimedia Appendix 2
Experts names and positions.
[DOCX File, 14 KB - formative_v8i1e50211_app2.docx ]

Multimedia Appendix 3
The video content validation form.
[DOCX File, 15 KB - formative_v8i1e50211_app3.docx ]

Multimedia Appendix 4
WhatsApp messages evaluation concepts and the corresponding survey questions used to assess them.
[DOCX File, 1738 KB - formative_v8i1e50211_app4.docx ]

Multimedia Appendix 5
The final version of health misinformation identification survey.
[DOCX File, 972 KB - formative_v8i1e50211_app5.docx ]

References
A Blended Intervention Targeting Emotion Dysregulation in Adults With Attention-Deficit/Hyperactivity Disorder: Development and Feasibility Study

Emilie S Nordby1,2, MSc; Frode Guribye3, PhD; Viktor Schønning4, MSc; Sander Lindholm Andersen1, MSc; Jonna Kuntsi5, PhD; Astrid J Lundervold2, PhD

1Division of Psychiatry, Haukeland University Hospital, Bergen, Norway
2Department of Biological and Medical Psychology, Faculty of Psychology, University of Bergen, Bergen, Norway
3Department of Information Science and Media Studies, Faculty of Social Sciences, University of Bergen, Bergen, Norway
4Department of Child and Adolescent Psychiatry, Division of Psychiatry, Haukeland University Hospital, Bergen, Norway
5Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, United Kingdom

Corresponding Author:
Emilie S Nordby, MSc
Division of Psychiatry
Haukeland University Hospital
Sandviksleiet 1
Bergen, 5036
Norway
Phone: 47 45440197
Email: Emilie.nordby@uib.no

Abstract

Background: Many adults with attention-deficit/hyperactivity disorder (ADHD) experience difficulties related to emotion regulation. Such difficulties are known to substantially impact quality of life and overall functioning. Yet, there is a lack of treatment interventions specifically designed to address these challenges.

Objective: This study aimed to describe the development and assess the feasibility, along with the initial clinical outcomes, of a novel blended intervention for adults with ADHD. The blended intervention combines both face-to-face and digital components and is specifically designed to address emotion dysregulation in ADHD.

Methods: This intervention was an 8-week blended intervention combining weekly face-to-face group sessions with a supplementary digital companion app. The intervention is based on elements from dialectic behavioral therapy skills training and positive psychology. To evaluate its feasibility, we performed a 10-week feasibility study with an uncontrolled pre-post study design, including 16 adults with ADHD and co-occurring emotion dysregulation. The feasibility measures encompassed adherence, satisfaction, and perceived credibility of the intervention. Clinical outcomes were evaluated by self-reported symptoms of emotion dysregulation, inattention, hyperactivity-impulsivity, executive function, depression, anxiety, and a measure of quality of life. Paired sample 2-tailed t tests were used to analyze clinical outcomes with a Bonferroni-corrected significance level.

Results: Both treatment credibility and treatment satisfaction were rated favorably by the majority of the participants. In particular, the participants emphasized meeting others with ADHD as beneficial. In terms of adherence, 3 participants withdrew before initiating the intervention, while another 4 participants did not complete the intervention. On average, the participants who enrolled in the intervention attended 6.2 of the 8 group sessions and completed 6.7 of the 8 skills training modules in the companion app. In terms of clinical outcomes, there was a reduction in symptoms of emotion dysregulation from before to after the intervention (d=1.1) and hyperactivity-impulsivity (d=0.9). However, no significant improvements were found in the domains of depression, anxiety, quality of life, and executive functioning.

Conclusions: The results are encouraging, both in terms of feasibility and the preliminary clinical results on emotion dysregulation. The blended format, combining digital and face-to-face elements, may also seem to offer some advantages: the group-based format was valued as it facilitated peer interaction, while a rather high completion of modules in the companion app highlights its potential to enhance skills training between the group sessions. Future randomized controlled trials are called for to further evaluate the clinical effectiveness of the intervention.
Introduction

Attention-deficit/hyperactivity disorder (ADHD) affects approximately 2.6% of the adult population [1]. Although symptoms of ADHD initially present themselves in childhood, most individuals diagnosed with ADHD as children continue to experience symptoms and associated impairments into adulthood [2]. Moreover, there is a rising trend of individuals first receiving their ADHD diagnosis in adulthood [3]. ADHD manifests through symptoms of inattention, hyperactivity, and impulsivity that significantly disrupt the individual’s daily functioning [4]. Beyond these core symptoms, ADHD is accompanied by an array of co-occurring symptoms and difficulties, both somatic and psychological, which can further complicate and amplify the challenges associated with the diagnosis [5].

Emotion dysregulation is a common deficit observed across many mental health conditions [6]. It is characterized by challenges in effectively managing and modulating one’s emotions, including the emotional experience (e.g., intensity and duration) and expression [7]. Emotion dysregulation commonly coexists with ADHD, impacting as many as 34%-70% of the adults with the diagnosis [8]. Emotion dysregulation also occurs in adults with ADHD without the presence of other comorbid diagnoses [9], giving support to the notion that it is a core component of ADHD [10-12]. The co-occurrence of emotion dysregulation with ADHD is linked to a range of negative outcomes, including occupational challenges, interpersonal conflicts, financial struggles, parenting stress, as well as tendencies for self-harm, illicit drug use, and suicidal ideation [13-15]. Moreover, research indicates that emotion dysregulation serves as an independent predictor of impairment in ADHD [16]. Its adverse influence on self-esteem and quality of life has further been observed to exceed the effects of inattention and hyperactivity [17]. Given the high prevalence of emotion dysregulation in ADHD and the associated negative outcomes, adults with ADHD should be offered treatment interventions that specifically aim to strengthen their emotion regulation skills.

While pharmacological treatment remains the main treatment approach for the management of ADHD in adults, the UK National Institute for Health and Care Excellence recommends supplementing with psychological interventions, such as psychoeducation or psychotherapy [18]. Psychological alternatives are especially critical in cases where the individual does not want to take medications or when medications either do not lead to sufficient clinical improvement or result in unwanted side effects. This may be particularly relevant for those with co-occurring emotion dysregulation, as ADHD medications appear to be less effective for these difficulties [19]. A systematic review and meta-analysis showed that traditional ADHD medications, including methylphenidate, atomoxetine, and lisdexamfetamine, only had small to moderate effects on emotion dysregulation among adults, which were significantly lower than the effect sizes reported for core symptoms of inattention, hyperactivity, and impulsivity [20]. As such, the authors of the abovementioned review emphasize that there is a need for more research on both pharmacological and psychological treatments targeting emotion dysregulation in adults with ADHD [20].

There is limited access to psychological treatments among adults with ADHD, and most available interventions mainly address the core symptoms of inattention and hyperactivity [21,22]. To the authors’ knowledge, there are currently 8 studies that have included emotion dysregulation as an outcome measure in studies of psychological interventions for adults with ADHD, including 2 randomized controlled trials [23-30]. However, only 2 of the studies reported emotion dysregulation to be the primary treatment target [24,26]. Carroll et al [26] developed and tested the psychological intervention “Group Therapy for Improving Emotional Acceptance and Regulatory Skills in Adults with ADHD” (GEARS), which consists of 14 weekly group sessions. In an uncontrolled pilot study with 226 participants, both treatment credibility and treatment satisfaction were rated as high, and preliminary clinical effects showed a reduction in emotion dysregulation symptoms with large effect sizes [26]. These findings are also supported by other studies [23,25,27-29]. On the other hand, Halmoy et al [24] did not find any significant difference in measures of emotion regulation between the control group and the treatment group in a randomized controlled trial of a 14-week dialectic behavioral therapy–based intervention. With the current evidence, it is thus premature to conclude whether psychological interventions are effective for adults with ADHD with co-occurring emotion dysregulation.

Digital psychological interventions have become increasingly popular in the past decade. The inclusion of digital tools in treatment interventions for adults with ADHD may be useful in addressing common challenges such as forgetfulness, nonadherence to treatment, and incomplete homework assignments [31,32]. For example, content from face-to-face therapy sessions, such as coping skills or psychoeducation, could be made available on the web through a website or a mobile app. The increased accessibility of such treatment elements may facilitate the generalization of therapeutic skills in everyday life for the clients, which is a central aim of most psychological interventions [33]. Blended interventions, where elements from face-to-face and digital treatment formats are combined, could be advantageous as they use the strengths of both treatment formats [33]. A systematic review of blended...
Interventions in mental health care suggest that these interventions may save clinician’s time, lead to lower dropout rates, and help to maintain positive changes made in psychotherapy over time [34]. However, there have been few studies examining a blended treatment format among adults with ADHD. To our knowledge, there has been 1 randomized controlled trial examining the effect of a mobile app to deliver psychoeducation in a group-based intervention for adults with ADHD [35]. In this study, it was found that the participants who used the mobile companion app had a greater reduction in ADHD symptoms and a higher completion rate of homework assignments as compared to participants who received a printed version of the psychoeducation [35]. As such, this study provides evidence that psychological interventions for adults with ADHD may be augmented by the implementation of digital tools.

The overall aim of this study was to describe the development of and assess the feasibility and preliminary clinical effects of a blended psychological intervention for adults with ADHD. This intervention integrates face-to-face group sessions with a digital companion app for skills training, designed to address emotion regulation. The developmental process and the core content of the intervention are outlined in the methods section. Through the feasibility study of the intervention, the following research questions were addressed:

1. How do participants with ADHD rate their satisfaction and the credibility of the intervention?
2. What is the participants’ adherence level to the intervention?
3. What are the preliminary clinical effects of the intervention on emotion regulation, inattention, hyperactivity-impulsivity, quality of life, anxiety, depression, and executive functioning?

**Methods**

### Phase 1: Development of Intervention

#### Participants

The participants who took part in the development phase of the intervention included 5 adults with ADHD, 2 clinical psychologists, 1 clinical psychology student, 2 human-computer interaction experts, and 2 user experience designers. The adults with ADHD were originally recruited from the local ADHD patient association. The adults with ADHD received a gift certificate worth NOK 400 (US $37) in total for their participation in the design workshop and evaluation meeting.

#### Methods of Development

The intervention was developed over 2 years through an iterative process. Multiple methods were used during this development (Table 1 shows an overview of methods and findings).
Table 1. Overview of the methods applied in the development of the intervention and key findings in the 2-year developmental process.

<table>
<thead>
<tr>
<th>Method</th>
<th>Content</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthesis of quantitative literature</td>
<td>A synthesis of previous studies examining psychological interventions targeting emotion dysregulation in ADHD was conducted. In this step, we examined the format, treatment approach, and common psychotherapeutic elements in the interventions.</td>
<td>- Various psychological frameworks have been used in interventions targeting emotion dysregulation in ADHD, including DBT, cognitive behavioral therapy, goal management training, and mentalization-based therapy.</td>
</tr>
<tr>
<td>Synthesis of qualitative literature</td>
<td>A synthesis of previous qualitative examining the experience of participating in psychological intervention for adults with ADHD was conducted. In this step, we sought to understand the needs and preferences of psychological interventions among adults with ADHD.</td>
<td>- In previous interventions, the opportunity to share experiences with peers was perceived as valuable by adults with ADHD.</td>
</tr>
<tr>
<td>Expert meetings</td>
<td>The development process included several expert meetings. In these meetings, the format and elements of the intervention were discussed and reviewed. Suggestions regarding intervention content and design were made based on clinical expertise and experience.</td>
<td>- Beneficial elements that should be included in the intervention based on clinical experience: group-based format, focus on positive aspects, option to choose from a variety of coping strategies, interactive in-class exercises and discussions, and short homework assignments.</td>
</tr>
<tr>
<td>Co-design workshop</td>
<td>The workshop included 5 adults with ADHD, 2 clinical psychologists, and 2 HCI experts. In the workshop, the adults were given information about the project and a general idea of the intervention. Following this, the adults were asked about challenging situations in terms of emotion dysregulation, common coping strategies, and their preferences for intervention content and features.</td>
<td>- Emotion dysregulation was a common challenge among adults with ADHD.</td>
</tr>
<tr>
<td>Design sprint of companion app</td>
<td>A design sprint lasting 5 days was conducted to create a prototype of the companion app. An HCI expert led the design sprint, with the inclusion of 2 UX designers and 1 clinical psychologist as participants.</td>
<td>- Perceived useful coping strategies by adults with ADHD: acceptance, self-compassion, distraction, reappraisal, time-out, and relaxation.</td>
</tr>
<tr>
<td>Co-creation of intervention manual</td>
<td>The first author created the intervention manual based on the previous findings. The protocol was then reviewed and revised by 2 other clinical experts, including 1 expert with ADHD.</td>
<td>- Useful features in companion app: reminders, note-taking, overview of coping strategies, calendar, peer support, and inclusion of videos.</td>
</tr>
<tr>
<td>Evaluation seminar with adults with ADHD</td>
<td>The evaluation seminar included the same participants as the co-design workshop. In the first part of the seminar, a clinical psychologist presented the content of 8 group sessions and the participants could give feedback. In the second part, an HCI expert presented a walk-through of the companion app, and participants could give feedback.</td>
<td>- A prototype of the companion app was completed, including design and features in the app.</td>
</tr>
<tr>
<td>Consultations with software company</td>
<td>We conducted several consultations with the software company that had the technical infrastructure for the companion app.</td>
<td>- A first version of the intervention manual was created.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Review of the intervention, both the group sessions and the companion app.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Revision of the intervention, including limiting the amount of text and clarifying instructions in companion app.</td>
</tr>
</tbody>
</table>

*aADHD: attention-deficit/hyperactivity disorder.

*DBT: dialectic behavioral therapy.

*HCI: human-computer interaction.

*UX: user experience.
The Emotion Regulation Intervention for ADHD

Overview

The development process led to the “Emotion Regulation Intervention for ADHD” (ERIA), which is a structured psychological intervention aimed at improving emotion regulation skills in adults with ADHD. The intervention is manual-based and includes components from dialectic behavioral therapy skills training and positive psychology. ERIA consists of 8 face-to-face group sessions and a digital companion app for skills training in between sessions. Table 2 contains an overview of the intervention content in ERIA.

Table 2. Overview of the intervention, including session themes, content for group sessions, weekly homework, and skills training assignments.

<table>
<thead>
<tr>
<th>Session</th>
<th>Content</th>
<th>Homework and skills training</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>• Introduction to group members and therapists</td>
<td>• Goal setting</td>
</tr>
<tr>
<td></td>
<td>• Overview of the program</td>
<td>• Identifying own strengths</td>
</tr>
<tr>
<td></td>
<td>• ADHD(^a) and emotion</td>
<td>• Complete plan for homework</td>
</tr>
<tr>
<td></td>
<td>• Discussion on challenges and strengths of ADHD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Strategies for homework completion</td>
<td></td>
</tr>
<tr>
<td>2. Mindfulness I</td>
<td>• Mindfulness practice</td>
<td>• Practice “what” skills in mindfulness</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>(observation, description, and participation)</td>
</tr>
<tr>
<td></td>
<td>• Mindfulness</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Introduction to “what skills” in mindfulness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “observe object in room”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “describe the pictures”</td>
<td></td>
</tr>
<tr>
<td>3. Mindfulness II</td>
<td>• Mindfulness practice</td>
<td>• Practice “how” skills in mindfulness</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>(being nonjudgmental, doing 1 thing at the time, and doing what works)</td>
</tr>
<tr>
<td></td>
<td>• Common barriers for homework completion</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Introduction to “how skills” in mindfulness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “judgmental vs nonjudgmental claims”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “multitasking”</td>
<td></td>
</tr>
<tr>
<td>4. Emotion regulation I</td>
<td>• Mindfulness practice</td>
<td>• Practice emotion regulation skills</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>(observation of emotions and naming and describing emotions)</td>
</tr>
<tr>
<td></td>
<td>• Emotions and emotion regulation</td>
<td>• Emotion diary</td>
</tr>
<tr>
<td></td>
<td>• Emotion regulation skills: part I</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Introduction to the emotion diary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “linking situations to emotions”</td>
<td></td>
</tr>
<tr>
<td>5. Emotion regulation II</td>
<td>• Mindfulness practice</td>
<td>• Practice emotion regulation skills</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>(check the facts, opposite action, and problem-solving)</td>
</tr>
<tr>
<td></td>
<td>• Adaptive vs maladaptive regulation strategies</td>
<td>• Emotion diary</td>
</tr>
<tr>
<td></td>
<td>• Emotion regulation skills: part II</td>
<td>• Skills training log</td>
</tr>
<tr>
<td>6. Emotion regulation III</td>
<td>• Mindfulness practice</td>
<td>• Practice emotion regulation skills</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>(covering basics needs and planning for challenging situations)</td>
</tr>
<tr>
<td></td>
<td>• Emotion regulation skills: part III</td>
<td>• Planning positive activities</td>
</tr>
<tr>
<td></td>
<td>• Tips for planning and organization</td>
<td>• Emotion diary</td>
</tr>
<tr>
<td></td>
<td>• Planning positive activities</td>
<td>• Skills training log</td>
</tr>
<tr>
<td>7. Crisis management</td>
<td>• Mindfulness practice</td>
<td>• Practice skills in crisis management</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>(stop and check-in, physical exercise, cold water, muscle relaxation, breath work, and distraction).</td>
</tr>
<tr>
<td></td>
<td>• Skills for crisis management and intense emotions</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Discussion on distraction strategies</td>
<td></td>
</tr>
<tr>
<td>8. Summary</td>
<td>• Mindfulness practice</td>
<td>• N/A(^b)</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Self-compassion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “supportive words”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Summary of program</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Road ahead: maintaining change and setbacks</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ADHD: attention-deficit/hyperactivity disorder.

\(^b\)N/A: not applicable.
Group Sessions
ERIA comprises 8 weekly group sessions. The group sessions are closed, and each group includes 6-8 adults with ADHD. In this study, the groups were led by 2 clinical psychologists with a minimum of 3 years of clinical experience. Each group session lasts approximately 1.5 hours, divided into 2 segments with a 15-minute break interlude. All sessions begin with a brief mindfulness exercise, led by 1 of the 2 psychologists. Afterward, group members are encouraged to share their reflections on the exercise. Following this, the group members share their experiences with the previous week’s homework assignment. After the break, the lead psychologist presents relevant theoretical and psychoeducational information and the new skills for the participants to practice at home. There are also discussion breaks and some practical exercises incorporated in the presentation of new skills.

Companion App
The participants are given access to a companion app, which they are asked to use for skills training at home between the group sessions. This app is organized with modules, where new skills training modules are released on a weekly basis following the course of the group sessions. The skills training modules include the weekly skills that the participants are to practice at home. In addition, the participants can use the companion app to message the group leader, log their skills training sessions, and access the PowerPoint presentations for the group sessions, as well as other relevant resources (eg, relevant websites or scientific papers on the weekly theme). The participants also receive an SMS text message reminder when the skills training modules are available and another SMS text message reminder if they have not accessed the module within 2 days. Given that this was a feasibility study, the app did not encompass all features originally present in its prototype. For instance, a peer-support feature suggested during the co-design workshop was excluded due to its resource-intensive nature. If the feasibility study yields encouraging outcomes, we intend to incorporate more sophisticated features in the app.

The app was a web-based application that could be accessed both through a computer and a smartphone. To access the app, the participants had to use the Norwegian web-based authentication platform, BankID, for a secure 2-factor log-in. BankID is a widely used platform for electronic identification for services within banking, health care, and education in Norway. The companion app was hosted on the eHealth platform developed by the software company “Youwell,” which is partnered with the Western Norway Regional Health Authority for clinical use and research. The platform has a patient and therapist portal, where participants can access the app through the patient portal, and the therapist can set up the program and monitor their patient’s progress in the therapist portal. The platform is also used for building the app and allows for the input of content, such as text, audio, videos, and images, to create modules. It has previously been used for other internet-delivered programs, including programs for ADHD [36], social anxiety [37], and cognitive residual symptoms after depression [38].

Figure 1 shows screenshots of the companion app. The first screen from the left shows the main page with an overview of the modules. The first box from the left reads, “Week 4: Emotion regulation I. This week we will practice skills in emotion regulation.” The second screen shows the instructions to a mindfulness skill with the text “Choose an option below and take 5 minutes to describe this in a nonjudgmental way.” The circles show 6 options they can choose from, which include describing (1) own emotions, (2) bodily sensations, (3) a tree or plant, (4) people walking by, (5) an object in the room, or (6) own thoughts. The third screen shows the side menu that participants can use to navigate within the module. This screen shows the mindfulness module and the side menu includes the specific module pages: “Welcome back,” “Skills,” “Observation,” “Description,” “Participate,” “Skills diary card,” and “Resources.”
Phase 2: Feasibility Study

Participants

Eligible participants for this feasibility study were adults with ADHD living in Bergen, Norway. The inclusion criteria for the study were as follows: (1) minimum age of 18 years; (2) a diagnosis of ADHD according to the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition]) criteria; (3) current problems with emotion dysregulation as indicated by a score of 280 on the Difficulties in Emotion Regulation Scale (DERS); (4) having a smartphone or computer to access the companion app; and (5) the ability to attend in-person group sessions in Bergen, Norway. Exclusion criteria were (1) a high risk of suicidality, as indicated by having attempted suicide within the past year, having previously attempted suicide and reporting current suicidal ideations, or reporting current suicidal ideations and having a plan and preferred method; (2) co-occurring severe mental illness, including substance abuse disorder, psychosis, and major depressive disorder; and (3) current participation in another psychological treatment intervention. However, participants could still partake in the study if they had less severe psychiatric conditions, such as mild to moderate anxiety or depression, and if they received pharmacological treatment for their ADHD or other conditions.

Recruitment

Participants were recruited through the ADHD patient advocacy group “ADHD Norge,” which shared information about the study with its members through email and social media. The study opened for participants to sign up on November 24, 2022, and closed within a week (November 30, 2022) due to a large number of individuals signing up for the study. Interested participants signed up through a website that contained a screening survey to examine eligibility as well as information about the study. Eligible participants were contacted for a face-to-face screening with a clinical psychologist. During this screening, eligibility in terms of psychiatric comorbidities and suicidality was examined using the Mini International Neuropsychiatric Interview [39]. Participants were also asked about the date, clinic, and diagnosing clinician for their ADHD diagnosis. In addition, the participants were asked open questions about ADHD symptoms (ie, could you tell me about your current ADHD symptoms?) and everyday functioning (ie, how do you experience that ADHD affects you in your daily life?). The participants had to report symptoms and functional impairments that were in accordance with the DSM-5 criteria for ADHD, as assessed by a clinical psychologist. Those who were deemed eligible for participation following this screening were invited to take part in the study and to sign an informed consent form through the companion app.

Outcome Measures

Overview

The participants were given the preassessment 1 week before the intervention started and the postassessment 1 week after the intervention ended. The assessments were completed on the internet, except for the Behavior Rating Inventory of Executive Functioning–Adult version (BRIEF-A), which was given in-person at the first and last group session.

Credibility

The third item of the Credibility and Expectancy Scale (CEQ) was used to examine treatment credibility of the intervention [40]. The item states, “Would you recommend this treatment to a friend with similar challenges?” and the responses are given on a scale from 1 (not certain at all) to 9 (very certain).
Treatment Satisfaction
The participants were asked about whether they thought they would continue using the coping skills they had learned in the future and responded on a 4-point scale from 1 (not very likely) to 4 (very likely). They were also asked about which elements of the intervention they considered to be most useful (multiple choice with option to add own text). The participants were also asked about negative experiences with the intervention (yes or no), which they could elaborate on in an open-text field.

Adherence
Adherence was assessed by the number of group sessions attended and the number of completed skills training modules in the companion app. Participants that attended at least 6 of 8 group sessions were defined as treatment completers. The participants were also asked about how many days per week they practiced the coping skills.

Difficulties in Emotion Regulation Scale
DERS is a self-report questionnaire that is commonly used to assess emotion dysregulation in clinical populations [41]. The scale includes 36 items rated on a 5-point Likert scale ranging from 1 (never) to 5 (almost always), yielding a total score between 36 and 180, with 180 indicating the most severe problems with emotion dysregulation.

The Adult ADHD Self-Rating Scale
The Adult ADHD Self-Rating Scale (ASRS) is a self-report questionnaire that is used to assess symptoms of inattention and hyperactivity-impulsivity [42]. The scale includes 18 items, with 9 items reflecting inattentive symptoms and 9 items reflecting hyperactive-impulsive symptoms. Responses are given on a 5-point scale with options 0 (never), 1 (rarely), 2 (sometimes), 3 (often), or 4 (very often), giving a total score between 0 and 72 and a score between 0 and 36 for the inattention and hyperactivity-impulsivity subscales.

The Adult ADHD Quality of Life Measure
The Adult ADHD Quality of Life (AAQoL) measure is used to assess quality of life among adults with ADHD [43]. The scale includes 29 items rated on a scale from 1 (not at all or never) to 5 (extremely or very often), yielding a total score between 0 and 100.

Hospital Anxiety and Depression Scale
The Hospital Anxiety and Depression Scale (HADS) is a self-report questionnaire used to assess symptoms of depression and anxiety [44]. The scale includes 14 items, with 7 items reflecting anxiety symptoms and 7 items reflecting depressive symptoms. The response options range from 0 to 3, with 3 being the most severe level. The scale yields a total score between 0 and 42 and a score between 0 and 21 for the anxiety and depression subscales.

The Behavior Rating Inventory of Executive Functioning
BRIEF-A is a self-report questionnaire used to assess executive functioning in everyday life [45]. The scale consists of 75 items, which are rated on a 3-point scale (1=never, 2=sometimes, and 3=often). For this study, we report the Global Executive Composite score, which is an overall summary score including 9 clinical BRIEF-A subscales.

Statistical Analysis
SPSS software (IBM Corp) was used for all statistical analyses [46]. The participant demographics, adherence measures, and treatment credibility measures were assessed using descriptive statistics, which include calculation of means, frequencies, ranges, and SDs. To evaluate preliminary clinical outcomes, paired sample t tests were used with an initial significance level set at .05. Due to the risk of family-wise error (type I error) associated with multiple t tests, a Bonferroni correction was included. This adjustment was achieved by dividing the α level by the number of conducted hypotheses tests, that is, t tests (.05/8), resulting in a corrected significance level of .006. The choice of analytic approach necessitated the inclusion of cases with both pre- and postassessment. The magnitude of treatment effect was quantified using standardized effect sizes, estimated through Cohen $d$, with the formula ($M_2 - M_1$)/SD pooled. The pooled SD was calculated by $\sqrt{[(SD_1)^2 + (SD_2)^2]/2}$. Effect sizes were interpreted according to conventions, where $d$=0.20, $d$=0.50, and $d$=0.80 are defined as small, moderate, and large, respectively.

Ethical Considerations
The study was reviewed and approved by the Regional Ethics Committee of Norway, Region West (494659). The participants were informed about the study and their rights both in-person and in writing. Before participation, all participants signed an informed consent form, acknowledging that they could withdraw their consent at any time without any repercussions. The data were pseudoanonymized and stored on a dedicated research server according to regulative standards at the university and hospital. The participants were compensated NOK 1000 (US $90) for their participation in the study.

Results
Participants
A total of 16 participants took part in this study (Table 3 shows participant characteristics). The majority were diagnosed with ADHD in adulthood, with the mean age for receiving the diagnosis being 30.4 (SD 11.7) years.

All participants were recruited within 1 week. During the recruitment period, 68 adults completed the digital prescreening, of whom 92.6% (n=63) met the inclusion criteria and provided contact information (Figure 2). However, due to the limited number of study places in this feasibility study, we only assessed the first 27 individuals who signed up for the phone screening, as this was sufficient to reach the maximum capacity of 16 included participants. We sent an SMS text message to the remainder of the people who signed up, informing them that the study was fully booked. To ensure a more balanced sex distribution in the study, 6 study places were reserved for men, as we wanted to make sure we received feedback from both sexes.
Table 3. Participant characteristics in the feasibility study (n=16).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10 (62.5)</td>
</tr>
<tr>
<td>Male</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>35-44</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>45-55</td>
<td>1 (6.2)</td>
</tr>
<tr>
<td>55-64</td>
<td>1 (6.2)</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>8 (50)</td>
</tr>
<tr>
<td>College or university level</td>
<td>8 (50)</td>
</tr>
<tr>
<td><strong>Occupational status</strong></td>
<td></td>
</tr>
<tr>
<td>Employed or student</td>
<td>9 (56.2)</td>
</tr>
<tr>
<td>Work assessment allowance or sick leave</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td>Disability pension</td>
<td>1 (6.2)</td>
</tr>
<tr>
<td><strong>ADHD medication status</strong></td>
<td></td>
</tr>
<tr>
<td>Daily medication usage</td>
<td>13 (81.3)</td>
</tr>
<tr>
<td>Weekly or monthly usage</td>
<td>2 (12.4)</td>
</tr>
<tr>
<td>Rarely, a few times a year</td>
<td>1 (6.2)</td>
</tr>
</tbody>
</table>
Treatment Adherence, Credibility, and Satisfaction

In terms of adherence, a total of 3 participants dropped out of the study before starting the intervention, making a pretreatment dropout rate of 18.8% (3/16). Among the participants who received the intervention (n=13), the average number of group sessions attended was 6.2 out of 8 sessions. A total of 4 (26.6%) participants were treatment dropouts, defined as participants who attended less than 6 group sessions. Consequently, the cumulative dropout rate, including the participants who dropped out before starting the intervention, reached 43.7% (n=7). Regarding the use of the companion app, the participants had a mean completion of 6.8 out of 8 modules. On average, the participants reported practicing the skills 4.3 days per week.

The intervention was generally well received in terms of treatment credibility with a mean rating of 7.1 (SD 2.6; range 1-9). Among the 10 participants who completed the postassessment, 7 participants reported that they were very certain or certain that they would recommend the treatment to a friend facing similar challenges as themselves; 2 were somewhat certain, whereas 1 was not at all certain about recommending it.

Feedback from participants highlighted overall satisfaction with the intervention, with a mean rating of 3.3 (SD 0.9; range 1-4). All but 1 participant planned to continue using the learned skills. The participants rated meeting others with ADHD (n=10), the in-person group sessions (n=7), the skills (n=7), and therapist support (n=4) as the most useful elements of the intervention. Their feedback also included suggestions for improvement, with 3 participants recommending more time for group interactions and discussions among the group members. Regarding this, 1 participant suggested that the group members should be able to interact between the sessions through the companion app. Other suggestions for improvements included an extension of the program by adding more group sessions, making the companion app available for direct download, incorporating more reminders, providing participants with a printed version of the skills, and involving an individual with ADHD as a group presenter for skills demonstration and experience sharing. While feedback was largely positive, 2 participants expressed negative experiences with the program: 1 found the skills in the crisis management module to be inadequate in emotional crises, and another felt that the intervention was not sufficiently tailored to ADHD. Importantly, there were no reports of clinical deterioration among the participants.
Clinical Outcomes

There was an overall statistically significant decrease in self-reported emotion dysregulation from before to after treatment, with a strong effect size ($d=2.0$). Table 4 shows the individual scores for emotion dysregulation, indicating a change in the positive direction for all participants.

Overall, the group of participants showed a significant decrease in secondary clinical outcome scores, reflecting the level of inattention and hyperactivity-impulsivity from before to after assessment. While initial analyses showed a change in quality of life and depressive symptoms, these changes were nonsignificant with a Bonferroni correction. No significant improvement was found in measures of executive functioning or anxiety symptoms (Table 5 contains an overview of all preliminary clinical outcomes).

Table 4. Individual scores for the Difficulties in Emotion Regulation Scale from before to after assessment (n=10).

<table>
<thead>
<tr>
<th>Participant</th>
<th>Preassessment scores</th>
<th>Postassessment scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>130</td>
<td>101</td>
</tr>
<tr>
<td>2</td>
<td>138</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>110</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>147</td>
<td>91</td>
</tr>
<tr>
<td>5</td>
<td>96</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>130</td>
<td>94</td>
</tr>
<tr>
<td>7</td>
<td>107</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>139</td>
<td>117</td>
</tr>
<tr>
<td>9</td>
<td>133</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>107</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5. Clinical outcomes of emotion dysregulation, inattention, hyperactivity-impulsivity, quality of life, depression, anxiety, and executive functioning from before to after assessment (n=10).

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Preassessment, mean (SD)</th>
<th>Postassessment, mean (SD)</th>
<th>$P$ valuea</th>
<th>Effect size, Cohen $d$</th>
<th>Difference, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DERSb</td>
<td>123.7 (17.2)</td>
<td>95.4 (9.6)</td>
<td>&lt;.001</td>
<td>2.0</td>
<td>15.8 to 40.8</td>
</tr>
<tr>
<td>ASRS full scalec</td>
<td>51.9 (9.6)</td>
<td>41.9 (8.9)</td>
<td>&lt;.001</td>
<td>1.1</td>
<td>6.2 to 13.8</td>
</tr>
<tr>
<td>ASRS inattention</td>
<td>26.9 (4.1)</td>
<td>22.3 (5.2)</td>
<td>&lt;.001</td>
<td>1.0</td>
<td>2.3 to 6.9</td>
</tr>
<tr>
<td>ASRS hyperactivity-impulsivity</td>
<td>25.0 (6.4)</td>
<td>19.6 (5.1)</td>
<td>.002</td>
<td>0.9</td>
<td>2.5 to 8.3</td>
</tr>
<tr>
<td>AAQoLd</td>
<td>44.4 (9.6)</td>
<td>56.7 (11.8)</td>
<td>.03</td>
<td>—e</td>
<td>−23.3 to −1.3</td>
</tr>
<tr>
<td>HADS Anxietyf</td>
<td>10.8 (4.3)</td>
<td>9.8 (3.6)</td>
<td>.24</td>
<td>—</td>
<td>−0.8 to 2.8</td>
</tr>
<tr>
<td>HADS Depression</td>
<td>6.0 (3.7)</td>
<td>4.3 (3.2)</td>
<td>.03</td>
<td>—</td>
<td>0.2 to 3.2</td>
</tr>
<tr>
<td>BRIEF-Ag GECfh</td>
<td>154.4 (15.9)</td>
<td>150.8 (12.6)</td>
<td>.20</td>
<td>—</td>
<td>−7.8 to 15.1</td>
</tr>
</tbody>
</table>

aSignificance level set to .006 with Bonferroni correction.
bDERS: Difficulties in Emotion Regulation Scale.
cASRS: Adult ADHD Self-Rating Scale.
dAAQoL: Adult ADHD Quality of Life.
eNot available.
fHADS: Hospital Anxiety and Depression Scale.
gBRIEF-A: Behavior Rating Inventory of Executive Functioning.
hGEC: Global Executive Composite.

Discussion

Principal Findings

The aim of this study was to describe the development and assess the feasibility of ERIA, a blended digital and face-to-face intervention targeting emotion dysregulation in adults with ADHD. Overall, the findings were promising and supported the feasibility of ERIA. Both treatment satisfaction and credibility were generally good, which aligned with findings from other psychological interventions targeting emotion dysregulation in adults with ADHD [26]. More specifically, the participants emphasized the group component and meeting others with
ADHD as useful aspects of the intervention. This is in line with previous studies showing that providing a forum where one can share experiences with peers is particularly valuable in treatment settings for this group of adults [47]. While the results are promising, there is room for refinements, including an even better tailoring of the intervention to ADHD. An important step forward would be to involve adults with ADHD in the review and refinement process to ensure that all key aspects of the interventions are well-adapted to the challenges and difficulties facing adults with ADHD. Nonetheless, the heterogeneous nature of ADHD requires finding a balance between including intervention components that fit most participants and including specialized strategies addressing the specific needs of a subset of individuals with ADHD.

This study highlighted concerns regarding treatment adherence, especially regarding pretreatment dropout. A total of 3 participants dropped out before starting the intervention, while another 4 participants did not complete the intervention. This finding aligns with well-known challenges related to treatment adherence among adults with ADHD [48,49]. The issues related to adherence in this study may be partially attributed to our community-based sample. Motivation for completing psychological treatment may thus have been lower than among clinic-recruited participants. However, there could also be intervention-specific factors related to dropouts, such as dissatisfaction with the intervention or finding the intervention too demanding. Yet it is also worth noting that although high levels of treatment adherence are generally seen as favorable for the patient, treatment dropout does not necessarily equate to clinical failure for the patients themselves. Still, when reviewing the current intervention, it may be necessary to include more strategies to prevent early termination of treatment. A literature review on the topic found that strategies such as pretherapy preparation, patient selection, time-limited treatment contracts, appointment reminders, motivation enhancement, facilitation of a therapeutic alliance, and facilitation of affect expression were specific strategies that could be applied to reduce premature termination of treatment across different psychiatric disorders [50]. As we progress, understanding the multifaceted factors associated with dropout will be crucial.

With regard to the companion app, it is interesting to note that the participants who initiated the intervention generally completed a high number of skills training modules in the companion app, with the majority completing all modules. This finding is in line with the results from Selaskowksi et al [35], which showed that the inclusion of a mobile app in a group-based intervention for adults with ADHD was linked to higher homework compliance [35]. Given the importance of homework for behavioral change as well as the common challenges related to homework compliance in psychological interventions, facilitating methods to ensure high attrition of homework may be particularly useful [51]. However, it is important to note that with the design of this study, we cannot determine whether the companion app resulted in higher homework compliance as opposed to not including the app; this would be an interesting topic for future studies.

The results were promising in terms of preliminary clinical findings, with participants showing a significant and large reduction from pre- to post treatment in emotion dysregulation. These findings are in accordance with findings from other psychological interventions targeting emotion dysregulation in individuals with ADHD [26]. Significant improvements were also observed in the secondary clinical outcome measures, including inattention and hyperactivity-impulsivity. However, given the study’s small scale, nonrandomized, and uncontrolled design, the clinical findings must be considered with caution. Moreover, we found no significant changes in the measures of anxiety, depression, quality of life, or executive functioning.

Regarding executive functioning, previous research has shown that measures of executive control remain stable in ADHD, regardless of remission or persistence [52]. More generally, the effect of interventions that are designed to improve cognitive abilities or executive functioning appears to be domain-specific and show mixed results [53,54].

Taken together, the results from this study were encouraging and call for further development of the intervention and a more extensive examination of clinical effects in a randomized controlled trial.

Limitations
This feasibility study has some limitations that should be noted, in particular the small sample size, the absence of a control group, and the lack of randomization. Due to these limitations, conclusions regarding the intervention’s clinical effects remain elusive, and we cannot rule out placebo or other random effects. Yet, the aim of this study was not to examine the effectiveness of the intervention but rather to assess its feasibility before paving the way for larger trials.

A further limitation is that 3 participants who took part in the intervention did not complete the postassessment. Therefore, it is possible that this could have impacted the treatment satisfaction and credibility scores of the interventions. Furthermore, the participants were recruited from the community and may therefore differ in some way from typical clinic-recruited adults with ADHD. The reliance on self-report scales to evaluate clinical outcomes in this study should also be considered a limitation. In future studies, it would be interesting to include other measures, such as those generated from sensor technology, clinician ratings, and performance on cognitive tests.

Conclusion
In conclusion, the results from this feasibility study support the potential of ERICA as a feasible intervention for addressing emotion dysregulation in adults with ADHD and call for further investigation in a randomized controlled trial. The blended approach, integrating digital and face-to-face elements, may offer some advantages compared to an exclusively digital or face-to-face treatment format. The in-person group sessions were especially valued because they provided opportunities to interact with peers. Meanwhile, the high completion rate of the companion app modules indicates their potential to facilitate skills training.
Acknowledgments
The study was funded by Helse Vest research funds. The authors would like to thank all the participants in this study and the experts to experience. We would also like to thank the patient advocacy group “ADHD Norge” for sharing information about this study. We would further like to thank Hedda Bakken, Sara Furuholmen, and Nanna Skram for their contributions to the project.

Authors’ Contributions
ESN, AJL, and FG contributed to the development, idea, and design of the study. ESN conducted the statistical analyses, while AJL, FG, JK, SLA, and VS contributed to the data collection for the feasibility study. ESN conducted the statistical analyses, while AJL, FG, JK, SLA, and VS contributed to the interpretation of the results. ESN was responsible for the drafting of the manuscript, while AJL, FG, JK, SLA, and VS critically reviewed and revised the draft. All authors approved the final version of the manuscript and take full accountability for the work.

Conflicts of Interest
JK has delivered talks at educational events sponsored by Medice; all funds were received by King’s College London and used for studies of ADHD. AJL has received speaker fees and travel expenses from Medice.

References


Abbreviations

AAQoL: Adult ADHD Quality of Life
ADHD: attention-deficit/hyperactivity disorder
ASRS: Adult ADHD Self-Rating Scale
BRIEF-A: Behavior Rating Inventory of Executive Functioning—Adult version
CEQ: Credibility and Expectancy Scale
DERS: Difficulties in Emotion Regulation Scale

https://formative.jmir.org/2024/1/e53931

Nordby et al JMIR FORMATIVE RESEARCH
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
ERIA: Emotion Regulation Intervention for ADHD
GEARS: Group Therapy for Improving Emotional Acceptance and Regulatory Skills in Adults with ADHD
HADS: Hospital Anxiety and Depression Scale

Please cite as:
A Blended Intervention Targeting Emotion Dysregulation in Adults With Attention-Deficit/Hyperactivity Disorder: Development and Feasibility Study
JMIR Form Res 2024;8:e53931
URL: https://formative.jmir.org/2024/1/e53931
doi:10.2196/53931
PMID:38231536

©Emilie S Nordby, Frode Guribye, Viktor Schønning, Sander Lindholm Andersen, Jonna Kuntsi, Astri J Lundervold. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Designing a Web-Based Navigation Tool to Support Access to Youth Mental Health Services: Qualitative Study

Alison L Calear¹, BAppPsych (Hons), PhD; Philip J Batterham¹, BSc Hons, MPH, PhD; Sonia M McCallum¹, GradCertPHC, BHlthSC (Hons), BSc, PhD; Michelle Banfield¹,², BA (Hons), BSc, PhD; Elizabeth Moore³, GradCertHPE, BS, MB; Natalie Johnson³, GradCertPM; Alyssa R Morse¹, AMusA, BPscy (Hons), PhD

¹Centre for Mental Health Research, The Australian National University, Canberra, Australia
²ALIVE National Centre for Mental Health Research Translation, Melbourne, Australia
³Office for Mental Health and Wellbeing, ACT Health, Canberra, Australia

Abstract

Background: Many young people with mental health problems do not readily seek help or receive treatment and support. One way to address low help-seeking behavior is to improve access to information on mental health services and how to navigate the mental health system via a web-based tool. Seeking input from the end users (young people and parents or caregivers) on key features of the tool is imperative to ensure that it is relevant, engaging, and likely to meet their needs and expectations.

Objective: This study aims to investigate young person and parent or caregiver views on the design, content, functioning, and user experience of a web-based mental health navigation tool to support connection to mental health services for children and young people aged up to 25 years.

Methods: A total of 4 online focus groups were conducted: 2 with young people aged 16 years and older (total n=15) and 2 with parents or caregivers (total n=13). Focus groups were structured around a series of guiding questions to explore participants’ views on content, features, user experience, and design of a mental health navigation website. Focus groups were audio recorded with detailed notes taken. In addition, 53 young people aged 16-25 years and 97 parents or caregivers completed an online survey, comprising closed- and open-ended questions; open-ended responses were included with the focus group data in the qualitative analysis. All qualitative data were analyzed using thematic analysis.

Results: A total of 2 topic areas and 7 themes were developed. The first topic area covered the types of information needs of young people and parents. Identified themes concerned the scope of the navigation website, as well as the provision of up-to-date and practical information on how to navigate the whole help-seeking process. The second topic area covered website features that would be beneficial and included the consideration of the website design; search engines; supported navigation; and forums, reviews, and user accounts.

Conclusions: This study provides important insights into the navigation needs of young people and parents or caregivers in seeking mental health services. Key findings identified through this research have directly informed the development of MindMap, a web-based youth navigation tool providing a searchable database of local services, including a clear description, their location, and potential wait times. The website can be navigated independently or with support.

doi:10.2196/48945

KEYWORDS

mental health services; youth; navigation tool; mental health; website; user experience; design; service; services; access; accessibility; health care system
Introduction

Globally, mental disorders are common and often emerge during childhood and adolescence [1,2]. The worldwide pooled prevalence of mental disorders in children and adolescents has been reported to be between 12.7% and 13.4% [2,3], while suicide is one of the leading causes of mortality in this age group [4]. Timely access to appropriate services and supports for mental disorders and suicidal distress can mitigate the persistence of poorer health, academic, and social outcomes into adulthood [5].

Research suggests, however, that many young people do not readily seek or receive treatment or support for psychological distress, suicidal ideation, or suicidal behavior [3,6,7]. A recent systematic review and meta-analysis of the prevalence of mental disorders in children and adolescents in high-income countries found that only 44.2% of young people with mental disorders received any services for their conditions [3]. Another study conducted in Australia reported slightly higher rates of service use among children and adolescents, with 56% accessing services for emotional and behavioral problems [7]. However, Sawyer et al [8] reported that only 12% of Australian young people aged 6-17 years with a mental health condition received what was considered adequate treatment.

Several possible drivers of low help-seeking behavior among young people, parents, and other caregivers have been identified. A recent review found that limited mental health knowledge, embarrassment and perceived social stigma, a lack of perceived confidentiality and trust in mental health providers, financial costs, logistical barriers, and limited availability of services were common barriers to accessing professional help for mental health problems among young people [5]. Similar barriers were identified among parents seeking treatment for mental health problems in their children and adolescents, with a lack of knowledge of where or how to seek help and a limited understanding of the mental health system identified as key barriers to service use [9].

One way to address some of these barriers may be to provide young people and parents or caregivers with a navigation tool they could use to identify available services and traverse often complex and disconnected mental health systems. Similar to other high-income countries, there are multiple entry points into the mental health care system in Australia, but access is often dependent on the individual knowing about them. As such, being able to access all service options in the same place would allow greater reach and timely updating of service information. In a survey assessing parents’ help-seeking for their adolescent’s mental health, over 75% of participants indicated that they would use the internet to find information about services [14]. In developing such a website, it is imperative that young people and parents or caregivers are involved, as there is clear evidence in the literature that the involvement of end users is essential to ensuring uptake and that the website meets the needs and expectations of users [15,16].

This paper reports the outcomes of focus groups and online surveys conducted with young people, parents, and other caregivers to identify the design, content, functioning, and user experience of a web-based mental health services navigation tool for children and young people aged up to 25 years in the Australian Capital Territory (ACT). This study was conducted in partnership with the ACT Office for Mental Health and Wellbeing and provides unique insights into the navigation needs of this population, with a dearth of previous research in this area, and the importance of the consultation process in producing an interactive website that is relevant, engaging, and likely to meet the needs of end users.

Methods

Participants

A total of 4 online focus groups (N=28), with 5-10 participants each, were conducted. Of these focus groups, 2 focus groups (15/28, 54%; 10/15, 66% female) were conducted with young people aged 16 years and older. These groups were conducted with members of 2 existing youth mental health reference groups: the ACT Youth Advisory Council and the ACT headspace Youth Advisory Group, who had experience representing the wider interests and views of their peers. The other 2 focus groups involved parents or caregivers (13/28, 46%; 12/13, 92% female). Participants for these were drawn from expressions of interest from the community and included parents with and without mental health service experience. To attract diverse representation in the focus groups, we advertised for recruitment through ACT mental health sector newsletters and on the social media website Facebook. Facebook advertising targeted adults in the ACT who identified as a parent or caregiver of a child aged 10-25 years. The online modality for the focus groups was chosen to maximize ease of attendance and minimize potential COVID-19 exposure.

Alongside the focus groups, 150 people participated in an online survey to further garner the perspectives of young people and parents or caregivers. Participants were young people aged 16-25 years (53/150, 35%; 44/53, 83% female; mean age 20, SD 2.8 y) and the parents or caregivers (97/150, 65%; 90/97, 93% female; mean age 45.5, SD 7.6 y) of young people aged 0-25 years (0-6 y: 20/97, 21%; 7-11 y: 38/97, 39%; 12-15 y: 47/97, 48%; 16-18 y: 30/97, 31%; and 19-25 y: 37/97, 38%). Participants for the survey were recruited through Facebook advertising. The Facebook advertisements were delivered to 21,600 accounts, with 1300 people engaging with the Facebook post and 374 clicking the link and commencing the survey. Of those who commenced the survey, 40.1% (150/374) went on to complete and submit the survey for analysis. The majority of participants reported English as their first language (47/53, 89% young people and 93/97, 96% parents or caregivers) and had accessed pediatric mental health services in the past (45/53, 85% young people and 86/97, 89% parents or caregivers). Just over half (28/53, 53%) of the young people who completed the survey identified as lesbian, gay, bisexual, transgender, intersex,
queer (LGBTIQ+), whereas 21% (11/53) reported having a disability.

**Ethical Considerations**

Ethical approval for this study was obtained from the Australian National University Human Research Ethics Committee (protocol 2020/200). All participants were provided with an information sheet outlining the study. Participants in the focus groups were required to return a signed consent form via email before attending the online focus group, whereas survey participants consented to participate in the survey by checking a box at the beginning of the survey. All survey data were collected anonymously, whereas all focus group data were deidentified before analysis and data storage. All focus group participants were provided with a small honorarium of AU $60 (US $40) in recognition of their time and contribution to the study.

**Materials**

The focus groups were structured, with a series of guiding questions and prompts that explored the views of young people and parents or caregivers on the potential design, content, functioning, and user experience (eg, optional phone and email support) of a web-based navigation tool (the full list of questions and prompts is given in Multimedia Appendix 1). Questions and prompts were developed in collaboration with staff members from the ACT Office for Mental Health and Wellbeing to gather information required to progress website development from an early concept and include youth and parent perspectives in the design process. Participants were encouraged to share their perspectives verbally with the group or to note any thoughts or ideas in the chat function.

The online survey assessed a range of topics relevant to the development of a youth navigation tool including website features (eg, information, links, quizzes, and service contact details), features and topics by which services could be searched (eg, cost, location, age, and gender), navigation support (eg, phone and online), and possible account functions (eg, acceptable account details and storage of information). Questions were developed with iterative feedback from the ACT Office for Mental Health and Wellbeing and the Youth Coalition of the ACT (peak body for youth affairs in the ACT) to ensure that they covered key issues for website design and were appropriate and accessible for young people. Participants were asked to share their preferences through a combination of close-ended and free-text responses.

**Procedure**

Because of the COVID-19 pandemic, all focus groups were conducted online in June 2020 using the Zoom videoconferencing platform (Zoom Video Communications). Given that this study was conducted in partnership with the ACT Office for Mental Health and Wellbeing, all focus groups were cofacilitated by a member of the research team and the ACT Office for Mental Health and Wellbeing. A second member of the research team was also present at all focus group sessions. Their role was to take notes, assist with linking participants with clinical support (if required), and monitor any discussion via the chat function, which was encouraged. All focus groups ran for approximately 90 minutes, online discussions were audio recorded, and any chat conversations were also saved for inclusion in the analysis. One author (ARM) listened to each focus group recording and took detailed point-by-point notes on the content of participant discussions. These were combined with researcher notes taken during the session and participants’ chat conversations for analysis. Because of this approach, all illustrative quotes were drawn from survey data. Participants were provided with help-seeking resources at the conclusion of each focus group, and clinical support or debriefing was available on request.

Two separate surveys were developed for young people and parents or caregivers and were administered online in June 2020 using Qualtrics online survey software (Qualtrics). Surveys took approximately 20 minutes to complete. All participants were provided with a list of help-seeking resources at the conclusion of the survey.

**Data Analysis**

Qualitative data were managed using the NVivo 12 software (QSR International). A thematic analysis [17,18] was conducted by 1 author (ARM) on focus group data notes, chat conversations, and free-text survey responses to identify and summarize the key topics and preferences raised by parents and young people within each area of questioning, while preserving the breadth and diversity of perspectives presented. ARM is a lived experience academic with personal experience of mental health service use and professional expertise in youth mental health and service evaluation. Analyses and developing categories were regularly discussed with other members of the research team, including researchers and ACT Office for Mental Health and Wellbeing staff members present at the focus groups, to test assumptions and clarify developing categories. A combination of deductive and inductive approaches to coding was applied. Before commencing data analysis, an initial broad coding framework was developed based on the key areas of questioning during the focus groups: content, features, user experience, and design. Focus group data were deductively coded within each category. As coding progressed, subcategories of describing specific areas of preference and new categories not represented by the deductive framework were developed inductively. A detailed coding framework was developed from the focus group data, including the original broad areas of questioning and inductively developed categories and subcategories. This framework was applied to free-text survey responses, adjusting the framework as necessary to adapt to new information. Finally, data were examined across categories to construct common themes in participants’ preferences. The key topic areas were summarized and grouped under 2 broad categories: the most important types of information for a service navigation tool and important website features.

**Results**

**Overview**

The overarching topic areas arising from the focus groups and surveys covered the types of information needs identified by young people and parents and the website features that would be beneficial to support the mental health of young people.
Themes collectively developed from the focus groups and surveys are described within these 2 areas. Data from the focus groups and surveys are integrated throughout this section. Where findings relate only to a specific group (parents or young people) or data collection format (focus group or survey), this is noted in the text.

**Types of Information**

**Up-to-Date Information About Services**

Preferences and needs for mental health service information were key topics of discussion in the focus groups and one of the most frequent topics raised in open-ended survey responses. Participants emphasized that it was important for the information on a mental health service navigation website to be regularly updated, with an indication of when updates last occurred, and for no information about listed services to be missing. Content should be relevant to the local region targeted by the website (ACT and surrounding region), as national-level websites can be difficult to navigate and regional content may better cover the full spectrum of services available in that area.

> Anything that is Canberra specific is great - wait times, contacts, prices, explaining which professionals do what. For general advice there are lots of national mh [mental health] services already. [Young person, survey participant]

> It would need to be regularly updated to match service changes. It would need to be all evidence based. [Parent, survey participant]

Current and accurate information about service wait times was particularly desirable for all participant groups. Families commonly experienced a lag between identifying that a young person had a problem and being able to access support services. Providing wait time information on a service navigation website could change the way families choose to access services and direct them to services with better availability. Knowing about wait times at the beginning of the help-seeking process would also inform young people and families about likely time frames for receiving support and enable them to plan ahead.

> I think the idea of including information about wait list times is very important - as this is usually critical at the time that you are seeking information and support. [Parent, survey participant]

> I think most people are primarily looking for immediate reassurance and advice. If they can get an idea on how long it might take to access different services this also gives them a time-frame so they can plan. [Parent, survey participant]

Young people highlighted the value of a navigation website in providing detailed information about services that would help them make informed choices. Information about accessibility was particularly important for young people. Specifically, cost and parental consent were identified as key barriers to service access, which could be partly addressed by providing detailed cost information (eg, service fees, rebates, and what benefits a child or family may be entitled to), indicating whether parental consent was required to use a service, and noting any relevant legal considerations. Young people also wanted to know if they would be able to physically access a service before they arrived and recommended that the website include information about wheelchair access (eg, presence of ramps and elevators), accessible and gender-neutral toilets, the languages spoken by staff members, whether tailoring was available for people who are vision- or hearing-impaired, and the age range targeted by a service. The content of a service navigation website should also specifically address issues faced by different minority groups, including LGBTIQ+ communities, people with disabilities, and migrant communities.

> Along with a list of mental health services, also ways of accessing them, with or without parental involvement and potential cost. [Young person, survey participant]

*Information specifically geared towards minority groups (BIPOC, LGBTQA+, people with disabilities etc.). Having information about different mental health issues is great but it would be a mistake not to factor in the relationship between mental health and other aspects of people’s identities.* [Young person, survey participant]

Parents wanted a navigation website that provided clear information about the kinds of services available in their area, including government and private services and general practitioners who specialize in youth mental health. The ideal website would provide specific information about how to contact services, referral pathways, referral requirements, and whether new referrals were being accepted. Details about service specializations and appointment options (eg, availability of telehealth appointments) were also indicated as helpful across both participant groups.

> Information about accessing public v [versus] private mental health services and explaining the difference. [Parent, survey participant]

> Make sure info on telehealth is clear on if it’s: via the internet text-based, internet video-call, via the telephone, via text message. Don’t just say “this place offers telehealth services.” [Young person, survey participant]

**How to Seek Help, and What Happens Next?**

Parents emphasized that a service navigation website should include practical information about how to navigate the whole help-seeking process—from recognizing signs and providing support for their child through to good questions to ask a general practitioner when seeking a mental health referral and how to advocate for and manage the ongoing care of their child. Real stories of how other families navigated mental health issues, practical step-by-step instructions, and diagrams were suggested as useful tools to facilitate access to appropriate services. Young people also wanted clear step-by-step instructions describing how to access help, including how to access services without parental involvement. They noted that finding and visiting a new location can be scary. Providing extra information about the service location, including a map, pictures of the building and the front door, and pictures of staff members, could make the process easier. Information about different types of therapies...
and how they work were also considered helpful for young people.

*Lists/help navigating processes - how to get a mental health plan [Mental Health Treatment Plan], how to see a specialist...who to see for gender dysphoria issues etc.* [Parent, survey participant]

*Maybe some little (short time, a few minutes only) video scenarios with intros to the various services to give a sense of what / who is involved, what to expect...to get a sense of what the service is about...* [Parent, survey participant]

Participants noted that help-seeking resources should foster hope, rather than disappointment, if a service access attempt did not work out. For example, a parent who completed the survey suggested that a service navigation website could illustrate what a good mental health service arrangement looks like and let people know it is okay to try different clinicians and services if the first referral was not a good fit. Parents also suggested that a navigation website should include easily accessible information on how to recognize, respond to, and seek urgent help during a crisis situation. One parent suggested including a function that could directly link people to a crisis service if needed (eg, a crisis telephone service like Lifeline); however, other participants felt that most people would already know about commonly advertised crisis hotlines.

*Information on what a good mental health service arrangement should look like - depending on individual need - and it is ok to not just stick with one person/service forever or choose none.* [Parent, survey participant]

*I think you need information clearly on the homepage about crisis care...and other urgent items that you don’t want to have to sift through a website for.* [Parent, survey participant]

**Defining Website Scope**

In all 4 focus groups and the online surveys, participants raised questions about the ideal scope of a service navigation website. Participants tended to agree that the website’s scope should not be too broad. However, there was uncertainty about whether a navigation website should provide information about mental health in general or only provide information about services. Some participants also expressed a preference for a broader focus on health and well-being, rather than limiting information to mental health services. Young people suggested that broader content could include information about career and employment guidance, coping with current events, mental health at school, and quick references for self-management strategies that could be used while waiting for services (eg, mindfulness, distress tolerance, and coping with panic attacks). Parents were interested in content about mental illnesses (including mood, anxiety, and eating disorders), how to recognize them in specific age groups, and information about common comorbidities and related issues like aggression and self-harm.

*This is far too mental illness and mental health service focused than I would be wanting. As a parent it’s helpful to have that information but I would also like resources that are tailor made for the site and have a focus on more on early intervention and mental well-being.* [Parent, survey participant]

Providing links to different websites or existing online programs was suggested as an acceptable option to prevent the scope of a service navigation website from becoming too broad. For example, a navigation website could provide links to existing early intervention and mental well-being resources to support young people’s and parents’ well-being and assist parents to provide support when issues were first identified. Young people suggested providing links to research papers; stories about other people’s lived experience; and different forms of media that represent mental health in a productive way, such as video games, books, and movies. Participants noted that this approach could also connect a website to resources that fall beyond what they would typically define as a “mental health service,” including information about physical health, disability services, drug and alcohol services, and community programs and events that support and empower young people.

*I would like to see it link not just mental health services, but other services to help a child overcome all the problems that may be adding to the mental health issue. For example, if the child is experiencing a lot of pain, if the child needs weight management help, if the child is being bullied or needs to develop resilience, if the child is on the autism spectrum, etc.* [Parent, survey participant]

*Stories from people who struggle with their mental well-being but have found support and renewed belief in themselves.* [Young person, survey participant]

**Website Features**

**Website Design**

Participants generally agreed that website design would be an important element of a service navigation website’s success. Elements of website design highlighted by young people included a quality user interface that was easy to use and attractive to the intended audience. Young people described a well-designed website as colorful (but not too colorful or gimmicky), engaging, private, welcoming, local, and informed by what we know about young people and how they think. A level of seriousness in the design was required to ensure that the website was viewed as a reliable resource. Government-branded websites were viewed as a trustworthy source of local information but could be off-putting for young people who had previous negative experiences with government mental health services. A list of organizations that support the website could also signal the trustworthiness of information. Young people were also very aware of accessibility issues, recommending that a navigation website be designed to work across different platforms, for people with slow internet connections, and that it met relevant accessibility standards (eg, for people with low vision or lower literacy levels). A parent suggested that the name of the website was also important and needed to be inviting, explanatory, and nonstigmatizing.

Easy website navigation was another key design issue. When trying to communicate about one’s own mental health in a
difficult time, participants felt that the most important thing was to access information easily. Participants indicated that they would be more likely to visit a service navigation website for a specific need or to find specific information. The website content should be set out plainly, with clear pathways to the kind of information the user is looking for. Website design should not be overly complicated, avoiding the need to navigate through tabs and the presence of too many distractions (e.g., moving images or videos). Young people were described as multitaskers; thus, a service navigation website needed to capture their attention quickly. To achieve this, it would be helpful to present information clearly and concisely, and to avoid walls of text that could be overwhelming. A frequently asked question section and fact sheets were seen as helpful, but only when presented as a dedicated webpage and not solely as a downloadable document.

Search Engines
Parents and young people agreed that a service navigation website should systematically connect people to relevant services quickly. A good quality search engine and filtering system was a particularly important aspect of young people’s user experiences; the search bar can be the first port of call for young people trying to find help. A good search engine was described as easy to navigate, with tags and search terms updated as service information changes. In 1 focus group, young people suggested that a service navigation website could be designed like a nice online shopping experience, with tabs and subtabs for different categories of services and filters that allow users to refine their search and locate the most relevant services. However, 1 parent survey participant noted that a filter system would make them feel terrible if it indicated that there were no services matching their child’s needs, indicating that a balance between detail and generalizability may be required. Participants suggested that if there were no services meeting a young person’s search criteria, the website could direct them to resources, fact sheets, or other information they may find helpful.

...a search engine that finds services that are relevant to you. E.g. you could put in the tags “stressed” “self harming” “aged 17” “don’t want to involve parents” and the search engine would suggest; things you could do to help yourself, services you could access, and a helpline. [Young person, survey participant] The filter would make me feel terrible if meant that my child had no services available so I think that you need to be careful regarding putting too many. [Parent, survey participant]

Focus group participants suggested having a quiz or questionnaire to help young people and parents navigate the website. For example, a pop-up box could appear when a person first accessed the website with some questions about what visitors are looking for. Answers could direct young people and parents to appropriate website sections, services, or self-help strategies. Parents suggested that a navigation website could include a symptom checklist, providing recommendations on whether a person needs to seek help from a health professional and within what time frame. Participants noted that any quizzes or checklists should be accompanied by a disclaimer stating that the website could not provide a diagnosis, and questions should be symptom or problem based, not diagnosis based. The results should be anonymous unless a young person chose to disclose them. However, some participants thought it would be important to determine if a young person was in immediate danger and requiring assistance. To facilitate help-seeking, a symptom checklist tool would ideally lead to an outcome, such as connecting users with a real person who can assist with identifying an appropriate service or next step.

Wouldn’t it be easier to complete a mental health survey upon entering the site that directs you to all the relevant pages? [Young person, survey participant] Adolescence is a tricky time. Parents don’t know when to worry, when to escalate to professional help, and when to leave kids to muddle through. A quiz that helps navigate that would be super helpful. [Parent, survey participant]

Supported Navigation
Participants were enthusiastic about having the option to contact a person, by phone or text-based chat, to help them gauge the seriousness of their issue, navigate the mental health system, connect with appropriate services, and answer questions about what to expect at an appointment. Parents emphasized the importance of creating a sense of trust, confidence, and reliability when a person makes a connection through a navigation website. Finding the right service could take a lot of time, research, and mistakes, particularly during times of stress. Some participants described help seeking as overwhelming, emphasizing the importance of positive experiences that could renew confidence in seeking help and support future service use. An interaction with a real person could be an opportunity to foster hope, positive regard, empathy, reassurance, and a sense of not being alone.

This is a great idea. I found navigating the system to be impossible to start with. I was googling everything, calling all these people, being passed from service to service and getting nowhere. None of the service providers knew what any of the other service providers could do. Someone to help you navigate that would be amazing - especially if you are out of your mind with worry and sleep deprivation like I was! [Parent, survey participant] When asking questions about mental health problems, some young people reported preferring to talk to a health professional, whereas some parents and young people suggested that this would be a good opportunity for peers to support young people and carers and that young adult peers may be better able to connect with younger website users. However, 1 young person noted that talking to a healthy peer may be intimidating for some young people. Across the surveys and focus groups, participants suggested that the person they contacted should be genuine, engaged, supportive, friendly, empathetic, unhurried, unscripted, and well trained, with appropriate counseling skills to support stressed or distressed callers. Their understanding of
mental health and local health services should be broader than the understanding provided by the caller’s own experiences.

*Having an advisory line (telephone or chat) which is supportive, anonymous if wanted, friendly and unhurried, staffed by a real person who is quietly supportive but knowledgeable would be great to help people navigate the system and get a sense of where they are best placed to use their energies in pursuing or connecting with services.* [Parent, survey participant]

Participants emphasized that the purpose of any phone or text-based contact options and the roles of the people running them need to be very clear. Contact information should be clearly stated and easy to find, and ideally some contact options would be available outside of normal business hours (eg, evenings, nights, and weekends). Participants had concerns around potential privacy issues, particularly related to data collection and storage, and the need to support people who disclose thoughts of suicide and self-harm. Parents recommended that all interactions end with some kind of closure, for example, actions such as making an appointment for the person with an appropriate service or taking a concrete step that progresses the issue, with timely outcomes.

*I’d also be more comfortable in knowing how any conversations via the website were recorded and stored since there is always the chance of personal stuff coming up in conversations, and I wouldn’t want that to be accessible by anyone except those who are directly working to help me.* [Young person, survey participant]

Further, participants discussed the option of receiving a follow-up call or text, after interacting with a person via a service navigation website. Some participants felt that a follow-up call could be comforting for people who were currently on a waitlist, for example, by providing updates on wait times and identifying alternative sources of support. Survey participants were primarily interested in receiving follow-ups related to service access, for example, checking in to see if services had connected with a young person, how effective the service has been, and if the young person’s needs are being met. Although some participants also wanted the opportunity to give feedback about services, some parents felt oversurveyed by mental health services. Other participants felt that a follow-up would not be helpful in all circumstances. For example, 1 parent felt that receiving a phone call may just be more frustrating if accessing services was not going well. Participants agreed that any follow-up from a service navigation website should be opt-in, the user should have control over how and when they are contacted, young people should decide whether their parents are contacted, and any promises made by a navigation service should be honored (ie, call if you say you will call). Choice and control over contact was seen as particularly important for young people living in high-risk situations, where receiving a phone call or message could potentially be unsafe.

*I think having options is good, especially for kids in potentially dangerous situations. And I do think texting is popular with kids.* [Young person, survey participant]

*I think it is also important to see how well people went with actually accessing the services and whether their child’s needs were met and what negatives there were.* [Parent, survey participant]

**Forums, Reviews, and User Accounts**

Website functions that could allow users to share information were met with a mixed reception. The options discussed included forums, service reviews, and user accounts. Forums received the most positive reception, but with important safety considerations for implementation. They were described as a positive tool for young people and parents to connect with peers who have had similar experiences and share coping techniques or to connect with health professionals. However, to be safe and useful, a forum or chat room would require careful moderation. Participants suggested that forums could be provided by a service navigation website itself or the website could instead provide links to external, good quality, moderated mental health social media pages or similar services.

*Particularly when I was younger, online resources were huge! Whether that was headspace’s online counselling, or online chat forums with other people. I think those chats were absolutely fantastic, however they definitely needed expert moderation.* [Young person, survey participant]

*I’m a little concerned about forums and/or chat rooms. They would need to be carefully monitored to make sure there are no trolls responding negatively or people using cyber bullying.* [Parent, survey participant]

User reviews of mental health services were raised as a possibility, but this option had both advantages and disadvantages. Reviews could provide information to help young people choose a suitable service and prepare for their own visit. However, mental health service needs and preferences were seen as highly individual; thus reviews may deter young people from accessing services that would actually suit them. One young person suggested that a government “check mark” (ie, accreditation), indicating that services were legitimate, could be an alternative to reviews, and a parent suggested including a feature where young people and parents could post questions to be answered by a website staff member.

User accounts were the least desirable function of a service navigation website. Young people’s responses to implementing user accounts were overwhelmingly negative, due to concerns around privacy, security, safety, stigma, and limiting access to website features for people without accounts. Young people thought their peers may also associate accounts with costs like subscription fees. A small number of survey participants thought it would be helpful to have a single digital record of the services they had tried and a “wish list” of services they would like to try in the future. However, most parents and young people were uncomfortable with the idea of their mental health information being stored online and believed that this could create a barrier to people using a service navigation website. Young people
emphasized the need to be able to access a service navigation website anonymously, particularly when living in high-risk situations where internet use was monitored, and some felt that an optional user account would overcomplicate a navigation website. Young people suggested a range of other, less invasive, ways to tailor user experiences, including features that made it easy to save pages to favorites in a browser, copy content to the clipboard, email a link to service contact information, and save records of service contact information into another application (eg, into the Notes or Photos app on a phone).

A record of interactions makes me feel insecure, because I know the website is keeping data on me. This is not just “not helpful,” but distinctly unhelpful. [Young person, survey participant]

I would be concerned about a portal where you sign in. While it might be good to have information all in one place, really mental health information belongs in a doctor’s office. If your child is displaying mental health problems, you don’t really want that recorded when you don’t know how that information will be used into the future. [Parent, survey participant]

Discussion

We conducted focus groups and online surveys to inform the design of a mental health services navigation website. Although the research activity was conducted with a specific region in mind, the findings may assist other organizations designing websites or apps to assist young people and parents in navigating mental health systems. Many of the themes represent common issues faced by young people in need of mental health support.

Participants emphasized the need for, and importance of, up-to-date and accurate information about local mental health services (public and private) and guidance on how to access them (including referral pathways). Participants were aware of existing lists and directories of services, but these were described as incomplete, out of date, and difficult to search. Navigation tools were seen to be most helpful if they could provide local, tailored information, including service information that could not be accessed elsewhere. Transparent information about service cost, wait times, and how to access services without parental permission was in particularly high demand. This information could support young people and parents to make informed choices about which services to select and pursue. Future research exploring the specific service information that is needed to adequately meet end-user needs and that translates into actual service contacts would be beneficial.

Similarly to previous research, participants also emphasized the need for information about referral pathways and how to navigate the mental health system [9,19]. Parents and caregivers shared that they often acted like case managers and advocates for their children, and they wanted access to information and resources that could support them. Step-by-step instructions, flow charts, and real-life stories of accessing mental health care, tailored to the local system, could all support young people and their parents on their journeys from first seeking help to accessing specialist services [19].

An effective search engine was an important aspect of the navigation tool’s design. All information within the navigation tool needs to be searchable, and users should be able to refine their search results using relevant filters. Young people and parents felt that they should be able to independently find service information more easily than they could with a web search engine and that a navigation tool should be broadly accessible, functional, easy to navigate, and tailored to its audience in content and appearance. This finding highlights again the importance of the co-design process, and end-user testing, to ensure that the tool developed is fit for purpose and fully meets the needs of those who will use it [15]. Relatedly, digital tools should also be designed in accordance with web accessibility standards to ensure that all users can effectively use them, including those with disabilities [20].

Participants also felt that it would also be helpful for the navigation tool to provide a phone, text, and/or web chat service. Specifically, participants wanted a phone line or web chat option that could help people to identify relevant and appropriate mental health services and support them to decide which service to contact first. Including this feature would reduce the burden on young people and parents to identify, research, and select services alone. Offering an option to receive a follow-up call or text from a phone line or web chat may also be beneficial.

Underlying many of these findings was the need to have control over the help-seeking process. Control over how and when to interact with content is an important concern for young people in the design of online mental health interventions [16,21]. Providing choices was important to participants in our study; both choice in how they could contact a support person through the navigation tool (eg, phone call, web chat, SMS text messaging, or email) and choice in if, how, and when they were contacted by the navigation tool or engaged with a service.

Other potential navigation tool elements, such as being able to create a user account, were not as desirable. Storing records of health information online, particularly mental health information, raised serious concerns around privacy. Generally, participants felt that people would primarily use the navigation tool anonymously and independently. This aligns with previous findings from the development of mental health interventions for young people and adults, which highlighted confidentiality, privacy, and trust in the organization delivering the service as key areas of importance for end users [21-23]. Some people felt that it could be helpful to be able to record their activity (eg, service wish lists) while using the navigation tool and/or service use in a single location, but participants emphasized that user account features should be opt-in. Such features may be off-putting for potential users.

The involvement of end users in the design of the navigation tool was highly valued by the commissioners of this research and resulted in the development of a youth navigation website called MindMap [24], which captures the key elements identified by participants. The findings of this study provided formative information for the development of MindMap, although there was additional development and testing involved to create the final website, which is beyond the scope of this paper. MindMap is an accessible web-based tool that provides a comprehensive
searchable database of local services and provides a clear description of the service, its location, and potential wait times. It is an initiative of the ACT Office of Mental Health and Wellbeing and delivered by a nongovernmental organization with strong connections into the local mental health sector, enabling frequent refreshing of service information. Users of MindMap can use the navigation tool independently or receive navigation support during the week or on weekends from members of the MindMap team via telephone, email, or web chat. Young people and parents or caregivers were involved in the iterative development and testing of MindMap to ensure that it continued to meet their needs.

There are some limitations to this research that should also be considered. First, the participants in the focus groups and survey may not have been representative of all young people and parents or caregivers in the community, and the study may have attracted people with a greater interest in mental health. Preferences for website appearance, content, and features may have varied by age and personal experience of mental health issues [16,25]. Future research would benefit from sampling more young men, and young people and parent or caregivers without mental health service experience, to ensure that the needs and preferences of all targets are adequately captured and met.

The scope of the questions in the focus groups may also have guided the discussion, placing more emphasis on the areas covered by the questions and consequently may have missed other issues. This issue was partly mitigated by providing time within focus groups for participants to identify issues not covered by the questions. The timeline of the project necessitated the use of point-by-point note-taking from focus group recordings, rather than verbatim transcripts, for data analysis. We acknowledge that this approach may have a higher risk of introducing research bias; however, this was mitigated through regular discussion of the analysis with the research team and the inclusion of qualitative survey data in the analysis. Lastly, the research was conducted with a specific region in mind, and thus not all findings may be relevant to other contexts.

Overall, this study provides important insights into the navigation needs of young people and their parents or caregivers seeking mental health services and how best to support them in this process. The focus groups and surveys identified the need for tailored local information, the provision of up-to-date service details, and the opportunity for users to navigate the site independently or with support. Ensuring that young people and their parents or caregivers can access mental health services in an efficient and timely manner is essential to the longer-term health and well-being of young people. Future research assessing the effectiveness of navigation tools in meeting this goal should be strongly encouraged.

Acknowledgments

We would like to thank and acknowledge the young people, parents, and carers who participated in this research. This research was funded by the Office for Mental Health and Wellbeing, ACT Health. ALC is supported by a National Health and Medical Research Council (NHMRC) fellowship (1173146). ARM is supported by a Suicide Prevention Australia Post-Doctoral fellowship.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ALC and PJB led the design of the study, with input from all other authors. ALC, SMM, EM, NJ, and ARM contributed to data collection. ARM and MB conducted the analyses. ALC and ARM drafted the paper. All the authors critically reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Focus group questions and prompts.

[DOCX File, 11 KB - formative_v8i1e48945_app1.docx ]

References


Abbreviations

ACT: Australian Capital Territory
LGBTIQ+: lesbian, gay, bisexual, transgender, intersex, queer
Development and Usability Testing of an mHealth Tool for Trauma-Informed Prevention of Substance Use, HIV Acquisition, and Risky Sexual Behaviors Among Adolescents: Mixed Methods Study

Carla Kmett Danielson¹, PhD; Angela Moreland¹, PhD; Austin Hahn¹, PhD; Devin Banks², PhD; Kenneth J Ruggiero³, PhD

¹Department of Psychiatry & Behavioral Science, Medical University of South Carolina, Charleston, SC, United States
²Department of Psychology, University of Missouri-Saint Louis, Saint Louis, MO, United States
³College of Nursing, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Carla Kmett Danielson, PhD
Department of Psychiatry & Behavioral Science
Medical University of South Carolina
67 President Street
Charleston, SC, 29425
United States
Phone: 1 8437922945
Email: danielso@musc.edu

Abstract

Background: Youth who experience traumatic events are at a substantially higher risk of engaging in substance use and sexual risk behaviors and problems (eg, HIV acquisition) than their non–trauma-exposed counterparts. Evidence-based substance use and risky sexual behavior prevention may reduce the risk of these outcomes. Trauma-focused mental health treatment provides a window of opportunity for the implementation of such preventive work with these youth. However, overburdened clinicians face challenges in adding prevention content while implementing evidence-based treatments. Mobile health (mHealth) tools can help reduce this burden in delivering prevention curricula. Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior (TIPS) is an mHealth app that was developed to aid trauma-focused cognitive behavioral therapy (TF-CBT) clinicians in the implementation of an evidence-based risk behavior prevention curriculum.

Objective: The goal of this paper is to describe the rationale for and development of the TIPS app and present the results of a mixed methods approach for the initial evaluation of its usability.

Methods: Participants included clinicians (n=11), adolescents (n=11), and caregivers (n=10) who completed qualitative interviews and an adapted version of the Website Analysis and Measurement Inventory.

Results: In total, 4 overarching themes emerged from the participants’ answers to the qualitative interview questions, demonstrating a generally positive response to the app. The themes were (1) strength of app content, (2) suggestions about app content, (3) esthetics and usability, and (4) benefits to the patient and session implementation. Clinicians, adolescents, and caregivers all agreed that the content was very relevant to adolescents and used examples and language that adolescents could relate to. All 3 groups also discussed that the content was comprehensive and addressed issues often faced by adolescents. All 3 groups of users made suggestions about the esthetics, which mostly comprised suggestions to change the font, color, or pictures within the app. Of all the groups, adolescents were most positive about the esthetics and usability of the app. Results from the Website Analysis and Measurement Inventory further illustrated the users’ favorable reaction to the TIPS app, with 100% (11/11) of clinicians, 100% (10/10) of caregivers, and most adolescents (7/11, 64%) selecting strongly agree or somewhat agree to the following statement: “This app has much that is of interest to me.” Adolescents generally found the app easier to use than did caregivers and clinicians.

Conclusions: The TIPS app shows promise as an mHealth tool for TF-CBT clinicians to integrate evidence-based substance use, risky sexual behavior, and HIV prevention during treatment. Future research, including a randomized controlled trial comparing
TF-CBT implementation with and without the inclusion of the app, is necessary to evaluate the feasibility and efficacy of the app in reducing the risk of substance use and risky sexual behavior among trauma-exposed adolescents.

**Trial Registration:** ClinicalTrials.gov NCT03710720; https://clinicaltrials.gov/study/NCT03710720

**KEYWORDS**
traumatic stress; prevention; substance use; HIV; qualitative methods; adolescents; mobile phone

**Introduction**

**Background**

Children and adolescents who experience traumatic events are at a substantially higher risk of engaging in sexual risk behaviors and substance use than their non-trauma-exposed counterparts [1,2]. Early traumatic experiences also have long-term effects on behavioral health as youth exposed to trauma are more likely than their nonexposed peers to develop substance use disorders [3,4] and experience unexpected pregnancy and sexually transmitted infections (STIs) [5,6], including increased risk of HIV acquisition [7,8]. Models addressing the link between trauma exposure and risk behavior have suggested that, although these behaviors are multidetermined, trauma exposure plays an important role by manifesting impairments in affect regulation, impulse control, identity development, and socialization [2,9].

Given the strong relationship between trauma exposure and problematic behavioral health outcomes, preventing morbidity related to substance use and sexual risk behavior may be best addressed from a trauma-informed lens. That is, an important prevention approach may involve addressing risk behavior in the context of trauma-focused assessment and treatment. However, data suggest that clinicians are often reluctant to integrate interventions related to substance use in the context of trauma treatment [10] because of clinical, systemic, and training barriers and limited resources (including limited support) relative to caseload demands [10-13]. Given the need for substance use and risky sexual behavior prevention for trauma-exposed adolescents—and to combat these barriers and demands for time and resources—clinicians may benefit from a structured tool to help efficiently implement evidence-based prevention strategies for risk behaviors common among trauma-exposed youth [14]. Such a tool may be best delivered via a mobile health (mHealth) approach, which can reduce the need for extensive clinician training in multiple new prevention intervention curricula, augment the formal training they do receive, and enhance or extend the effectiveness of the traditional clinical encounter [15,16]. As such, clinicians would be significantly more likely to deliver substance use prevention if a ready-made mHealth tool were available that helped them deliver it efficiently and accurately. This study describes the development and feasibility evaluation of an mHealth app designed to supplement trauma treatment among adolescents, targeting prevention of substance use, sexual risk behavior, and associated health consequences (eg, HIV).

**Adolescent Trauma Treatment as an Opportunity for Prevention**

The gold-standard treatment for addressing adolescent trauma is trauma-focused cognitive behavioral therapy (TF-CBT) [17,18]. With >20 completed randomized controlled trials (RCTs)—including international studies—supporting its effectiveness in addressing a range of mental health problems and improving functioning among trauma-exposed youth, TF-CBT has received the highest ranking for empirical support from professional organizations and federal agencies [19-38]. TF-CBT has achieved widespread dissemination through numerous implementation efforts, including comprehensive training programs (eg, community-based learned collaboratives), educational materials, and collaborations with mental health organizations, ensuring its accessibility and use by therapists and clinicians across various regions and populations. In fact, the widespread utility of TF-CBT is exemplified by the staggering number of clinicians who have accessed TF-CBT Web 1.0 and 2.0 (original and updated versions of the TF-CBT web-based learning course); as of August 31, 2023, a total of 496,061 clinicians worldwide are registered users of the web-based training in the TF-CBT model, and 305,120 clinicians have completed the training [39]. Modules of TF-CBT, including psychoeducation and enhancing safety, may include general psychoeducation about sex and sexual revictimization risk reduction; however, information and skill development specific to HIV and STIs, pregnancy, and healthy dating and sexual decision-making are not detailed in the model manual or training. Similarly, skill development for preventing substance use problems that extends beyond psychoeducation and helps translate this knowledge into skills is not formally or typically incorporated into TF-CBT, limiting the potential impact of these psychoeducation modules. In other words, these clinicians do not systematically receive the training and support needed to feel confident in their delivery of psychoeducation content or in the ways to translate this education into skills (eg, realistic refusal skills) [10].

Although clinicians can informally incorporate risk behavior topics into their implementation, the data suggest that this is uncommon. A national survey of mental health clinicians found that providers feel ill-equipped to address topics of substance use and sexual risk behavior when treating trauma-exposed adolescents even when trained in trauma-informed treatment models [10]. Most clinicians did not report receiving formal training for addressing substance use disorder (54%) or sexual risk behavior (67%), suggesting that this reluctance may be related to deficits in training on these topics [10].

These data reflect a global lack of training in evidence-based practice for adolescent prevention, which results in limited translation and
accessibility despite the availability of numerous efficacious preventive interventions [40]. In summary, although effective sexual risk and substance use prevention interventions are available, there is a significant gap in the implementation of these interventions even among those at high risk of such behaviors because of trauma exposure.

**mHealth as a Viable Approach to Address Prevention**

The availability of mobile technology has increased dramatically over the past decade, with 85% of Americans reporting smartphone ownership in 2022 compared with 35% in 2011 [41,42] and 95% of American adolescents reporting smartphone access [43,44]. As mobile technology has rapidly developed, so have health care approaches that leverage mHealth—the use and development of mobile technology, including mobile apps, to improve health care [16,45]. Within mental and behavioral health care, mHealth interventions are diverse, targeting different stages of treatment from education and engagement to the maintenance of treatment gains [16]. mHealth administrations are also diverse and may be stand-alone, client-led interventions or supplements to traditional clinician-facing treatments [16].

Most of the extant mHealth approaches to behavioral health among adolescents are stand-alone interventions rather than supplements to face-to-face treatment [46]. Several stand-alone treatments have shown strong feasibility and efficacy for both primary and secondary prevention of substance use and sexual risk behavior [47-50]. However, most stand-alone behavioral interventions for adolescents lack a theoretical framework in their design and show inconsistent efficacy [46,51]. Furthermore, although data suggest that digital mental health tools delivered in real-world contexts are more likely to be accessed than professional services, they are less likely to be delivered with a sufficient therapeutic dose [52]. These limitations are compounded by clinician- and patient-reported barriers to stand-alone treatments, including the lack of personalization [51,53], lack of privacy and security with regard to sensitive behavioral topics [53,54], limited follow-up [51], and attrition and low completion rates [54,55]. Researchers have suggested that these barriers could be overcome with more support and involvement from clinicians [53]. Accordingly, behavioral health treatments that blend face-to-face and technology-based approaches have been found to save clinician time, demonstrate lower dropout rates, and lead to better treatment outcomes among adolescents and young adults [56,57].

With regard to trauma treatment, 3 popular mHealth apps have emerged for use with adults: Posttraumatic Stress Disorder (PTSD) Coach, Cognitive Processing Therapy Coach, and Prolonged Exposure Coach [58]. The most widely used of these is PTSD Coach, a stand-alone treatment that suffers from limitations similar to those mentioned previously [58], including high attrition and inconsistent efficacy [59]. Conversely, Cognitive Processing Therapy Coach and Prolonged Exposure Coach are supplementary to existing evidence-based treatments for PTSD. Although these apps have not been evaluated for efficacy via RCTs, clinician perceptions of such apps have been favorable, particularly regarding their relative advantage over exclusively face-to-face practices and compatibility with clinicians’ needs [60]. In addition, supplementary mHealth approaches have demonstrated effectiveness in the treatment of comorbid trauma conditions (eg, panic disorder, anxiety, and depression) [61], including comorbid PTSD and substance use [62].

Despite evidence of the feasibility and effectiveness of mHealth interventions in trauma treatment among adults, mHealth approaches to trauma treatment among adolescents are limited. In total, 3 stand-alone prevention approaches for adolescents exposed to acute trauma have demonstrated effectiveness in reducing persistent trauma symptoms, depressive symptoms, and behavioral problems [63-65]. Although no studies have examined the effectiveness of supplementary mHealth approaches for adolescent trauma treatment, the perspectives of trauma-focused clinicians indicate that these approaches would be feasible and useful. For example, Oreno-Aguayo et al [66] found that 96% to 100% of surveyed TF-CBT providers reported that an mHealth supplement would be helpful for enhancing TF-CBT components, extending coping skill development, and improving out-of-session practice among adolescents and families. Clinicians, patients, and families have also responded favorably to pilot versions of a supplemental Apple iPad–based app designed to improve patient engagement and provider fidelity in-session during TF-CBT [67,68]. Thus, mHealth supplements to adolescent trauma treatment may be a feasible and acceptable way to extend and improve behavioral health prevention among adolescents.

**This Study**

The purpose of this paper is to report on the development and perceived usability of Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior (TIPS), a novel mHealth tablet-based app designed to supplement trauma treatment targeting the prevention of adolescent risk behavior (trial registration: ClinicalTrials.gov NCT03710720). We used a mixed methods approach based on qualitative interviews and quantitative ratings to assess the usability of the TIPS app with clinicians, adolescent patients, and their caregivers engaged in TF-CBT in a community-based outpatient clinic.

**Methods**

**TIPS App Structure**

The first author, a national trainer in the TF-CBT model, led a small team of TF-CBT clinicians and trainees in the development of the content of the TIPS mHealth app. This included 7 total topics (Figure 1) designed to be used as psychoeducational tools for TF-CBT clinicians to implement with adolescents and caregivers throughout the TF-CBT treatment process. The tool helps clinicians assess their clients’ and caregivers’ current knowledge and comfortability surrounding topics related to risky sexual behaviors, STIs, drug use, and healthy relationships. Thus, there are 3 intended users of the app: clinicians, adolescents, and caregivers. Each user sets up a unique log-in, and the content displayed on the app is tailored to the user type. For example, Family Check-Up example videos demonstrating parenting skills are displayed for the caregiver but not for the youth, and a sexting decision-making activity is displayed for the youth but not for...
the caregivers; clinicians view both caregiver and youth content but also have unique introduction videos for each app topic tailored to a clinician audience and have additional drop-down menus on the app, such as suggested homework that the clinician can assign.

**Figure 1.** Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior (TIPS) home screen (after log-in) listing the 7 topics and activities.

In particular, the app was designed to map onto the psychoeducation, parenting, and enhancing safety treatment components of TF-CBT. Each topic of the app is introduced by a short video that explains the aims of the section as well as the directions needed to successfully complete the section. After the completion of each component activity (briefly described in this section), users are presented with feedback relevant to their performance on the current activity as well as resources (links to websites) to help further their education on the topic. Specifically, the **What Do You Know?** component is formatted as a multiple-choice trivia game that serves as a psychoeducational tool to educate trauma-exposed teenagers and their caregivers on the effects of drugs and alcohol. The **Myth Busters** component educates users on the common facts and myths about HIV and other STIs. This section was designed as a drag-and-drop game and includes a video that demonstrates how to perform an at-home HIV test kit. The inclusion of the video aimed to reduce negative connotations that may currently exist related to getting tested for HIV (eg, scary, involving needles, and time-consuming). The **Hot or Not** component was designed as a psychoeducational tool to aid adolescents in recognizing unhealthy behaviors within romantic relationships. This section comprises engaging videos, multiple-choice trivia, and feedback that is presented after response submission. A main goal was to highlight warning signs that may precede more obvious unhealthy behavior as trauma-exposed teenagers may be more likely to stay in unhealthy romantic relationships. **Make a Play** helps guide adolescents through tough yet common situations that teenagers may experience. This component consists of 5 activities that educate users on choices involving sending nude pictures, having unprotected sex, consuming alcohol, and being offered different types of drugs. These activities are presented as choose-your-own-adventure screenplays that allow adolescents to make different choices for each situation that is presented and view the results for their different decisions. **Figure 2** illustrates an adolescent presented with the choice to accept or decline the offer of marijuana by a peer. The **Family** topic of the app presents the empirically supported **Family Check-Up** [69] in an engaging way, educating caregivers on the different components involved in positive parenting (eg, communication, encouragement, and supervision). This section consists of videos, text information, and quizzes. **Cyber Life** was created to aid in the education regarding safe web behaviors and choices. This component presents users with possible situations that they may encounter on the web and the choices they can make in a multiple-choice format. Each choice is followed by feedback on its positive aspects and the risks involved. The **Vision 25** topic is to be used as a guide in goal setting. This section helps users think about their current goals, what choices they can make to help them successfully accomplish their goals, and what choices they can make that may result in their goals becoming harder to reach. This section presents users with a road map (**Figure 3**) where they can choose different ages at which to set goals, providing guidance for both short- and long-term goals.
The TIPS platform was developed using AppBuilder, a content management system developed by our digital health team under our institution’s Clinical and Translational Science Award. AppBuilder includes a wide array of design templates and features to ensure that native iOS and Android apps can be created by researchers, program staff, and other team members who have no formal coding experience. Investigators and innovators use it to wireframe, pilot test, change and add content, and launch and evaluate mHealth apps without the extensive involvement of a technical development team. Mobile app developers often become involved in AppBuilder-based initiatives only on a strategic, limited basis, consistent with our goal of significantly shortening the timeline and costs associated with building mHealth apps.

**Usability Testing Overview**

Usability testing is incorporated into technology development to improve user experience by measuring whether the user can successfully and effectively use the tool. Usability testing incorporating both qualitative and quantitative methods can also help identify barriers to task completion and examine areas that take the user off topic, create confusion, or decrease satisfaction [70]. The purpose of usability testing in this study was to obtain objective metrics and refine the final TIPS product before formal efficacy testing.

**Ethical Considerations**

All procedures were approved by the Medical University of South Carolina Institutional Review Board (Pro00041527). Participants (ie, clinicians, caregivers, and teenagers) were given an information sheet that detailed the purpose of the research; procedures; risks, discomforts and benefits; costs and compensation; alternatives; and confidentiality. Participants provided verbal consent in lieu of written consent. Participant data were deidentified to preserve confidentiality. In addition,
participants were compensated for their participation with US $30 in the form of a gift card, cash, or money order.

**Participants**

Participants included mental health clinicians (11/32, 34%; 26/32, 82% female), trauma-exposed adolescents (11/32, 34%; 26/32, 82% female), and caregivers of adolescents (10/32, 31%; 25/32, 80% female). More specifically, with regard to inclusion criteria, clinicians were master’s or doctoral-level mental health providers in the area local to the study who were fully trained and experienced in TF-CBT and carried active child trauma caseloads. Adolescents were aged between 13 and 18 years, had experienced at least one traumatic event, were in the process of completing or had recently completed TF-CBT, and assented to participate (with caregiver consent). To be included, caregivers needed to have served for at least the previous 2 months in the role of primary caregiver of a teenager in treatment for PTSD who was in the process of completing or had recently completed TF-CBT. Among the clinicians, 72% (8/11) were aged between 25 and 34 years, whereas 27% (3/11) were aged between 35 and 44 years. The adolescents ranged in age from 13 to 18 (mean 15.25, SD 1.90) years, with most (9/11, 82%) falling between the ages of 13 and 17 years. Finally, caregivers ranged in age from 34 to 53 (mean 44.2, SD 6.65) years.

**Procedures**

Participants (clinicians, adolescents, and caregivers) were recruited from community-based mental health centers that specialized in the treatment of traumatic stress and served multiple counties (including urban and rural areas) in a large city in the Southeast United States. There were 2 primary methods through which participants were recruited: (1) flyers were posted in and around the clinics with contact information for the study project coordinator and (2) potential participants were informed by clinic staff that they may qualify for a study and asked whether they would be interested in learning more. When potential participants responded positively, the study coordinator contacted them and provided a full description of the study, screened them to ensure that the inclusion criteria were met (see the previous section), and consented them to participate in a qualitative interview regarding the TIPS app and completion of the Website Analysis and Measurement Inventory (WAMMI; see the following section).

**Measures**

The semistructured qualitative interview consisted of a study team member providing the participant with an iPad loaded with the TIPS app and walking the participant through each section. Following each section of the app, the interviewer asked several open-ended questions (eg, “First, tell me, when did you first hear of the app?” “What do you think of the app?” “What don’t you like about the app?” and “What changes would you like to see?”). Follow-up probes were used to clarify the information provided whenever necessary.

The WAMMI [71] is a standardized 20-item assessment measure that captures users’ personal opinions on a given website’s ease of use. The items were slightly revised to refer to the application rather than the website. In the measure, users are asked to rate various aspects of their experience with the app (eg, content, navigation, and design) on a 5-point Likert scale from strongly agree to strongly disagree. Items are then scored to produce 5 subscales measuring the app’s attractiveness (level of visual interest of the app in terms of both function and information provision), control (app navigation ease), learning (users’ ability to easily understand the content of the app and learn what they expect to learn), helpfulness (usefulness and expected content and structure of the app), and efficiency (users’ ability to quickly find and do what interests them on the app). The WAMMI was developed using latent variable analysis, has high reliability, and reports standardized scores (eg, 50 = average; 100 = perfect) for the 5 aforementioned themes based on a reference database [72].

**Data Analysis Plan**

**Quantitative Approach**

All descriptive analyses were performed using Stata (version 17; StataCorp LLC) [73]. The data were screened for outliers and impossible values. Group differences in the WAMMI items were compared using 2-tailed paired t tests.

**Qualitative Approach**

Data analysis consisted of a qualitative content analysis [74] informed by grounded theory [75], which is used to explore participants’ unique perspectives via the identification of themes and patterns that naturally emerge from the data and the systematic classification of these themes. Specifically, a 3-step inductive approach was used, which involves collecting and analyzing data without preconceived categories or theories. To analyze using this approach, each participant’s interview responses (ie, raw data) were carefully examined to develop a comprehensive codebook to capture all possible themes emerging from the data. The codebook was then used by 2 independent coders to code and analyze each participant’s responses to the interview questions [74,76]. Coders were able to apply more than one code to the participant responses if applicable. The interrater reliability for the double-coded interview responses was 86% and ranged from 82% to 93%. Interrater discrepancies were discussed and resolved by the 2 independent coders. Finally, themes were refined, merged, or divided into subthemes via collaborative discussions in multiple in-person meetings until a comprehensive codebook was developed. The NVivo software (version 11.1; Lumivero) was used for data management and analysis. The interviews were approximately 45 minutes in length and were audio recorded and transcribed.

**Results**

**Technology and App Use Descriptive Information**

Clinicians answered some descriptive questions about comfort using smartphones and apps and the benefits and drawbacks of using websites or apps in treatment. All clinicians (10/10, 100%) reported (1) being comfortable using smartphones and tablets and (2) that the primary benefit of using websites and mHealth app tools is that they engage teenagers in treatment and that they are readily accessible. Other benefits reported by clinicians
included that apps are free and that teenagers feel comfortable using them. Drawbacks reported by clinicians included that it is difficult to get teenagers to stop using apps and that not all homes have Wi-Fi access.

Adolescents and caregivers were asked about their use of tablets and cell phones. Most adolescents (8/11, 73%) and caregivers (8/10, 80%) reported personally owning a tablet, and all adolescents (11/11, 100%) and caregivers (10/10, 100%) reported owning a cell phone. The primary uses of tablets by adolescents included accessing social media, playing games, watching television shows or movies, SMS text messaging, and completing schoolwork. Adolescents reported that they primarily used their cell phones for talking to friends, SMS text messaging, accessing social media, playing games, and listening to music. Caregivers reported that they primarily used their tablets for playing games, paying bills, surfing the internet, watching television shows and movies, school and work, accessing social media, and checking email. Caregivers primarily used their cell phones for talking to people, SMS text messaging, accessing social media, and surfing the internet.

**Qualitative Results**

**Overview**

Through the individual interviews, valuable information about the usability and perceived effectiveness of TIPS was obtained from clinicians, adolescents, and caregivers. Four overarching themes, each with its own subthemes, emerged from the participants’ answers to the interview questions: (1) strength of app content, (2) suggestions about app content, (3) esthetics and usability, and (4) benefits to the patient and session implementation. Table 1 shows the app themes that emerged. Each is described in greater detail in the following sections, with representative quotes provided throughout for illustrative purposes.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Responses, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strengths of app content</strong></td>
<td></td>
</tr>
<tr>
<td>Comments regarding appropriateness of content for adolescents</td>
<td>27 (13.9)</td>
</tr>
<tr>
<td>Content of modules</td>
<td>10 (5.2)</td>
</tr>
<tr>
<td>Pictures or videos</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Content is entertaining or would keep adolescents’ attention</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td><strong>Changes in or suggestions for app content</strong></td>
<td></td>
</tr>
<tr>
<td>Make content more engaging</td>
<td>17 (8.8)</td>
</tr>
<tr>
<td>Revise the video or picture content</td>
<td>9 (4.6)</td>
</tr>
<tr>
<td>Suggestions for content topics to add</td>
<td>20 (10.3)</td>
</tr>
<tr>
<td>Revise the language to make it more relatable to teenagers</td>
<td>22 (11.3)</td>
</tr>
<tr>
<td>Allow content to be more individualized</td>
<td>15 (7.7)</td>
</tr>
<tr>
<td><strong>Esthetics and usability</strong></td>
<td></td>
</tr>
<tr>
<td>Strengths regarding esthetics of the app</td>
<td>8 (4.1)</td>
</tr>
<tr>
<td>Suggestions for esthetics of the app</td>
<td>17 (8.8)</td>
</tr>
<tr>
<td>Dislikes regarding esthetics of the app</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td><strong>Benefits to the patient and session implementation</strong></td>
<td></td>
</tr>
<tr>
<td>Increases comfort of the adolescent in the session</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td>Would be more likely to use the TIPS app</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

*Themes that emerged from the responses of all users during the qualitative interviews (32 participants and 194 responses).*

**Theme 1: Strengths in the App Content**

A total of 62% (20/32) of the participants discussed the strengths of the content of the TIPS app. Most strengths mentioned within this theme included comments about the appropriateness of the content for adolescents (21/32, 66%), the content of the modules (17/32, 53%), pictures or videos (13/32, 41%), or that the content was entertaining and would keep an adolescent’s attention (11/32, 34%). Other comments centered on language content, diversity of content, positive framing of information, or that the content differed by activity. Clinicians, adolescents, and caregivers all agreed that the content was very relevant to adolescents and used examples and language that adolescents could relate to. All 3 groups also discussed that the content was comprehensive and addressed issues often faced by adolescents. For example, clinicians commented the following:

*I really like the feedback sessions when you give the examples. They’re very detailed, but not overly detailed.*

*I think the language is really accessible to most kids.*
Examples of strengths stated by adolescents include the following:

The videos are funny. I definitely like those.
I actually think that the text itself, that was pretty cool how y’all made it like the iPhone with all the emojis and stuff. And it is kind of relatable to people my age, at least, text. And the responses to it were...like sending the cat picture, that’s definitely something people would do if they don’t feel comfortable and they’re trying to make it funny.

Strengths mentioned by caregivers included the following:

I think the Let’s talk about Sext [video] needs to come as soon as you put the phone in their hand. Plain and simple, right off the rail. It’s not like there’s going to be a certain age, it’s like having the talk about the birds and the bees.

Man, I liked the points they’re making. You know, I think they’re absolutely making the right points with, you know, being present, you know, when your kids are around. Knowing your kids’ friends, knowing where they are, surprise phone calls. I like all that stuff.

**Theme 2: Changes in or Suggestions for App Content**

A total of 59% (19/32) of the participants discussed changes or suggestions regarding the TIPS app content. Most comments within this theme included making content more engaging (22/32, 69%), revising the video or picture content (22/32, 69%), suggestions for content topics to add (17/32, 53%), revising the language to make it more relatable to teenagers (15/32, 47%), and allowing the content to be more individualized (8/32, 25%).

Other comments included language suggestions, adding more outside resources, increasing cultural sensitivity, clarifying content, improving instructions, and removing potentially triggering content. Although the specific suggestions differed among the clinicians, adolescents, and caregivers, the feedback overlapped in that adding more interaction and individualizing some of the content would help improve the app. Some examples of comments made by clinicians include the following:

I’ve seen a lot of kids get through school that can’t read, so try to look at all factors. Maybe some more visual aids along with the words.

Some statements made by adolescents included the following:

Add a game or something. I don’t know what type of game, but a game that you can play for the answer.
Maybe even include something about rape in a relationship, because a lot of people that I’ve talked to, they don’t think that you can be raped in a relationship.

Comments made by caregivers included the following:

You know, make it a little bit more interesting. If it could be made more interactive.
A lot of kids don’t have parents. They’re already living with a family member, or in foster care. Another line to put in there somewhere is, talk to someone you trust...doesn’t have to be a parent.

**Theme 3: Esthetics and Usability**

A total of 25% (8/32) of the participants discussed the esthetics and usability of the TIPS app. The most common comments within this theme involved strengths regarding the esthetics of the app (15/32, 47%), suggestions for the esthetics of the app (13/32, 41%), and dislikes regarding the esthetics of the app (9/32, 28%).

Participants also made a few comments about usability, including suggestions and strengths regarding the functionality of the app. All 3 of the groups (clinicians, adolescents, and caregivers) made suggestions about the esthetics, which mostly comprised suggestions to change the font, color, or pictures within the app. Of all the groups, adolescents were the most positive about the esthetics and usability of the app. Some specific comments made by clinicians included the following:

Maybe use highlight and change the color or something to make it clear. I think that the progress bar at the bottom is so simple. It stays out of the way, so you’re not sacrificing real estate.

I like the progress bar on the bottom. It’s a nice touch because then I don’t have to keep wondering how many more questions. I can see that I’m almost done, so that helps me not get frustrated, especially for a teen.

Some examples of statements made by adolescents include the following:

The activities are very easy to operate.
I think it’s a good font size, I wear glasses, and I didn’t have to squinch my eyes to see. I also like how the page is light blue, and it’s dark blue for the words to stick out. So you’ll be able to look at it. That’s one thing I like about it.

Specific comments made by caregivers included the following:

I like the subtitles on the bottom, because I feel like you sometimes get lost in the words, so I like that they’re there.

I like the bright colors. The bright colors catch my eye.

**Theme 4: Benefits to the Patient and Session Implementation**

A total of 19% (6/32) of the participants discussed how the TIPS app benefits the patient or made comments about session implementation. Most comments within this theme included that the TIPS app increases the comfort of the adolescent in the session (10/32, 31%) or that clinicians would be more likely to use the TIPS app with particular patients (10/32, 31%), such as female individuals or sexually active teenagers. Other comments in this theme included that the app makes the session more interactive, increases engagement for teenagers, is easy to incorporate into the session, and allows the clinician to modify or select content based on patient needs. Some examples of comments made by clinicians include the following:
This is good for me, too, the way I’m able to give feedback to the patient, or at least talk about it, cuz once it’s there, okay, I get it, and I can explain it to them, so I really like it.

Comments made by adolescents included the following:

The app makes it easier because it’s just awkward to talk about these things.

Some statements from caregivers included the following:

I think for this exercise as far as getting them to wrap their brain around where they really have established rapport with you, it’s probably a nice way of doing that where they don’t really have to look you in the eye and tell you about what they want.

Quantitative Results

The results of the WAMMI across the 3 groups are shown in Table 2. Responses across the groups indicated that all participant groups viewed the app favorably. In general, clinicians tended to report the most critical responses to the app. Overall, adolescents found the app (relatively) easier to use and understand compared with clinicians and caregivers. Clinicians had a greater propensity to report the app having annoying features compared with caregivers and adolescents. All clinicians (11/11, 100%) and caregivers (10/10, 100%) and 64% (7/11) of adolescents selected somewhat agree or strongly agree for the following statement: “This app has much that is of interest to me.” All adolescents (11/11, 100%) somewhat agreed or strongly agreed with the following statements: “I can quickly find what I want on this app,” “Using this app for the first time is easy,” and “Everything on this app is easy to understand.”
Table 2. Website Analysis and Measurement Inventory (WAMMI) responses regarding the Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior app by group\(^a\).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Clinicians (n=11)</th>
<th>Caregivers (n=10)</th>
<th>Adolescents (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median (range)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median (range)</td>
</tr>
<tr>
<td>This app has much that is of interest to me.</td>
<td>1.73 (1.19)</td>
<td>1 (1-5)</td>
<td>1.20 (0.42)(^b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-2)</td>
</tr>
<tr>
<td></td>
<td>2.36 (0.81)(^b)</td>
<td>2 (1-4)</td>
<td>1 (1-2)</td>
</tr>
<tr>
<td>It is difficult to navigate this app.</td>
<td>2.91 (1.04)(^c)</td>
<td>2 (2-4)</td>
<td>3.90 (1.79)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (1-5)</td>
</tr>
<tr>
<td></td>
<td>4.64 (0.92)(^c)</td>
<td>5 (2-5)</td>
<td>1 (1-2)</td>
</tr>
<tr>
<td>I can quickly find what I want on this app.</td>
<td>2.36 (1.03)(^d)</td>
<td>2 (1-4)</td>
<td>1.40 (0.70)(^d)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-3)</td>
</tr>
<tr>
<td></td>
<td>1.45 (0.52)</td>
<td>1 (1-2)</td>
<td>1 (1-2)</td>
</tr>
<tr>
<td>This app seems logical to me.</td>
<td>1.55 (1.21)</td>
<td>1 (1-5)</td>
<td>1.50 (1.27)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-5)</td>
</tr>
<tr>
<td></td>
<td>1.36 (0.67)</td>
<td>1 (1-3)</td>
<td>1 (1-3)</td>
</tr>
<tr>
<td>This app needs more introductory explanations.</td>
<td>3.45 (0.93)</td>
<td>4 (2-4)</td>
<td>3.80 (1.40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 (1-5)</td>
</tr>
<tr>
<td></td>
<td>3.82 (1.54)</td>
<td>4 (1-5)</td>
<td>3.82 (1.54)</td>
</tr>
<tr>
<td>This app is very attractive.</td>
<td>2.55 (1.37)</td>
<td>2 (1-5)</td>
<td>1.90 (1.29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-4)</td>
</tr>
<tr>
<td></td>
<td>2.36 (1.03)</td>
<td>2 (1-4)</td>
<td>2.36 (1.03)</td>
</tr>
<tr>
<td>I feel in control when I’m using this app.</td>
<td>2.09 (1.04)</td>
<td>2 (1-4)</td>
<td>1.50 (0.97)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-4)</td>
</tr>
<tr>
<td></td>
<td>1.73 (0.79)</td>
<td>2 (1-3)</td>
<td>1.73 (0.79)</td>
</tr>
<tr>
<td>This app is too slow.</td>
<td>4.0 (1.0)</td>
<td>4 (2-5)</td>
<td>3.90 (1.60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (1-5)</td>
</tr>
<tr>
<td></td>
<td>4.18 (0.98)</td>
<td>4 (2-5)</td>
<td>4.18 (0.98)</td>
</tr>
<tr>
<td>This app helps me find what I am looking for.</td>
<td>2.09 (0.94)</td>
<td>2 (1-4)</td>
<td>1.70 (1.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-4)</td>
</tr>
<tr>
<td></td>
<td>1.73 (0.65)</td>
<td>2 (1-3)</td>
<td>1.73 (0.65)</td>
</tr>
<tr>
<td>Learning to find my way around this app is a problem.</td>
<td>3.45 (1.04)(^b)</td>
<td>4 (2-5)</td>
<td>4.30 (1.34)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (1-5)</td>
</tr>
<tr>
<td></td>
<td>4.73 (0.47)(^c)</td>
<td>5 (4-5)</td>
<td>4.73 (0.47)</td>
</tr>
<tr>
<td>I don’t like using this app.</td>
<td>4.09 (1.58)</td>
<td>5 (1-5)</td>
<td>4.80 (0.42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (4-5)</td>
</tr>
<tr>
<td></td>
<td>4.09 (1.04)</td>
<td>4 (2-5)</td>
<td>4.09 (1.04)</td>
</tr>
<tr>
<td>I feel efficient when I’m using this app.</td>
<td>2.27 (1.19)</td>
<td>2 (1-4)</td>
<td>1.60 (0.97)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-4)</td>
</tr>
<tr>
<td></td>
<td>2.09 (1.14)</td>
<td>2 (1-5)</td>
<td>2.09 (1.14)</td>
</tr>
<tr>
<td>It is difficult to tell if this app has what I want.</td>
<td>3.18 (1.47)</td>
<td>3 (1-5)</td>
<td>3.90 (1.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (1-5)</td>
</tr>
<tr>
<td></td>
<td>3.72 (1.19)</td>
<td>4 (2-5)</td>
<td>3.72 (1.19)</td>
</tr>
<tr>
<td>Using this app for the first time is easy.</td>
<td>2.64 (1.21)(^c,d)</td>
<td>2 (1-4)</td>
<td>1.50 (1.27)(^d)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-5)</td>
</tr>
<tr>
<td></td>
<td>1.45 (0.52)(^c)</td>
<td>1 (1-2)</td>
<td>1.45 (0.52)</td>
</tr>
<tr>
<td>This app has some annoying features.</td>
<td>2.73 (1.49)(^c,d)</td>
<td>2 (1-5)</td>
<td>4.40 (1.07)(^d)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (2-5)</td>
</tr>
<tr>
<td></td>
<td>4.18 (1.17)(^c)</td>
<td>5 (2-5)</td>
<td>4.18 (1.17)</td>
</tr>
<tr>
<td>Remembering where I am on this app is difficult.</td>
<td>3.18 (1.08)(^c)</td>
<td>3 (2-5)</td>
<td>3.70 (1.70)(^b,c)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.5 (1-5)</td>
</tr>
<tr>
<td></td>
<td>4.82 (0.40)(^b,c)</td>
<td>5 (4-5)</td>
<td>4.82 (0.40)</td>
</tr>
<tr>
<td>Using this app is a waste of time.</td>
<td>4.45 (0.93)</td>
<td>5 (2-5)</td>
<td>4.90 (0.32)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (4-5)</td>
</tr>
<tr>
<td></td>
<td>4.64 (0.92)</td>
<td>5 (2-5)</td>
<td>4.64 (0.92)</td>
</tr>
<tr>
<td>I get what I expect when I click on things on this app.</td>
<td>2.00 (1.10)</td>
<td>2 (1-4)</td>
<td>1.80 (1.48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-5)</td>
</tr>
<tr>
<td></td>
<td>1.55 (0.52)</td>
<td>2 (1-2)</td>
<td>1.55 (0.52)</td>
</tr>
<tr>
<td>Everything on this app is easy to understand.</td>
<td>2.45 (1.13)(^c,d)</td>
<td>2 (1-4)</td>
<td>1.30 (0.48)(^d)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (1-2)</td>
</tr>
<tr>
<td></td>
<td>1.36 (0.50)(^c)</td>
<td>1 (1-2)</td>
<td>1.36 (0.50)</td>
</tr>
</tbody>
</table>

\(^a\)WAMMI responses among all users interviewed, including clinicians, adolescents, and caregivers. Each item was scored on a Likert scale (1=strongly agree, 2=somewhat agree, 3=neutral, 4=somewhat disagree, and 5=strongly disagree). Lower numbers indicate greater agreement with the statement.

\(^b\)Significant differences between the caregiver and adolescent groups (P<.05).

\(^c\)Significant differences between the clinician and adolescent groups (P<.05).

\(^d\)Significant differences between the clinician and caregiver groups (P<.05).

Discussion

Principal Findings

As of October 2021, a national state of emergency in child mental health has been jointly declared by the American Academy of Pediatrics, American Academy of Child and Adolescent Psychiatry, and Children’s Hospital Association because of the rising rates of behavioral health problems coupled with a limited and overburdened mental health workforce [52]. To address this public mental health crisis, it is critical to enlist a broad and creative range of approaches, including those that leverage mHealth tools, to implement empirically supported content and interventions that target the significant drivers of this state of emergency. Also critical to this crisis is the need for more robust prevention of substance use problems (including opioid overdose and opioid use disorders) and HIV and other STI acquisition (particularly among young people at the highest...
risk for new HIV acquisition) [77,78], which is signaled by every public health indicator. The goal of this paper was to describe the development and usability testing of an mHealth app tool that collectively targets the aforementioned problems (ie, high prevalence of trauma-related mental health problems, substance use disorder, and opioid overdose and new HIV acquisition risk, as well as limited resources and opportunities dedicated to implementing substance use prevention and training clinicians in doing so).

The primary overall finding from the usability testing of the app leveraging a mixed methods approach was that the TIPS app was perceived by all 3 user types to be a highly usable mHealth tool to be implemented during the course of TF-CBT. Qualitative data collected from the usability testing of the app yielded positive feedback from clinicians, adolescents, and caregivers, and the quantitative data (ie, responses to the WAMMI) concurred with the qualitative findings (eg, 27/32, 84% of all users agreed or strongly agreed that “The app has much of interest to me”). However, more generally, the results of this study provide valuable insights into the use of technology and apps in adolescent mental health treatment. Clinician responses illustrated that they generally felt comfortable using smartphones and tablets and believed that these mHealth tools yielded benefits of engagement and accessibility for their adolescent clients. Moreover, the fact that all clinicians reported comfort with using these devices suggests a high level of digital readiness among this workforce in need of resources and support, which is a shared sentiment among other clinicians [79]. Despite this positive feedback, clinicians also identified potential drawbacks related to apps and mHealth tools, including the challenge of teenagers potentially overusing or being distracted by apps and issues related to limited wireless internet access in some homes. Indeed, equitable access to technology is essential when developing, evaluating, and implementing mHealth tools as augmentations to mental health interventions.

Specific to the interview responses, 4 overarching themes related to the TIPS app’s usefulness and effectiveness emerged. The first theme, strengths of the app content, highlights the positive aspects of the app, such as content tailored well to an adolescent audience, modules that are viewed as engaging, and relatable language. These strengths are crucial for maintaining adolescent clients’ attention and fostering meaningful interactions during trauma treatment, particularly when talking about what may be perceived as sensitive topics (eg, sexual decision-making). The second theme, changes in or suggestions for app content, points to the need for interactivity, revisions in video and picture content, and the inclusion of important topics such as consent and rape in relationships. Echoing findings of previous studies, these suggestions underscore the importance of user input and continuous improvement and adaptation of app content to meet evolving needs [80]. This also highlights the importance of using tools that allow for efficient and inexpensive minor edits (eg, to language and images) to mHealth apps when possible, as was done using the AppBuilder platform for the TIPS app. The third theme, esthetics and usability, emphasizes the significance of the app’s design and functionality, with users offering both praise and suggestions for improvement. Beyond further underscoring the importance of having the capacity to revise content within an app as part of the iterative app development process, these praise interview responses also help inform implementation strategies, indicating what end users like most and find most engaging about the app. For example, regarding TIPS, the Make a Play activity emerged as a favorite (eg, the opioid pill activity and Let’s talk about Sext), and it may be an activity to highlight when first explaining the app to a clinician or adolescent client (eg, engaging). Finally, the fourth theme, benefits to the patient and session implementation, highlights the app’s potential to enhance adolescent and caregiver comfort during sessions and improve engagement, particularly for sensitive topics. It also suggests that clinicians are more likely to use the app with specific patient demographics, such as sexually active teenagers.

Regarding one of the clinicians’ comments that it can be difficult to get teenagers to stop using apps, it is important to note that a detailed implementation manual for the TIPS app is provided to clinicians when they are trained in how to use the app. Specifically, they are guided on how to structure the time spent on activities on the app, which occurs in the context of a TF-CBT treatment session. Thus, TF-CBT clinicians are able to contain adolescents’ use of the TIPS app in sessions.

The quantitative results, presented in Table 1, reinforce the positive reception of the TIPS app across all participant groups. Adolescents in particular found the app less difficult to use and more understandable compared with clinicians and caregivers, but clinicians expressed some critical feedback, including minor annoyance with certain features. Notably, all clinicians and caregivers, along with most adolescents, expressed a strong interest in the app’s content. In addition, adolescents found the app easy to navigate and understand, suggesting that user-friendliness is a key factor in their engagement with digital mHealth tools [81]. Overall, these quantitative findings align with the qualitative feedback, highlighting the promising utility of the TIPS app for clinician implementation among adolescents and caregivers receiving TF-CBT while also emphasizing areas for improvement.

A strength of the TIPS tool is that it offers a platform that is telehealth compatible. The past several years have underscored the value of ensuring access to mental health treatment and risk behavior prevention through telehealth strategies. In addition to the benefits of breaking down geographical barriers, the integration of mHealth tools such as TIPS with telehealth helps promote confidential and convenient interactions for patients with mental health clinicians, fostering a sense of privacy and comfort that is crucial for therapy [82,83]. This may be particularly helpful when addressing sensitive topics such as trauma, substance use, and HIV. Ultimately, the synergy between mHealth tools and telehealth in behavioral care represents a transformative shift toward prevention-focused holistic and inclusive mental well-being support.

Limitations

The primary limitation of this study is that it is limited in scope to usability testing with a small sample of clinicians, adolescents, and caregivers. To establish the feasibility and efficacy of the app in reducing the risk of substance use and risky sexual behavior, a fully powered RCT is necessary to compare TF-CBT...
implementation with and without the inclusion of the app, including assessment time points that follow youth and their caregivers over time. Although this study informs possible revisions to the app and suggests that clinicians, adolescents, and caregivers will respond positively to its inclusion in trauma-focused treatment, the efficacy trial will ultimately reveal whether indeed the app is able to make a dent in the youth mental health state of emergency and help eliminate some of the burden on clinicians in implementing a prevention curriculum.

Conclusions

In conclusion, this study demonstrates the positive reception of technology—and the TIPS app in particular—in adolescent trauma-focused treatment, with clinicians, adolescents, and caregivers recognizing the benefits of engagement, accessibility, and user-friendliness of this novel mHealth tool. The qualitative themes shed light on the strengths of the app’s content, areas for improvement, esthetics, and usability as well as its potential to enhance adolescents’, caregivers’, and clinicians’ experiences during TF-CBT sessions. These findings underscore the importance of the ongoing development and refinement of digital tools in mental health care—including those that can be integrated into telehealth mental health care delivery—to better meet the evolving needs of trauma-affected adolescents and their caregivers.

Acknowledgments

This study was supported by K24DA039783 (principal investigator [PI]: CKD) and in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under grant UL1 TR001450 (multiple PIs [MPIs]: Kathleen Brady and Patrick Flume). The manuscript preparation was also supported in part by grants from the National Institute of Mental Health (R01MH112209 [PI: CKD] and T32MH018869 [MPIs: CKD and Dean Kilpatrick]) and the National Institute on Drug Abuse (R01DA031285 and 1R01DA03288 [MPIs: CKD and Paula Riggs]; K23DA050800 [PI: AH]).

Data Availability

The data are available from the study principal investigator (CKD) upon request.

Conflicts of Interest

None declared.

References


39. Home page. Medical University of South Carolina. URL: https://tcfbc2.musc.edu/ [accessed 2023-12-10]


72. Claridge N, Kirakowski J. Website analysis and measurement inventory questionnaire. WAMMI. URL: http://www.wammi.com/index.html [accessed 2023-09-06]


77. Coady JA. Youth substance use prevention month. Substance Abuse and Mental Health Service Administration. 2022. URL: https://www.samhsa.gov/blog/youth-substance-use-prevention-month [accessed 2023-12-10]


Abbreviations

mHealth: mobile health
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
STI: sexually transmitted infection
TIPS: Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior
TF-CBT: trauma-focused cognitive behavioral therapy
WAMMI: Website Analysis and Measurement Inventory

©Carla Kmett Danielson, Angela Moreland, Austin Hahn, Devin Banks, Kenneth J Ruggiero. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Virtual Exercise in Medicine: A Proof of Concept in a Healthy Population

Barbara Le Roy1,2, PhD; Charles Martin-Krumm1,2,3, Prof Dr; Charlotte Poupon4, PhD; Raphaëlle Richieri5, Prof Dr Med; Eric Malbos5, MD, PhD; Fanny Barthélemy6; Eric Guedj6, Prof Dr Med; Marion Trousselard1,2,7,8, Prof Dr Med

1Unité neurophysiologie du stress, Institut de recherche biomédicale des armées, Brétigny-sur-Orge, France
2Adaptation, mesure et évaluation en santé. Approches interdisciplinaires, Metz, France
3Vulnérabilité, capabilité, rétablissement, Unité religion, culture et société, École des psychologues praticiens, Paris, France
4École nationale supérieure de création industrielle, Paris, France
5Département psychiatrie, Hôpital de la Conception, Institut Fresnel, Marseille, France
6Département de médecine nucléaire, Centre européen de recherche en imagerie médicale, Institut Fresnel, Marseille, France
7Service de santé des armées, Paris, France
8Réseau ABC des Psychotraumas, Montpellier, France

Corresponding Author:
Barbara Le Roy, PhD
Unité neurophysiologie du stress
Institut de recherche biomédicale des armées
Place Général Valérie André
Brétigny-sur-Orge, 91220
France
Phone: 33 123456789
Email: barbara.m.le.roy@gmail.com

Abstract

Background: Science is beginning to establish the benefits of the use of virtual reality (VR) in health care. This therapeutic approach may be an appropriate complementary treatment for some mental illnesses. It could prevent high levels of morbidity and improve the physical health of patients. For many years, the literature has shown the health benefits of physical exercise. Physical exercise in a VR environment may improve the management of mild to moderate mental health conditions. In this context, we developed a virtual environment combined with an ergocycle (the augmented physical training for isolated and confined environments [APTICE] system).

Objective: This study aims to investigate the impact of physical exercise in a VR environment.

Methods: A total of 14 healthy participants (11 men and 3 women; mean age 43.28, SD 10.60 years) undertook 15 minutes of immersive physical exercise using the system. Measures included mindfulness and immersion disposition, subjective perceptions of sensory information, user experience, and VR experience (ie, psychological state, flow, and presence).

Results: First, the APTICE system appears to be a useful tool because the user experience is positive (subscales in the AttrakDiff questionnaire: pragmatic quality=0.99; hedonic quality–stimulation=1.90; hedonic quality–identification=0.67; attractiveness=1.58). Second, the system can induce a positive psychological state (negative emotion, P=.06) and an experience of flow and presence (P values ranging from <.001 to .04). Third, individual immersive and mindful disposition plays a role in the VR experience (P values ranging from <.02 to .04). Finally, our findings suggest that there is a link between the subjective perception of sensory information and the VR experience (P values ranging from <.02 to .04).

Conclusions: These results indicate that the device is well accepted with positive psychological and exteroceptive outcomes. Overall, the APTICE system could be a proof of concept to explore the benefits of virtual physical exercise in clinical medicine.

(JMIR Form Res 2024;8:e45637) doi:10.2196/45637

KEYWORDS
countermeasures; mental health; physical activity; virtual reality; user experience
**Introduction**

The Roman poet Juneval wrote “mens sana in corpore sano” (a healthy mind in a healthy body) [1]. Still relevant today, it has never made as much sense. The body and the mind seem to be indivisible, truly part of a whole [2].

**Virtual Reality**

In recent years, virtual reality (VR) has been recognized as a new approach to health [3-6] (that seeks to connect the body and mind [7]. The term was first used by Jaron Lamier in 1986 to refer to an advanced technological interface in which the user interacts with a 3D environment that is generated by a computer to simulate real-world experiences [8,9]. The tool can simulate reality and stimulate the body’s senses in ways that are only limited by our imagination. It creates a new space-time that is halfway between the real and the unreal, pushing back the boundaries of reality and experimenting with new paradigms that we would not otherwise have access to [10,11]. Thus, VR goes beyond a simple simulation of the external world. The modulation of interoceptive, exteroceptive, and vestibular information leads the participant to create a representation of their own body. This conceptualization is described as the body matrix, which refers to the multisensory representation of the body in the brain and the space directly around the body [12]. Through VR, it is possible to induce the illusion of being and moving in a fake body. This interstice allows individuals to perceive, interpret, and interact with their environment through an internal representation of the world [13]. Repeated VR use may stimulate changes in the brain based on neuroplasticity mechanisms [14]. Riva et al [7] noted that the effects may be heightened by immersive VR systems and the induced sense of presence in the surrounding virtual environment.

**Immersion** is a characteristic of VR systems and is created when the virtual environment replaces the user’s sensory stimuli with virtual sensory stimuli. Through immersion, it is possible to induce the sense of presence. Multisensory integration generates a feeling of being there and can sometimes lead to the illusion of being in an alternative body [15,16]. Slater [17] defined presence as “the strong illusion of being in a place despite the sure knowledge that you are not there.” Thus, participants have the strong illusion of being in the virtual environment and being able to perceive what is happening in it such as the virtual precipice. However, they consciously know that this is only a perceptual illusion not a reality [18]. Presence is related to flow, which refers to “the holistic sensation that people feel when they act with total involvement” [19]. It is a psychological state corresponding to enjoyment, cognitive absorption, and distortions in time perception. The literature on VR highlights the influence of immersion, induced by VR systems, on both presence and flow in the virtual environment [20,21].

Nevertheless, interindividual differences have been noted regarding both presence and flow. One relevant factor is mindfulness disposition (MD). MD is characterized by the awareness that emerges when paying purposeful attention to the present moment and responding nonjudgmentally to the unfolding experience [22,23]. It is associated with a protective function in both a healthy population and among patients [24,25]. A recent study by Lefranc et al [26] highlighted that high MD is associated with better positive emotions, interoception, and subjective extrasensory acuity. Top-down conceptual representations and bottom-up multisensory inputs contribute to body awareness. Moseley et al [12] suggested that these representations be integrated with exteroceptive data in the body matrix.

Over the years, VR has become increasingly accessible. It has been particularly beneficial in the field of medicine, whether in the context of medical training, surgery, the treatment of certain neurodegenerative diseases, rehabilitation, pain management, or cognitive and psychological disorders [11,27-34]. The literature shows the value of using VR as a therapeutic tool to treat mental disorders such as anxiety, depression, posttraumatic stress disorder, and phobias [8,31,33-38]. Antidepressants, such as selective serotonin-norepinephrine reuptake inhibitors, or benzodiazepines are the first-line treatment for anxiety symptoms in patients while cognitive behavioral therapy has been found to be effective in reducing them [39-41]. VR interventions such as exposure therapy have been shown to be effective as a coadjuvant in mental illness and appear to have the same effects as drug treatments, although the results take longer to become apparent [34]. Used as a complementary therapy, VR may have many advantages, including the ability to recreate a realistic traumatic environment under controlled conditions, which can be complex in vivo [33,42,43]. Most studies show that participants have a high degree of acceptance, and VR use is consistent with postintervention improvements in symptom awareness; a decrease in depressive symptoms; greater motivation to exercise; and better enjoyment, engagement, and affect, particularly in clinical populations [35,37]. VR therapy can stimulate emotion (notably fear), as the participant has the feeling of being present in the unreal environment [18,44]. Thus, it appears to be an innovative nondrug supplement to other treatments that can be demanding for the patient and may have side effects. Although the quality of the technology may play a role in positive outcomes [45], it appears to be an interesting new tool that poses no serious threat to participants [46].

**The Potential of Immersive Physical Activity**

In recent years, an increasing body of the literature has investigated the power of immersive physical activity. Physical activity preserves health and protects individuals from many pathologies [47-49]. It can be defined as “any bodily movement produced by skeletal muscles that results in the expenditure of more energy than the resting metabolism” [50]. One of the components of physical activity is physical exercise, understood as “planned, structured, repetitive physical activity whose objective is to improve or maintain one or more components of physical fitness” [50]. For many years, the literature has shown the benefits of physical activity on health, not only physical (ie, reduced mortality, reduced risk of cardiovascular pathologies, reduced incidence of cancer, or weight maintenance) but also cognitive (ie, improved cognitive function, improved sleep, or reduced risk of dementia) and psychological (ie, reduced signs of anxiety and depression or reduced risk of depression), both in the general population (ie, adults, children, and older adults) and in the context of various chronic diseases and health...
conditions [47-49,51]. However, it is only recently that the scientific community has begun to take an interest in the biological and physiological mechanisms underlying these outcomes [52,53]. People with mental illness often exhibit disrupted sensory processing and perception [54]. Thus, physical activity therapy can be both a physical and psychological countermeasure. However, compliance is a key issue as regular practice is necessary for optimal mental illness management.

Few studies have examined the use of VR in this context, although the pioneering work of Plante et al [55-57] seems to indicate real benefits in terms of well-being, particularly in women [56]. The addition of VR has been found to enhance mood, increase enjoyment and energy, reduce tiredness, enhance motivation and confidence, and increase compliance [57,58]. Enjoyment may play an important role in the benefits gained from exercise [58].

In recent years, there has been an increase in the number of studies that encourage the practice of sports to prevent anxiety disorders and protect against anxiety and depression [59,60]. A recent study demonstrated its importance in the context of the COVID-19 pandemic, where it was able to improve well-being through improved physical and cognitive outcomes and limit psychological disorders related to isolation and confinement [61]. Thus, the literature suggests that VR coupled with physical activity may be a useful way to improve the symptomatology of individuals with anxiety disorders, posttraumatic stress disorder, and depression [61]. Furthermore, many studies have highlighted the ability of natural environments to induce positive emotions, promote well-being, reduce anxiety, improve self-esteem, and reduce negative emotions (ie, fatigue, confusion, tension, depression, and anger-hostility) compared with urban or indoor environments [62,63]. The same observation has been made in VR environments [64]. A virtual environment that offers physical activity in a natural setting seems to have the potential to improve the benefits of VR, especially for people with mental illness [65,66].

Gaps in the Literature and Objectives of the Study

Many of the systematic reviews and meta-analyses that have been carried out have important limitations, notably related to differences in technology. There is also a lack of longitudinal studies on the long-term effects of VR. Most studies are one-shot experiments that evaluate its benefits before and after the intervention. Evaluation itself is problematic as subjective measures (questionnaires) are typically used and few studies have measured physiological effects (ie, heart rate variability, heart rate, and electrodermal activity). As it can be complex to overcome these gaps, caution is advised in interpreting any results or conclusions [33,35,62,65,67,68]. Given these gaps in the literature, there is a need for more rigorous testing. Any evaluation should be based on three assessment criteria: (1) the activity does not duplicate other countermeasures; (2) it must improve the experience of sport and thus increase its attractiveness (especially relevant for patients with depression) [55,56]; and (3) immersion must provide a multimodal sensory input to the user [69-72]. The benefits of multisensory stimulation have been demonstrated in the context of cognitive and sensorimotor rehabilitation [73] and emotion regulation [10,74].

Thus, the aim of this preliminary proof-of-concept study was to investigate the association between VR and physical exercise in a virtual natural environment to improve the psychological state of healthy participants and the underlying processes, before evaluating its benefits in clinical medicine. We measured the user experience (UX) and evaluated 3 hypotheses:

- **Hypothesis 1**: positive changes in psychological state are associated with flow and presence during the session in the VR environment.
- **Hypothesis 2**: both MD and immersion disposition are positively related to change in the participant’s psychological state, flow, and presence.
- **Hypothesis 3**: there is a relation between the subjective evaluation of sensory information, immersive disposition, and mindful disposition and psychological change.

### Methods

#### Ethical Considerations

This study was approved by the Minarm Ethical Committee (N 125 132/MIP/DGA/MINARM). Written consent was obtained from all participants in accordance with the Declaration of Helsinki and subsequent amendments.

#### Participants

In total, 14 volunteers (3 women and 11 men), who were declared medically fit, were recruited during the 3 innovation open days at the French Armed Forces Biomedical Research Institute in 2019. They ranged in age from 22 to 59 (mean 43.28, SD 10.60) years and were either working for the French Armed Forces Biomedical Research Institute (n=9) or the French Football Federation (n=5). See Table 1 for the demographic information. The participants were recruited by email and contacted to determine whether they met the inclusion and exclusion criteria. If eligible, they were assigned an appointment for the laboratory session. All participants were asked to abstain from exercise on the day of their participation to ensure that the results were due to the experiment. The inclusion criteria were based on the following: affiliation to a health care system (social security); age between 18 and 75 years; and no history of neurological or cardiovascular disease, diabetes, or medications that could affect the response. Exclusion criteria included pregnancy, the presence of a contraindication to VR (people who had experienced anxiety or nausea during a VR experience, photosensitive epilepsy, vestibular disorder, or severe myopia >3.5 diopters).
Table 1. Sociodemographic characteristics of participants (N=14).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>43.28 (10.60)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>11 (78.57)</td>
</tr>
<tr>
<td>Women</td>
<td>3 (21.42)</td>
</tr>
<tr>
<td>Screen time, mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Professional</td>
<td>300.00 (164.73)</td>
</tr>
<tr>
<td>Personal</td>
<td>111.42 (63.95)</td>
</tr>
<tr>
<td>Physical activity, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (71.42)</td>
</tr>
<tr>
<td>No</td>
<td>4 (28.57)</td>
</tr>
<tr>
<td>Video games, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4 (28.57)</td>
</tr>
<tr>
<td>No</td>
<td>8 (57.14)</td>
</tr>
<tr>
<td>Ocular correction, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (71.42)</td>
</tr>
<tr>
<td>No</td>
<td>4 (28.57)</td>
</tr>
</tbody>
</table>

Participants who practiced a physical activity or engaged in video games completed the Addictive Intensity Evaluation Questionnaire (AIEQ). The analysis found that 10 out of 14 (85%) participants engaged in physical activity (mean 31.00, SD 6.20) and 4 (35%) played video games (mean 28.20, SD 14.34). No addictive behaviors were found among the participants in either of these modalities.

Augmented Physical Training for Isolated and Confined Environments

This proof-of-concept study is based on the augmented physical training for isolated and confined environments (APTICE) system. The aim of the system is to use physical exercise in a VR environment to improve the well-being of patients with depression. It is composed of a VR - enabled cycle ergometer (VirZOOM Bike Controller) and a VR-based head-mounted display (Oculus Rift CV1, Oculus VR), which provides visual and auditory inputs. The VR application was developed by GAMIT (Petit-Quevilly) and ran on an Asus A15 TUF566IU-HN326T laptop with an AMD Ryzen 5 4600H 16 GB processor, a 512 GB solid state drive, and an Nvidia GeForce GTX1660 Ti 6 GB graphics card. The VR environment consisted of natural areas of forests and mountain plains (Figure 1). See Multimedia Appendix 1 for further images of the APTICE device.

Figure 1. Natural virtual environment images. (A) Forest with stretches of water. (B) Mountain plain with sheep.
Data Collection

Population Overview

A 7-item sociobiographical questionnaire was developed to collect standard sociodemographic data such as gender, age, hobbies, physical activity, video game use, and VR experience. The AIEQ evaluated addictive practices [75]. Two versions were used: the 14-item AIEQ-g that measures the intensity of video game playing and the risk of its problematic use and the 14-item AIEQ-s that measures sports practice and the risk of its problematic use.

UX of the APTICE Device

The UX of the APTICE device was assessed using the 10-item AttrakDiff questionnaire, which evaluates the hedonic and pragmatic qualities of interactive systems [76]. It is divided into 4 subscales: pragmatic quality, hedonic quality–identity, hedonic quality–stimulation, and attractiveness. Values close to the mean (from 0 to 1) are considered standard values. They indicate that the device meets its objectives with no negative impacts on the user.

Psychological Questionnaires

Two questionnaires were used to evaluate psychological dispositions. The 14-item Freiburg Mindfulness Inventory was used to measure MD [77]. It is divided into 2 subscales: acceptance and presence. Immersion disposition was assessed using the 18-item immersive tendencies questionnaire, which is divided into 4 subscales: focus, involvement, emotions, and games [78]. Two questionnaires were used to evaluate psychological state. First, the 12-item Scale of Positive and Negative Experience (SPANE) questionnaire assessed subjective feelings of well-being [79]. The overall scale is divided into 2 subscales: positive and negative emotions. Second, the 20-item Activation-Deactivation Adjective Checklist (AD-ACL) evaluates the level of awareness and emotional disposition [80]. This is divided into 2 dimensions: energetic arousal (from energy to tiredness) and tense arousal (from tension to calmness). The energetic arousal is further divided into 2 subscales—general activation and deactivation—while the tense arousal is subdivided into general tenseness and calmness.

Subjective Evaluation of the Quality of Sensory Information

We developed the Personal Evaluation of Six Senses questionnaire to assess subjective perceptions of vision, sound, touch, olfaction, taste, and equilibrium. Participants evaluated the accuracy of their perceptions from each of their 6 senses using a ranked scale running from 1 to 6.

The VR Experience

The VR experience was assessed using the 12-item Educational Flow Questionnaire (EduFlow2), which measures flow [81]. It is divided into 4 dimensions: cognitive control, immersion and time transformation, loss of self-consciousness, and autotelic experience. Cognitive absorption (a summary of the first 3 dimensions) was added as the fourth scale. The 24-item Presence Questionnaire assessed presence [82]. It is divided into 7 subscales: realism, possibility of action, quality of interface, possibility of examination, self-evaluation of performance, sounds, and haptic. APTICE device sickness was assessed using the 16-item Simulator Sickness Questionnaire [83]. It is divided into 2 subscales: nausea and oculomotor.

Procedure

The experimental protocol is illustrated in Figure 2. Upon arrival, the participants were asked a few questions to ensure they met the inclusion criteria and signed the consent form. They then completed a series of questionnaires in the following order: sociobiographical questionnaire, AIEQ-g, AIEQ-s, Freiburg Mindfulness Inventory, Personal Evaluation of Six Senses, 18-item immersive tendencies questionnaire, SPANE, and AD-ACL. Then, they engaged in a moderate-intensity bout of exercise in a natural environment for 15 minutes while wearing the VR headset. They could choose their trajectory along various predefined paths and, by turning their head, obtain a 360° view of the virtual environment. At the end of the session, they were asked to complete another series of questionnaires in the following order: SPANE, AD-ACL, EduFlow2, 24-item Presence Questionnaire, AttrakDiff, and Simulator Sickness Questionnaire.
Figure 2. The augmented physical training for isolated and confined environments (APTICE) experimental protocol. AD-ACL: Activation-Deactivation Adjective Checklist; FMI: Freiburg Mindfulness Inventory; ITQ-f: Immersive Tendencies Questionnaire; SPANE: Scale of Positive and Negative Experience; SSQ: Simulator Sickness Questionnaire; VR: Virtual Reality.

Statistical Analysis
Statistical analyses were performed using the RStudio software (version 1.2.5001). Descriptive statistics are expressed as mean (SD). The Shapiro-Wilk test was used to determine whether the data were normally distributed. The effects of the APTICE device experience on emotional and activation-deactivation states were assessed as follows: a t test (2-tailed) for pre-post comparisons and parametric data or the Mann-Whitney U test for nonparametric data. Kendall correlations were run to explore the relationship among virtual exercise, subjective sensory accuracy, and VR experience. For all analyses, significance was set at $P<.05$. Trends were considered when $.05<P<.10$. Deltas were calculated to compare the temporal impact of the experience measured using the SPANE questionnaire and the AD-ACL.

Results
The UX
The APTICE tool was assessed in terms of UX. Participants reported a positive experience measured as pragmatic quality, hedonic quality–stimulation, hedonic quality–identification, and attractiveness (Figure 3). The scores were particularly high for hedonic quality–stimulation and attractiveness. No participant reported any cybersickness.
Figure 3. AttrakDiff’s subscales. Values close to the mean (from 0 to 1) are considered standard and indicate that the device meets its objectives with no negative impacts on the user. However, they also suggest that improvements could be made to the system to obtain high positive values. Values outside this neutral zone are considered positive (1-3) or negative (−1 to −3). X1: pragmatic quality; X2: hedonic quality (stimulation); X3: hedonic quality (identification); X4: attractiveness.

Relationships Between Psychological Assessments, Exteroception, and VR Experience

Overview

Multimedia Appendix 2 summarizes the significant correlations between the tested variables.

Hypothesis 1: The VR Experience (Change in Psychological State, Flow, and Presence)

The analysis of emotional and arousal states only identified a trend for negative emotions. Participants tended to have fewer negative emotions after the APTICE experiment ($t_{12}=2.06$, $P=.06$).

There were significant positive and negative correlations between flow and presence. Participants who scored high for possibility to examine also scored high for flow cognitive control ($τ=0.45$, $P<.001$), flow cognitive absorption ($τ=0.67$, $P=.001$), and flow immersion and time transformation ($τ=0.55$, $P=.01$). Participants who scored high for possibility to act also scored high for flow cognitive absorption ($τ=0.58$, $P=.004$), flow cognitive control ($τ=0.76$, $P<.001$), flow immersion and time transformation ($τ=0.58$, $P=.006$), and flow-autotelic experience ($τ=0.58$, $P=.001$). As realism increased, flow cognitive control also increased ($τ=0.52$, $P=.01$). However, as haptic increased, flow loss of self-consciousness ($τ=-0.52$, $P=.02$) decreased.

Concerning change in psychological states related to flow and presence, our results suggest that there is no correlation between change in emotional state (measured with the SPANE questionnaire) and either flow or presence. However, there were significant negative correlations between flow and changes in activation-deactivation states (measured using the AD-ACL). An increase in tense activation (positive delta) was associated with lower scores for flow immersion and time transformation ($τ=-0.46$, $P=.04$) and flow-autotelic experience ($τ=-0.52$, $P=.01$). No correlation was found between presence and flow, and there were no changes in activation-deactivation.

Hypothesis 2: Disposition and the VR Experience (Change in Psychological State, Flow, and Presence)

No relationship was observed between immersive disposition and MD for any subscale.

The analysis found a significant positive correlation between MD and presence. More precisely, higher MD-acceptation was associated with a higher score for possibility to examine ($τ=0.49$, $P=.02$). There was also a significant positive correlation between MD and flow. High scores for MD-acceptation were associated with high scores for flow cognitive control ($τ=0.45$, $P=.03$).

Finally, there was a significant positive correlation between immersion and flow. Specifically, high scores for flow loss of self-consciousness were slightly associated with high scores for involvement ($τ=0.54$, $P=.01$).

Concerning disposition and the VR experience, the analysis found no correlation between change in emotional state and either immersive or mindful disposition. Significant positive and negative correlations were found between immersion and change in activation-deactivation. An increase in tense activation (positive delta) was associated with higher scores for games ($τ=0.45$, $P=.04$). However, an increase in general activation...
(positive delta) was associated with lower scores for involvement \( (\tau = -0.52, P = .02) \).

**Hypothesis 3: Subjective Exteroceptive Accuracy, Disposition, and the VR Experience**

The analysis found no relation between the subjective exteroceptive evaluation and changes in emotional and activation states, presence, or MD. However, significant positive and negative correlations were observed between immersion and subjective acuity.

Correlation matrices for immersion and subjective acuity variables are shown in Figure 4.

**Figure 4.** Correlation matrices for immersion and subjective acuity variables. Distributions are shown on the diagonal. Trend curves are shown at the bottom of the diagonal scatter plots. The top diagonal shows correlation coefficients and significance levels. A1: olfaction; A3: vision; I1: focus; I2: involvement. * > .99, ** .10, *** .05, **** .01.
Increased involvement was associated with higher subjective visual acuity ($\tau=0.48, P=0.03$). In contrast, an increase in focus was associated with lower subjective smell acuity ($\tau=-0.43, P=0.04$). Low scores for subjective hearing were associated with high scores for flow cognitive absorption ($\tau=-0.43, P=0.04$) and flow immersion and time transformation ($\tau=-0.47, P=0.03$). Similarly, low scores for subjective taste were associated with high scores for flow cognitive control ($\tau=-0.49, P=0.02$).

**Discussion**

**Principal Findings**

The main aim of this proof-of-concept study was to investigate the effect of VR associated with physical activity on the psychological state of healthy participants before further evaluation of patients with depression in a randomized controlled trial. This exploratory study evaluated a new device, named APTICE, which couples physical exercise with a VR headset. This pilot feasibility study proposed variables of interest, which will form the basis for our next randomized controlled trial. The latter will investigate clinical and neurofunctional changes in a population affected by depression using VR associated with physical activity. Results from this study provide new insights into the benefits of this type of technology when used in clinical medicine to improve health.

**A Positive UX Experience**

As Hassenzahl et al [84] demonstrated, the evaluation of the hedonic and pragmatic qualities of a system is known to influence overall perceptions of its attractiveness. Understanding the UX is crucial in the design of a new device, which is often ignored. The participants in our study were very positive regarding both the hedonic quality–stimulation and the attractiveness of the device. However, pragmatic quality and hedonic quality–identification scores were lower. Furthermore, the responses were the most disparate for these 2 dimensions. Hedonic quality–stimulation was associated with ideas such as outstanding, impressive, exciting, or interesting. Although the response to the UX appears to be positive, there is room for improvement. The relatively low scores for hedonic quality–identification are not surprising, as this aspect relates to the ability of the system to reflect the user’s identity. Similarly, pragmatic quality needs to be improved with a focus on usability. Both the appropriateness of the functionality and its accessibility need further attention. However, this short 15-minute experiment allowed us to conclude that the APTICE system meets its development and quality objectives—specifically, to design a device that supports physical exercise in VR. In the longer term, we will need to consider how to improve it, particularly in light of the technological development that has taken place since its creation.

**Psychological Changes Induced by the APTICE Device**

Our main hypothesis was that physical exercise in VR environment could create a positive experience, measured as psychological and sensory feedback from the participants.

Consistent with the literature, our initial results suggest that the APTICE device experience decreases negative emotions [44,85-89]. However, our first hypothesis (that the APTICE device would induce a positive psychological state and an experience of flow and presence) was only partially confirmed. The literature [7,17] notes that presence and flow are usually positively linked, although a negative correlation has been found between haptic presence and loss of self-consciousness in flow experiments. In the absence of a meaningful haptic system, interactions with objects in the VR environment can widen the gap between actual and virtual realities [90]. In our experiment, haptic feedback from the interaction with the ergocycle did not reflect reality, which suggests its key role in inducing flow. For example, there was no body movement when going around the bends and almost no return on effort. The poor quality of the correspondence between the virtual exercise environment and reality could explain the absence of a change in positive emotions.

Our initial results suggest a close relationship between the quality of the technology and the VR experience. This is all the more important as flow (characterized by a deep involvement and absorption in an activity) promotes a state of inner well-being and positive emotions [91,92]. Overall, our results suggest that practicing a physical activity in a VR setting could be used to improve psychological outcomes. According to previous studies [60,61,66], the APTICE device may have potential benefits for patients, especially those with mental illness. The literature also shows that natural scenes support a positive psychological state both in general [65,86] and in the treatment of mental illness [67,93]. This is in line with the reduction in negative emotions in individuals following our study’s APTICE session. Although APTICE needs improvement, both the positive response to the UX and its effect on the user’s psychological state suggest that regular use may have a positive impact on mental health.

**Relationships Between Disposition and VR Experience**

Our results partially confirm our second hypothesis, which focused on the impact of immersive disposition and MD on the VR experience. We found no relationship between immersive disposition and MD in our sample. Immersive disposition is used to evaluate the potential to immerse a subject in a situation, whereas MD is characterized by the ability to be in the here and now. Therefore, it is possible that these 2 dimensions are unrelated. Our experiment showed that the involvement subscale of immersive disposition was associated with a loss of self-consciousness in terms of flow effect. An individual’s interest in a target object [94] or their motivational state in relation to a target object [95] has been described as a condition for flow experience in VR [96]. Furthermore, our experiment found that immersion was unrelated to presence, which conflicts with the literature [15,37]. A key difference compared with earlier work is that our participants were asked to make a physical effort. It is possible that this effort counteracted their immersive disposition. If we turn to the relationship among MD, presence, and flow, acceptance seems to be the most relevant dimension. Acceptance consists of accepting inner events such as emotions, thoughts, or beliefs as they are felt [97]. It does not mean resignation but rather perceiving one’s own experience with an attitude that acknowledges it, rather than judging it as either good or bad. Thus, the ability to accept what is happening now may be a more useful way to examine...
presence and cognitive control than simply being in the here and now. Collectively, these results suggest that physical exercise in VR may be improved by acceptance, which enhances the feeling of presence.

**APTICE Device and Exteroceptive Modulations**

Our final hypothesis concerning the relationship between subjective exteroceptive perceptions of sensory information and physical exercise in the VR experience was exploratory. On the one hand, our results show that there is an assumption that information provided by all 5 senses may help the user to become immersed in the experience of where they are, whom they are with, and what they are doing. The feeling of a real experience gives rise to presence. On the other hand, mindful participants pay more attention to information from their bodies, leading to better adaptation to the environment [98]. Using functional magnetic resonance imaging, Farb et al [99] identified several brain regions associated with mindfulness. In particular, they found that deactivation of the medial prefrontal cortex and increased activation of parietal areas were associated with proprioception and sensory–motor body experiences. Mehling [100] reported the use of external stimulation when attempting to understand how felt sensations are used internally to regulate stress or attention. Such information is integrated and linked to the person’s emotional state as a function of whether the body is experienced as safe [98,100].

Our results suggest that subjective preferences in exteroception-perception are linked to the experience of physical exercise in the VR environment. Furthermore, they show that immersion is correlated with subjective visual acuity. The participants in our experiment were cycled in a virtual environment based on natural visual information. Unsurprisingly, high scores for subjective visual acuity were associated with flow. Many studies have highlighted the potential of external sensory information to enrich the lived experience [101-104]. Exteroception information can generate intense emotional processes [105] and flavor manipulation within VR [101-104]. However, the evidence is weak, and it is also possible that such an environment may inhibit VR experiences because of its limited capacity to provide wider sensory inputs [105]. Another outcome of our study was that individual preferences may play a role in the VR experience. Our findings showed that this experience is negatively associated with all forms of external sensory stimulation (ie, hearing, taste, and smell) except vision. This suggests that other senses are partially inhibited, and only vision is recruited on a large scale. Vision is an essential component of the APTICE experience.

In this context, Slater and Usoh [106] suggested that an individual’s experience is encoded by visual, auditory, and kinesthetic systems of representation. Depending on the context, the person will naturally tend to favor one system over another. However, the latter authors noted that the visual system predominates in individuals who report a higher sense of presence and those who process information in the first person. Thus, individual characteristics may be a key factor in any experiment. Overall, our study suggests that the APTICE system may alter multisensory representations during physical exercise.

Future studies should address this issue, which remains unexplored.

**Future Clinical Applications**

VR technologies appear to complement established approaches to mental health care. Its association with physical activity makes it an interesting new approach that merits further investigation. Furthermore, the use of VR in health care is expanding rapidly. There are many new opportunities in clinical medicine, including mental illness, where VR may be an alternative treatment [4,5,107,108]. Our findings validated the impact of physical exercise in a VR environment on negative emotions in a healthy population. Although our results should be interpreted with caution, because of the small sample size, they highlight the importance of better understanding the processes involved in healthy participants. Beyond the efficacy of interventions to determine which populations might benefit from VR combined with physical activity, it is important to understand the processes that predispose this state in healthy individuals. Further studies with larger sample sizes are required to evaluate the role of these processes in clinical research. Thus, the next step is to study clinical and neurofunctional subtracts in a population with depression before proposing the tool as a countermeasure (ID-RCB: 2020-A03415-34) for this population and other people in health care. There is an untapped opportunity to use VR as a prevention tool and to target the processes that make an individual poorly adapted to the environment. This is particularly the case for people who work in challenging confined and isolated environments or extreme and unusual environments [65].

**Limitations**

This study has 4 main limitations. The first and most important factor is the small sample size. This study was intended to be a pilot feasibility study that will support a future controlled randomized trial. In this context, it validated the usefulness of the APTICE system and highlighted the interaction between the variables of interest. In the next phase of our work, we will launch a larger clinical study of participants with depression. The second limitation relates to the use of subjective self-report measures. An objective sensory evaluation needs to be developed for healthy participants, which would help researchers to better investigate the human-body relationship. Subjective variables should be combined with physiological measures, such as heart rate variability. Third, our results cannot be generalized because the study population was recruited from among armed forces personnel and footballers, who are usually different from the general population in terms of fitness and psychological state. Finally, the last limitation concerns the VR equipment used in our experiment, which is becoming dated. A new version of the Oculus headset is already available, with a better graphics interface.

**Conclusions**

This exploratory proof-of-concept study investigated some of the processes implicated in physical exercise in a VR environment with the aim of better understanding their relationship with psychological state in a sample of healthy individuals. It represents the first step in a larger randomized
controlled trial that will investigate clinical and neurofunctional
subtractions in a population with depression. Our results suggest
that the APTICE environment can change negative emotional
states, consistent with the experiences of flow and presence.
Moreover, our findings demonstrate that immersive and mindful
disposition play an important role in the VR experience. Finally,
they also suggest that the subjective exteroceptive perception
of sensory information may be a key aspect and seems to
indicate that one sense may prevail over another at the level of
the individual. Our study has several implications for clinical
medicine: (1) VR can help enhance and reinforce the beneficial
actions of physical activity; (2) APTICE is a promising system
and may be effective in improving mental health; and (3)
APTICE has the potential to be used as an alternative treatment
to drugs and to improve quality of life. However, many
questions remain unanswered, and further work is needed to
exploit the potential of VR associated with physical activity
both as prevention and treatment.

Acknowledgments
The authors wish to thank Eric Malbos. The authors thank the French Armed Forces Biomedical Research Institute and French
Football Federation for their involvement. They would like to thank Florent Samat and Nicolas Greverie from GAMIT France
for the development of the APTICE system. Finally, the authors would like to thank the Centre Nationales des Études Spatiales
and the Direction Générale de l’Armement for their support.

This research was funded by the Direction Générale de l’Armement (N 2016/017/S).

Data Availability
The data sets generated and analyzed during this study are not publicly available because they are the property of the French
Armed Forces Health Service. Although data are not available to the public, they are available from author MT upon reasonable
request.

Authors’ Contributions
BLR, CMK, CP, RR, EM, FB, EG, and MT conceptualized the study and developed the methodology. All authors actively
contributed to the experimental process. BLR, CMK, and MT wrote the original manuscript. BLR conducted the formal analysis.
All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Augmented Physical Training for Isolated and Confined Environments (APTICE) system description.
[DOCX File, 2387 KB - formative_v811e45637_app1.docx ]

Multimedia Appendix 2
Significant correlations between tested variables as a function of the 3 hypotheses.
[DOCX File, 31 KB - formative_v811e45637_app2.docx ]

References
Res 2017;9(9):3867-3880 [FREE Full text] [Medline: 28979666]
33401596]


Québecoise pour la Recherche en Psychologie (SQRP); November 1-3, 2002; Trois-Rivières, Canada. [doi: 10.1037/h0081306]


Abbreviations

- AD-ACL: Activation-Deactivation Adjective Checklist
- AIEQ: Addictive Intensity Evaluation Questionnaire
- APTICE: augmented physical training for isolated and confined environments
- MD: mindfulness disposition
- SPANE: Scale of Positive and Negative Experience
- UX: user experience
- VR: virtual reality

©Barbara Le Roy, Charles Martin-Krumm, Charlotte Poupon, Raphaëlle Richieri, Eric Malbos, Fanny Barthélény, Eric Guedj, Marion Trousselard. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Experiences Among Patients With Cystic Fibrosis in the MucoExocet Study of Using Connected Devices for the Management of Pulmonary Exacerbations: Grounded Theory Qualitative Research

Maxime Morsa¹,², PhD; Amélie Perrin³; Valérie David³, Dr med; Gilles Rault⁴, Dr med; Enora Le Roux⁵,⁶, PhD; Corinne Alberti⁵,⁶, Prof Dr; Rémi Gagnayre², Prof Dr; Dominique Pougheon Bertrand², PhD

¹Adaptation, Resilience and Change Research Unit, University of Liège, Liège, Belgium
²Laboratory Health Promotion and Education (UR3412), Sorbonne Paris North University, Bobigny, France
³Paediatrics CF Centre, Nantes University Hospital, Nantes, France
⁴Laboratoire Educations et Pormotion de la santé, Université Sorbonne Paris Nord, Bobigny, France
⁵Institut National de la Santé et de la Recherche Médicale (UMR 1123 ECEVE), Université de Paris, Paris, France
⁶Unit of Clinical Epidemiology (CIC-EC 1426), Hôpital Universitaire Robert Debré, Assistance Publique des Hôpitaux de Paris, Institut National de la Santé et de la Recherche Médicale, Paris, France

Corresponding Author:
Maxime Morsa, PhD
Adaptation, Resilience and Change Research Unit
University of Liège
1 Place des Orateurs
Liège, 4000
Belgium
Phone: 32 4 3662272
Email: maxime.morsa@uliege.be

Abstract

Background: Early detection of pulmonary exacerbations (PEx) in patients with cystic fibrosis is important to quickly trigger treatment and reduce respiratory damage. An intervention was designed in the frame of the MucoExocet research study providing patients with cystic fibrosis with connected devices and educating them to detect and react to their early signs of PEx.

Objective: This study aims to identify the contributions and conditions of home monitoring in relation to their care teams from the users’ point of view to detect PEx early and treat it. This study focused on the patients’ experiences as the first and main users of home monitoring.

Methods: A qualitative study was conducted to explore patients’ and professionals’ experiences with the intervention. We interviewed patients who completed the 2-year study using semistructured guides and conducted focus groups with the care teams. All the interviews were recorded and transcribed verbatim. Their educational material was collected. A grounded analysis was conducted by 2 researchers.

Results: A total of 20 patients completed the study. Three main categories emerged from the patients’ verbatim transcripts and were also found in those of the professionals: (1) task technology fit, reflecting reliability, ease of use, accuracy of data, and support of the technology; (2) patient empowerment through technology, grouping patients’ learnings, validation of their perception of exacerbation, assessment of treatment efficacy, awareness of healthy behaviors, and ability to react to PEx signs in relation to their care team; (3) use, reflecting a continuous or intermittent use, the perceived usefulness balanced with cumbersome measurements, routinization and personalization of the measurement process, and the way data are shared with the care team. Furthermore, 3 relationships were highlighted between the categories that reflect the necessary conditions for patient empowerment through the use of technology.

Conclusions: We discuss a theorization of the process of patient empowerment through the use of connected devices and call for further research to verify or amend it in the context of other technologies, illnesses, and care organizations.

Trial Registration: ClinicalTrials.gov NCT03304028; https://clinicaltrials.gov/ct2/show/results/NCT03304028
Introduction

Background

Pulmonary exacerbations (PEx) are the main cause of decline in lung function in patients living with cystic fibrosis (CF), representing the leading cause of death. Recommendations emphasize the importance of diagnosing PEx early to treat patients effectively and for them to have the best chance of regaining their previous baseline lung function after treatment [1]. Identifying warning signs of PEx requires many skills from patients daily—studies have shown that they must be able to monitor a combination of physiological parameters and patient-reported perceptions, such as weight loss, decreased spirometry, increased coughing, or increased sputum production reported daily, to diagnose PEx episodes and put in place the appropriate treatment [2,3]. Nevertheless, patients living with CF do not systematically monitor these warning signs, as few are equipped with devices to monitor variations in their physiological parameters or their perceptions over time, with the exception of patients who have received a lung transplant, who may be equipped with spirometers to detect a decrease in their respiratory function, which is a warning sign of acute rejection. However, patients need to access accurate and reliable measurements to monitor their lung function.

In recent years, a contemporary trend has emerged in health improvement and disease prevention: the “quantified self.” It refers to the quantitative measurement of various parameters linked to the state of one’s health (eg, heart rate and weight) or to lifestyle (eg, diet and physical activity) to monitor a disease or improve well-being. The premise is that one cannot improve what they cannot quantify. This quantification, which was still difficult to achieve a few years ago, has become more accessible through the development of new technologies and connected devices. These devices are connected to the internet and can collect, store, process, and transmit health-related data through sensors [4].

Connected devices can help patients gain a better understanding of disease and treatment and increase their levels of satisfaction and adherence to treatment when combined with patient education interventions [5,6]. Patient education is an empowerment approach for patients with chronic diseases aiming to improve their understanding and adherence to treatment by transferring knowledge from health care providers to patients through educational workshops and also by using patients’ experiential knowledge, which helps them adjust their management of the disease in their daily lives [7]. Patient education is known to have a significant positive impact on bioclinical indicators and on the well-being of patients [8]. Connected devices would act as a learning aid for patients by promoting real-life behavioral experimentation thanks to quick (or even immediate) access to objective data and to the development of knowledge about oneself anchored in one’s memory [9]. The use of connected devices by patients in their daily lives allows them to transfer what they learned during the workshops provided by health care providers to real-life situations, thus expanding on patient education. Experiential and continuous learning is facilitated when it is supported by health care providers to learn to interpret real-life data and compare them with the data collected at the hospital.

This way, connected devices could promote the process of empowerment, a concept that is understood as the development of patients’ ability to identify and meet their own needs, solve their own problems, mobilize the necessary resources to take action, and feel that they are in control of their health and their own lives [10]. According to Funnell and Anderson [11], empowerment is a process that is facilitated by counseling, educational, or psychological techniques to help the individual take control of the day-to-day management of their illness.

Currently, data are scarce on how connected devices are used in real-life situations by patients with chronic diseases and on how they influence knowledge of oneself and of one’s body, health, and disease [12]. However, we know that the dropout rate of connected devices can be high because of how cumbersome their use may be or the fact that they are too pressing a reminder of the person’s disease in their daily life [13,14], whereas adherence is mainly observed in young people and high-income socioprofessional categories who are more familiar with new technologies [15]. People’s experiences of using such connected devices vary depending on the person, the context, and their care environment. Therefore, the assessment of health technology is now moving toward a contextualized, patient-based evidence approach. According to this approach, the evaluation of eHealth devices is based on knowledge that originates directly from patients about their experiences of health, quality of life, and health services [16]. This approach is represented internationally by the work of the Warwick Patient Experiences Framework or the National Institute for Health and Care Excellence Patient Experience Guideline Development Group [17].

Drawing from the humanities and social sciences, it is now recommended for qualitative studies to be centered on patients’ feedback to understand the processes through which connected devices facilitate their acquisition of knowledge (of the body, risks, and diseases), in particular through the intimate and empirical experiences of the quantified body translated into data [18].

Objectives

Therefore, we conducted a qualitative study with patients living with CF and with specialized CF centers in metropolitan France to explore the processes through which connected devices become an essential part of patients’ knowledge to allow them to self-manage their health and to contribute to a theory of individual patient empowerment through technology. The aim
of the study was to understand how patients and health care providers lived and perceived this new intervention based on connected devices associated with patient education workshops to identify the contributions and conditions of home monitoring. The work is focused on stakeholders’ experience with the intervention. This study is part of an interventional project based on the hypothesis that an intervention that combines the provision of connected devices set up with personalized alert thresholds and a patient education intervention by health care providers can enable patients with CF to detect early signs of PEx and begin managing it themselves in a timely manner. For this self-management process to lead to the implementation of appropriate patient behavior, it is assumed that the educational intervention teaches patients to identify and respond appropriately to alerts.

**Methods**

**Overview**

The MucoExocet (from the French for “Cystic Fibrosis Exacerbation Connected Devices Therapeutic Education”) study, a pilot interventional study, was conducted from 2018 to 2021 and involved 22 adults and 14 adolescents (aged >12 years) with CF to assess whether the use of connected devices was feasible and useful to detect and treat PEx early (trial registration: ClinicalTrials.gov NCT03304028). As part of the overall research project, this qualitative study explored the users’ experiences at the end of the intervention. The intervention and protocol have been extensively described previously [19]. We used the EQUATOR (Enhancing the Quality and Transparency of Health Research) standards for reporting qualitative research elaborated by O’Brien et al [20] to present our study design.

**Summary of the Intervention in Its Context**

Since 2005, a national organization associating health care providers from CF centers and patients and parents in France has been working to define the patient and parent competency framework (in pediatrics) and the associated set of educational tools. A therapeutic education tool named “React to PEx” (“Réagir en cas d’exacerbation”) was used to support patients’ and parents’ self-management of PEx episodes at home (Multimedia Appendix 1).

The intervention designed for the MucoExocet study combined the provision of connected devices with an educational program based on the React to PEx tool. It was renamed “React with CDs” and incorporates measurements from connected devices and personalized alerts (Figure 1). The goal of the intervention was to develop the patients’ (or parents’) ability to take action at the first signs of exacerbation identified through measurement deviations by connected devices. For this study, connected devices were used to collect 13 parameters, including 6 physiological parameters measured by the devices (forced expiratory volume in 1 second [FEV1], cardiac frequency, arterial hemoglobin oxygen saturation, weight, sleep duration [min/night], and physical activity [step count/d]) and 7 patient-reported perceptions described using emoticons in a journal provided by the spirometer application (trouble breathing, need for more airway clearance, increased symptoms at night, difficulty performing usual activities, greater fatigue, loss of appetite, and change in sputum [color or quantity]). At the request of both physicians and patients, the option chosen in the study was to not send the data collected via connected devices to the physician but only to the patient. However, the data could be shared during a consultation at the center or during a phone call or email exchange if the patient (or parent) wished to do so.

**Figure 1.** Design of the study intervention. CD: connected device; EW: educational workshop; M0: month 0; M3: month 3; M6: month 6; M9: month 9; M12: month 12; M15: month 15; M18: month 18; M21: month 21; PEx: pulmonary exacerbation.
Thresholds of irregular variation for patients’ parameters were calculated using the cumulative sum control chart method based on the data collected during phase 1 of the implementation of connected devices (3 months); this allowed physicians to set alert thresholds for each parameter and each patient during the first educational workshop with patients. Thus, alerts could be sent by email to the patients or parents throughout the period of routine monitoring using connected devices (12 months). An action plan was agreed upon by the physician and the patient or parent during this educational workshop to respond to alerts.

Development of the Educational Tool and Educational Intervention in the Centers

One of the centers played a leading role in the implementation of the study because of its leading role in therapeutic education for patients with CF in France (the CF center in Nantes). During phase 2, the physician and therapeutic patient education nurse developed the different educational tools for the 3 stages of the program (educational workshops 1-3) based on the statistical analyses of patient data and with the participation of an adult patient expert and tested these tools with a parent and an adolescent patient from their center. The tools and educational program are described in the publication cited previously [19].

Beyond the patient recruitment process, a physician and a health care provider (nurse or physiotherapist) from each investigating center were involved in handing the connected devices to the patients or parents, setting them up with the patients or parents and explaining how to use them, solving technical problems with the patients or parents with the help of the device suppliers, and participating in the interpretation of statistical analysis of the data collected during the implementation phase (3 months) to define personalized alert thresholds for their patients. They were trained in the use of the React with CDs educational tool to agree with their patients on an action plan in the event of PEx; they educated their patients in the 3 sessions provided as part of the educational program (“Your impressions of using connected devices during the implementation phase” and “Your action plan for responding to exacerbation warning signs, and Review of your action plan after 6 months of routine monitoring”). At the end of the study, health care providers participated in a focus group to report on their experience with the study and this monitoring method.

Study Population and Study Centers

The centers were selected by the research group on a voluntary basis among centers who had participated in the quality improvement program. They were familiar with the educational tool “React to an exacerbation.” The 7 CF centers were selected to include patients and their families of various conditions of life, economic statuses, and geographic areas (either urban or rural). Finally, the 7 centers were located in 4 different geographical areas; 3 (43%) were pediatric centers (4 patients per center), and 4 (57%) were adult centers (6 patients per center). None had previous experience with connected devices for their patients at the initiation of the study. In total, 36 patients, adults or adolescents, were included in the MucoExocet study. The sample was defined according to the recruitment capacities of the centers and the possibility of observing a saturation phenomenon in the qualitative study [21].

The recruitment process conducted by physicians in the centers was carried out based on patients’ voluntary participation and their interest in using new technologies. The participant inclusion criteria were as follows: age of ≥12 years, clinically stable condition (no PEx requiring intravenous antibiotics within the previous 4 weeks), at least one PEx within the previous 12 months, current follow-up at a participating CF center (and no plans to change centers during the course of the study), no history of having undergone solid organ transplants, prescription of at least one pulmonary medication (eg, inhaled mucolytic, inhaled or oral antibiotic therapy, or hypertonic saline), ability to speak French, ability to connect a tablet to Wi-Fi, and provision of written informed consent.

The number of individuals to be recruited took into account an estimated dropout rate of 20%. A form was offered to the patients leaving the study to identify the main reasons for their withdrawal (Multimedia Appendix 2).

Data Collected by the Patients Using the Connected Devices

The 13 parameters were collected using 5 different connected devices, and a few of them collected more than one parameter. During the first 3 months, data were to be collected twice a week. During the routine phase, the frequency was agreed upon between the patient and the care team from twice a week to once every 2 weeks depending on the patient’s health outcomes and life conditions. The procedure used to collect the data was explained in a document given to the patients at inclusion (Textbox 1).

Owing to the great variety of measurements taken, the time spent on the measurements was not recorded.

Textbox 1. Procedure to collect the data.

- Data collected without any participation from the patient (sleep, steps, and cardiac frequency): the duration of sleep was measured by the sensor under the mattress, and the step count and cardiac frequency were measured by the watch.
- Clinical data (spirometry and oxygen) required patient participation; at the end of the spirometry measurement, emoticons were presented for each of the 7 perceptions.
- Weight was expected to be measured in the morning (naked) the same day as the clinical data.

Qualitative Data Collection

Patients’ experiences were collected through semistructured interviews using an interview guide with 8 open-ended questions (Textbox 2), derived and adapted from validated protocols for patient narrative elicitation in outpatient care experiences [22]. The experience and workload of the care teams were explored...
Three sources of data were collected: (1) data collected during patient or parent interviews using an interview guide with open-ended questions (Textbox 2), (2) data regarding the educational program with the physician (the documents completed by the patient and the clinician, including the personalized action plan in case of PEx), and (3) data from the focus groups with care teams at the end of the study using a semistructured guide (Textbox 3).

Textbox 2. Guide for the semistructured interviews with patients or parents.

1. For you, what are the most important aspects in the management of your respiratory exacerbations in your daily life?
2. How do you rate the conditions for managing exacerbations during the study (based on what is most important to you)?
3. Can you tell us about a positive experience you had during this study concerning the management of your exacerbations? What happened and how did it make you feel? Did you do anything in particular after this positive experience (eg, change your attitude or behavior)?
4. Can you tell us about an experience that turned out differently than you expected? What happened and how did you feel at the time?
5. Regarding this last experience where you wished things had turned out differently, did you or your doctor do anything to rectify the situation?
6. Did your participation in the study change your outlook on the way you manage your exacerbations?
7. What do you think would be the best way to integrate this type of long-term follow-up so that it addresses the aspects that are most important to you in the management of your exacerbations?
8. Is there anything else you wish to tell us about?

Textbox 3. Guide for the focus groups with care teams.

1. From the point of view of the health care team, what are the most important aspects in the management of patients’ respiratory exacerbations, particularly in their daily lives?
2. In your opinion, how have the proposed monitoring methods, including connected devices and patient education, addressed these priorities or with what limitations?
3. During this research project, what changes have you noticed in the way the team works or in its workload with regard to monitoring patients for the management of their exacerbations? Have you noticed a change in your relationship with the patients’ out-of-hospital physiotherapist?
4. What difficulties or bad experiences have you had in the process of managing patient exacerbations using connected devices?
5. Do you feel that you had positive experiences during this study with the management of patient exacerbations? How would you rate these experiences in relation to the most important aspects of the management of respiratory exacerbations?
6. In your opinion, should this type of long-term patient follow-up be included in the management of exacerbations or in other aspects of their management? If so, what would be the best way to integrate it and for which patients and with which objectives?
7. Is there anything else you wish to tell us about?

Analysis Framework

All the interviews were transcribed verbatim and subjected to a descriptive qualitative analysis. The analysis framework used was grounded theory [23].

Grounded theory is a qualitative research method with an inductive approach aimed at constructing a theory on a cultural, social, or psychological phenomenon by proceeding with the progressive and valid conceptual representation and mapping of qualitative empirical data [24]. In this study, the phenomenon explored was learning and empowerment in health management through the use of connected devices. Grounded theory is relevant as this phenomenon is currently sparsely studied. Studies on connected devices in patients with chronic conditions, and especially in patients with CF, are mostly intended to demonstrate the efficacy of the use of connected devices on various health outcomes. The theories mostly reported in the literature, such as digital behavior change interventions or the theory of reasoned action, are mainly focused on compliance with connected devices. However, the concept of empowerment includes other dimensions, such as understanding, the ability to decide, and self-assessment. Using grounded theory, we aimed to complete the current knowledge by eliciting the various dimensions of empowerment from the patient experiences with the use of connected devices for remote monitoring of their symptoms and by identifying elements that could enrich the theories in the field of remote monitoring.

According to the constructivist grounded theory method by Charmaz [25], which focuses on social processes or actions and the meaning of human actions, we adopted a social psychological approach to explore how and in which context individuals feel that connected devices have an impact on their learning to take care of themselves and on their empowerment. In grounded theory, verbatim transcripts are analyzed using codes to highlight what was stated by the participants in the study and derive meaning from it. We applied the standard steps of grounded theorizing. In initial coding, we generated as many ideas as possible inductively from the initial data. In focused coding, we relied on the most prevalent and important codes to select the central codes for analysis. In theoretical coding, we refined the final categories of the theory by connecting them to
each other, thus allowing for the integration of the categories into a model and the construction of a theory on the phenomenon studied [26].

The grounded dimensional analysis of patients’ or parents’ and health care providers’ data was conducted by 2 researchers (MM and DPB) using NVivo (QSR International) taking into account their evolution over the course of the study and the various natures and production conditions of the collected material while constantly comparing the data within and across patients or parents and health care providers. A constant comparison between the verbatim transcripts of patients and health care professionals was carried out to bring out the invariant elements as “the essence of the phenomenon,” which elaborate “conceptual categories” remaining as close as possible to the lived realities of patients [27]. The 2 researchers who analyzed the verbatim transcripts were a psychologist and the parent of a child with CF, and both had PhDs in public health and great experience in qualitative research.

Analysis of the Educational Documents

The educational documents filled in by the physician and the patient during the educational session (educational workshop 2) were collected by the research team and reviewed globally but were not analyzed in connection with the patient interview. The aim was to understand which actions had been agreed upon between the patient and the physician when symptoms of a PEx were detected by the patient at home (central column) and whether they could resolve the PEx episode and prevent deterioration through their actions.

Ethical Considerations

The research project was submitted for evaluation by the Committee for the Protection of Persons designated randomly under conditions provided for in the Code of Public Health (Article L. 1123-14). The study was approved by the Committee for the Protection of Persons (CPP North West III) on June 10, 2017 (2017-A00723-50). Free and informed consent was obtained before any act related to research was undertaken.

Results

Population Interviewed and Dropout Rates During the Study

A total of 56% (20/36) of the study participants were interviewed. The population interviewed in relation to the population included in the study and who benefited from the different stages of the educational program (educational workshops 1-3) is listed by center in Table 1. The dropout rate at the end of the first phase of intensive data collection (3 months) was 25% (9/36). In total, 3% (1/36) of the patients died during the study. The death was unrelated to the study. A total of 67% (24/36) of the patients were educated in the first 2 workshops (educational workshops 1 and 2), allowing them to enter the routine monitoring phase using connected devices. Only 39% (14/36) of the patients attended the third educational workshop held at the midpoint of the routine monitoring phase using connected devices. At the end of the study, the nonresponse rate to interview solicitations compared with the number of patients who entered the routine monitoring phase was 25% (6/24). These results differed from one center to another. The gender, age, and geographic area characteristics of the patients interviewed (presented in Table 2) were similar to those of the entire study population. However, the patients interviewed had a higher level of education and employment rate than the entire study population.

A total of 12 health care providers from 7 hospitals participated in focus groups between May 2020 and February 2021 (Table 3).

Table 1. Number of patients interviewed per center (n=20).

<table>
<thead>
<tr>
<th>Patients included (n=36), n (%)</th>
<th>Patients educated, n (%)</th>
<th>Patients interviewed (n=20), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EW¹ 1 (n=30)</td>
<td>EW 2 (n=24)</td>
</tr>
<tr>
<td>Pediatric CFb centers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4 (11)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>2</td>
<td>5 (14)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>3</td>
<td>5 (14)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Adult CF centers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8 (22)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>5</td>
<td>6 (17)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>6</td>
<td>3 (8)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>7</td>
<td>5 (14)</td>
<td>4 (13)</td>
</tr>
</tbody>
</table>

¹EW: educational workshop.

bCF: cystic fibrosis.
Table 2. Characteristics of the study participants (patients; n=20).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>Adolescents, n (%)</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Adults, n (%)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Age of adolescents (years; n=8), median (SD)</td>
<td>14.5 (1.1)</td>
</tr>
<tr>
<td>Age of adults (years; n=12), median (SD)</td>
<td>29.6 (7.7)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Female</td>
<td>11 (55)</td>
</tr>
<tr>
<td><strong>Geographical area, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Living in a city</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Living near a city</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Living in the countryside</td>
<td>3 (15)</td>
</tr>
</tbody>
</table>

Table 3. Focus group participant characteristics.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Participants (n=20), n (%)</th>
<th>Date</th>
<th>MD^a (n=3), n (%)</th>
<th>Nurse (n=4), n (%)</th>
<th>Physiotherapist (n=3), n (%)</th>
<th>Other (n=2), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 (10)</td>
<td>November 6, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>2</td>
<td>3 (15)</td>
<td>February 5, 2021</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>3</td>
<td>3 (15)</td>
<td>November 9, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>4</td>
<td>3 (15)</td>
<td>October 11, 2020</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td>1 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>5</td>
<td>4 (20)</td>
<td>December 9, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
<td>1 (50; coach in physical activities)</td>
</tr>
<tr>
<td>6</td>
<td>3 (15)</td>
<td>June 23, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>7</td>
<td>2 (10)</td>
<td>May 19, 2020</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (50; clinical research assistant)</td>
</tr>
</tbody>
</table>

^aMD: doctor of medicine.

The forms completed by patients exiting the study reported technical difficulties with certain connected devices, in particular with the tablet computer provided to synchronize data before sending them to the research server. These difficulties were also reported by health care providers in the focus groups, who mentioned that a member of the team (nurse, physiotherapist, or clinical research associate) had spent a significant amount of time solving technical problems with the device suppliers, sometimes unsuccessfully. The reasons for dropping out of the study were multiple and are listed in Multimedia Appendix 3.

Results From the Educational Documents

The educational documents collected after educational workshop 2 show that the first column (“normal state of health, routine activities and treatments actually followed”) was filled with detailed information on the treatments and activities of the patient in their daily life, unlike the central column, which contained little information. The agreed upon actions in case of signs of exacerbation were mainly “increase physiotherapy” or “try to do more physical activity” and always “call or send a message to the center team.” The actions were aimed more at the diagnosis of the exacerbation by the physician, who then decided what the patient should do, than at the actions that the patients should take by themselves. Most of the physicians added the following comment—“They already know what to do”—meaning that they had not delegated new actions to the patients. One pediatrician decided to give conditional prescriptions of oral antibiotics to the parents after the educational session, thus delegating to them the decision to start the treatment and asking them to inform the team that they had started the treatment.

Descriptive Results From the Interviews

Stage 1: Initial Coding

A total of 12 codes emerged from the patients’ verbatim transcripts. In total, 10 codes emerged from the health care providers’ verbatim transcripts. The analysis allowed for the assignment of a name to each code that identified its area of interest (Table 4).
Table 4. Codes and categories of transcripts from patient interviews and focus groups with health care providers.

<table>
<thead>
<tr>
<th>Category and codes from patient interview transcripts</th>
<th>Codes from transcripts of caregiver focus groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1: patient empowerment</strong></td>
<td></td>
</tr>
<tr>
<td>• Learnings</td>
<td>• Learnings</td>
</tr>
<tr>
<td>• New knowledge mentioned by patients that helps them understand alerts and manage PEx&lt;sup&gt;a&lt;/sup&gt;</td>
<td>• Confirmation of the patients’ perceptions of symptoms using measurements; better understanding of their state of health at the first signs of exacerbation</td>
</tr>
<tr>
<td>• Empowerment</td>
<td>• Patient-physician relationship</td>
</tr>
<tr>
<td>• Impression of being more capable of self-managing their treatments, their health, and their life projects</td>
<td>• Better understanding by the physician of the patient’s situation, their life circumstances, and their care; better understanding by the physician of the treatments carried out and the patient’s behavior in the event of exacerbation or in life in general</td>
</tr>
<tr>
<td>• Loss of control</td>
<td>• Remobilization of the team to manage PEx</td>
</tr>
<tr>
<td>• Impression of being less capable of self-managing their care, health, and life projects</td>
<td>• Renew the motivation of the teams to focus on the main objective of jointly managing PEx through a different approach with the patient</td>
</tr>
<tr>
<td><strong>Category 2: TTF&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
</tr>
<tr>
<td>• Perceived usefulness</td>
<td>• Usefulness of monitoring using CDs</td>
</tr>
<tr>
<td>• Needs expressed by patients to monitor PEx and expectations of the use of CDs&lt;sup&gt;c&lt;/sup&gt; to help them self-monitor</td>
<td>• Depending on the patient’s health status (unstable or stabilized), on the caregiver’s previous experience with telemonitoring, and on the patient’s ability to use devices and keep them in good operating condition</td>
</tr>
<tr>
<td>• Perceived reliability</td>
<td>• Technical reliability and accuracy of measurements</td>
</tr>
<tr>
<td>• Patients’ level of trust in the reliability of the data collected by the devices during the study</td>
<td>• Checking the accuracy of the measurements taken using CDs in comparison with hospital standards and reliability over time</td>
</tr>
<tr>
<td>• Negative experiences</td>
<td>• Negative experiences</td>
</tr>
<tr>
<td>• Problems encountered using CDs; negative consequences described by patients</td>
<td>• Problems encountered using CDs and negative consequences described by people—one death that was not related to the study but that CDs did not prevent</td>
</tr>
<tr>
<td><strong>Category 3: use of technology by patients and health care providers</strong></td>
<td></td>
</tr>
<tr>
<td>• Conditions for a favorable use of CDs</td>
<td>• Conditions of integration of the use of CDs into the organization of care</td>
</tr>
<tr>
<td>• Technical, human, and environmental conditions of CD use considered favorable for the optimal management of PEx</td>
<td>• Technical, human, and organizational conditions for the health care team to integrate the support of the use of CDs by patients—resources and time needed for education and remote support of patients</td>
</tr>
<tr>
<td>• Motivation</td>
<td>• Factors of motivation in health care providers</td>
</tr>
<tr>
<td>• Personal and contextual factors that motivate patients to use CDs</td>
<td>• Monitoring method that cannot be overlooked considering the current demographic increase in the number of adult patients; necessary monitoring method (using telecommunications) in case of a crisis (COVID-19)</td>
</tr>
<tr>
<td>• Hindrances</td>
<td></td>
</tr>
<tr>
<td>• Personal and contextual factors negatively affecting the use of CDs</td>
<td></td>
</tr>
<tr>
<td>• Support from health care providers in the use of CDs</td>
<td>N/A&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>• Support provided by health care providers in the use of data from CDs for the management of PEx that helps promote the use of CDs by patients</td>
<td></td>
</tr>
<tr>
<td>• Use of CDs</td>
<td>N/A</td>
</tr>
<tr>
<td>• Modalities of CD use reported by patients</td>
<td></td>
</tr>
</tbody>
</table>
of patients’ and health care providers’ perspectives. These categories are independent from one another and do not include the same codes. At this stage of analysis, some codes (“Patient recommendations” and “Implementation approach”) were set aside as they did not correspond to the modeling purpose. They will be considered later in the model. Each category was defined with a general title, a description, and detailed transcripts.

**Category 1: Patient Empowerment**

During the process of analysis, empowerment was defined as individual empowerment, characterized by the learning achieved during the intervention, the decisions and actions implemented by the patient for their care or health, and their sense of control over their health. Connected devices allow patients to access data on their health status daily to monitor episodes of PEx, prevent them, and adjust the course of action when they happen. They contribute to making some patients more autonomous in the early management of PEx by supporting their decision-making and ability to take action without seeing or contacting primary health care providers:

> CDs allow for a better assessment of one’s health status, and to take better care of oneself. It helps to be more autonomous and to avoid waiting until we are very sick to go to the doctor’s. It also helps to complement one’s care with extra physiotherapy, more sports, things like that... [Adult patient]

The use of connected devices in the management of PEx not only allows for the adoption of preventive behaviors or better adherence to medical recommendations. Through their use of connected devices, patients also learn to manage their health with new data about themselves that confront the objective evolution of their health status with the way they feel and the effects of their lifestyle and their attitude toward their care. This process is characterized by the acquisition of new knowledge of one’s state of health, the validation of subjective perceptions, a better understanding of what happens on a physiological level, and focusing more attention on certain monitoring indicators. These learnings can be observed as early as during adolescence:

> Well, I found that the fact that I could make my own measurements... allowed me to understand better... to be able to compare and to feel when I was not doing so well. And for example, I found it interesting when I thought I was doing less well but still had good results. I waited a little while to see if I should rely more on the results or more on how I felt, and I actually relied more on my results. And, yes, I thought it was good, because it’s mostly meant for prevention, and it helped me a few times. [Teenage patient]

However, it has been found that measurements reflecting a deterioration can lead to higher stress levels and a loss of empowerment when no action plan has been put in place in advance with health care providers. A lack of patient education to support the understanding of data, including the meaning of alerts, can cause a feeling of helplessness in patients, especially if the caregiver also appears to be confused by the new monitoring method. Obtaining data on one’s health status on a near-continuous basis only enhances patient empowerment if the patient possesses the skills to interpret and act on the data. Similarly, connected devices lead patients to think almost constantly about their health despite not always being in the right mental state to do so:

> In the past, I probably used to desaturate without really realising it. I probably had headaches, but there you go... But now, I’m constantly stressed out, because I check my measurements pretty much all the time. [Adult patient]

> Well, no. I can’t see the results of the measurements on the connected devices, and when the nurse called customer services, she was told that it was me who had the data anyway. But I don’t understand the data, and neither does she. So, perhaps we need to be taught how to interpret them better or to get clearer explanations in the alerts we receive. [Adult patient]

Health care providers focused more on the concordance between the perceptions reported by patients and the data collected than on patients learning to manage their PEx. Empowerment was seen by health care providers as patients’ ability to detect PEx and respond to it to limit its effects. Health care providers’ objective was to ensure the **effectiveness of the device in**
improving patients’ state of health. Some health care providers emphasized the beneficial relational change fostered by the educational intervention that accompanied the implementation of connected devices, owing to which they communicated better with their patients. This allowed them to better understand their living conditions with the disease and how they cared for themselves daily. From their point of view, the data collected using connected devices increased patients’ level of information and of awareness of their condition. This gave the team the feeling of having a new tool to involve patients in the management of PEx and, thus, the capacity to influence the evolution of the disease:

*Behind the word “anticipation,” we mean they should know how to spot the early signs and manage to put things in place and then call us. They should know not to wait for one or two weeks before calling to tell us they haven’t been feeling well for two weeks. So, for me, the study had an aspect of therapeutic education, thanks to the information panel (“React with CDs” Tool) that allowed us to sit down with the patients and have them think a little bit about what exactly they were doing.* [Health care provider]

**Category 2: Adequacy of Technology (Combined With Education) to the Needs of Patients**

Often found in the literature as “task technology fit” (TTF) [28], this category includes aspects related to the reliability of the devices, the accuracy of the data measured (in comparison with a standard), the ease of access to the data, and how adequate the educational program is, all of which shape patients’ perception of how well this “technology” fits their monitoring needs. Patients expressed concern that the devices should accurately reflect their condition. The adequacy of the devices for monitoring purposes can be assessed based on several criteria over time as patients experience the use of connected devices. The first criterion is the perceived reliability of connected devices over time and the accuracy of the measurements compared with measurements taken at the hospital:

*When I took several measurements, I sometimes got very contradictory results. I sometimes wasn’t sure whether it was reliable.* [Adult patient]

The second criterion is the ergonomics and ease of use to connect to the tablet to access data and send them to the research server, which enables the sending of alert notifications:

*But the fact is that the spirometer...it does not save the results. So, I could get a good score at the beginning, but I tried again and because I coughed a little bit the result wasn’t so good, so I started again from the start, but it’s a bit difficult. Results should be saved automatically.* [Teenage patient]

The third criterion is the technical support provided for the implementation of the devices:

*What bothered me was that the curves on the graph—there were two curves—I never knew what they represented. And I even asked the nurse, and the nurse replied: “Indeed, it’s weird, what does it mean?” Even she searched for an explanation. To*

*this day, I don’t actually have the answer... [Adult patient]*

These aspects were supposed to be controlled in the context of interventional research, but some patients had a disappointing experience even though they were aware that they were participating in a pilot study that would sometimes involve “teething problems”:

*We told them that they were the first ones to go through all the steps and that everything was not necessarily perfectly set up for them...I told them that future patients would have fewer difficulties because we would manage to solve some things with them in the study. They tested the tools from beginning to the end and therefore experienced all the computer bugs.* [Health care provider]

The reliability of the devices used to monitor patients was mainly assessed by health care providers in comparison with the measurements taken using standard hospital equipment. This reliability was, from their point of view, guaranteed by the research context. Some patients felt that their health care providers did not have answers to the technical problems they faced:

*I received several emails from the CF centre telling me that they were not getting the data. But I assured them that I was sending the results. I managed to show them that I had uploaded the data...I went onto HealthMate as I was getting an update every Sunday by email for the Withings devices. So, I forwarded it to them, and in fact, they said that the data were loading, but not on the research server.* [Adult patient]

Patients’ interest in technology may vary according to the connected devices proposed, the need they feel to monitor certain health indicators, and the attractiveness of the device. Moreover, patients may not wish to use them for fear of being confronted with poor results on certain critical measurements for the patient (or for the physician):

*So, I found the sleep analysis option rather useful. Because I do sleep well at night, but I cough without realising it. I was either a little tired when I woke up in the morning, or even not at all tired, while it turned out that I had exacerbations at night. So, I could see that from two criteria: the first one was the decibel peak levels at night, and then the second one was when I didn’t have a restful night’s sleep. So, these were two rather useful criteria, I think. And then...yes, there also was a third one...It is my heart rate, which increased as soon as I coughed.* [Adult patient]

The integration of technology and patient education into the care process was seen as an additional workload by health care providers. Although dealing with technical problems took more time than expected for those in charge of the study (nurse or clinical research associate), physicians mainly mentioned the time spent on the patient education workshop (education workshop 2). Patient education undertaken by physicians in the adult patient care pathway is new for some adult centers, and those centers hope to benefit from a “return on investment”
from it in the future. From the point of view of the care team, taking measurements using connected devices adds to the time already spent by patients managing their disease daily:

For us, it takes time, but obviously, for the patients it represents a lot of time too. In patients' daily lives, it clearly adds minutes to their basic treatment. In terms of the team's workload, it obviously adds work, and the therapeutic education workshops linked to the protocol were particularly cumbersome. It's a lot of work at the time, but it clearly is really beneficial for the future. [Health care provider]

Although the educational tool proposed in the study (educational workshop 2) was generally appreciated by adult patients, it may have seemed complicated to the adolescent audience although it was developed by a pediatric team and tested with several teenagers before releasing it to be used for research:

Therapeutic education went well too...The information pane ("React with CDs" tool) was really well done, and it allowed us to look into many habits that we didn't have, well at least that I didn't have. [Adult patient]

The dashboard was not bad, but super complicated to use for a teenager. There is too much stuff on it. And clearly, too much information on the same page. You can't go straight to what you're looking for...I mean, you really need to look for it. In that sense, I think this table needs to be more legible, because there was a lot of data on it. And reading a lot of data in a table with many columns, it's...it's not appealing. [Parent]

Personalized alert thresholds were set for each patient based on data collected during the first phase of the study following the statistical analysis (cumulative sum control chart). However, these alerts were rarely used by patients to manage their exacerbation episodes as reading measurement results alone allowed them to understand their health status or the lack of updates to thresholds rendered the alerts irrelevant:

At the end of the year, my FEV1 had increased by quite a bit, so when I started the new year with a new secondary infection, my FEV1 didn’t drop lower than the year before. As a result, I never received any alerts. So, I think in this case, we need to update the thresholds, because things can really fluctuate. [Adult patient]

Questions emerged among health care providers on the profile or profiles of patients for whom it is more relevant to introduce self-monitoring measures via connected devices. The inclusion criteria of the study targeted patients with good to moderate lung function (FEV1 >50%) so as to limit the risk of patients leaving the study because of lung transplantation, which is considered as soon as FEV1 decreases to <40%. Some physicians who followed adult patients believed that stabilized patients are good candidates for this follow-up through connected devices, whereas others pointed out that very unstable patients could benefit from this reactive warning system to manage decompensation. In such a critical situation, physicians emphasized the importance of systematically transmitting patient data to the center to help monitor the patients using alerts. Although most physician investigators wanted the study not to send patient data to the center as they felt that they did not have the resources to treat them, other physicians considered it not to be viable for patients who were critically ill. The fear of widening existing social inequalities in health was also mentioned by the care teams:

I think it is useful to integrate the use of such devices with severely ill patients who have frequent exacerbations, who are hospitalised...It can really have a positive impact by confirming the patient’s perception that they are not doing so well, and that they may need to begin an intravenous treatment. It can help patients and us, health care providers, for patients who are severely ill, by providing objective data on exacerbations.... But at the same time, we must not delude ourselves. It is with these severely ill patients that it will be more difficult to set up a monitoring process with CDs. Because they often are in complicated situations socially, psychologically, and so on. So, I don’t know whether it will really be possible with these patients. There are biases and inequalities that will remain true with CDs. Whereas patients who are already autonomous and stabilised will more easily appropriate the CDs. [Health care provider]

However, some patients want to maintain control over their data and make decisions themselves as they feared that connected device monitoring would increase the control of the care team:

I don't need a doctor's supervision to tell me to be careful and that today's measurement was not good. Because on the contrary, I find it more worrying than anything else. But then, it depends on the CF centre. For example, some CF centres will use the measurements and overprescribe antibiotics, while others will want to see the patient in consultation.... It should be up to us, it's our responsibility. [Adult patient]

Category 3: Device Use by Patients (and Health Care Providers)

In the context of this study, device use refers to the ways in which patients used connected devices, whether continuously or intermittently, which may have evolved during the course of the intervention according to factors linked to the patients’ life circumstances, what they experienced during the study, and the conditions of integration of the new monitoring process into the organization of the care team’s work. These uses reflect patients’ perceptions of the benefit-risk balance of the technology and its evolution during the study. Patients adapted the frequency of their connected device use to their need to self-monitor between quarterly visits to the center or, instead, to their need to “let go” slightly on disease management. This need for monitoring increases in periods such as the introduction or cessation of treatment, and it fluctuates depending on life circumstances (work), events related to the environment (high pollen count), or symptoms linked to the disease.
The following is an example of patients’ need for self-monitoring in between consultations at the hospital:

_This allows us to watch the evolution of our data. The problem is that we go to the hospital once a month, or even every three months. So, we don’t have a regular follow-up as such. Whereas with these devices, for example, if I do a spirometry test once a week, I get a score every week, and I will check quite regularly, either it is effective or it is not. It’s complementary to my usual care and it could perhaps help patients be more autonomous._ [Adult patient]

The following is an example of adopting connected device monitoring in specific situations or for particular diagnoses:

_I am planning to get pregnant, and therefore, I think connected devices will be very useful during that time. Indeed, I may not be able to take all the treatments that I can usually take when I am not pregnant. So, I think the devices will be useful then and I also think I’ll be more conscientious in such circumstances._ [Adult female patient]

The use of connected devices also depends on the way measures are integrated into the patient’s personal organization, also known as the routinization process, which, when compatible with their lifestyle, can alleviate the feeling of burden related to the use of devices and contribute to making the collected data more reliable. In the absence of a routine, the use of connected devices can also be taught through therapeutic education sessions and become part of a self-normative approach connected to the patient’s perceptions of their health status:

_I do it when I have a quiet moment before leaving in the morning, before physiotherapy, and that’s it. I always tried to do it in the same conditions, so that it wouldn’t skew the data._ [Adult patient]

In the particular case of adolescents monitored using connected devices, their use was regulated by the parents, which adds to the burden of preparation and control of certain treatments. The collaboration with an out-of-hospital physiotherapist in this monitoring was seen as a relief for the parent caregiver, and it emphasizes the importance (credibility) of the follow-up for the adolescent patient. The question of maturity related to patients’ age was raised regarding the implementation of monitoring using connected devices in adolescence. Conversely, a parent mentioned the help that these connected devices could bring for the empowerment of young patients. Additional notification functionalities inspired by other applications could also support their use of connected devices:

_The greatest thing that could happen for kids would be that the watch sent them a notification if the scores were low and told them what to do. For example, we would set up some instructions onto the app, and as a result, they would receive notifications with the steps to follow on their watch. It would really make them autonomous then. Some apps allow the creation of a schedule and then send out notifications. Youngsters just have to look at their watch and it reminds them they have to bring a check on Monday at 10 AM to the school secretary to pay for the canteen. So, it doesn’t replace the parents, but it would relieve them of the task of always repeating things like a parrot, which causes a lot of conflicts in families._ [Parent]

Sometimes, connected devices reactivated conflicts between parents and adolescents regarding the fear of addiction to the tablet for uses other than health monitoring or because they give parents access to data on the adolescent’s behavior:

_There’s a very intrusive aspect to it. It feels quite overbearing for teenagers to know that they have lost 200 grams and that mum and dad want them to eat more to get the weight back on. Parental monitoring of sleep also creates conflict, and it was the case for almost all teens, with the parents saying: “You’re going to bed too late, that’s why you’re tired, it’s not healthy for you...” Some parents decided not to look at the data for that reason._ [Health care provider]

The use of connected devices by patients is also determined by the interest and attention that health care teams pay to discussions on these data during consultations, phone calls, or teleconsultations with patients, which we will refer to as “patient support”:

_We talked about it, but then, we didn’t focus the consultations on it at all...I expected there would be more guidance in terms of therapeutic education...we did it once about the information panel (“React with CDs” Tool), it took a very long time, it lasted almost two and a half hours. But I expected it would be that way during consultations, precisely to teach us to manage it ourselves...Sometimes, I wonder how it would be like if we had a chat every month, just for five minutes, just to ask me if things were going well, if there were any problems, or if I thought something was wrong._ [Adult patient]

Conversely, when health care providers fail to take into account information from connected devices in patient monitoring, it can make patients doubt the importance and usefulness of such data, which, in addition to the burden of taking measurements, can lead to a lack of interest in these devices:

_We need to talk about the consultations on it at all...I expected there would be more guidance in terms of therapeutic education...we did it once about the information panel (“React with CDs” Tool), it took a very long time, it lasted almost two and a half hours. But I expected it would be that way during consultations, precisely to teach us to manage it ourselves...Sometimes, I wonder how it would be like if we had a chat every month, just for five minutes, just to ask me if things were going well, if there were any problems, or if I thought something was wrong._ [Adult patient]

_We talked about it, but then, we didn’t focus the consultations on it at all...I expected there would be more guidance in terms of therapeutic education...we did it once about the information panel (“React with CDs” Tool), it took a very long time, it lasted almost two and a half hours. But I expected it would be that way during consultations, precisely to teach us to manage it ourselves...Sometimes, I wonder how it would be like if we had a chat every month, just for five minutes, just to ask me if things were going well, if there were any problems, or if I thought something was wrong._ [Adult patient]

Beyond the use of connected devices for monitoring purposes between consultations, some patients suggested that this follow-up could allow them to space out their visits to the center, particularly when they lived far from the center or to limit the risk of contamination at the hospital:

_Not on a regular basis, but sometimes when needed, to avoid going to the CF centre, because I live a little over an hour away. So, to avoid the journey, especially if I’m going to the hospital just to do a spirometry test to analyse FEV1, then yes, I might as_
well do it with the device at home...it allows me to do the measurements myself. At the time of the consultation, we can either have a video call or talk over the phone, and then, we just give the results...in addition, we can extract our data, so we can even send them by email and the doctor can look at them beforehand. [Adult patient]

**Stage 3: Theoretical Coding**

The comparison of verbatim transcripts in the 3 categories revealed the relationships between them, as shown in Figure 2. These bidirectional relationships can be explained as follows. First, the TTF–empowerment relationship: this is reflected in patients’ trust in technology as a necessary condition to consolidate their learnings, which in turn strengthens their trust in the support they receive from technology. Second, the empowerment-use relationship: patients’ capacity to take action and their feeling of control over their health condition with the way the technology is used, which in turn strengthens their capacity to act on their health. This relationship is mediated by the support provided by the care team to help patients adjust the use of technology for their daily management of the disease and, therefore, improve the PEx diagnoses and the suitability of the prescription. Third, the TF-use relationship: the adequacy of the technology to patients’ needs influences its use by patients, reflecting the perceived advantage for patients of being monitored using devices compared with their “standard” follow-up at the CF center. Patients use devices more if they seem adapted to their needs and if they are reliable and easy to use. This relationship is mediated by the care team’s appropriation of the technology, which translates to their coordination of the remote monitoring, the use of real-life data in patient education, and them learning to master the use of devices for patient care.

**Figure 2.** Category modeling and mapping. CD: connected device.

Connecting all the results leads to theorization, the final stage of grounded theory analysis, which can be formulated as follows. The use of connected devices by patients results in an increase in their ability to take action over their health (empowerment) through the continuous adjustment of this use to their degree of autonomy, which influences and is also influenced by the conditions in which the technology is integrated into the organization of the care teams and the patient educational program. The motivation of patients with CF to use connected devices to prevent and manage PEx is dynamic. It depends on the patients’ priorities and specific concerns as well as triggers that will increase the feeling of usefulness related to the connected devices. The data provided by connected devices become a source of new knowledge (eg, about their disease and health) and capacities (eg, to prevent and manage) if a learning process to use them in daily life is implemented. This learning process can be supported by patient education.

This theory accounts for the conditions under which the implementation of connected devices for the management of PEx in patients living with CF can increase their capacity to act on their health.

**Discussion**

**Principal Findings**

Our study theorized the conditions that favor individual empowerment in patients living with CF in the management of PEx using connected devices as part of the MucoExocet study. This study took place in the context of a rare disease, proposed by the health care team to patients who could be interested in the technology. This theorization of the individual empowerment of patients through the use of connected devices is provisional, similar to any theory derived from the grounded theory approach, and remains subject to verification.
Our study places the concept of individual empowerment through the use of technology at the center of our research. Indeed, empowerment is an important mechanism of eHealth self-management, but validated assessment tools are rare [29]. In our theorizing, we viewed empowerment as the ability of individuals to identify and meet their own needs, solve their own problems, and mobilize the necessary resources through connected devices and education provided by health care providers to feel in control of the management of their health [10]. This leads us to consider the success of the implementation of connected devices from the perspective of the patient empowerment outcome, weighted by a typology of the intensity of connected device use by the patient deduced from our results. First, low use: patients doubt the feasibility and usefulness of this continuous monitoring; they are in favor of a “standard” follow-up at the CF center using data collected from the clinical examinations in situ. Second, advanced use: patients know that this monitoring can be useful in case symptoms appear or when a follow-up appointment at the CF center is not easily accessible. The use of connected devices remains optional and selective between the “standard” clinic visits at the center. Third, high use: patients experiment with the use of connected devices when in particular situations or because of particular behaviors, enabling them to consolidate or develop skills in relation to their health and the factors that influence it. This use is connected to their desire to improve their health and control its evolution and to the belief that the use of connected devices can support them in doing so. Depending on the patient’s situation (eg, developmental, emotional, and environmental), the level of motivation to engage in less or more intense connected device use will vary.

We suggest that empowerment, adherence to treatment, and quality of life be favored as primary outcomes of remote digital follow-up. In a recent study using a randomized trial that compared 2 groups of patients (using the tracker device against not using the device), Wildman et al [30] highlighted that the health improvement objectives were not achieved but that the intermediate objective of improving adherence to treatment was exceeded. These findings tend to confirm that, when assessing how effective the implementation of technology is with patients, the improvement of health indicators may not be the first outcome to be expected. This strengthens the case for a patient-based evidence evaluation approach.

In addition, the identification of “opposite cases” encountered in our study, for which patient empowerment was compromised, supports the theory stated—cases in which the devices were unreliable (TTF) or no action plan was defined in response to alerts or variations in measurements or cases of difficulties reaching health care providers (lack of support or difficulty integrating the use of technology into their organization) all led to lower levels of patient empowerment. Our study questions eHealth-backed education models, for which data are currently scarce. Following the work of Greenhalgh et al [31], our results highlight the interaction among the patient, the device, and the organizational and social system as the cornerstone of the learning process in patients. This interdependence underscores the systemic approach to connected device implementation, wherein connected device adoption and use and the positive experience with them cannot be attributed to the patient’s lack of motivation alone. Indeed, connected devices introduce a technopedagogical transformation among health care providers, which pushes them to rethink organizational and educational activities to support a new relationship with patients.

This study shows that connected devices may have enabled health care providers to gain a new understanding of patients thanks to the quality and novelty of the information obtained via connected devices. In this sense, connected devices could help bridge the gap that is sometimes observed between theoretical models based on medicine that is “centered on the person and their family” and the practice of care that lacks understanding of patients’ experiences in daily life [32]. Health care providers are made to understand the daily lives of patients living with a chronic disease in physical, psychological, and social terms, thereby creating a more symmetrical relationship of information sharing [33]. Our study shows that caregiver-patient interactions are modified by the introduction of connected devices. They are enriched by a new outlook on patients’ daily lives mediated by technology, which leads to a new understanding by health care providers.

In addition, this study confirmed that the implementation of connected devices should be considered based on patients’ health goals and not simply focused on education on the device [34]. Patient empowerment depends on the connected devices’ capacity to meet the needs of patients’ health project. Patients then enter a learning process supported by the connected device and with educational support from health care providers structured in 4 phases, as described by Almalki et al [35]: identification of an area of interest (the patient is focused on a specific health goal that requires the collection of data about themselves), personal analysis (analyzing one’s behavior in light of the objective data collected), self-experimentation (structuring a reasoning based on the trends identified in support of the experimentation carried out with the connected device), and activation (confirmation of the hypotheses made during the experimentation phase and development of personal knowledge). This process must be structured and accompanied to unfold properly.

Although a recent review of the literature [36] on the use of mobile devices (phones, patient monitoring devices, digital assistants, and other wireless devices) by patients with CF has shown medical, psychological, and behavioral benefits as well as benefits in terms of level of satisfaction with care, the psychological aspect has thus far received little attention, as is the case with the educational dimension of technological devices. In adolescence, although the disease significantly influences the development of one’s body image and self-concept [37], the integration of new technologies into self-care leads to a new understanding of oneself and, therefore, to a potentially modified relationship with one’s body, health, and illness. This process is an integral part of the use of new technologies. Therefore, it is a potential topic for future research, which is necessary to understand the use of new technologies in care and their effects on people.
Limitations of the Study

This pilot study was based on an interventional research protocol. On the one hand, this protocol was implemented differently depending on the centers and the devices selected for the research—the teams applied the educational program differently, the elaboration of self-management action plans in the event of exacerbations was done differently (educational workshop 2), or the midterm review session of the routine follow-up of the patient was different (educational workshop 3). These differences in the implementation of the protocol were noted when collecting the experience of patients and data from the focus groups; they contributed to enriching the definitions of the categories and the relationships between them. In contrast, within the framework of this research protocol, we could not modify the tools that proved to be unreliable (which had been selected in 2016 via a market analysis while planning for the study), adjust the formatting of the data, or more generally adapt the intervention according to the results collected throughout the early phases of the study. Thus, having a protocol that is too fixed is probably a mistake to avoid in health technology research if we wish to adapt the intervention during its implementation to explore the best way of using health technology. In the context of the intervention, the choice of connected devices and the setup chosen did not allow patients to access a dashboard displaying the data collected by all connected devices. Some more motivated patients created their own dashboards separately. Furthermore, the pulmonologists in charge of patient follow-up were not always involved in the study, and this dichotomy made routine monitoring more complicated for patients. Similarly, when the clinical research associate in charge of the study was not the patient’s coordinating nurse, the latter was unable to answer patients’ questions during consultations or phone calls. Eventually, the study included people interested in technology, which could have biased the results based on the experience of using technology.

If this had been a descriptive pilot study, a quality improvement approach would have allowed for adjustments and improvements to the intervention over the course of the study and would have been directly driven by the care team. This format has been used to introduce connected devices into the patient care process for CF in the United States [38], which enabled patients to be equipped and monitored remotely when CF centers were closed during the COVID-19 pandemic by using a connected spirometer coupled with teleconsultation. The quality approach allowed for the evaluation of the results during the course of implementation, and adjustments were made to the intervention to improve its impact. The results were convincing:

In March 2020, the beginning of the pandemic, 37% (49/131) of patients owned a HS (home spirometer) and around 50% (9/20) of patients seen via telemedicine performed spirometry at home. By September 2020, 97% (127/131) of adult patients at UVA owned a HS, and by October 2020, 96% (24/25) of patients provided spirometry results during their telemedicine encounters.

Prospects for Transferability

Assessing how transferable the theory could be outside the context of its development would require studying the introduction of connected devices in other circumstances: with patients living with different diseases, using different devices, or with a different organization of care.

Two opposite contexts could be studied in terms of patient empowerment through technology: (1) a context of patient dependence on self-regulated or caregiver-driven technology, whether it is telemonitoring, implantable devices for which the use is predetermined (dependence on technology and on health care providers making the care decision in the event of an alert or emergency), or protocolized treatment with little margin for adaptation or action because of side effects (eg, protocol dependence in cancer treatments); and (2) a context of patient-developed technologies [39] made available to patients living with the same condition in open source, as is the case with type 1 diabetes mellitus (T1DM). T1DM has the highest degree of patient empowerment and has recently led to the publication of an international consensus for the guidance of professionals caring for patients who use such devices. The case of T1DM is also interesting as research was conducted on the transition of patients from devices that allow for the management of glycemia and insulin delivery in a semiautomated way to a closed-loop insulin delivery system, which is designed to “free” the patient from self-management by automating the process of insulin delivery. However, this specific case might also lead patients to feel that they lose control over their glucose levels before they take back control over some other parameters of the automated process.

Contribution to an Extended Theory of Empowerment From Remote Monitoring for Health Symptom Tracking

A recent publication by White et al [40] reports on a systematic review to help define engagement with remote monitoring for health symptom tracking (RMT) and how to measure it. Engagement is seen as a mediating factor that eventually explains the impact of RMT on patient health outcomes. Their analysis is of most interest to our own work and shows that concepts still need to be clarified in the context of RMT. They propose a definition of engagement through a remote monitoring protocol (dropouts), objective engagement, subjective engagement, and interactions between objective and subjective engagement. Although objective engagement (with remote monitoring itself, with symptom tracking compliance, and with app use of statistics) is clearly measurable, subjective engagement appears to gather a wide range of concepts, some of them from the technology acceptance model literature (usability, TTF, satisfaction with the technology, utility for symptom management, ease of use, and intention for future use; Davis [41] revised by Venkatesh et al [28] and Chang et al [42]).

In a further extended theory, we would rather build on certain determinants of the technology acceptance model and distinguish them from the concept of patient engagement. These determinants leading to the “behavioral intention of use” would be the personal characteristics (age and sex, expectations, social influence, hedonic motivations, and previous experiences with
information and communication technologies), the facilitating conditions over time, TTF (over time as technologies are continuously refined), and the mediating factors (perceived ease of use and perceived usefulness). We would propose to include the “engagement with the research protocol” by White et al [40] as a determinant, renamed as “conditions for the RMT introduction/intervention” (either research or routine care or self-care). Our study aimed to add elements to modulate the “behavior use” in the RMT context, which is not explained by the previous theories and not necessarily consistent with the “behavioral intention of use.” From our study, these elements could refer to patient empowerment, such as their learnings about their own body, their trust in the technology, and the relationship and support they receive from their care team. We agree with the conclusion of White et al [40] to explore the RMT field in its own right as separate from Digital Behavior Change Interventions or general eHealth literature.

Conclusions
Our study allowed us to propose a theory on individual patient empowerment through the use of connected devices based on patients’ and health care providers’ experiences in the context of an interventional pilot study. This theory needs to be validated with a larger sample and verified in the context of different diseases, different devices, and a different organization of care.

It implies that, if the empowerment of patients with chronic diseases is indeed a desirable goal for all parties involved (patients, health care providers, and the health care system), the necessary conditions for the successful implementation of connected devices cannot be looked at separately for each party (health care providers, patients, and health care system). On the contrary, these conditions must be adjusted to the overall collaboration among these stakeholders, who cooperate toward patient empowerment. Only if all these conditions are met can patient empowerment be the outcome of the use of technology.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The “React to PEx” educational tool.
[DOCX File, 180 KB - formative_v81e38064_app1.docx]

Multimedia Appendix 2
Exit interview questionnaire.
[DOCX File, 108 KB - formative_v81e38064_app2.docx]

Multimedia Appendix 3
Reasons for leaving the study.
[DOCX File, 70 KB - formative_v81e38064_app3.docx]

References


Abbreviations
CF: cystic fibrosis
EQUATOR: Enhancing the Quality and Transparency of Health Research
FEV1: forced expiratory volume in 1 second
PEX: pulmonary exacerbation
RMT: remote monitoring for health symptom tracking
T1DM: type 1 diabetes mellitus
TTF: task technology fit
Culturally Adapting the World Health Organization Digital Intervention for Family Caregivers of People With Dementia (iSupport): Community-Based Participatory Approach

Anna Messina, PhD; Rebecca Amati, PhD; Anna Maria Annoni, MS; Beatrice Bano, MS; Emiliano Albanese, MPH, MD, PhD; Maddalena Fiordelli, PhD

Institute of Public Health, Faculty of Biomedical Sciences, Università della Svizzera italiana, Lugano, Switzerland

Corresponding Author:
Anna Messina, PhD
Institute of Public Health
Faculty of Biomedical Sciences
Università della Svizzera italiana
Via Buffi 13
Lugano, 6900
Switzerland
Phone: 41 0782104055
Email: anna.messina@usi.ch

Abstract

Background: Informal caregivers of people with dementia are at high risk of developing mental and physical distress because of the intensity of the care provided. iSupport is an evidence-based digital program developed by the World Health Organization to provide education and support for the informal everyday care of people living with dementia.

Objective: Our study aims to describe in detail the cultural adaptation process of iSupport in Switzerland. We specifically focused on the participatory strategies we used to design a culturally adapted, Swiss version of iSupport that informed the development of the desktop version, mobile app, and printed manual.

Methods: We used a mixed methods design, with a community-based participatory approach. The adaptation of iSupport followed the World Health Organization adaptation guidelines and was developed in 4 phases: content translation, linguistic and cultural revision by the members of the community advisory board, validation with formal and informal caregivers, and refinement and final adaptation.

Results: The findings from each phase showed and consolidated the adjustments needed for a culturally adapted, Swiss version of iSupport. We collected feedback and implemented changes related to the following areas: language register and expressions (eg, from “lesson” to “chapter” and from “suffering from” dementia to “affected by” dementia), resources (hyperlinks to local resources for dementia), contents (eg, from general nonfamiliar scenarios to local and verisimilar examples), graphics (eg, from generalized illustrations of objects to human illustrations), and extra features (eg, a glossary, a forum session, and a read-aloud option, as well as a navigation survey).

Conclusions: Our study provides evidence on how to culturally adapt a digital program for informal caregivers of people living with dementia. Our results suggest that adopting a community-based participatory approach and collecting lived experiences from the final users and stakeholders is crucial to meet local needs and to inform the further development, testing, and implementation of digital interventions in a specific cultural context.

(JMIR Form Res 2024;8:e46941) doi:10.2196/46941

KEYWORDS
informal caregivers; iSupport; dementia; digital interventions; mHealth; community-based participatory research; community; caregiver; mental distress; physical distress; support; development
Introduction

Background
Approximately 55 million people are currently living with dementia worldwide [1]. Switzerland accounts for >150,000 cases of dementia, with an expected doubling by 2050 [2]. In Switzerland, as in most countries, the majority of people living with dementia live at home assisted by an informal caregiver, who is usually a family member who provides daily support and coordinates care delivery [3]. There are positive outcomes that may be associated with the caring role, such as the perception of a better relationship and closeness with the care recipient [4]. Nonetheless, the increasing complexity of taking care of a person affected by dementia exposes informal caregivers to psychological distress and increases the risk of loneliness and developing symptoms of anxiety and depression [5,6]. Caregivers’ psychological distress is also associated with a lower quality of care provided [7] and with the worsening of behavioral and psychological symptoms in the care recipients [8].

Providing guidance and support to informal caregivers is one of the priority areas identified by the World Health Organization (WHO) to reduce the global impact of dementia and to improve the quality of life of caregivers and their families [9]. In the last 2 decades, digital educational and psychosocial interventions for caregivers have bloomed [10,11]. Internet-based interventions are more easily accessible [12,13] and adaptable to the time and geographic constraints of caregivers [14]. Some reviews suggest that multiple components of digital interventions can contribute to reducing the burden and improving the quality of care and be even more beneficial if tailored to caregivers’ specific needs and contexts [10,15,16]. The active involvement of the final users and relevant stakeholders in the design and local adaptation as well as the testing and piloting of interventions is crucial for need-centered interventions in terms of their uptake, integration, and scalability at the community level [17,18]. However, more evidence is required to understand the most effective methods and strategies needed to involve participants in the design and adaptation of digital interventions [19,20].

iSupport is an evidence-based digital training intervention developed by the WHO to provide support and education to informal caregivers of people with dementia [21]. The original program consists of 22 thematic lessons distributed across 5 modules (Figure 1). Each lesson covers a specific topic associated with care that ranges from the daily assistance of the care recipient (eg, toileting, personal care, and nutrition) to the self-care of the carer (eg, reducing stress and involving others in care duties). All lessons include theoretical and informative sections and case scenarios with interactive multiple-choice questions. The WHO provides guidelines to culturally adapt iSupport contents to the local language, culture, and context before implementation [22]. The reporting of adaptation processes of complex interventions is limited but extremely important [23]. Knowledge exchange of methodologies and approaches as well as evidence on barriers and facilitators to local adaptation are crucial preliminary steps to inform the implementation of interventions and their mid- to long-term uptake and sustainability [24].
Objectives
This study aims to describe in detail the cultural adaptation process of iSupport in Switzerland. We specifically focused on the participatory strategies we used to design a culturally adapted, Swiss version of iSupport that informed the development of the desktop version, mobile app, and printed manual. Our purpose is to inform the implementation of not only iSupport but also other complex health interventions, specifically in the context of the cultural adaptation process.

Methods
Study Setting
The study took place in the Italian-speaking part of southern Switzerland, namely the canton of Ticino.

Ethical Considerations
Before initiating the study, we sought ethics oversight by submitting our project to the cantonal ethics committee, and we obtained a waiver of ethics approval and official authorization to proceed with the study (project ID: 2020-02030 / CE 3731).

Specifically, the ethics committee determined that our project did not fall within the scope of the Swiss Federal Human Research Act [25], thereby granting us permission to proceed.

Theoretical Approach
The overall process of culturally adapting iSupport in Switzerland was based on principles from the community-based participatory research (CBPR) framework, which can be defined as “an approach to research that involves collective, reflective and systematic inquiry in which researchers and community stakeholders engage as equal partners in all steps of the research process” [26]. In intervention research, adopting CBPR has the advantage of facilitating knowledge exchange between the community and the researchers, reducing potential power imbalances, and increasing the likelihood of intervention uptake and success [27].

We based the specific phases and procedures of the adaptation process on the WHO iSupport adaptation guidelines [22], which, in turn, are based on the ecological validity framework proposed by Bernal et al [28] that is widely used for developing culturally sensitive interventions and strengthening their ecological validity [29-31].
Study Procedure

The Community Advisory Board

At the outset, we established a community advisory board (CAB) comprising community members and organization representatives who shared a common identity, geography, language, culture, and other values and principles [32]. We identified potential members of the CAB through a structured stakeholder analysis and mapping that accounted for the different levels of power, importance, and interest of the stakeholders in the project. We included representatives of the project’s funding agencies and other collaborating partners, caregivers of people with dementia, and members of the IT service in charge of developing the iSupport web platform and app. Once consensus among researchers was reached, we contacted and informed the identified members via email using a brief description of the project, the scope of the CAB, and their expected roles and responsibilities.

In the context of iSupport adaptation, the specific roles of establishing a CAB were to (1) help researchers to identify the needs and legitimate interests as well as the expectations of the different stakeholders and the final users and (2) inform the development of the intervention throughout a purposely co-designed process.

The adaptation process of iSupport consisted of four phases: (1) content translation, (2) linguistic and cultural revision, (3) validation with formal and informal caregivers, and (4) refinement and final adaptation. Each phase was based on, and adapted from, the WHO guidelines. Any change or proposed addition was discussed with, and approved by, the WHO. The members of the CAB were constantly informed and updated on the progress of the study. A flowchart of the phases is summarized in Figure 2.

Figure 2. Flowchart of the adaptation process of iSupport in Switzerland. CAB: community advisory board; WHO: World Health Organization.

Phase 1: Content Translation

The first step in the cultural adaptation of iSupport was the translation of the contents (approximately 60,000 words) from English, the original language of the program, into Italian, the local language in southern Switzerland. The process started in May 2020 and ended in August 2020. According to the WHO guidelines, the translation should be accurate while recognizing the local culture and its people. In line with this, we conducted a preliminary adaptation of culturally sensitive terms, including (1) personal names of the characters used in the case studies, (2) available information materials and local services, and (3) reference to cultural habits and leisure activities in the region [22].

One member of the research team fluent in English, AM, a psychologist with previous expertise in the dementia field, translated the original contents of the iSupport program into Italian. Subsequently, a senior member of the team, MF, with expertise in the field of health communication, checked the translations and proposed changes and modifications. All disagreements or doubts about the translation of sensitive terms
and expressions were documented and discussed within the research team in meetings until a consensus was reached. We sought the support of an external professional translator to resolve some specific language locutions and terms.

Throughout the process, translators applied the international standards and available dementia guidelines to avoid stigmatizing expressions and to use language that promotes the inclusion and dignity of people living with dementia and their carers [33]. During this phase, we did not apply any changes to the meanings of the original structure of the iSupport program, including case studies or activities. All translations were copied into secure Microsoft Word files and stored in a dedicated Microsoft Teams workspace to optimize efficiency.

**Phase 2: Linguistic and Cultural Revision**

In September 2020, the first CAB meeting took place with the main goals of introducing the members of the CAB to the iSupport program and the research team and clarifying their roles and involvement throughout the research process. During the meeting, we answered all questions and proposed an interactive activity where participants were asked to provide the translation from English into Italian of a selection of sensitive terms and expressions used in iSupport that were noted by researchers during phase 1. At the end of the introductory meeting, participants were asked to sign a letter of intent that summarized the functioning of the CAB and their role and commitment as members of the local iSupport CAB. We explicitly specified the structure and definition of the CAB; goals, roles, and responsibilities (of both CAB members and the research team); and duration (Multimedia Appendix 1). All 9 invited participants agreed to join the iSupport CAB: 4 (44%) were informal caregivers, and the remaining 5 (56%) included representatives of the government (1/5, 20%), the local Alzheimer association (1/5, 20%), a health care service provider (1/5, 20%), the IT service (1/5, 20%), and the University of Applied Sciences and Arts of Southern Switzerland (1/5, 20%).

In October 2020, we shared the translated contents of iSupport with the members of the CAB and asked them to evaluate, and provide feedback on, each chapter and module of the program by the end of December 2020.

On the basis of the work of previous adaptations of iSupport [34], participants were asked to carefully go through the 23 thematic lessons and assess the translation and preliminary adaptation of iSupport considering six main parameters: (1) familiarity, (2) sensitivity, (3) comprehensibility, (4) precision, (5) cultural adequacy, and (6) overall evaluation. In addition, they were asked to assess the extent to which (1) the terms used were familiar to the target group (eg, the use of idioms and figures of speech), (2) the language used respected and promoted the dignity of people living with dementia and their carers (eg, the use of stigmatizing terms), (3) the contents were intelligible and easy to understand (eg, minimal use of technical jargon), (4) the contents were presented in an accurate way (eg, they were in accordance with the facts, and there were no mistakes), (5) the contents were appropriate and reflected the experiences of local people (eg, case studies), and (6) the content of each chapter was overall culturally appropriate. At the end of each chapter, participants were asked to fill out a digital survey via Research Electronic Data Capture (REDCap; Vanderbilt University) [35,36] to evaluate each of the aforementioned parameters using a Likert scale ranging from 1 = requiring an extensive revision to 4 = no additional revision needed. We also invited participants to provide additional comments about individual chapters through a dedicated open-ended question in the survey or to provide free feedback on the overall program via email. The survey was specifically designed for the purpose of this phase and was based on the work of Teles et al [34] to evaluate the cultural adequacy of the contents, as recommended by the WHO adaptation guidelines [22].

After the data collection period, AM and BB (a research assistant with a degree in psychology and health communication) performed a descriptive analysis of the quantitative data and a thematic analysis of the qualitative data. For the quantitative analysis, we used SPSS statistical software (version 25.0; IBM Corp) [37] for Windows to compute mean scores for each program module and survey parameters. For the qualitative analysis, we performed a thematic content analysis of open comments [38]. The maintenance of scientific rigor was ensured through regular meetings among research team members, particularly involving MF and RA, both experts in qualitative research methods.

In January 2021, the main findings of this phase were summarized in a report shared across, and approved by, all CAB members.

**Phase 3: Validation With Formal and Informal Caregivers**

We adopted a qualitative descriptive design, and we used focus groups (FGs) as a data collection method [39]. Between June 2021 and August 2021, we conducted FGs with formal and informal caregivers to explore their attitudes toward, and impressions of, the adapted version of iSupport. We prompted and collected suggestions for improvement, as also recommended by the WHO guidelines. In addition, we decided to expand our inquiry to caregivers’ attitudes toward support measures and help-seeking behaviors, which we have previously reported in detail elsewhere [40].

From April to May 2021, we crafted an invitation letter and a flyer presenting the project, the main purpose of the FGs, the eligibility criteria, and contact information. We disseminated these materials in a local newspaper, to members of the CAB and their associations and institutions, to daycare centers for people with dementia, and to participants of other ongoing research projects who had consented to be informed about further research activities.

Eligibility criteria for both formal and informal caregivers included (1) having (at present or in the past) experience in caring for a person living with dementia, (2) being fluent in Italian, and (3) living in the canton of Ticino. Caregivers who met the inclusion criteria could contact us via email or telephone. Those who contacted us were given an overview of the iSupport program, with excerpts from the same translated material used in phase 2, and the informed consent form (Multimedia Appendix 2). The FGs, which lasted approximately 2 hours, were audio recorded and took place either digitally via the Zoom
Familiarity With Terms and Expressions

The unfamiliarity with the terms referred especially to some expressions that were largely used throughout the text to designate caregivers and health care workers or dementia health and social care facilities, and the comments by the CAB members allowed us to improve the translations; for instance, “informal carers” and “paid in-home helpers” were newly translated using local terms that were easily identifiable and familiar to participants (e.g., “informal carer” was replaced with “familiare curante,” which literally means “family carer”). Importantly, the term is also used at an institutional level [42] to refer to people who take care of a loved one (relative or friend) affected by a chronic disease.

Table 1. Mean scores of the 6 parameters for the linguistic and cultural revision of iSupport.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Module 1, mean (SD)</th>
<th>Module 2, mean (SD)</th>
<th>Module 3, mean (SD)</th>
<th>Module 4, mean (SD)</th>
<th>Module 5, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>3.60 (0.55)</td>
<td>3.59 (0.62)</td>
<td>3.00 (0.00)</td>
<td>3.87 (0.52)</td>
<td>3.89 (0.47)</td>
</tr>
<tr>
<td>Familiarity</td>
<td>3.60 (0.55)</td>
<td>3.65 (0.61)</td>
<td>3.00 (0.00)</td>
<td>3.67 (0.62)</td>
<td>4.00 (0.00)</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>4.00 (0.00)</td>
<td>3.94 (0.24)</td>
<td>4.00 (0.00)</td>
<td>3.93 (0.26)</td>
<td>4.00 (0.00)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>3.75 (0.50)</td>
<td>3.76 (0.44)</td>
<td>3.67 (0.58)</td>
<td>3.73 (0.70)</td>
<td>4.00 (0.00)</td>
</tr>
<tr>
<td>Cultural adequacy</td>
<td>3.60 (0.55)</td>
<td>3.59 (0.62)</td>
<td>4.00 (0.00)</td>
<td>3.80 (0.56)</td>
<td>3.89 (0.47)</td>
</tr>
<tr>
<td>Overall assessment</td>
<td>3.20 (0.45)</td>
<td>3.65 (0.49)</td>
<td>3.00 (0.00)</td>
<td>3.80 (0.41)</td>
<td>4.00 (0.00)</td>
</tr>
</tbody>
</table>

Results

A detailed description of all final adaptations made to the original iSupport program, resulting from the 4 phases of the adaptation process, is presented in Multimedia Appendix 4. The results of phases 2 and 3 are summarized in the subsections that follow.

Adaptations From Phase 2: Linguistic and Cultural Revision

Overview

All 9 members of the iSupport CAB revised ≥1 modules of iSupport and provided feedback, with each module revised by at least 1 CAB member. Module 5, which dealt with behavior changes, was the most revised and received the most comments (7/9, 78%). Descriptive analysis showed that all modules were generally positively evaluated with an overall mean evaluation score of 3.72 (SD 0.00) out of 4. Mean scores of the parameters across all modules ranged from 3 (SD 0.00) to 3.89 (SD 0.47) for sensitivity, from 3 (SD 0.00) to 4 (SD 0.00) for familiarity, from 3.93 (SD 0.26) to 4 (SD 0.00) for comprehensibility, from 3.67 (SD 0.58) to 4 (SD 0.00) for the accuracy of the information, and from 3.59 (SD 0.62) to 4 (SD 0.00) for cultural adequacy (Table 1).

We conducted a qualitative thematic analysis of the open comments and identified 7 potential areas for the improvement of iSupport (for more details, refer to Multimedia Appendix 4).

platform (Zoom Video Communications, Inc) or in person at the Università della Svizzera italiana in Lugano, Ticino. AM moderated all FGs, with the supervision of MF and RA. The discussions were transcribed verbatim and pseudonymized by EB, an independent research assistant. AM, RA, and MF performed a thematic content analysis to identify key themes [38]. Initially, the researchers familiarized themselves with the data through repeated reading of notes and transcripts to get an idea of the overall meaning and begin discerning key themes. Subsequently, each researcher independently identified codes within each FG (vertical analysis) and across the whole data set (horizontal analysis) to uncover variations and patterns within the data. Themes were progressively refined and consolidated through discussion in weekly meetings over 4 months (from November 2021 to February 2022) and until a consensus was reached. Data management and coding processing were facilitated by NVivo 12 software (Lumivero) [41]. Additional methodological details have been previously reported [40].

Phase 4: Refinement and Final Adaptation

All data collected during phases 2 and 3 were collated to generate a set of proposed changes and adaptations to the iSupport program. AM and BB familiarized themselves with the data and differentiated the feedback between cross-cutting and general comments and specific chapter–related comments and arranged them across 5 dimensions: language, resources, contents, graphics, and extra features. Each comment was then discussed between AM and BB and categorized as (1) rejected/not applicable, (2) possibly applicable, and (3) applicable. The categorization was based on the number and contents of suggestions received as well as in accordance with the WHO adaptation guidelines [22]. The feedback data that were considered possibly applicable and applicable were then discussed with the other members of the team to agree on their potential implementation.

Subsequently, all specific and applicable comments were charted using the iSupport WHO adaptation spreadsheet (Multimedia Appendix 3). All proposals of change were then supported by a rationale and by the source of the proposal: the research team (preliminary adaptation during phase 1), the members of the CAB (results from the linguistic and cultural adaptation during phase 2), and FG participants (data collected during phase 3). Attached to the adaptation spreadsheets, we also sent a list of general suggestions and feedback. The material was sent for revision to the authors of the WHO iSupport program in October 2021, and the results of their final fidelity check were received in January 2022. Subsequently, the local research team implemented all approved changes and uploaded the new adapted contents on the beta version of the iSupport Swiss web platform.

https://formative.jmir.org/2024/1/e46941

JMIR FORMATIVE RESEARCH

Messina et al

(2024)
Sensitivity of the Language
Participants also suggested improving the sensitivity of the language by removing expressions potentially stigmatizing such as “suffering from dementia,” which was replaced with “affected by dementia” (affetto da demenza), or “show compassion,” which was replaced with “show comprehension” (mostrare comprensione).

Scientific Accuracy of the Terms
The jargon used was generally perceived as comprehensible and easy to understand. However, some of the participants (2/9, 22%), especially those working in the field of dementia, reported the need to use scientific terms to improve the accuracy of the language and ultimately the users’ literacy. As a result, the expression “memory loss” was, for example, replaced with “memory impairment” (difficoltà di memoria), and “helpful/unhelpful thoughts” was replaced with “functional or dysfunctional thoughts” (pensieri funzionali e disfunzionali).

Educational Approach
The educational approach referred to the use of terms considered scholastic and potentially belittling by participants, such as “lesson” and “learn,” which were replaced with “chapter” (capitolo) and “know more about” (conoscere di più), respectively.

Use of English
Some English terms that were retained during the content translation because they are normally used in spoken Italian were translated into Italian, including “focus” (obiettivo) and “relax” (rilassa).

Use of Numbers
To make the reading smoother, some of the participants (3/9, 30%) suggested replacing numbers with sentences (eg, from “20%-30%” to “approximately one-third”; circa un terzo).

Language Register
Finally, almost all participants (8/9, 89%) found that the tone and prose were at times informal or even childish; therefore, for example, the original sentence at the end of each chapter “you finished the lesson, well done” was replaced with “you finished the chapter, let’s go to the next!” (hai completato il capitolo, passa al successivo!).

Adaptation From Phase 3: Validation With Formal and Informal Caregivers

Overview
Between May 2021 and August 2021, we conducted 6 FGs: 1 (17%) with formal caregivers and 5 (83%) with informal caregivers. Most of the participants (16/19, 84%) were female, and the formal caregivers (6/19, 32%) had longer years of caring experience than the informal caregivers (13/19, 68%). The main characteristics of the caregivers are reported in detail in Tables 2 and 3.

Table 2. Sociodemographic characteristics of formal caregivers.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Age (y)</th>
<th>Employment status</th>
<th>Years of professional caring experience</th>
<th>Years of personal caring experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>52</td>
<td>Housewife or retired</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>54</td>
<td>Housewife or retired</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>28</td>
<td>Housewife or retired</td>
<td>6-10</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>45</td>
<td>Employed</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>29</td>
<td>Housewife or retired</td>
<td>1-2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Female</td>
<td>59</td>
<td>Employed</td>
<td>&gt;10</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3. Sociodemographic characteristics of informal caregivers.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Employment status</th>
<th>Relationship with the person with dementia</th>
<th>Living situation of the person with dementia</th>
<th>Years of caring experience</th>
<th>The person with dementia has passed away</th>
<th>Focus group attendeda</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>58</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>55</td>
<td>Housewife or retired</td>
<td>Daughter</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>1, 3, and 4</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>59</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>67</td>
<td>Employed</td>
<td>Son</td>
<td>Own residence</td>
<td>1-2</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>58</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Carer’s residence</td>
<td>1-2</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>57</td>
<td>Employed</td>
<td>Son</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>1 and 4</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>74</td>
<td>Employed</td>
<td>Son</td>
<td>N/Ab</td>
<td>&gt;10</td>
<td>Yes</td>
<td>1 and 2</td>
</tr>
<tr>
<td>8</td>
<td>Female</td>
<td>55</td>
<td>Employed</td>
<td>Daughter</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>2 and 4</td>
</tr>
<tr>
<td>9</td>
<td>Female</td>
<td>75</td>
<td>Housewife or retired</td>
<td>Daughter</td>
<td>N/A</td>
<td>&gt;10</td>
<td>Yes</td>
<td>1, 2, 4, and 5</td>
</tr>
<tr>
<td>10</td>
<td>Female</td>
<td>76</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Female</td>
<td>82</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Carer’s residence</td>
<td>1-2</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>Female</td>
<td>55</td>
<td>Employed</td>
<td>Daughter</td>
<td>Own residence</td>
<td>1-2</td>
<td>No</td>
<td>3 and 4</td>
</tr>
<tr>
<td>13</td>
<td>Female</td>
<td>81</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Carer’s residence</td>
<td>3-5</td>
<td>No</td>
<td>2 and 4</td>
</tr>
</tbody>
</table>

aNumber of the focus group attended.

bN/A: not applicable.

FG With Formal Caregivers

Formal caregivers are professionals who are trained, hired, and paid to provide care to a person living with dementia. In our study, all formal caregivers (n=6) actively participated in the digital discussions. All participants were female. Their mean age was 44 (range 28-59) years. All participants had professional experience in taking care of a person living with dementia. Of the 6 participants, 3 (50%) reported having >10 years of experience in dementia care. In addition to the professional caregiving experience, 3 (50%) of the 6 participants also reported taking care, or having taken care in the past, of a relative affected by dementia (Table 2). The main findings of the FG are summarized in the following paragraphs.

Participants agreed that an intervention aiming to support and improve the knowledge of informal caregivers of people with dementia was much needed. A caregiver compared information learning to a safeguard not only for the carer but also for the care recipient:

> I hope that this program will spread because information protects all of us: the carer, and especially the person who is cared for. [ID 6]

iSupport was generally appreciated and acknowledged by participants as a useful tool. The contents were found appropriate and sufficiently comprehensive. The difficulty regarding accepting the disease and the changes in the relationship with the care recipient were found to be the main challenges and contents to cover in the program:

Relatives find it extremely difficult to accept the disease and the change...I believe a very strong support is needed...also at a social level because the disease is often associated with shame. [ID 1]

Similarly, a participant also suggested adding to the program specific resources for social and psychological support:

You could mention [the existence or the opportunity for family members] to benefit from psychological support because they need it, always. [ID 2]

This quote underscored the recognition by formal caregivers of the potential emotional and psychological strain on family members as they witness the progression of the disease of their care recipients.

In light of the participants’ perspectives, an important feature to add to the original iSupport format was the inclusion of a platform for caregivers to engage with each other and that facilitated the caregivers’ interactions with each other (this adaptation was also needed to differentiate iSupport from another repository of information or digital available resources on dementia):

There are a billion guides on dementia...I think people need to interact. [ID 4]
Regarding case scenarios, the caregivers generally found that the examples were appropriate and consistent with their experiences. However, the answer options often did not reflect the variety of, and differences in, caregiving situations and experiences, including the age of the person affected by dementia, the severity of dementia, the living situation, or the type of dementia (e.g., Alzheimer disease and frontotemporal dementia). A participant suggested adding general guidelines to the examples to include more answers:

> If the examples aim to increase knowledge, they should give general indications that can apply to different caring situations. [ID 6]

**FG With Informal Caregivers**

Of the 20 informal caregivers who contacted us, 13 (65%) joined the FGs. Reasons for nonparticipation were lack of time and geographic distance. Most of the caregivers (10/13, 77%) were female; nearly half were spouses (6/13, 46%) of the persons living with dementia, and more than half were children (7/13, 54%; daughter: n=4, 57%; son: n=3, 43%) of the persons living with dementia. Their age ranged from 55 to 82 years. Most of the participants (10/13, 77%) reported a caregiving experience of at least 3 years, and most of them (8/13, 62%) cared for a relative who lived at their own residence. Of the 13 participants, 2 (15%) reported that the person they cared for had passed away.

The number of caregivers attending each FG ranged from 2 to 7: of the 13 caregivers, 6 (46%) attended FG 1 on June 14, 2021; a total of 7 (54%) attended FG 2 on July 12, 2021; a total of 2 (15%) attended FG 3 on July 15, 2021; a total of 6 (46%) attended FG 4 on August 18, 2021; and 2 (15%) attended FG 5 on August 24, 2021. Of the 13 participants, 7 (54%) attended >1 FG (Table 3). The main findings of the 5 FGs are summarized in the following paragraphs.

Participants generally believed that iSupport holds the promise to be useful, to increase dementia knowledge, and provide information about available services and support measures for people living with dementia and their families:

> The idea is brilliant because everything can be useful...In my opinion, the most interesting thing is the overview of what is locally available to support caregivers. [ID 9; daughter]

The need for guidance and orientation to services was felt owing to a perceived lack of support and direction, likely stemming from the uncertainty and confusion that frequently followed the diagnosis. This feeling of bewilderment was echoed by a participant:

> It's confusing outside, you don’t know where to go, whom to turn to...there are no guidelines, no support. [ID 2; daughter]

Regarding the contents, participants reported familiarity with most of the case scenarios.

A participant commented as follows on a scenario (module 3, chapter 3) involving a person affected by dementia who cannot find the house keys and does not want the carer to leave him alone at home:

> It happened to me many times, not always with the keys though. [ID 12; daughter]

However, despite the familiarity reported and the need to obtain information and increase knowledge to cope with difficult situations, the original exercise format was seen as a limitation by some of the participants. A participant reported feeling diminished when choosing between wrong and right answers:

> It’s almost guilt-inducing...There is the best solution and if you guess wrong you are doing your role wrong. [ID 8; daughter]

In addition, some of the answer options were considered to be so wrong as to be offensive to the carer; for example, in module 5, chapter 9, a case scenario described a situation where the person affected by dementia (Matteo) makes sexual remarks toward a domestic worker, and the user is asked what they would do in this situation. A participant commented on the option “shout at Matteo and shame him for his conduct” as inconceivable:

> Shout?!...We do know what we’re doing! [ID 9; daughter]

Similar to what formal caregivers reported about case scenarios, participants also highlighted the risk of generalizing solutions that may not be appropriate for all caregiving situations:

> It should be clear that each user has to transpose his or her situation by taking cues from the scenario, but unfortunately it isn’t black and white. [ID 13; spouse]

Finally, the informal caregivers too suggested adding interactive features to the digital version of iSupport to minimize the risk of the caregivers isolating themselves; for instance, a participant commented as follows:

> For me, the biggest utility is in connecting people...there should be people behind the app. [ID 5; spouse]

**Discussion**

**Overview**

This study described in detail the main steps taken to culturally adapt the WHO iSupport program for informal caregivers of people living with dementia in Switzerland. Our results demonstrate the complexity as well as the necessity of adapting an evidence-based complex intervention to a specific cultural context and population. We collected feedback and implemented changes, in accordance with the WHO authors of the program, to the original iSupport version in the areas of the language, resources, contents, graphics, and features used in the program. In the following paragraphs, we summarize and comment on the main lessons learned.

**Valuing Experiential Knowledge**

One of the main messages we took away during the adaptation process was the importance placed by informal caregivers on being recognized for their role and expertise. This finding is consistent with the findings of other studies, including the works reporting on iSupport adaptation processes in other countries [31,34,43-45]. Our participants suggested that the learning
approach used in the original iSupport program was too scholastic and recommended the removal of expressions that likely resulted from a top-down approach to content and compilation. Referring to case scenarios, some of the informal caregivers (5/13, 38%) felt that the simplicity of certain answer options was offensive. Informal caregivers claimed to be recognized because of their lived experience as experts in the field who could contribute to not only locally adapting iSupport but also integrating and shaping it. This echoes the inclusive procedures used to develop iSupport in the first place [21] and the work done for the iSupport adaptation process in Portugal and the United Kingdom [34,43]. Informal caregivers can spend on average 170 hours a month providing care to a loved one affected by dementia [46]. In our study, more than half of the caregivers (10/13, 77%) reported a caregiver experience of at least 3 years and up to 10 years. Although one may argue that caregivers acquire and improve their learning by doing, it is undeniable that they can become experts in caring; surely, they provide a unique perspective of the person with dementia and their own needs. However, besides the years of personal experience, caregivers’ knowledge of dementia and caring may also depend on other factors and may be influenced by their educational level and sociocultural background. Similar to any complex health intervention [47], it is important to ensure that the final version of iSupport is adapted to the real user’s experience and preexisting abilities. An early, timely, and active involvement of caregivers is needed [48,49]. The adoption of a language register and skills training techniques that promote preexisting abilities, rather than replace them, may enhance the acceptance and use of the intervention.

Enhancing Social Contacts

According to participants, iSupport could benefit from the inclusion of interactive features (eg, chat and forum) that allow the user to communicate with other caregivers and share experiences and problem-solving strategies. This finding is consistent with a recent study [50] that found that peer support can be complementary to professional support and beneficial in reducing social isolation, as well as in connecting patients and caregivers to others with similar issues. Similarly, Greenwood et al [51] found that, besides providing psychosocial support, peer support interactions for caregivers of people with dementia can offer practical information and guidance in managing difficult situations and gaining new perspectives on their caring role.

The adoption of peer support programs for informal caregivers of people with chronic diseases and disabilities is well established in the literature [52]. A recent scoping review [53] found that peer support was often part of multicomponent interventions that also addressed information sharing, skills development, personal coping skills, and self-management. Despite the difficulty in identifying what component may or may not be beneficial for the carers, the authors concluded that peer support, particularly if delivered digitally, could represent a cost-effective medium and opportunity to meet caregivers’ needs and preferences.

Importantly, digital meets among peers seem more promising, usable, and potentially effective for caregivers when embedded in digital interventions [10] such as iSupport.

Facilitating Access to, and Navigation of Local Services

Another suggested feature to implement in the program was the inclusion of contacts of local resources for dementia, such as health care services and facilities, charities, or other relevant organizations. Consistent with what our participants reported, informal caregivers often experience a lack of information and support, especially at the beginning of the caregiver journey, when it is best to establish fruitful contacts and interactions with local health and social care services and offers in general [1]. According to the latest World Alzheimer Report [3], <50% of informal caregivers are advised to contact the local Alzheimer association or receive postdiagnostic support information. The navigation of the services and various offers for both people living with dementia and informal caregivers is taxing, often ineffective, and can be frustrating. The lack of information about existing services and support is associated with caregiver burden and distress [54]. A recent review on the needs of family caregivers revealed that information provided on available support services and measures was one of the main needs reported by caregivers after their loved one was diagnosed with dementia [55]. Caregivers may seek support autonomously, mainly digitally. However, the variety of information and sources available on the internet about dementia may contribute to creating feelings of bewilderment and difficulties in finding relevant and reliable information [56]. Hence, digital interventions that also include contacts with external and local resources may help users to access and navigate the health care system and find the most appropriate service or information for their situation.

Limitations

We acknowledge that our study has limitations. First, we included only a few participants for each phase of the adaptation process. Because of their pressing needs and duties, informal caregivers are a challenging population to reach and involve in research [57]. However, the number of caregivers and experts that we included in our study was adequate for the qualitative methods used and is higher than the minimum recommended by the WHO guidelines to adapt iSupport to local contexts [22]. In addition, we set up a CAB that included both stakeholders and caregivers who worked continually and with great dedication through the adaptation process of iSupport. Second, the discrepancy in FG size between formal and informal caregivers and the attendance of informal caregivers in >1 FG may have contributed to reaching data saturation, but this may have reduced social desirability bias, thanks to both the progressive cementing of positive small group dynamics among participants and the variety of the contents discussed. Third, our study was conducted in Switzerland, a high-income country, equipped with a National Dementia Strategy that aims to improve the quality of life of people affected by dementia and to promote awareness and education on dementia [58]. Therefore, the feedback and experiences that we collected may not be easily generalized to all contexts. However, the adaptation strategies and phases described in our study may be useful for
all countries interested in adapting digital interventions for caregivers of people with dementia, not only iSupport. Our findings suggest that digital interventions benefit from a community-based participatory approach and the involvement of caregivers to ensure that the final program meets the needs and preferences of users [17].

Future Research

The recommendations and feedback that we collected during this study allowed us to adapt the original contents of the iSupport program to the Swiss context and to inform the development of the iSupport desktop version, mobile app, and printed manual. Following the Medical Research Council guidelines for the development of complex health interventions [59], we will proceed to assess the usability and feasibility of iSupport before its implementation. Evidence not only on the effectiveness but also on the ease of implementation and scalability of caregivers’ interventions is still rare in our country. We are determined to design and conduct good-quality studies to address these gaps and to promptly disseminate our findings and experience widely through peer-reviewed publications, the WHO knowledge exchange platform [60], and the global WHO iSupport network coordinated by the Brain Health Unit at the WHO.

Finally, the iSupport original program was developed by the WHO based on evidence related to carer training and support interventions and in collaboration with experts and caregivers [21]. Therefore, the program can be adapted to the extent that it maintains the original aims and structure [22]. During the study, we collected recommendations and feedback that would have required a consistent change in terms of resources and digital infrastructure to be implemented. These included, for instance, contents based on the type of dementia and stage of the disease, a comprehensive map of all digital and local resources available, and consultation from professionals as well as legal and financial assistance. Therefore, further development of iSupport could focus on supporting specific groups of caregivers, such as young carers or caregivers of people with rare dementia, and on providing personalized support tailored to the stage of the caregiver journey and the care needs of the care recipient.

Conclusions

Despite the recognized importance of culturally adapting interventions to implement them in real-world settings, the evidence on how to conduct this process is still limited. Our study enriches this landscape by underscoring that an active engagement of the final users and stakeholders allows to adapt an intervention to their culture, values, and needs. In addition, this study provides examples of concrete strategies and methods to involve community members and stakeholders across different phases of the intervention. Indeed, despite the emerging importance of coconstructing research together with people as collaborators, rather than as simply subjects of traditional research, there is limited evidence regarding the modalities of this practice.

Our experience confirms that the adoption of a CBPR approach is necessary to identify and address criticisms and potential barriers to the use and acceptance of a digital educational intervention before its implementation. In conclusion, we envision this study as a potential driver for enhancing a more robust dialogue between researchers and communities. We firmly believe that CBPR represents a transformative research opportunity where the needs of academics and community members can be met and where both groups can find opportunity for mutual knowledge exchange and growth.

Acknowledgments

The authors thank all informal and formal caregivers of people with dementia for their time and precious contribution to this work. The authors also wish to acknowledge the support provided by the members of the community advisory board, the funders, and collaborators for the realization of this project.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Community advisory board agreement.

[PDF File (Adobe PDF File), 187 KB - formative_v8i1e46941_app1.pdf ]

Multimedia Appendix 2

Informed consent.

[PDF File (Adobe PDF File), 246 KB - formative_v8i1e46941_app2.pdf ]

Multimedia Appendix 3

World Health Organization adaptation spreadsheet.
References


Abbreviations

CAB: community advisory board
CBPR: community-based participatory research
FG: focus group
REDCap: Research Electronic Data Capture


60. WHO launches new platform for knowledge exchange on dementia. World Health Organization. URL: https://www.who.int/news/item/05-05-2021-who-launches-new-platform-for-knowledge-exchange-on-dementia [accessed 2023-12-13]
A Bluetooth-Enabled Device for Real-Time Detection of Sitting, Standing, and Walking: Cross-Sectional Validation Study

Reza Daryabeygi-Khotbehsara¹, MSc, PhD; Jonathan C Rawstorn¹, PhD; David W Dunstan², PhD; Sheikh Mohammed Shariful Islam¹, MD, PhD; Mohamed Abdelrazeq², PhD; Abbas Z Kouzani³, PhD; Poojith Thummala³, MSc; Jenna McVicar¹, MSc, PhD; Ralph Maddison¹, PhD

¹Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne Burwood, Australia
²Baker-Deakin Department of Lifestyle and Diabetes, Melbourne Burwood, Australia
³School of Information Technology, Deakin University, Melbourne Burwood, Australia
⁴School of Engineering, Deakin University, Geelong, Australia

Corresponding Author:
Reza Daryabeygi-Khotbehsara, MSc, PhD
Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences
Deakin University
221 Burwood Hw
Melbourne Burwood, 3125
Australia
Phone: 61 3 924 45936
Email: reza.d@deakin.edu.au

Abstract

Background: This study assesses the accuracy of a Bluetooth-enabled prototype activity tracker called the Sedentary behaviOR Detector (SORD) device in identifying sedentary, standing, and walking behaviors in a group of adult participants.

Objective: The primary objective of this study was to determine the criterion and convergent validity of SORD against direct observation and activPAL.

Methods: A total of 15 healthy adults wore SORD and activPAL devices on their thighs while engaging in activities (lying, reclining, sitting, standing, and walking). Direct observation was facilitated with cameras. Algorithms were developed using the Python programming language. The Bland-Altman method was used to assess the level of agreement.

Results: Overall, 1 model generated a low level of bias and high precision for SORD. In this model, accuracy, sensitivity, and specificity were all above 0.95 for detecting sitting, reclining, standing, and walking. Bland-Altman results showed that mean biases between SORD and direct observation were 0.3% for sitting and reclining (limits of agreement [LoA]=–0.3% to 0.9%), 1.19% for standing (LoA=–1.5% to 3.42%), and –4.71% for walking (LoA=–9.26% to –0.16%). The mean biases between SORD and activPAL were –3.45% for sitting and reclining (LoA=–11.59% to 4.68%), 7.45% for standing (LoA=–5.04% to 19.95%), and –5.40% for walking (LoA=–11.44% to 0.64%).

Conclusions: Results suggest that SORD is a valid device for detecting sitting, standing, and walking, which was demonstrated by excellent accuracy compared to direct observation. SORD offers promise for future inclusion in theory-based, real-time, and adaptive interventions to encourage physical activity and reduce sedentary behavior.

(JMIR Form Res 2024;8:e47157) doi:10.2196/47157

KEYWORDS
activity tracker; algorithms; deep neural network; machine learning; real-time data; Sedentary behaviOR Detector; sedentary behavior; SORD; standing; validation; walking; wearables

Introduction

Sedentary behavior (SB) is defined as “any waking behavior characterized by an energy expenditure of less than 1.5 metabolic equivalents while in a sitting, reclining, or lying posture” [1,2]. SB is an independent risk factor for many noncommunicable diseases, with the risk being most pronounced in those who are also physically inactive (ie, not meeting physical activity [PA] guidelines) [3–5]. Reducing SB for all
people, including those who are physically active, can assist in producing health benefits [6]. Interrupting SB with standing or light or moderate intensity PA can improve chronic risk factors including glucose homeostasis, insulin sensitivity, blood lipid concentrations, and diastolic blood pressure [7-10]. Recent World Health Organization guidelines on PA and SB explicitly state the importance of reducing sedentary time in addition to promoting PA for adults and older adults, including those with chronic conditions [11]. This has subsequently led to the development of interventions targeting SB reduction, although interventions to date have been compromised by the lack of a tool that can capture SB accurately and in real time. Accurate measurement of sitting, standing, and walking in real time will enable the design of interventions that can adapt to changes in the activity state and can be delivered at times when an individual is most responsive to the intervention, therefore maximizing the potential opportunity for reducing SB and increasing PA [12].

To date, the majority of interventions to reduce SB and promote PA have relied on subjective measurement of these behaviors, which are subject to self-report bias [13,14] and may underestimate daily sitting time by up to 2 hours compared with objective measurement [15]. Few activity trackers, including research-grade (eg, activPAL) and commercial (eg, GENeActive and Fitbit One), measure sedentary time with reasonable precision [16-22], but they are not optimal for SB change interventions [23]. The 2 main issues involve technical difficulty in using support software for real-time interventions and concerns about device accuracy in distinguishing postural states (sitting, standing, etc) [24,25]. Most activity trackers use similar technologies, including accelerometers, magnetometers, and gyroscopes, to detect posture and activity [26]. However, the placement of devices on the body can considerably influence accuracy [27]. Commercial wrist-worn devices such as the Garmin Vivofit are unable to detect sit-to-stand transition [23,28]. Other thigh-worn devices, such as activPAL and SitFIT, are capable of detecting sitting and standing due to their horizontal placement [23]. In terms of behavioral intervention, activPAL does not offer any real-time prompts or feedback to participants [29]. The SitFit device provides real-time feedback to the user, and its accuracy, although acceptable, was lower when compared to the activPAL, which is considered the preferred device for research purposes [30]. However, SitFit is pocket-worn, which limits its use for those not wearing suitable clothing (eg, trousers) or garments without pockets (eg, dresses) [30]. More importantly, SitFit does not distinguish standing from walking [31] and therefore cannot be used to assess standing as a unique outcome both for real-time and adaptive interventions. It should be noted that these devices (SitFit and Fitbit One) are no longer available on the market and were included in our discussion to provide historical context and illustrate the evolution of activity-tracking technology. Evidence on the positive impact that standing may have on health outcomes in different population groups is emerging from short-term and small-scale studies [32,33], although real-time assessment and behavior change interventions are missing. This, in turn, suggests a need for a platform to momentarily evaluate both sedentary and standing outcomes to study their exclusive health effects and intervene accordingly.

In summary, despite the presence of activity tracker devices, few have included evidence- and theory-based interventions or strategies to promote PA and reduce SB (eg, self-monitoring and goal setting), and the use of some other devices is restricted due to a lack of real-time assessment of outcomes (eg, standing). In response, we designed and developed a new wearable platform called “Sedentary behaviOR Detector” (SORD), which collects real-time sedentary data, including lying, reclining, sitting, and standing, as well as walking activity time. Therefore, this study aimed to assess the validity of the SORD device in detecting sedentary and walking activities among adult participants.

Methods

Overview

A cross-sectional, laboratory-based study was conducted to assess the criterion validity (SORD vs direct observation) and convergent validity (SORD vs activPAL). Adults were recruited to take part in this laboratory-based study through print and email advertisements at a university campus. Adults aged 18 years or older, without gait abnormalities, able to walk on a treadmill easily, with no skin sensitivity to plasters or tapes, and able to communicate in English were included. Upon arrival, participants completed a demographic questionnaire including age, sex, ethnicity, job status, marital status, education, and the Physical Activity Readiness Questionnaire [34] for safe exercise. Anthropometric measures, including height to the nearest 0.1 cm and weight to the nearest 0.1 kg, were taken using a stadiometer (Seca 213) and Tanita scale (Tanita Innerscan 50), respectively.

Participants were given a printed activity protocol to help familiarize them with the required activities and the order in which they were to be performed. Textbox 1 presents a range of different states of activities included in the study protocol to mimic typical postures that may be encountered during everyday life.

Hypoallergenic retention dressing tape (Hypafix) was used to attach the SORD and activPAL devices on the midline of the right thigh. Participants were then instructed to engage in a combination of activities in the order of sitting, reclining, sitting, standing, walking, standing, sitting, lying, and walking on a treadmill. Each activity variation lasted for a minimum of 2 minutes and a maximum of 3 minutes and 30 seconds, except walking, which involved participants walking at their regular walking pace along a 10-m-long path. Participants had 2 minutes of optional resting to break up the activities if needed. Ground truth, or the true time spent on each of the activities, was measured by a researcher with the help of a video camera for direct observation.
### Sedentary behaviOR Detector

#### Overview
The SORD is a wearable electronic device (Figure 1A) that collects and provides real-time data associated with sitting, reclining, lying, and PA. Data provided by the device can be used to separate sitting versus standing versus ambulation. To separate sitting time from lying time, 2 same devices will be attached to 2 different locations of the body.

The SORD device includes a number of internal components (Figure 1B): a low-power processor and transceiver, inertial measurement unit, voltage regulator, battery charger, battery, antenna, micro-USB connector, LEDs, motherboard, and an enclosure. These components have been described below.

#### Figure 1.
(A) Sedentary behaviOR Detector (SORD). (B) Internal components of the SORD device. The SORD is a small device with the following dimensions: 0.9 mm (height), 37 mm (width), and 68 mm (length). It is also lightweight, with a weight of 23.5 g. The device can operate for about 45 hours on a single charge. The SORD device measures 3-axis orientation using the accelerometer that gives acceleration signals for 3 axes, the gyroscope that provides rotation along 3 axes, and the magnetometer that gives motion in the magnetic field in 3 axes. It hosts an embedded C firmware that continuously reads from the sensors, records their data at 25-Hz frequency, preprocesses the data, and transmits the data wirelessly. No initialization is required for the SORD device, as the data are captured and transmitted through the 2.4-GHz Bluetooth Low Energy 5.0 transceiver in real time.

### Processor and Transceiver
The ATSAMB11-ZR210CA is used that includes a low-power ARM Cortex M0 32-bit processor, 128 KB of RAM, 128 KB of stacked flash memory, a 2.4 GHz Bluetooth Low Energy 5.0 transceiver and modem, a power management unit, a ceramic high-gain antenna, and a printed circuit board with a small footprint.

### Inertial Measurement Unit
The BNO055 is used that includes a single-chip integrated circuit incorporating an intelligent inertial measurement unit with a triaxial 14-bit accelerometer, a triaxial 14-bit gyroscope, a triaxial geomagnetic sensor, an I2C communication interface, and an ARM Cortex M0+ 32-bit processor executing a sensors data fusion algorithm.

### Voltage Regulator
The XC9264B755MR-G is used which includes a synchronous step-down DC/DC voltage regulator. It operates within the voltage range of 3-18 V and provides a 500 mA output current. It has a selectable switching frequency of 500 kHz, 1.2 MHz, or 2.2 MHz. It also features overcurrent protection as well as thermal shutdown.

---

**Textbox 1. Details of the Sedentary behaviOR Detector phase 1 activities.**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lying</strong></td>
<td>• Face up, on the right shoulder, face down, or on the left shoulder</td>
</tr>
<tr>
<td><strong>Reclining</strong></td>
<td>• Normal (135 slope chair), left leg over right, or right leg over left</td>
</tr>
<tr>
<td><strong>Sitting</strong></td>
<td>• Upright, ankle-on-knee (left-right and right-left), right foot move, left foot move, both feet move, elbows on legs, or sitting with outstretched legs</td>
</tr>
<tr>
<td><strong>Standing</strong></td>
<td>• Stand normal, casual standing (more weight on the right foot), casual standing (more weight on the left foot), right shoulder on the wall, or left shoulder on the wall</td>
</tr>
<tr>
<td><strong>Walking</strong></td>
<td>• Normal on level, on treadmill at 4 km/h, or on treadmill at 6 km/h</td>
</tr>
</tbody>
</table>

---

https://formative.jmir.org/2024/1/e47157

---

JMIR Form Res 2024 | vol. 8 | e47157 | p.393

(page number not for citation purposes)
Battery Charger
The BQ25101YFPR is used which includes a linear Li-Ion and Li-Pol battery charger with a very small footprint. It has a single power output that charges a battery in 3 steps: conditioning, constant current, and constant voltage. The junction temperature of the device is monitored to control the charge current.

Micro-USB Connector
A micro-USB connector is used for programming the processor and also for establishing serial communications as well as charging the onboard battery.

LEDs
A total of 2 multicolor LEDs are used to illuminate different functional states of the device to the user.

Battery
A 3.7-V, 700-mAh, 303759 Lithium Polymer rechargeable battery is used. Its height, width, and length are 3 mm, 37 mm, and 59 mm, respectively, and its weight is 14 g.

Antenna
A Freedom 2.4-GHz flex circuit PCB antenna is used.

Motherboard
A printed circuit motherboard is designed and fabricated to host all the electronic components of the SORD device.

Enclosure
A small enclosure for the SORD device is designed and 3D printed. It hosts all the components of the device.

ActivPAL
ActivPAL is a thigh-worn triaxial accelerometer that classifies an individual’s activity into periods of time spent sedentary (lying or sitting), standing, and walking, as well as the number of steps and stepping speed [29,35]. ActivPAL devices were initialized before the data collection and date-time stamped 1-second epoch files were used for comparative analysis.

Direct Observation
True time spent engaging in activities was logged by a trained researcher (RDK). This was assisted by a video camera positioned in the room and checked by another researcher (JM). If there was any discrepancy, RDK and JM reviewed the camera data together to achieve consensus. No formal intra- or interrater reliability was conducted.

Data Handling and Analysis
SORD data were transmitted to a computer through Bluetooth Low Energy. A program was developed in MATLAB (MathWorks) and run on a Microsoft Windows (Microsoft Corp.)-based computer to receive data from the SORD devices in real time and store it into a Microsoft Excel (Microsoft Corp.) file. The program starts by initializing relevant variables and a communications port, creates a file name based on the current date and time, continuously receives data from the SORD devices, and stores the incoming data in the Microsoft Excel file in real time. Each data packet received from the SORD devices includes values obtained from the onboard sensors at the current time. For a data packet, the following information is then stored in the file in real time: date, time, angle, accelX, accely, accelZ, gyroX, gyroY, gyroZ, magX, magY, magZ, and battery voltage. To avoid potential Bluetooth transmission package loss, this study used the time-stamp data from the SORD device instead of the computer receiver (ie, the sending time stamp rather than the receiving time stamp). Thus, we had computer receiver and accelerometer data, along with their timestamps. Based on the real sampling rate, the computer calculated the time stamp difference between each data point (ΔT). In this research, the number of missing data points was defined by missing = ΔT/(1/25Hz) – 1. The values of these data points were filled by the average of the 2 data points before and after the missing data points (eg, Vi[missing] = [Vi – 1 + Vi + 1]/2). Before sending data to the server for inference, the phone app waits until all required data have been received (processing buffer length). Using the VANE (standard) classification algorithms, activPAL data were processed and collected using proprietary software (activPAL Professional Research Edition, PAL Technologies). The software-generated event file was used. This file contains a chronological list of all episodes of sedentary, standing, and stepping (ie, walking) activities recorded at 1-second intervals. The frequency of the recorded signals from SORD was subsequently reduced to 1 Hz (ie, 1-second epochs) for comparative analysis. This reduction in frequency simplifies data processing and facilitates direct comparison with activPAL, which was also sampled at 1 Hz. Furthermore, outliers or irregular data points were identified and removed. Once individual data sets were cleaned, they were combined for subsequent comparative analysis. The combining process involved aligning the data sets temporally so that corresponding data points from both devices were synchronized for direct comparison.

Due to multiple limitations, we did not use the available open-source activity recognition algorithms. These limitations include (1) inconsistency in the data format and ranges, (2) differences in the frequency of raw data assumed by these algorithms compared to SORD (which is 28 measurements per second), and (3) the variations of activities considered by these algorithms were not exactly the same as the ones we wanted to address in this research at this stage and in the future. Thus, we developed the data engineering and activity recognition models. Although ensemble learning techniques outperform deep learning, they demand higher computation resources and have longer processing latency [36,37]. Therefore, for practical reasons and real-world applications of SORD, this study used deep neural network models—a combination of convolutional neural network and recurrent neural network—to develop algorithms. Deep neural network can learn features automatically from the raw data, therefore performing better than statistical and basic machine learning methods, and they are suitable for recognizing complex activities [38].

A data scientist developed deep learning algorithms to classify activity type and postural states from preprocessed motion sensor data using the Python programming language [39]. First, machine learning classifiers were developed, trained, and tested for the SORD device. A dynamic sliding window approach was used for machine learning [40], where each window was related...
to a particular activity and multiple variables were examined within each window to identify patterns. When a particular activity was detected in the sensor readings, features were extracted to classify activities between the previous one and the current one (further details are provided below). Then, criterion validity (against direct observation) and convergent validity (against activPAL micro) were evaluated. Using Python, the Bland–Altman method was used to assess the level of agreement between SORD and each reference measure (criterion agreement=directly observed time and convergent agreement=activPAL). Mean difference represents the systematic bias, and the limits of agreement (LoA) show the range of agreement between SORD and reference methods, where a positive value indicates underestimation and a negative value indicates overestimation by SORD. For all activity states, we predefined the acceptable LoA between ±10%.

Classification Algorithms
A single data set included SORD, activPAL, and direct observation data for 1 participant. Deep learning was used to randomly select 6 data sets for training, 1 for validation, and 7 for testing. In the training set, similar patterns were identified for the previous 35 data points to specify an activity. Confusion matrices were used to visualize the model’s performance. In a confusion matrix, each row represents the instances in the predicted activity, and each column represents the instances in the actual activity.

Ethical Considerations
Ethics approval was granted by the Deakin University Human Research Ethics Committee’s Human Ethics Advisory Group (HEAG-H 109_2019). All participants provided written informed consent. All research data were anonymized before cleaning and analysis. Participants were remunerated with an Aus $20 (US $14) gift voucher.

Results
Overview
In total, 15 adults (12 female adults) aged between 20 and 62 years completed the experimental study. Table 1 presents the demographic characteristics of the participants.
Table 1. Demographic information of study participants.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Mean (SD) 35.2 (11.6)</td>
</tr>
<tr>
<td></td>
<td>Range 20-62</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>Mean (SD) 70.4 (10.5)</td>
</tr>
<tr>
<td></td>
<td>Range 55.2-84.8</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>Mean (SD) 168.1 (9.6)</td>
</tr>
<tr>
<td></td>
<td>Range 147.0-186.5</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td>Mean (SD) 24.9 (3.0)</td>
</tr>
<tr>
<td></td>
<td>Range 20.1-29.4</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td>Female 12 (80)</td>
</tr>
<tr>
<td></td>
<td>Male 3 (20)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td>Australian 4 (27)</td>
</tr>
<tr>
<td></td>
<td>European 6 (40)</td>
</tr>
<tr>
<td></td>
<td>Asian 4 (27)</td>
</tr>
<tr>
<td></td>
<td>South American 1 (7)</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td>Degree higher than bachelor’s (bachelor’s with honors, masters, or PhD) 7 (47)</td>
</tr>
<tr>
<td></td>
<td>Bachelor’s degree 5 (33)</td>
</tr>
<tr>
<td></td>
<td>Technical and further education or university course below a bachelor’s degree 2 (13)</td>
</tr>
<tr>
<td></td>
<td>Other school qualifications (eg, overseas school, Cambridge examination, or A level) 1 (7)</td>
</tr>
<tr>
<td>Job status, n (%)</td>
<td>Full-time salary or wage earner 6 (40)</td>
</tr>
<tr>
<td></td>
<td>Part-time salary or wage earner 2 (13)</td>
</tr>
<tr>
<td></td>
<td>Student 7 (47)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td>Married or living with a partner 8 (53)</td>
</tr>
<tr>
<td></td>
<td>Single or never married 1 (7)</td>
</tr>
<tr>
<td></td>
<td>Separated, divorced, or widowed 6 (40)</td>
</tr>
</tbody>
</table>

Deep Learning Results

A total of 4 models were presented for SORD. Model 1 classified 3 activities, including sedentary (lying, reclining, or sitting), standing, and walking separately. As illustrated in Figure 2, model accuracy, sensitivity, and specificity for detecting sedentary time were 0.92, 0.99, and 0.87; for standing, they were 0.95, 1.00, and 0.91; and for walking, they were 0.96, 0.92, and 1.00, respectively.

Model 2 included 4 activities: sitting, reclining, standing, and walking; lying was excluded (ie, lying moments observed by video camera were omitted from the data set). As illustrated in Figure 3, model accuracy, sensitivity, and specificity for detecting sitting and reclining were 1.00, 1.00, and 1.00; for standing, they were 0.99, 0.99, and 1.00; and for walking, they were 0.98, 1.00, and 0.95, respectively.

Model 3 included 3 activities: sitting, standing, and walking; reclining and lying were excluded. Respectively, model accuracy, sensitivity, and specificity for detecting sitting were
and specificity for detecting lying were 0.70, 0.54, and 1.00; for sitting and reclining, they were 0.85, 1.00, and 0.75; for standing, they were 0.75, 0.63, and 0.93; and for walking, they were 0.99, 1.00, and 0.98 (Multimedia Appendix 2).

Model 4 included all 5 activities: lying, sitting, reclining, standing, and walking. Respectively, model accuracy, sensitivity, and specificity for detecting lying were 0.70, 0.54, and 1.00; for sitting and reclining, they were 0.85, 1.00, and 0.75; for standing, they were 0.75, 0.63, and 0.93; and for walking, they were 0.99, 1.00, and 0.98 (Multimedia Appendix 1).

Figure 2. Confusion matrix for model 1 classification algorithms. Sedentary (lying, sitting, and reclining), standing, and walking were included in the model.

<table>
<thead>
<tr>
<th>True activity</th>
<th>Sedentary</th>
<th>Standing</th>
<th>Walking</th>
<th>Total data points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sedentary</td>
<td>393013</td>
<td>27674</td>
<td>5041</td>
<td>425728</td>
</tr>
<tr>
<td>Standing</td>
<td>553</td>
<td>182017</td>
<td>1800</td>
<td>184370</td>
</tr>
<tr>
<td>Walking</td>
<td>1</td>
<td>58</td>
<td>65540</td>
<td>65599</td>
</tr>
</tbody>
</table>

Figure 3. Confusion matrix for model 2 classification algorithms. “Sitting and reclining,” standing, and walking were included in the model.

<table>
<thead>
<tr>
<th>True activity</th>
<th>Sitting and reclining</th>
<th>Standing</th>
<th>Walking</th>
<th>Total data points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting and reclining</td>
<td>317479</td>
<td>534</td>
<td>1061</td>
<td>319074</td>
</tr>
<tr>
<td>Standing</td>
<td>0</td>
<td>182211</td>
<td>2159</td>
<td>184370</td>
</tr>
<tr>
<td>Walking</td>
<td>0</td>
<td>46</td>
<td>65553</td>
<td>65599</td>
</tr>
</tbody>
</table>

**Predicted activity**

**Total data points**
Agreement

For models 1 and 2, results of the Bland-Altman analysis comparing second-by-second data on sedentary, standing, and walking time between direct observation versus SORD and activPAL versus SORD are presented in Figures 4 and 5. Multimedia Appendices 3 and 4 illustrate Bland-Altman for the other models. Percentage values are presented in the text (see Figures 4 and 5 and Multimedia Appendices 3 and 4 for true values).

Figure 4. Bland-Altman plot comparing seconds of sedentary behavior, standing, and walking between direct observation and activPAL against the Sedentary behaviOR Detector (SORD) activity tracker (model 1).
Mean differences (biases) between SORD model 1 and direct observation were 6.4% for sedentary (LoA=–4.3% to 17.1%), –8.7% for standing (LoA=–23.5% to 6.1%), and –8.9% for walking (LoA=–22.2% to 4.4%). Results of model 1 show wide limits, although the mean biases were below 10% for all activities. Relative to total activity durations, mean biases between SORD model 1 and activPAL were –2.5% for sedentary (LoA=–15.0% to 9.9%), 1.7% for standing (LoA=–23.3% to 26.9%), and 7.4% for walking (LoA=–2.3% to 17.1%). Results of model 1 comparing SORD to activPAL show wide limits.

Mean biases between SORD model 2 and direct observation were 0.3% for sitting and reclining (LoA=–0.3% to 0.9%), 1.19% for standing (LoA=–1.05% to 3.42%), and –4.71% for walking (LoA=–9.26% to –0.16%). Model 2 showed the narrowest LoA for “sitting and reclining,” standing, and walking, denoting excellent agreement with direct observation. All the mean biases were within ±10%. Relative to total activity durations, mean biases between SORD Model 2 and activPAL were –3.45% for sitting and reclining (LoA=–11.59% to 4.68%), 7.45% for standing (LoA=–5.04% to 19.95%), and –5.40% for walking (LoA=–11.44% to 0.64%). Results of model 2 comparing SORD to activPAL show a wider LoA, although mean biases are relatively low for “sitting and reclining” and walking.

Mean biases between SORD model 3 and direct observation were –6.4% for sitting (LoA=–18.6% to 5.7%), 12.4% for standing (LoA=–6.6% to 31.5%), and –4.9% for walking (LoA=–12.5% to 2.5%). Therefore, sitting and walking were...
overestimated, while standing was underestimated. The mean bias was acceptable for sitting and walking but not standing. A narrow LoA were observed for walking.

Mean biases between SORD model 4 and direct observation were 52.2% for lying (LoA=−6.9% to 111.4%), −32.0% for sitting and reclining (LoA=−78.2% to 14.1%), 48.8% for standing (LoA=−13.7% to 111.4%), and −2.2% for walking (LoA=−6.7% to 2.2%). Therefore, “sitting and reclining” and walking were underestimated while lying and standing were overestimated. Model 4 shows the broadest LoA for “sitting and reclining” and standing, while the narrowest LoA were observed for walking in this model.

Discussion

This laboratory-based study assessed the criterion and convergent validity of a prototype activity tracker (ie, SORD). A high level of accuracy in detecting sitting, standing, and walking for the SORD device among adults was confirmed. Based on the Bland-Altman plots, high levels of agreement with direct observation demonstrated high criterion validity.

ActivPAL is a triaxial accelerometer that has been validated for detecting sitting, standing, and walking activity [29,41,42] and has been widely used in previous intervention studies [43-47]. However, a recent review found that activPAL has lower accuracy during fidgeting [48]. In this study, the agreement between SORD and activPAL was not ideal. The discrepancy observed might result from the inclusion of various fidgeting states. In addition, since activPAL does not enable real-time transmission of data to external devices or networks [29], it cannot be used for real-time or adaptive interventions. SitFit [30] is among the few devices that provide real-time feedback on SB. SitFit (PAL Technologies Ltd) is a pocket-worn device that requires appropriate clothing (eg, trousers with a front pocket), which is a barrier to its usability [30]. SitFit has an embedded screen to provide visual feedback to users and is also Bluetooth-enabled for connectivity to smartphones, tablets, and PCs. However, outputs generated by SitFit include sedentary time (sitting or lying), upright time, and step count [30]. The upright time includes both quiet standing and stepping [30], meaning that SitFit alone is not suitable for measuring standing as an outcome. Measuring standing and its variations (eg, fidgeting while standing) in real time will enable future intervention studies to identify distinct behavioral determinants of standing and to study its long-term clinical implications. As described in this study, SORD accurately measures sedentary (sitting and reclining), standing, and walking time. Other deep learning models (eg, model 4) examined whether the algorithms could distinguish lying from other sedentary states. A lower accuracy was observed for SORD in distinguishing lying from other sedentary activity states. Since the thigh is horizontal during lying posture, distinguishing sitting and lying postures with thigh-worn devices would be difficult. Methods that include rotational angle thresholds to determine the orientation of the thigh have been able to distinguish lying from sitting [49], even though these techniques require validation against direct observation to produce robust evidence.

A strength of this study is the exclusion of several variations of activity states (eg, sitting with outstretched legs, sitting while ankle-on-knee, and standing while shoulder on the wall), allowing more robust testing of the device accuracy and improving the generalizability of findings. For example, detecting standing as it appears in real-life situations and distinguishing from walking will enable the design of interventions measuring standing as a behavioral or clinical outcome. There are also limitations with this study, including the laboratory-based nature of the study. As with any laboratory-based experiment, it is possible that participants behave differently (eg, sit tall and neat and not as they would do normally). Moreover, a comparison between devices in terms of walking intensities was not conducted. This work is the first step in the validation of SORD, and longer-term studies in free-living environments would be necessary future steps to assess its practicality and accuracy under diverse conditions. The majority of participants in this study were female, and that might be considered a source of bias, that is, sex bias. However, evidence suggests that there are no significant differences between female individuals and male individuals in terms of posture, including sitting, standing, and walking [50]. Most participants were younger adults, and therefore the findings may not be generalizable to older adults. Investigating the usability of SORD in populations beyond young adults can help determine its broader applicability. Finally, we observed errors in the raw data from 2 participants for SORD and 3 others for activPAL.

In this study, we did not intend to compare or advance the activity recognition models; rather, the goal was to use the best approach for real-world applications of SORD for real-time intervention. The future development of SORD will include exploring other models (eg, ensemble learning).

In conclusion, SORD accurately detected sitting, standing, and walking activities among healthy young adults, and measurement accuracy was excellent compared to direct observation. While the current iteration of SORD displays promising levels of accuracy, it requires more work and real-world testing in an intervention to assess its applicability. Therefore, SORD holds potential for future integration into evidence- and theory-driven, real-time adaptive interventions to promote activity and reduce sedentary time.

Conflicts of Interest

None declared.

Multimedia Appendix 1

https://formative.jmir.org/2024/1/e47157

JMIR Form Res 2024 | vol. 8 | e47157 | p.400

(page number not for citation purposes)
Supplemental Figure 1. Confusion matrix for Model 3 classification algorithms. Sitting, standing and walking were included in the model.

[ PNG File, 231 KB - formative_v8i1e47157_app1.png ]

Multimedia Appendix 2
Supplemental Figure 2. Confusion matrix for Model 4 classification algorithms. Lying, sitting, reclining, standing and walking were included in the model.

[ PNG File, 352 KB - formative_v8i1e47157_app2.png ]

Multimedia Appendix 3
Supplemental Figure 3. Bland-Altman plot comparing seconds of sitting, standing and walking between the direct observation and SORD activity tracker (Model 3).

[ PNG File, 99 KB - formative_v8i1e47157_app3.png ]

Multimedia Appendix 4
Supplemental Figure 4. Bland-Altman plot comparing seconds of lying, sitting, reclining, standing and walking between the direct observation and SORD activity tracker (Model 4).

[ PNG File, 129 KB - formative_v8i1e47157_app4.png ]

References


34. Physical Activity Readiness Questionnaire PAR-Q and YOU. CSEP. 2002. URL: https://www.qns.org/sites/default/files/par-q.pdf [accessed 2023-12-12]


Abbreviations

- LoA: limits of agreement
- PA: physical activity
- SB: sedentary behavior
- SORD: Sedentary behaviOR Detector
Demographics and Social Factors Associated With Persistent Nonuse of Video Appointments at a Multisite Health Care Institution: Cross-Sectional Study

Pravesh Sharma¹, MD; Celia Kamath², PhD; Tabetha A Brockman³, MA; Anne Roche⁴, PhD; Pamela Sinicrope⁵, MPH, DrPH; Ruoxiang Jiang², BS; Paul A Decker², MS; Vanessa Pazdernik², MS; Christi Patten⁵, PhD

¹Psychiatry and Psychology, Mayo Clinic Health System, Eau Claire, WI, United States
²Quantitative Health Sciences, Mayo Clinic, Rochester, MN, United States
³Health Equity and Community Engagement Research, Mayo Clinic, Rochester, MN, United States
⁴Psychology, Mayo Clinic, Rochester, MN, United States
⁵Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Pravesh Sharma, MD
Psychiatry and Psychology
Mayo Clinic Health System
1221 Whipple Street
Eau Claire, WI, 54703
United States
Phone: 1 7158385369
Email: Sharma.Pravesh@mayo.edu

Abstract

Background: During the COVID-19 outbreak, video appointments became a popular method for health care delivery, particularly in the early stages of the pandemic. Although Mayo Clinic aimed to reduce face-to-face (F2F) appointments to prevent the spread of the virus, some patients continued seeing their health care providers in person. In the later stages of the pandemic, many patients became comfortable with video appointments, even if they were initially hesitant. However, a subset of patients continued to avoid video appointments. It is not yet clear what sociodemographic factors may be associated with this group of patients.

Objective: This cross-sectional study aimed to examine demographic and social determinant of health (SDoH) factors associated with persistent nonusers of video appointments among a sample of patients within a multistate health care organization. We also explored patient beliefs about the use of video for health care appointments.

Methods: We conducted a 1-time cross-sectional paper survey, mailed between July and December 2022, of patients matching the eligibility criteria: (1) aged ≥ 18 years as of April 2020, (2) Mayo Clinic Midwest, Florida, or Arizona patient, (3) did not use video appointment services during April-December 2020 but attended F2F appointments in the departments of primary care and psychiatry/psychology. The survey asked patients, “Have you ever had a video appointment with a healthcare provider?” “Yes” respondents were defined as “users” (adapted to video appointments), and “no” respondents were defined as “persistent nonusers” of video appointments. We analyzed demographics, SDoH, and patient beliefs toward video appointments in 2 groups: persistent nonusers of video appointments and users. We used chi-square and 2-tailed t tests for analysis.

Results: Our findings indicate that patients who were older, lived in rural areas, sought care at Mayo Clinic Midwest, and did not have access to the patient portal system were likely to be persistent nonusers of video appointments. Only 1 SDoH factor (not having a disability, handicap, or chronic disease) was associated with persistent nonuse of video appointments. Persistent nonusers of video appointments held personal beliefs such as discomfort with video communication, difficulty interpreting nonverbal cues, and personal preference for F2F appointments over video.

Conclusions: Our study identified demographic (older age and rural residence), sociodemographic factors (not having a disability, handicap, or chronic disease), and personal beliefs associated with patients’ decisions to choose between video versus F2F appointments for health care delivery. Health care institutions should assess patients’ negative attitudes toward technology prior to introducing them to digital health care services. Failing to do so may result in its restricted usage, negative patient experience,
and wasted resources. For patients who hold negative beliefs about technology but are willing to learn, a “digital health coordinator” could be assigned to assist with various digital health solutions.

(JMIR Form Res 2024;8:e50572) doi: 10.2196/50572

KEYWORDS
digital health; telemedicine; telehealth; video visits; appointments; SDoH, social determinants of health; social determinants; appointment; users; sociodemographic; prevention; discomfort; video communication; communication; willingness; mobile phone

Introduction

Since the COVID-19 pandemic, video appointments have been widely implemented for remotely delivered health care [1,2]. Both prepandemic and pandemic literature suggest that video visits improve provider access to patients, reduce patient travel and wait times, and provide health care quality comparable to face-to-face (F2F) appointments [3-6]. Despite these positive associations with telemedicine, video appointments for nonemergent care have not been as widely used by patients as expected. Studies show that individuals who could not adapt to digital health care delivery have faced significant health care access barriers during and since the pandemic [7]. In the current post-COVID-19 era, digital health care services are a new standard of care [8], and patients who need to be connected to the health care system digitally but are struggling to adapt to telemedicine may experience suboptimal health care [9]. Therefore, factors associated with nonengagement with video visits, especially in patients who have persistently not engaged in video appointments, require further exploration. While patients’ attitudes to telemedicine, especially in the COVID-19 era [10-12], have been explored, gaps remain in understanding social and individual characteristics associated with the persistent nonuse of video appointments for health care.

A large body of evidence suggests that older age, low education, poor digital access (broadband [BB] internet and smart devices), [13] and personal preferences [14] are independently and interactively associated with lower engagement with digital health care [14-17]. This is ironic, given that a critical reason behind the embarkation of digital health care technology was to provide uninterrupted health care access to those who live in remote areas where access to health care providers is limited, those who experience low socioeconomic status and associated transportation challenges and those with poor mobility due to old age and other constraints [18,19]. Evidence also shows that if the individual digital barriers are addressed [20-22], people are willing to engage in technology and participate in telemedicine programs. Preliminary public and institutional efforts to mitigate patient-related barriers to telemedicine are in their infancy but may include brief verbal and printed technology instructions, digital navigation programs for those who have poor digital literacy (comfort and ease of using technology), use of public Wi-Fi and “to-go kits” (smart devices with written instructions on connecting for a visit) [18,23]. Despite nationwide initiatives that accelerated after COVID-19 to encourage patients to use digital health care, many still chose to engage in F2F appointments [14,24]. This phenomenon was observed with and without social distancing associated with the COVID-19 pandemic.

Social determinants of health (SDoH) such as economic stability, access to quality education and health care, neighborhood safety and housing, community and social contexts, and experiences of racism and discrimination [25] significantly impact people’s well-being. The adverse outcomes associated with SDoH inequitably impact marginalized groups and prevent them from accessing quality health care. SDoH not only impact how easily and efficiently people can access health care, but also how they will access it (video vs F2F vs both).

For example, evidence shows that patients who identified as Black, indigenous, or people of color, and were non-English speaking patients and lived in neighborhoods with low socioeconomic status were less likely to engage in digital health care [17,26]. However, it has not been investigated yet which SDoH factor is linked to the persistent use and nonuse of video technology for health care appointments.

This cross-sectional study aimed to examine demographic and SDoH factors associated with no video use (self-reported persistent nonusers of video appointments) among a sample of patients within a large multisite health care organization. We also explored patient beliefs about the use of video for health care appointments. We hypothesized that certain demographic factors, including older age, being a woman, low education, rural residence, and SDoH, such as financial constraints and limited transportation options, may be associated with the persistent nonuse of video appointments.

Methods

Setting

Mayo Clinic consists of a large academic medical center and associated health system spanning the United States in 5 states (Minnesota, Wisconsin, Iowa, Florida, and Arizona). Mayo Clinic’s main campuses are located in Rochester, Minnesota; Phoenix, Arizona; and Jacksonville, Florida. Mayo Clinic Health System (MCHS) consists of clinics, hospitals, and other health care facilities in 4 regions in southern Minnesota, western Wisconsin, and northern Iowa. The Mayo Clinic Midwest (MN, WI, and IA) serves patients that are predominantly White, older people, and living in rural areas. In contrast, Mayo Clinic in Florida and Arizona serve a more diverse patient population.

Ethics Approval

This study was approved by the Mayo Clinic Institutional Review Board (21-004523).

Study Overview and Design

During earlier phases of COVID-19, in early 2020, the US government and the Centers for Disease Control and Prevention
recommended social distancing measures, including stay-at-home orders and video appointments with health care providers [27]. Despite Mayo Clinic’s attempts to minimize F2F appointments to prevent the spread of the virus, many patients requested F2F appointments with their health providers. We were therefore interested in examining whether demographic and SDoH, including area-based metrics (where patients live), were associated with F2F visits. This study used a cross-sectional design with data collected from a 1-time survey administered to Mayo Clinic and MCHS patients.

Survey Instrument and Measures

The survey was designed using results from a prior qualitative study detailed elsewhere [14]. Guided by the qualitative results and informed by a scoping literature review, the survey items were developed to address existing gaps in the literature. The finalized paper survey was pretested with study staff with an estimated 10-15 minutes to complete. The survey included 21 items querying patient’s digital access such as BB internet connection and smart devices, digital literacy (the ease and comfort of using digital technology), use of the patient portal (Mayo Clinic patient online messaging system), use of video appointments, attitudes, and beliefs toward F2F versus video appointments and barriers to engaging in video appointments. The SDoH-related questions included in our survey were adapted from the Social Needs Screening Tool [28] (Multimedia Appendix 1).

Demographic characteristics (age, gender, and race or ethnicity), education status (highest during this study’s period), patient portal status (yes or no), and residence zip codes were extracted from the electronic health record (EHR). Rurality was ascertained from patient zip codes to identify corresponding rural-urban commuting area (RUCA) codes based on the University of Washington classification C method classification [29].

The dependent (outcome) variable was a dichotomous response (yes or no) to the question, “Have you ever had a video appointment with a healthcare provider?”

Data Collection or Procedure

We extracted data from the EHR of adult patients with this study’s eligibility of (1) being aged ≥18 years as of April 2020, (2) being a Mayo Clinic Midwest (Rochester or MCHS), Florida or Arizona patient, (3) not using video appointment services during the time frame of April-December 2020 but attending F2F appointments for nonemergent outpatient clinical care in the departments of primary care and psychiatry/psychology. The Mayo Clinic Survey Research Center mailed eligible patients a survey in a prelabeled return envelope in early July 2022. By that time, a significant number of patients were oriented and made aware of video appointment procedures through self-learning and efforts by our health care institutions. Thus, the following survey item: “Have you ever had a video appointment with a healthcare provider?” with dichotomous responses “Yes/No,” provided valuable cross-sectional information distinguishing patients in this cohort in terms of their ability to adapt or not to evolving remote health care delivery appointments through video appointments for nonemergent care after April 2020. The respondents who marked “no” were defined as “persistent nonusers.” In contrast, those who responded “yes” were defined as “users” who, despite not having used video appointments between April and December 2020, adapted to the changing digital landscape, using them later.

Surveys were mailed to eligible patients stratified by departmental visit type (psychiatry/psychology versus primary care), demographic characteristics (gender, race, and Mayo Clinic location), and if the patient has an active patient portal account. The Survey Research Center mailed reminder letters to nonresponders in August 2022 and then conducted phone call reminders to nonresponders in October-December 2022. Survey participation was closed in January 2023. Survey respondents received a sheet of forever stamps valued at US$5.

Statistical Analysis

Demographics, SDoH, and patient beliefs about video encounters were compared across groups, persistent nonusers of video appointments, and users groups, using the chi-square (exact) test for categorical variables and the 2-sample t test (rank sum) for continuous variables. In all cases, P values <.05 were considered statistically significant.

Results

Overview

Respondent sociodemographic characteristics (N=321) are described in Table 1 overall and by use of video appointments. The survey response rate was 11% (321/3000). In the total respondent sample, 172 (54%) were women, 217 (68%) were White, 169 (53%) had bachelor’s or higher education degrees (persistent nonusers vs users; 84, 52.5% vs 85, 52.8%), and 282 (88%) were urban dwelling (persistent nonusers vs users; 133, 83.1% vs 149, 92.5%; $P=.01$). In addition, 266 (83%) had access to an online patient portal account (persistent nonusers vs users; 122, 76.2% vs 144, 89.4%; $P=.002$).

https://formative.jmir.org/2024/1/e50572

JMIR Form Res 2024 | vol. 8 | e50572 | p.407

(page number not for citation purposes)
Table 1. Demographic factors associated with the using and not using video appointments.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Total (N=321), n (%)</th>
<th>Persistent nonusers of video (n=160), n (%)</th>
<th>Users of video (n=161), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>57.440 (16.400)</td>
<td>60.434 (16.849)</td>
<td>54.465 (15.425)</td>
<td>.001a</td>
</tr>
<tr>
<td>Range</td>
<td>18.111-90.010</td>
<td>18.111-90.010</td>
<td>21.791-86.867</td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td>.05b</td>
</tr>
<tr>
<td>Women</td>
<td>172 (53.6)</td>
<td>77 (48.1)</td>
<td>95 (59)</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>149 (46.4)</td>
<td>83 (51.9)</td>
<td>66 (41)</td>
<td></td>
</tr>
<tr>
<td><strong>Specialty</strong></td>
<td></td>
<td></td>
<td></td>
<td>&gt; .99c</td>
</tr>
<tr>
<td>Community medicine</td>
<td>309 (96.3)</td>
<td>154 (96.2)</td>
<td>155 (96.3)</td>
<td></td>
</tr>
<tr>
<td>General internal medicine</td>
<td>6 (1.9)</td>
<td>3 (1.9)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>Psychiatry and psychology</td>
<td>6 (1.9)</td>
<td>3 (1.9)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Site</strong></td>
<td></td>
<td></td>
<td></td>
<td>.02b</td>
</tr>
<tr>
<td>Arizona</td>
<td>134 (41.7)</td>
<td>67 (41.9)</td>
<td>67 (41.6)</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>96 (29.9)</td>
<td>38 (23.8)</td>
<td>58 (36)</td>
<td></td>
</tr>
<tr>
<td>Mayo Clinic Midwest</td>
<td>91 (28.3)</td>
<td>55 (34.4)</td>
<td>36 (22.4)</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td>.38b</td>
</tr>
<tr>
<td>Non-White</td>
<td>89 (27.7)</td>
<td>47 (29.4)</td>
<td>42 (26.1)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>15 (4.7)</td>
<td>5 (3.1)</td>
<td>10 (6.2)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>217 (67.6)</td>
<td>108 (67.5)</td>
<td>109 (67.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td>.54b</td>
</tr>
<tr>
<td>Less than equal to 12th grade</td>
<td>30 (9.3)</td>
<td>16 (10)</td>
<td>14 (8.7)</td>
<td></td>
</tr>
<tr>
<td>Some college, no degree</td>
<td>22 (6.9)</td>
<td>9 (5.6)</td>
<td>13 (8.1)</td>
<td></td>
</tr>
<tr>
<td>Associate degree</td>
<td>36 (11.2)</td>
<td>18 (11.2)</td>
<td>18 (11.2)</td>
<td></td>
</tr>
<tr>
<td>Bachelors</td>
<td>84 (26.2)</td>
<td>36 (22.5)</td>
<td>48 (29.8)</td>
<td></td>
</tr>
<tr>
<td>Higher education</td>
<td>85 (26.5)</td>
<td>48 (30)</td>
<td>37 (23)</td>
<td></td>
</tr>
<tr>
<td>Decline to answer</td>
<td>64 (19.9)</td>
<td>33 (20.6)</td>
<td>31 (19.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
<td>.36c</td>
</tr>
<tr>
<td>Married</td>
<td>238 (74.1)</td>
<td>117 (73.1)</td>
<td>121 (75.2)</td>
<td></td>
</tr>
<tr>
<td>Single, separate, divorced, or widowed</td>
<td>71 (22.1)</td>
<td>39 (24.4)</td>
<td>32 (19.9)</td>
<td></td>
</tr>
<tr>
<td>Unknown or chose “N”</td>
<td>12 (3.7)</td>
<td>4 (2.5)</td>
<td>8 (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Rural versus urban</strong></td>
<td></td>
<td></td>
<td></td>
<td>.01b</td>
</tr>
<tr>
<td>Rural</td>
<td>39 (12.1)</td>
<td>27 (16.9)</td>
<td>12 (7.5)</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>282 (87.9)</td>
<td>133 (83.1)</td>
<td>149 (92.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Portal (online patient messaging system)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.002b</td>
</tr>
<tr>
<td>No</td>
<td>55 (17.1)</td>
<td>38 (23.8)</td>
<td>17 (10.6)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>266 (82.9)</td>
<td>122 (76.2)</td>
<td>144 (89.4)</td>
<td></td>
</tr>
</tbody>
</table>

aTwo-sample 2-tailed t test.
bChi-squared test.
cFisher exact test.
Demographic Correlates to Persistent Nonuse of Video Appointments

Persistent nonusers of video appointments were older than users ($P=.001$). In addition, patients living in rural residences ($P=.01$) were more likely to be persistent nonusers of video appointments. Other demographic factors, such as gender, education, and race, were not significantly different between persistent nonusers and users of video appointments (Table 1).

Institution Site Correlates to Persistent Nonuse of Video Appointments

Patients who sought care at Mayo Clinic Midwest, comprising Mayo Clinic, Rochester, and MCHS, were more likely to be persistent nonusers of video appointments ($P=.02$; Table 1).

Social Correlates to Persistent Nonuse of Video Appointments

Patients without any disability, handicap, or chronic disease were more likely to be persistent nonusers of video appointments than users ($P=.01$; Table 2). Other SDoH-related factors were not statistically significant.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=321)</th>
<th>Persistent nonusers of video (n=160)</th>
<th>Users of video (n=161)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within the past 12 months, did you worry that your food would run out before you got money to buy more?</strong>&lt;br&gt;Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.12</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>311 (97.2)</td>
<td>157 (98.7)</td>
<td>154 (95.7)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>7 (2.2)</td>
<td>1 (0.6)</td>
<td>6 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>2 (0.6)</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, did the food you bought just not last, and you did not have money to get more?</strong>&lt;br&gt;Missing</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>.45</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>311 (97.5)</td>
<td>156 (98.7)</td>
<td>155 (96.3)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>7 (2.2)</td>
<td>2 (1.3)</td>
<td>5 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Do you have housing?</strong>&lt;br&gt;Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.10</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>158 (98.1)</td>
<td>152 (95.6)</td>
<td>158 (98.1)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>2 (1.2)</td>
<td>4 (2.5)</td>
<td>5 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.6)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Are you worried about losing your housing?</strong>&lt;br&gt;Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>309 (96.6)</td>
<td>154 (96.9)</td>
<td>155 (96.3)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>9 (2.8)</td>
<td>4 (2.5)</td>
<td>5 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>2 (0.6)</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, have you or your family members you live with been without utilities?</strong>&lt;br&gt;Missing</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>.62</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>311 (98.4)</td>
<td>155 (99.4)</td>
<td>156 (97.5)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>4 (1.3)</td>
<td>1 (0.6)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, lack of transportation?</strong>&lt;br&gt;Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.28</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>312 (97.5)</td>
<td>154 (96.9)</td>
<td>158 (98.1)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>7 (2.2)</td>
<td>5 (3.1)</td>
<td>2 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, did you have trouble paying your bills?</strong>&lt;br&gt;Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.89</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>304 (95)</td>
<td>152 (95.6)</td>
<td>152 (94.4)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>14 (4.4)</td>
<td>6 (3.8)</td>
<td>8 (5)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>2 (0.6)</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Does any disability, handicap, or chronic disease make it difficult for you to engage in your typical activities?</strong>&lt;br&gt;Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.01</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>279 (87.2)</td>
<td>147 (92.5)</td>
<td>132 (82)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>36 (11.2)</td>
<td>11 (6.9)</td>
<td>25 (15.5)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>5 (1.6)</td>
<td>1 (0.6)</td>
<td>4 (2.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Are you currently working for pay?</strong>&lt;br&gt;Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.61</td>
</tr>
</tbody>
</table>
Users of video (n=161)  
Persistent nonusers of video (n=160)  
Total (N=321)  

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=321)</th>
<th>Persistent nonusers of video (n=160)</th>
<th>Users of video (n=161)</th>
<th>P value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>145 (45.6)</td>
<td>74 (46.8)</td>
<td>71 (44.4)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>172 (54.1)</td>
<td>83 (52.5)</td>
<td>89 (55.6)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>1 (0.6)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Fisher exact test.

**Video Encounter-Related Beliefs Correlate to Persistent Nonusers of Video Appointments**

Scenario 1: “Imagine you are having a video appointment with a Mayo Clinic doctor for a general medicine health check-up that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.” A significantly lower proportion of persistent nonusers responded “agree,” while a significantly higher proportion of persistent nonusers of video appointments responded “somewhat disagree” or “disagree,” respectively, to the following statements in response to this scenario: “I am confident my doctor would be able to address any medical concerns effectively” \(P=.047\), “I am confident I would be able to express all my concerns clearly” \(P=.04\) and “I am confident I would feel comfortable enough to talk openly” \(P<.001\) compared to users (Table 3). No other responses were statistically significantly associated with the comparison groups.
Table 3. Patients’ beliefs about video encounters and their association with the use of video appointments.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=321)</th>
<th>Persistent nonusers of video (n=160)</th>
<th>Users of video (n=161)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario #1: “Imagine you are having a video appointment with a Mayo Clinic doctor for a general medicine health check-up that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.”</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am confident my doctor would be able to address any medical concerns effectively</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=miss</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>186 (58.7)</td>
<td>80 (51)</td>
<td>106 (66.2)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>91 (28.7)</td>
<td>52 (33.1)</td>
<td>39 (24.4)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>19 (6)</td>
<td>12 (7.6)</td>
<td>7 (4.4)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>21 (6.6)</td>
<td>13 (8.3)</td>
<td>8 (5)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would be able to express all my concerns clearly</td>
<td></td>
<td></td>
<td></td>
<td>.047a</td>
</tr>
<tr>
<td>N=miss</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>222 (70.3)</td>
<td>99 (63.1)</td>
<td>123 (77.4)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>64 (20.3)</td>
<td>41 (26.1)</td>
<td>23 (14.5)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>16 (5.1)</td>
<td>9 (5.7)</td>
<td>7 (4.4)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>14 (4.4)</td>
<td>8 (5.1)</td>
<td>6 (3.8)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would feel comfortable enough to talk openly</td>
<td></td>
<td></td>
<td></td>
<td>.0101b</td>
</tr>
<tr>
<td>N=miss</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>247 (78.4)</td>
<td>106 (67.9)</td>
<td>141 (88.7)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>46 (14.6)</td>
<td>33 (21.2)</td>
<td>13 (8.2)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>11 (3.5)</td>
<td>8 (5.1)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>11 (3.5)</td>
<td>9 (5.8)</td>
<td>2 (1.3)</td>
<td></td>
</tr>
<tr>
<td>I feel video appointments should cost the same and are of equal value to face-to-face appointments</td>
<td></td>
<td></td>
<td></td>
<td>.09a</td>
</tr>
<tr>
<td>N=miss</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>66 (21)</td>
<td>25 (16)</td>
<td>41 (25.9)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>80 (25.5)</td>
<td>45 (28.8)</td>
<td>35 (22.2)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>91 (29)</td>
<td>43 (27.6)</td>
<td>48 (30.4)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>77 (24.5)</td>
<td>43 (27.6)</td>
<td>34 (21.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Scenario #2: Imagine you are having an appointment with a Mayo Clinic psychiatrist or psychologist that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am confident my doctor would be able to address any medical concerns effectively</td>
<td></td>
<td></td>
<td></td>
<td>.09a</td>
</tr>
<tr>
<td>N=miss</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>179 (57)</td>
<td>79 (50.6)</td>
<td>100 (63.3)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>81 (25.8)</td>
<td>43 (27.6)</td>
<td>38 (24.1)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>32 (10.2)</td>
<td>21 (13.5)</td>
<td>11 (7)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>22 (7)</td>
<td>13 (8.3)</td>
<td>9 (5.7)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would be able to express all my concerns clearly</td>
<td></td>
<td></td>
<td></td>
<td>.038a</td>
</tr>
<tr>
<td>N=miss</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>193 (61.9)</td>
<td>84 (54.2)</td>
<td>109 (69.4)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>72 (23.1)</td>
<td>41 (26.5)</td>
<td>31 (19.7)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>27 (8.7)</td>
<td>19 (12.3)</td>
<td>8 (5.1)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>20 (6.4)</td>
<td>11 (7.1)</td>
<td>9 (5.7)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would feel comfortable enough to talk openly</td>
<td></td>
<td></td>
<td></td>
<td>.001a</td>
</tr>
<tr>
<td>Variable</td>
<td>Total (N=321)</td>
<td>Persistent nonusers of video (n=160)</td>
<td>Users of video (n=161)</td>
<td>P value</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-------------------------------------</td>
<td>------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>N=miss</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>202 (64.7)</td>
<td>84 (54.2)</td>
<td>118 (75.2)</td>
<td></td>
</tr>
<tr>
<td>2=somewhat agree, n (%)</td>
<td>67 (21.5)</td>
<td>42 (27.1)</td>
<td>25 (15.9)</td>
<td></td>
</tr>
<tr>
<td>3=somewhat disagree, n (%)</td>
<td>24 (7.7)</td>
<td>15 (9.7)</td>
<td>9 (5.7)</td>
<td></td>
</tr>
<tr>
<td>4=disagree, n (%)</td>
<td>19 (6.1)</td>
<td>14 (9)</td>
<td>5 (3.2)</td>
<td></td>
</tr>
</tbody>
</table>

I feel video appointments should cost the same and are of equal value to face-to-face appointments

| N=miss                                                                   | 9             | 5                                   | 4                      | .26 a   |
| 1=agree, n (%)                                                           | 94 (30.1)     | 39 (25.2)                           | 55 (35)                |         |
| 2=somewhat agree, n (%)                                                  | 84 (26.9)     | 46 (29.7)                           | 38 (24.2)              |         |
| 3=somewhat disagree, n (%)                                               | 70 (22.4)     | 38 (24.5)                           | 32 (20.4)              |         |
| 4=disagree, n (%)                                                        | 64 (20.5)     | 32 (20.6)                           | 32 (20.4)              |         |

Video encounter-related beliefs not specific to any discipline

I am confident I would be able to understand when the doctor explains my symptoms or health

| N=miss                                                                   | 7             | 5                                   | 2                      | .046 b  |
| 1=agree, n (%)                                                           | 206 (65.6)    | 90 (58.1)                           | 116 (73)               |         |
| 2=somewhat agree, n (%)                                                  | 79 (25.2)     | 47 (30.3)                           | 32 (20.1)              |         |
| 3=somewhat disagree, n (%)                                               | 17 (5.4)      | 10 (6.5)                            | 7 (4.4)                |         |
| 4=disagree, n (%)                                                        | 12 (3.8)      | 8 (5.2)                             | 4 (2.5)                |         |

I am confident I would be able to read my doctor's facial expressions or nonverbal cues

| N=miss                                                                   | 6             | 5                                   | 1                      | .05 a   |
| 1=agree, n (%)                                                           | 140 (44.4)    | 57 (36.8)                           | 83 (51.9)              |         |
| 2=somewhat agree, n (%)                                                  | 114 (36.2)    | 62 (40)                             | 52 (32.5)              |         |
| 3=somewhat disagree, n (%)                                               | 41 (13)       | 25 (16.1)                           | 16 (10)                |         |
| 4=disagree, n (%)                                                        | 20 (6.3)      | 11 (7.1)                            | 9 (5.6)                |         |

I am confident I would be able to hear my doctor clearly

| N=miss                                                                   | 7             | 6                                   | 1                      | .004 b  |
| 1=agree, n (%)                                                           | 200 (63.7)    | 85 (55.2)                           | 115 (71.9)             |         |
| 2=somewhat agree, n (%)                                                  | 79 (25.2)     | 43 (27.9)                           | 36 (22.5)              |         |
| 3=somewhat disagree, n (%)                                               | 22 (7)        | 16 (10.4)                           | 6 (3.8)                |         |
| 4=disagree, n (%)                                                        | 13 (4.1)      | 10 (6.5)                            | 3 (1.9)                |         |

I would enjoy connecting with my doctor as much as if the appointment were face-to-face

| N=miss                                                                   | 4             | 3                                   | 1                      | .009 b  |
| 1=agree, n (%)                                                           | 108 (34.1)    | 40 (25.5)                           | 68 (42.5)              |         |
| 2=somewhat agree, n (%)                                                  | 88 (27.8)     | 46 (29.3)                           | 42 (26.2)              |         |
| 3=somewhat disagree, n (%)                                               | 68 (21.5)     | 38 (24.2)                           | 30 (18.8)              |         |
| 4=disagree, n (%)                                                        | 53 (16.7)     | 33 (21)                             | 20 (12.5)              |         |

I would feel comfortable talking with a doctor I have met before in-person

| N=miss                                                                   | 7             | 5                                   | 2                      | .16 b   |
| 1=agree, n (%)                                                           | 221 (70.4)    | 101 (65.2)                          | 120 (75.5)             |         |
| 2=somewhat agree, n (%)                                                  | 64 (20.4)     | 36 (23.2)                           | 28 (17.6)              |         |
| 3=somewhat disagree, n (%)                                               | 15 (4.8)      | 8 (5.2)                             | 7 (4.4)                |         |
residences, those who sought care at Mayo Clinic Midwest and observed that patients of older age, those living in rural video appointments in our institution (April 2020). We further persistently have not engaged with video appointments for nonemergent primary and psychiatric care since the start of appointments for health care in a multisite medical institution. This cross-sectional study demonstrated demographic and SDoH Principal Findings Discussion No other responses were statistically significantly associated with the comparison groups. Scenario 2: “Imagine you are having an appointment with a Mayo Clinic Psychiatrist or Psychologist that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.” A significantly lower proportion of persistent nonusers of video appointments responded “agree,” while a significantly higher proportion of persistent nonusers of video appointments responded “somewhat disagree” or “disagree,” respectively, to the following statements in response to this scenario: “I am confident I would be able to express all my concerns clearly” (P=.03), and “I am confident I would feel comfortable enough to talk openly” (P=.001) compared to users. No other responses were statistically significantly associated with the comparison groups. Video Encounter-Related Beliefs as a Correlate to Persistent Nonuse of Video Appointments A significantly lower proportion of persistent nonusers of video appointments responded “agree,” while a significantly higher proportion of persistent nonusers of video appointments responded “somewhat disagree” or “disagree” to the following statements: “I am confident I would be able to read my doctor’s facial expressions or non-verbal cues” (P=.05), “I am confident I would be able to hear my doctor clearly” (P=.004), “I would enjoy connecting with my doctor as much as if the appointment were face-to-face” (P=.009), and “I would feel comfortable talking with a doctor I have never met before in-person” (P=.01) compared to users. No other responses were statistically significantly associated with the comparison groups. Discussion Principal Findings This cross-sectional study demonstrated demographic and SDoH factors associated with persistent nonusers of video appointments for health care in a multisite medical institution. We observed that about 50% (161 of 321) of respondents persistently have not engaged with video appointments for nonemergent primary and psychiatric care since the start of video appointments in our institution (April 2020). We further observed that patients of older age, those living in rural residences, those who sought care at Mayo Clinic Midwest and those who did not have access to the patient portal system were more likely to be persistent nonusers of video appointments. Only a single SDoH-related factor (not having a disability, handicap, or chronic disease) was associated with persistent nonuse of video appointments. We also observed that individuals held certain personal beliefs about video appointments that were associated with their decision to use versus not use video appointments for health care. The persistent nonusers of video appointments held beliefs that included being potentially uncomfortable communicating with their doctor through video, difficulty reading their doctor’s facial expressions or nonverbal cues, struggle to hear the doctor clearly, and overall better comfort with F2F appointments over video appointments. Much evidence has demonstrated that older patients have limited engagement with telemedicine, including using video appointments for their health care needs [30-32]. Our study observed an analogous association with older age correlated with persistent nonuse of video appointments for health care. Given that few individuals in our sample experienced substantial limitations in SDoH (Table 2) and that most respondents lived in urban dwellings and had access to the online patient portal (which requires smart devices and internet BB connection), we speculate that factors other than just digital access barriers should be considered when approaching older patients for increasing digital engagement. One possible factor is limited interest in digital health care due to negative personal beliefs toward video appointments. Given that older adulthood is a period when many individuals experience a decline in physical and cognitive health and could lose interest in exploring newer concepts (technology in this case), it is essential for health care providers and health care systems to take a patient-centered approach to understand the reasoning behind an older adult patient’s preference for in-person versus video appointment and provide the appropriate support and develop barrier mitigating strategies tailored to age to engage these individuals with needed care. This study found that patients who lived in rural areas were more likely to be persistent nonvideo users. This finding has been established by many studies. A key reason for the rural-urban digital health disparity is unequal access to BB connections. Additionally, individuals living in rural areas tend to be older, have limited education, and lack the financial resources to invest in BB connections and smartphones. Overall, our research effectively collected information on demographic
indicators associated with not using video appointments that parallels the geographic demographics of Mayo Clinic, Rochester and various MCHS locations in rural areas with mostly older White residents.

As part of this study, we also aimed to evaluate the social factors or SDoH-related concerns that contributed to the persistent nonusers of video appointments. In our sample, the only factor associated with not using video appointments was not having a disability, handicap, or chronic disease. It is possible that their mobility or health allowed for greater flexibility in choosing an F2F visit or that they simply had fewer visits overall and were, therefore, less likely to choose video visits as an alternative. On the other hand, video appointments could be specifically beneficial for patients with disabilities or chronic illnesses who may have challenges with physical energy or mobility, be at higher risk for contracting illness when in public or have more health care appointments to attend overall. Given that there was limited variation in SDoH within our sample, this may have limited our ability to identify potential correlations between SDoH factors and selecting video or F2F appointments. Large-scale studies with socially diverse patients are required to fully understand the extent to which SDoH factors play in patients’ decision-making in choosing health care delivery methods. This understanding will further enhance patient outreach efforts and strengthen high-impact population health and research initiatives.

Since the pandemic, a significant public effort has been made to increase patients’ digital access at state, federal, and institutional levels [33]. Still, some patients may be unenthusiastic about attending video appointments [14,34]. We found that persistent nonusers of video appointments feel that they may not be able to “express” their concerns and are not able to “feel comfortable enough to talk openly” when having video appointments with primary care and psychiatry practicing physicians. Enjoying F2F encounters better than video appointments and potentially being unable to hear doctors clearly during a video appointment were 2 other beliefs that persistent nonusers of video appointments cited in high proportion. These patients conveyed these beliefs despite evidence of never engaging in video appointments in our institution. It is possible that individuals who consistently do not use video appointments have formed their opinions based on information obtained from sources other than their personal experience. These sources may include internet forums or the opinions of their peers. Another potential explanation could be poor digital experience when they attempted to engage in video appointments due to limited digital access (low-speed internet), language barrier, and low digital and health literacy [35]. It has been widely understood that poor digital experience could trigger patients’ anxiety regarding existing and emerging technology used in health care and may lead to its avoidance. This problem could be solved by appointing a “digital health coordinator” at the institutional level whose sole responsibility should be assisting patients with digital health solutions. This could overcome the perceived reluctance of patients to use digital services for health consultations. In addition, health care institutions should take into account the strong negative attitudes of this group toward video appointments when introducing them to digital technology for health care delivery.

Overall, our study results may inspire researchers to initiate a conversation about video adoption that goes beyond digital access and literacy. Our research examined the impact of SDoH and confidence or belief in video appointments adoption. Previous studies have not investigated which SDoH is most closely associated with video use. Furthermore, individuals with digital access and digital literacy may still choose to refrain from using video appointments. Therefore, objective measures should consider patient beliefs. Health care institutions should assess and evaluate patient preferences when implementing digital health care, especially those with digital competencies. We have yet to identify any digital literacy (validated) scales that have assessed an individual’s digital belief as one of the variables (negative vs positive) to assess overall digital literacy. From the perspective of behavior change theories [36], it is widely accepted that targeting one’s beliefs is essential for behavior change (digital adoption in this case). Therefore, our study adds novelty to the literature by informing researchers about understanding digital beliefs as a confounder in digital literacy and adoption. We suggest that through the community-based participatory research (CBPR) approach, researchers should attempt to identify facilitators to expedite behavior change. In a subsequent study with a larger sample size, it would be worth exploring if patients with limited interest in video appointments have sufficient digital access and literacy.

Limitations

Our study has several limitations, including the low survey response rate, which may have led to selection bias, resulting in a study population that does not accurately represent the target population, and respondents may differ systematically from nonrespondents. We used the self-reported data and the possibility of recall bias. To help alleviate such a concern, we did verify eligibility and the existence of an F2F appointment via EHR. Additionally, though we tried to enroll participants from diverse backgrounds, the majority of patients in our sample were White, lived in urban areas, and did not experience major social challenges, limiting the generalizability of our findings. Our demographic variables were not extensive due to lack of availability or missing values in the data extracted from EHR. In addition, the survey did not include factors related to the health care system, such as whether patients requested a video appointment, if video appointments were encouraged and offered to patients, or if video appointments were offered but declined by the patient. The results of our study may also lack generalizability because the sample was derived from Mayo Clinic patients and there was no feasible way to assess if patients sought care outside Mayo Clinic and used video visits. However, we enrolled patients who have their primary care providers (PCPs) at Mayo Clinic (ie, paneled patients), reducing the likelihood of video visits being done outside of our health care system. According to FAIR Health [37], a national database of private and Medicare claims data, only 0.1% of all claims nationally in 2019 were related to telehealth. This percentage was even lower in rural areas. Based on these data, it is highly unlikely for patients with a PCP at Mayo Clinic to seek outside video-based care. It’s important to note that FAIR Health data
include not just video visits, but also other telehealth technologies such as mobile health, remote patient monitoring, and store and forward technologies. We aimed to gather diverse data by including Mayo Clinic, Arizona, and Mayo Clinic, Florida. Despite a larger number of responses from Florida, the participants who responded were not from a diverse population. The studies show that there are differences in participation rates based on race, including lower rates of completing consent forms and responding to surveys, with Blacks and Hispanics being the most underrepresented [38]. Future work should explore the patterns of video use in more diverse patient populations, especially those who may be more likely to face barriers to health care (eg, patients living in rural areas and patients experiencing challenges with transportation). Our study also had several strengths, including that our sample was drawn from a multistate institution spanning rural and urban settings, the use of a validated measure of SDoH and the inclusion of scenario-based questions to better understand patients’ beliefs about video encounters.

Conclusions

Our study identified sociodemographic factors and personal beliefs about video appointments that influenced patients’ decisions to choose between video versus F2F appointments for health care delivery. The patients who were older, lived in rural residences, sought care at Mayo Clinic Midwest, and who did not have access to the patient portal were more likely to be persistent nonusers of video appointments. We observed a single SDoH factor, that is not having a disability, handicap, or chronic diseases associated with persistent nonusers of video appointments. Not being able to adequately “express” their medical concerns and not “feel comfortable enough to talk openly” and adequately listen to their provider were notable beliefs held by persistent nonusers of video appointments. We recommend that health care institutions consider and evaluate patients’ strong negative beliefs toward video appointments when introducing them to digital technology for health care delivery. Conducting large-scale studies that encompass a diverse range of social and demographic backgrounds is imperative to comprehend why patients prefer video or in-person appointments. Only through such research can we identify the factors that influence their decision-making process.

Acknowledgments

This work was supported by a Mayo Clinic Clinical Practice Committee Eradicating Racism Award and Clinical and Translational Science Awards (CTSA) from the National Center for Advancing Translational Science (grant UL1 TR002377). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The funding source had no role in this study’s design or in the collection, analysis, and interpretation of data; in writing the paper; or in the decision to submit the paper for publication. We thank the Mayo Clinic Survey Research Center for providing assistance with the survey design and data collection.

Authors’ Contributions

PS and CP conceptualized this study, performed the investigation, acquired funding, provided resources, designed the methodology, wrote the original draft of the paper, and reviewed and edited the paper. RJ and PAD performed the formal analysis and reviewed and edited the paper. TB, CK, AR, PS, and VP reviewed and edited the paper.

Conflicts of Interest

PS is a recipient of the Robert A Winn Diversity in Clinical Trials Career Development Award, funded by the Bristol Myers Squibb Foundation.

Multimedia Appendix 1

Novel Strategies to Increase Telehealth Engagement (NSITE) Survey.

References


Abbreviations

BB: broadband  
EHRR: electronic health record  
F2F: face-to-face  
MCHS: Mayo Clinic Health System  
PCP: primary care provider  
SDoH: social determinants of health
provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Transfer of Knowledge on Pneumoconiosis Care Among Rural-Based Members of a Digital Community of Practice: Cross-Sectional Study

Brian Soller¹, PhD; Orrin Myers², PhD; Akshay Sood³,4, MPH, MD

¹Department of Sociology, Anthropology, and Public Health, University of Maryland Baltimore County, Baltimore, MD, United States
²Department of Family and Community Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, United States
³Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, United States
⁴Miners Colfax Medical Center, Raton, NM, United States

Corresponding Author:
Akshay Sood, MPH, MD
Department of Internal Medicine
University of New Mexico Health Sciences Center
1 University of New Mexico MSC 10 5550
Albuquerque, NM, 871310001
United States
Phone: 1 5052724751
Fax: 1 5052728700
Email: asood@salud.unm.edu

Abstract

Background: Given the re-emergence of coal workers’ pneumoconiosis in Appalachia and Mountain West United States, there is a tremendous need to train rural professionals in its multidisciplinary management. Since 2016, the Miners’ Wellness TeleECHO (Extension for Community Health Outcomes) Program held by the University of New Mexico, Albuquerque, and Miners’ Colfax Medical Center, Raton, New Mexico, provides structured longitudinal multidisciplinary telementoring to diverse professionals taking care of miners by creating a digital community of practice. Program sessions emphasize active learning through discussion, rather than didactic training. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. Rural-urban differences in knowledge transfer in such a community of practice, however, remain unknown.

Objective: We aim to evaluate the role of the rurality of the patient or client base in the transfer of knowledge to professionals caring for miners using the digital community of practice approach.

Methods: This is a cross-sectional study of 70 professionals participating in the Miners’ Wellness TeleECHO Program between 2018 and 2019. Drawing insights from social network analysis, we examined the association between the rurality of participants’ patient or client base and their self-reported receipt of knowledge. Our focal independent variable was the respondent’s self-reported percentage of patients or clients who reside in rural areas. We measured knowledge transfer sources by asking participants if they received knowledge regarding the care of miners by creating a digital community of practice. Program sessions emphasize active learning through discussion, rather than didactic training. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. Rural-urban differences in knowledge transfer in such a community of practice, however, remain unknown.

Objective: We aim to evaluate the role of the rurality of the patient or client base in the transfer of knowledge to professionals caring for miners using the digital community of practice approach.

Methods: This is a cross-sectional study of 70 professionals participating in the Miners’ Wellness TeleECHO Program between 2018 and 2019. Drawing insights from social network analysis, we examined the association between the rurality of participants’ patient or client base and their self-reported receipt of knowledge. Our focal independent variable was the respondent’s self-reported percentage of patients or clients who reside in rural areas. We measured knowledge transfer sources by asking participants if they received knowledge regarding the care of miners by creating a digital community of practice. Program sessions emphasize active learning through discussion, rather than didactic training. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. Rural-urban differences in knowledge transfer in such a community of practice, however, remain unknown.

Results: Respondents, on average, identified 4.46 (SD 3.16) unique knowledge sources within the community, with a greater number of cross-stakeholder knowledge sources (2.80) than same stakeholder knowledge sources (1.72). The mean knowledge source range was 2.50 (SD 1.29), indicating that, on average, respondents received knowledge sources from roughly half of the 5 stakeholder groups. Finally, the mean heterogeneity of knowledge sources, which can range between 0 and 0.80, was near the midpoint of the scale at 0.44 (SD 0.30). Multivariable analyses revealed that as the rurality of patient or client bases increased, participants reported more knowledge sources overall, more knowledge sources from outside of their stakeholder groups, a higher knowledge source range, and greater heterogeneity of knowledge sources (P<.05 for all comparisons).

Conclusions: Our findings suggest that participants who serve rural areas especially benefit from knowledge transfer within the TeleECHO community of practice. Additionally, the knowledge they receive comes from diverse information sources,
emphasizing its multidisciplinary nature. Our results underscore the capacity of the TeleECHO model to leverage technology to promote rural health equity for miners.

**(JMIR Form Res 2024;8:e52414)** doi:10.2196/52414

**KEYWORDS**

community of practice; knowledge transfer; pneumoconiosis; telementoring; rural health care; transfer; information; rural; virtual community; lung diseases; lung disease; rural professionals; rural professional; multidisciplinary management; multidisciplinary; miners; miner; health equity

**Introduction**

Recent studies reveal an increasing prevalence and severity of pneumoconiosis (ie, dust-related lung diseases) among US coal workers since the late 1990s [1-7]. Data from the US Coal Workers Health Surveillance Program indicated that the 2017 prevalence of radiographic pneumoconiosis for coal miners with over 25 years of underground mining experience was greater than 10%, which was double the prevalence from the late 1990s. Similarly, the 2014 rate of complicated pneumoconiosis (a particularly deadly form) among long-tenured underground coal miners was 1.1%, compared to 0.3% at its lowest point in the late 1990s [7,8].

This re-emergence of pneumoconiosis presents unique challenges for rural communities. US counties with the highest mortality rates for pneumoconiosis are concentrated in rural contexts with long histories of mining, such as the Rocky Mountain states and central Appalachia [9]. The prevalence of radiographic pneumoconiosis and complicated pneumoconiosis in rural central Appalachian miners, in particular, is much higher than the national average [6]. While the number of miners requiring specialized care has increased, multidisciplinary expertise and access to complex care for pneumoconiosis have decreased in rural areas [10]. Compared with urban residents, residents of rural areas have less access to outpatient pulmonary rehabilitation [11] or pulmonologist services [12]. Rural practitioners also face unique challenges, including professional isolation and complex patient profiles [13], and describe multiple barriers to knowledge acquisition, such as resources and personal costs, physical distance, and time [14]. Such challenges amplify health inequities and mandate innovative approaches to enhance the health and well-being of rural miners, who constitute an underserved, geographically isolated, medically vulnerable, and often underinsured population [10].

Increasing access to education and mentoring of rural professionals involved in the multidisciplinary care of miners can ameliorate the current dearth of skilled expertise in mining-related diseases in rural areas. The multidisciplinary skills required include medicolegal, clinical, and “soft” skills, the latter including the interpersonal and communication skills needed to navigate highly collaborative work in the care of miners. Insufficient expertise among rural providers in these diverse skills demands innovative education and mentoring solutions.

The Miners’ Wellness TeleECHO (Extension for Community Health Outcomes) Program was established in 2016 to provide structured longitudinal multidisciplinary telementoring to members of professional groups caring for miners who reside in pneumoconiosis mortality hotspots in the United States [9]. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. These members from various stakeholder groups constitute a digital community of practice, or a group of people who “share a concern or a passion for something they do and learn how to do better as they interact regularly” [15]. This approach facilitates knowledge transfer and translation among participants. Knowledge transfer refers to the transmission of information and insights between people or groups [16,17], and knowledge translation involves enhancing users’ awareness of multidisciplinary knowledge and its use in day-to-day work and decision-making in the “real world” [16-18]. Importantly, little is known about how digital communities of practice transfer knowledge across professional stakeholder groups that tend to be geographically isolated, such as rural home health care workers and clinicians or specialists. Thus, examining the patterns of knowledge transfer in such communities of practice can provide insight into how technology can be leveraged to enhance care of complex disease in rural settings and how to promote shared objectives within communities of practice.

Preliminary studies indicate a favorable impact of the ECHO telementoring strategies on providers’ self-efficacy in the care of miners [19], adding to the knowledge base about how the ECHO model can enhance the management of other chronic diseases [20-22].

The rurality of the patient bases for those serving miners limits professionals’ capacity to seek and obtain specialized knowledge concerning the care of pneumoconiosis. This specialized knowledge tends to be concentrated within groups from urban areas, where academic health centers are located. Conversely, urban and suburban practitioners may have limited knowledge concerning the day-to-day challenges of rural patients undergoing treatment for complex diseases. The complexity and interdisciplinary nature of care for pneumoconiosis, coupled with the decline of multidisciplinary knowledge sources within rural areas, underscores the need for specialized knowledge transfer to underserved rural areas. Digital communities of practice are well-equipped to transfer multiple kinds of knowledge [23] across stakeholder groups. First, digital meetings help counteract large geographic distances, thereby providing opportunities for transmitting knowledge concerning facts (eg, know-what) and practical knowledge, skills, and expertise (eg, know-how) among otherwise isolated community members. Perhaps equally important, digital communities of practice help members make social connections and leverage their social
networks to gain more access to practical skills and best practices and to adapt to the evolving needs of patients [23]. Understanding knowledge transfer between the community of practice participants from urban and rural patient or client bases is, therefore, essential but has not been fully evaluated—in turn, constituting a critical gap in knowledge. Addressing this knowledge gap can inform evidence-based interventions to enhance future efforts aimed at providing interdisciplinary care for rural miners.

This study evaluates the transfer of knowledge to professionals caring for miners using the digital community of practice approach. We integrate methods from social network analysis to examine patterns of knowledge transfer within and across stakeholder groups within a digital community of practice. We consider the association between the rurality of professionals’ patient or client base and (1) the number of knowledge sources from within the community of practice, (2) the number of knowledge sources outside one’s stakeholder group, (3) the number of knowledge sources within one’s stakeholder group, and (4) the range and heterogeneity in knowledge sources across stakeholder groups. Our study represents a crucial step in assessing the potential to reduce health inequity through greater investment in workforce diversity and interprofessional telementoring efforts that promote collaborative health care in medically underserved mining communities. Our study thus has important implications for understanding how technology fused with specialized expertise can be used to address complex health issues within rural, remote, and medically underserved communities and begin to address health inequities rooted in unequal access to medical care, more broadly. This approach may help rural communities counter the re-emergence of the pneumoconiosis epidemic.

Methods

Study Design

This is a cross-sectional study of professionals participating in the Miners’ Wellness TeleECHO Program, a community-university partnership between a small rural hospital—Miners’ Colfax Medical Center—and its academic partner—University of New Mexico School of Medicine, together constituting the “hub” site of experts. Stakeholder groups include clinical professional groups (clinicians, respiratory therapists, and home health professionals) and nonclinical professional groups (ie, benefits counselors, lawyers or attorneys, and others, including policymakers, administrators, and mine safety officers), constituting the “spoke” sites located across the United States. The hub and spoke partners regularly engage in telementoring and together form a digital community of practice.

Recruitment

This study was based on a convenience sample of 70 participants who volunteered to complete this study’s survey, among all program attendees invited, during the 1-year study duration from September 12, 2018, to September 18, 2019. Core program faculty did not participate in the survey.

Program Description

As detailed in a previous publication [19], TeleECHO sessions have a uniform format and are held at the same time twice every month, lasting 75 minutes. Program sessions begin with 10-minutes of introductions and announcements, followed by a 15-minute didactic delivered by an invited expert and a 20-minute facilitated question-answer session. Next, the program director facilitates a 30-minute interactive case discussion. Program sessions emphasize active learning through discussion, rather than didactic training. Participants earn continuing medical education (CME) credits without charge, upon completing a CME survey. A multidisciplinary curriculum committee follows a structured curriculum that is continually adapted based on the needs of the learning community, which are identified through review of the CME feedback reports. Attendance at ECHO sessions is open and voluntary, which allows those not presenting a case to view the didactic, partake in case discussions, contribute insight from professional experiences, and learn from the expert panel. Participants can also access experts at hub or spoke sites for urgent consultation outside of program sessions through telephone or electronic correspondence. Recorded didactic sessions are made available through a web-based archive.

Program Development

Since July 2016, our program has used the ECHO model to provide long-term and structured telementoring in the care of miners. This approach deviates from traditional telemedicine where providers assume short-term direct care of individual patients [24]. Further, unlike webinars or traditional didactic lectures, the ECHO model provides an interactive discussion of cases with expert panels in real-time that is highly contextualized and adheres to key learning theory principles. As detailed in a previous publication [19], the ECHO model is based on the following five key principles. (1) The model uses internet-based technology for multipoint videoconferencing, to leverage scarce resources. (2) It uses an established disease-management model associated with best evidence for that disease that has been demonstrated to improve outcomes by reducing variation in processes of care and sharing best practices [21,22,25,26]. (3) It uses the principle of case-based learning for participants to learn with guidance from mentors, based on discussion, questions, and investigation of patient cases under their care. Over time, with iterative practice and feedback, participants gain knowledge and skills and progressively become more independent. (4) It creates a digital community of practice, which emphasizes reciprocity in the sharing of skills and information, and acknowledges that all participants bring useful expertise in the care of miners. Through regular interaction, community members increase their own expertise and that of other participants. As a result, the program aims to increase the ability of individual participants to (a) refer miners appropriately to other experts, (b) accept miner referrals from other experts, and (c) to serve as local experts for less experienced community professionals, thereby improving the care of miners. (5) Finally, it uses an internet-based database (ie, iECHO software) to monitor participant outcomes.
Outcomes

We conceptualized knowledge transfer as the transmission of “facts, experiences, and insights” between people or groups [16,17]. We used a social network approach to examine knowledge transfer among community members by measuring respondents’ number of unique knowledge sources. We also considered the stakeholder group where knowledge originates, which allows us to examine the extent to which participants receive interdisciplinary knowledge from others outside of their own stakeholder groups as well as the overall distribution of knowledge sources across stakeholder groups.

We measured knowledge transfer sources by asking participants if they received new and important knowledge regarding the care of miners during and outside of TeleECHO sessions from each of the other participants. To measure knowledge transfer, respondents were given rosters that included names of all registered participants, with the option of providing additional names not on the roster. Rosters were arranged by stakeholder groups to reduce respondent burden and assist recall. We used these nominations to measure our dependent variables that capture unique dimensions of knowledge transfer. Our first dependent variable, number of knowledge sources, is the count of other participants from whom respondents received new and important knowledge (regardless of the source’s stakeholder group).

Apart from the number of knowledge sources, we tested whether rural participants report greater numbers of knowledge sources from outside of their primary stakeholder group. We thus measured the number of cross-stakeholder knowledge sources, which captures the number of participants from whom respondents received knowledge who were outside of respondents’ stakeholder group. We also analyzed the number of same stakeholder knowledge sources, with a measure capturing the number of participants from whom respondents received knowledge that were in the same stakeholder group as the focal respondent.

We also consider 2 dimensions of diversity in the sources of knowledge transfer among respondents. Range captures the extent to which individuals are connected to others from different social systems or interpersonal environments (eg, employers, associations, and schools) [27,28]. Importantly, a higher range level translates to greater access to nonredundant information [29]. We measure knowledge sourcerange by calculating the number of unique stakeholder groups from which respondents reported receiving knowledge. This variable ranges from 0 to 5, with 0 indicating respondents reported no knowledge sources to 5 indicating respondents received knowledge from at least one member from each of the 5-stakeholder groups.

Our second measure capturing the diversity of knowledge sources is heterogeneity of knowledge sources. Our measure of heterogeneity of knowledge sources taps the distribution of knowledge sources across stakeholder groups for each respondent and is calculated as follows [30]:

Here, \( A_j \) is the number of knowledge sources that belong to a stakeholder group \( j \), \( ks \) is the number of knowledge sources, and \( n \) is the total number of stakeholder groups from which the focal respondent reported receiving knowledge (ie, knowledge sourcerange). Heterogeneity increases when respondents receive knowledge from a larger number of different stakeholder groups (ie, have high knowledge source range) and the knowledge sources are equally distributed across the stakeholder groups. In our study, this measure potentially ranges from 0 to 0.8, with higher values indicating greater heterogeneity in knowledge sources. Note, heterogeneity is undefined for respondents reporting 0 knowledge sources, which was the case for 2 of our respondents, who were excluded for analyses of heterogeneity.

Independent Variables

Our focal independent variable captures the level of rurality among a respondent’s patient or client base. The measure is based on the percent of patients or clients who reside in rural areas, as self-reported by the participant. Initial responses were ordinal and included five categories: 1 (0% to 20%), 2 (21% to 40%), 3 (41% to 60%), 4 (61% to 80%), and 5 (81% to 100%). For this study, we collapsed the ordinal variable into a binary variable indicating rural patient or client base, which equals 1 if 41% to 100% of a respondent’s patient or client base resided in rural areas, and 0 if only 0% to 40% of their patients or clients lived in rural areas. We collapsed the categories for 2 reasons. First, exploratory analyses revealed that only 5 respondents reported serving a 21% to 40% rural patient or client base. Second, comparisons of the means of the knowledge source variables across levels of patient or client rurality suggested a threshold effect, with only minimal differences in the outcomes for those serving 0% to 20% versus 21% to 40% rural patient or client base but large differences between these combined categories and those serving a 41% or greater rural patient or client base. The results based on the original 5-category ordinal variable were similar to those presented in this study.

Covariates

Multivariable models include control variables to account for potential confounding between the association between patient or client rurality and our outcomes. Experienced versus fresh participant: fresh participants were defined as those who first attended the community of practice in or after the summer of 2018 (defined as from May 9, 2018, onwards) versus experienced participants (defined by those who had first attended any time between July 1, 2016, and May 8, 2018). Experienced participants had greater cumulative participation and therefore, experience with the TeleECHO Program than fresh participants (11.4, SD 9.8 vs 4.6, SD 4.6 total sessions attended before or during this study’s timeframe; \( P=.03 \)). This cutoff date was chosen based on the date of funding by the sponsor, which allowed the frequency of the TeleECHO Program to be raised from monthly to twice a month. Respondents’ length of care for miners taps the number of years each participant reported having served miners. Initial responses were measured in years. To aid in the interpretation of our regression results, we divide the reported number of years cared for miners by 10, so that the variable captures the number of decades respondents reported having cared for miners.
We also control for participant demographics. We control for age with 2 dummy variables indicating 51 to 60 years old and older than 60 years (1=yes, 0=no) with younger than 50 years old serving as the reference category. Male sex is binary and indicates the respondent reported a male sex identity (1=yes, 0=no). Respondents reported their race and Hispanic ethnicity status. Based on the responses to these questions, respondents were initially categorized as either Asian, non-Hispanic-Black or African American, Hispanic, multiracial or some other race, or non-Hispanic White. We report the percentages of respondents in each race or ethnic category but collapsed categories into a binary variable indicating non-White (1=yes, non-Hispanic White is the reference) in our regression analyses due to the small sizes of non-White racial or ethnic groups in the sample. Alternative methods of collapsing race or ethnic categories resulted in similar findings as those presented here.

**Data Collection**

The program monitored the number of sessions, learners, unique learners, geographical sites of learners, and patient cases presented (using the iECHO software). Survey data were collected using the REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web app for building and managing online surveys and databases.

**Analytic Strategy**

All analyses were conducted in Stata/MP (version 16.0; StataCorp LLC). We used negative binomial regressions to analyze the total number of knowledge sources, number of cross-stakeholder knowledge sources, and number of same stakeholder knowledge sources, which were discrete counts and were over dispersed. Ordinary least squares regression was used to analyze knowledge source diversity and knowledge source heterogeneity. Model coefficients ($b$) and SE were used to summarize effect sizes. Data missingness due to nonresponse was minimal, with 2 respondents declining to report their age, 1 respondent declining to report on length of care for miners, and 1 respondent declining to report on rurality (for a total of 3 respondents having missing data on at least 1 variable). Missing values on these measures were imputed using the Stata `ice` procedure [31], and models were estimated with 10 imputed data sets using the `mi` command suite in Stata 16. The results based on unimputed data using listwise deletion were nearly identical to those presented here.

**Ethics Considerations**

Approval was obtained from the institutional review board, Human Research Protections Office, at the University of New Mexico Health Sciences Center (18-386). Anonymized consent was obtained from all participants. Study data were deidentified for analysis to maintain confidentiality. All participants were provided an electronic merchandise card of US $50 upon survey completion.

**Results**

Table 1 shows the descriptive characteristics of the 70 ECHO participants caring for pneumoconiosis in a cross-sectional study during the timeframe of 2018-2019.
Table 1. Descriptive characteristics of study participants (N=70).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge source variables, mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Number of knowledge sources (N=70)</td>
<td>4.46 (3.16)</td>
</tr>
<tr>
<td>Same stakeholder knowledge sources (n=61)</td>
<td>1.72 (1.46)</td>
</tr>
<tr>
<td>Cross-stakeholder knowledge sources (n=61)</td>
<td>2.80 (2.63)</td>
</tr>
<tr>
<td>Knowledge source range (N=70)</td>
<td>2.50 (1.29)</td>
</tr>
<tr>
<td>Heterogeneity of knowledge sources (n=68)</td>
<td>0.44 (0.30)</td>
</tr>
<tr>
<td>Age group (y), n (%)</td>
<td></td>
</tr>
<tr>
<td>50 or younger</td>
<td>36 (51)</td>
</tr>
<tr>
<td>51 to 60</td>
<td>15 (21)</td>
</tr>
<tr>
<td>Older than 60</td>
<td>19 (27)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>45 (64)</td>
</tr>
<tr>
<td>Male</td>
<td>25 (36)</td>
</tr>
<tr>
<td>Race or ethnicity, n (%)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5 (7)</td>
</tr>
<tr>
<td>Non-Hispanic Black or African American</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>55 (79)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Primary stakeholder group, n (%)</td>
<td></td>
</tr>
<tr>
<td>Clinician</td>
<td>20 (29)</td>
</tr>
<tr>
<td>Respiratory therapist</td>
<td>12 (17)</td>
</tr>
<tr>
<td>Lawyer or attorney</td>
<td>7 (10)</td>
</tr>
<tr>
<td>Benefits counselor</td>
<td>8 (11)</td>
</tr>
<tr>
<td>Home health professional</td>
<td>14 (20)</td>
</tr>
<tr>
<td>Others</td>
<td>9 (13)</td>
</tr>
<tr>
<td>Rurality of patient or client base, n (%)</td>
<td></td>
</tr>
<tr>
<td>Nonrural patient or client base</td>
<td>19 (27)</td>
</tr>
<tr>
<td>Rural patient or client base</td>
<td>51 (73)</td>
</tr>
<tr>
<td>Participant experience, n (%)</td>
<td></td>
</tr>
<tr>
<td>Fresh participant</td>
<td>40 (57)</td>
</tr>
<tr>
<td>Experienced participant</td>
<td>30 (43)</td>
</tr>
<tr>
<td>Decades serving miners (N=70), mean (SD)</td>
<td>0.76 (0.72)</td>
</tr>
</tbody>
</table>

**Knowledge Source Variables Among all Participants**

Respondents, on average, identified 4.46 (SD 3.16) unique knowledge sources within the community. Respondents, on average, reported greater numbers of cross-stakeholder knowledge sources (2.80, SD 2.63) than same stakeholder knowledge sources (1.72, SD 1.46). The mean knowledge source range was 2.50 (SD 1.29), indicating that, on average, respondents received knowledge sources from roughly half of the 5 stakeholder groups. Finally, the mean heterogeneity of knowledge sources, which can range between 0 and 0.80, was near the midpoint of the scale at 0.44 (SD 0.30).

**Knowledge Source Variables Among Participants Serving Rural Versus Nonrural Bases**

We explain the means of the knowledge source measures among those serving rural versus nonrural patient or client bases. Those serving rural patient or client bases, on average, reported 5.00 (SD 3.13) unique knowledge sources compared to 3.00 (SD 2.88) among those primarily serving nonrural patients or clients. There was only a minor difference in the mean number of same
stakeholder knowledge sources for those serving rural (1.68, SD 1.42) versus nonrural (1.86, SD 1.66) patients or clients. However, rural providers, on average, identified 3.30 (SD 2.61) cross-stakeholder knowledge sources, whereas nonrural providers, on average, identified 1.14 (SD 1.99) cross-stakeholder knowledge sources. Finally, comparing the measures of diversity of knowledge sources, those serving rural patients or clients had a higher mean knowledge source range 2.78 (SD 1.22) versus 1.74 (SD 1.19) and mean knowledge source heterogeneity 0.52 (SD 0.25) versus 0.23 (SD 0.31) than those serving primarily nonrural patients or clients.

**Multivariable Results**

Table 2 presents results from multivariable regression models of the different dimensions of knowledge transfer.

### Table 2. Multivariable regression analyses of knowledge transfer among digital community of practice members caring for pneumoconiosis in a cross-sectional study during the timeframe of 2018-2019.  

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Model 1: number of knowledge sources (N=70)</th>
<th>Model 2: same stakeholder knowledge sources (n=61)</th>
<th>Model 3: cross-stakeholder knowledge sources (n=61)</th>
<th>Model 4: knowledge source range (n=70)</th>
<th>Model 5: heterogeneity of knowledge, sources (n=68)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>P value</td>
<td>b (SE)</td>
<td>P value</td>
<td>b (SE)</td>
</tr>
<tr>
<td>Rurality</td>
<td>0.50 (0.22)</td>
<td>.02</td>
<td>0.04 (0.26)</td>
<td>.89</td>
<td>0.91 (0.37)</td>
</tr>
<tr>
<td>Age (y; ≤50 y is the reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51-60</td>
<td>-0.05 (0.24)</td>
<td>.84</td>
<td>0.67 (0.30)</td>
<td>.03</td>
<td>-0.33 (0.34)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>-0.35 (0.28)</td>
<td>.22</td>
<td>0.14 (0.34)</td>
<td>.68</td>
<td>-0.52 (0.40)</td>
</tr>
<tr>
<td>Male sex (female sex is the reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-White (non-Hispanic White is the reference)</td>
<td>0.32 (0.22)</td>
<td>.14</td>
<td>-0.24 (0.30)</td>
<td>.41</td>
<td>0.62 (0.34)</td>
</tr>
<tr>
<td>Non-White (non-Hispanic White is the reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory therapist</td>
<td>0.15 (0.27)</td>
<td>.58</td>
<td>-0.50 (0.36)</td>
<td>.17</td>
<td>0.54 (0.37)</td>
</tr>
<tr>
<td>Lawyer or attorney</td>
<td>-0.04 (0.32)</td>
<td>.89</td>
<td>-0.50 (0.41)</td>
<td>.23</td>
<td>0.36 (0.45)</td>
</tr>
<tr>
<td>Benefits counselor</td>
<td>0.29 (0.29)</td>
<td>.31</td>
<td>-0.82 (0.43)</td>
<td>.06</td>
<td>0.84 (0.39)</td>
</tr>
<tr>
<td>Home health professional</td>
<td>0.34 (0.25)</td>
<td>.18</td>
<td>0.20 (0.30)</td>
<td>.50</td>
<td>0.38 (0.36)</td>
</tr>
<tr>
<td>Others</td>
<td>0.20 (0.29)</td>
<td>.50</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Experienced participant (fresh participant is the reference)</td>
<td>0.37 (0.17)</td>
<td>.03</td>
<td>-0.17 (0.21)</td>
<td>.42</td>
<td>0.74 (0.25)</td>
</tr>
<tr>
<td>Decades serving miners</td>
<td>0.18 (0.16)</td>
<td>.24</td>
<td>0.08 (0.19)</td>
<td>.68</td>
<td>0.17 (0.24)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.68 (0.30)</td>
<td>.02</td>
<td>0.66 (0.35)</td>
<td>.06</td>
<td>-0.58 (0.45)</td>
</tr>
</tbody>
</table>

*a*P values obtained from 2-tailed tests.

*b*In these models, our dependent variables are knowledge sources, that is, other participants from whom respondents (ie, targets) received new and important knowledge (within and outside the source’s stakeholder group).

*c*Not available.

**Number of Knowledge Sources**

Model 1 examines the number of knowledge sources using a negative binomial regression. The results indicate that rural patient or client base is positively associated with the number of knowledge sources (b=0.50; P<.05). This finding suggests that providers serving rural clients or patients identify greater numbers of knowledge sources within the community of practice than participants whose patients or clients reside in nonrural areas, even after accounting for key confounders. Compared with fresh participants, experienced participants report greater numbers of knowledge sources (b=0.37; P=.03). No other participant characteristics were significantly associated with the number of knowledge sources (all P>.05).

**Number of Same Stakeholder and Cross-Stakeholder Knowledge Sources**

Models 2 and 3 examine the number of same stakeholder and cross-stakeholder knowledge sources, respectively. These models provide insight into whether participants tend to identify
knowledge sources from within or outside of their primary stakeholder groups. Note, participants from “Other” stakeholder groups were dropped from Models 2 and 3, as they by definition have all different stakeholder ties and 0 same stakeholder ties. The results from Model 2 indicate that rural patient or client base has a nonsignificant association with the number of same stakeholder knowledge sources ($b=0.04; P=.89$). Compared with those aged 50 years or younger, participants between the ages of 51 and 60 years report more ties to members of the same stakeholder group ($b=0.67; P=.03$). Turning to Model 3, which examines the number of cross-stakeholder knowledge sources, the rurality of the patient or client base was positively associated with the number of cross-stakeholder ties ($b=0.91; P=.01$). This indicates that participants serving larger proportions of rural patients or clients reported larger numbers of cross-stakeholder knowledge sources than those serving smaller proportions of rural patients or clients. Additionally, experienced participants report larger numbers of cross-stakeholder knowledge sources than fresh participants ($b=0.74; P=.003$) and benefits counselors report larger numbers of cross-stakeholder knowledge sources than clinicians ($b=0.84; P=.03$).

**Range and Heterogeneity of Knowledge Sources**

The final models in Table 2 examine the range and heterogeneity of participants’ knowledge sources. These models provide insight into the number of different stakeholder groups from which participants received knowledge, and the extent to which participants’ knowledge sources are equally dispersed across different stakeholder groups. Turning to Model 4, which is a linear regression of knowledge source range, we found that a rural patient or client base has a positive and significant coefficient ($b=0.92; P=.01$). Model 4 also indicates that experienced participants reported higher knowledge source range than fresh and new participants ($b=0.70; P=.03$).

Finally, Model 5 examines the association between rurality and participants’ knowledge source heterogeneity. Whereas range is the count of the number of unique stakeholder groups from which participants receive knowledge, knowledge source heterogeneity also assesses whether stakeholders from which participants receive knowledge tend to be concentrated in 1 stakeholder group (low heterogeneity) versus equally distributed across multiple groups (high heterogeneity). Note, that because knowledge source heterogeneity can only be measured among participants with at least one knowledge source, Model 5 excludes 2 respondents who reported 0 knowledge sources. Patient or client rurality was positively associated with participants’ heterogeneity of knowledge sources ($b=0.25; P=.003$), indicating knowledge sources are more equally distributed across stakeholder groups as the rurality of their patient and client bases increased.

**Discussion**

**Principal Findings**

Community of practice participants with higher proportions of rural patient or client base, on average, report more knowledge sources overall, more knowledge sources from outside of their stakeholder groups, a higher knowledge source range, and greater heterogeneity of knowledge sources than those with a lower proportion of rural patient or client base. These findings were confirmed after adjustment for potential confounders in regression analyses. More broadly, these findings suggest participants who serve rural areas especially benefit from knowledge transfer within the TeleECHO community of practice. Additionally, the knowledge they receive comes from diverse information sources, emphasizing its multidisciplinary nature.

Further, 1 primary objective of Project ECHO is to decentralize knowledge for the care of patients through exchanging insights and information. Knowledge transfer is key to enhancing the care of complex disease by timely, evidence-based information shared by experts who have used, amplified, and applied this knowledge with interested professionals who (1) are seeking knowledge to assist their patients or clients and (2) through its application, increase access to complex disease care for patients in rural and underserved communities. Project ECHO supports knowledge transfer within the community of practice, through experts sharing and discussing evidence in association with challenging questions with which professionals at program spoke sites are wrestling. Our study suggests this knowledge transfer may be particularly effective among professionals with longer experience with the program.

Professionals in rural mining communities often lack access to traditional knowledge sources. This disparity results from professional isolation; challenges with continuing professional education that requires travel to a distant site for participation with resultant closure of their practices, often without adequate coverage available; and unavailability of specialists with more in-depth knowledge about the clinical, medicolegal, and interpersonal aspects of care of miners. The need to increase access to information for rural professionals is, therefore, obvious. To this end, information technology has come to the fore. However, research suggests that even when electronic information services are provided to rural practitioners, they may not be well used [32]. The lack of information handling skills, lack of time, and perceived peripherality to the job are all seen as major constraints [33,34]. However, our study challenges this belief by demonstrating that professionals serving rural areas especially benefit from access to knowledge through the innovative TeleECHO model, which would otherwise remain siloed within stakeholder groups. Further, the knowledge source range and heterogeneity that the TeleECHO model promotes may allow greater access to thought-provoking ideas that foster learning and other growth-enhancing actions [27,35]. To the best of our knowledge, our approach of studying patterns of knowledge transfer, using social network analysis tools, has never been used previously.

**Strengths**

Our study has multiple strengths. It involves an innovative intervention that addresses the barriers to the care of miners by using the TeleECHO model, which provides a multidisciplinary community of practice approach, using internet-based technology, an approach that has been well studied in other diseases [21,22,25,26]. This study is topical and significant because it addresses a critical gap related to the emerging pneumoconiosis epidemic in the rural United States. Since the
ECHO model has been adopted nationally and globally to improve rural access to care for patients with numerous chronic diseases, there already exists infrastructure to allow for rapid scaling of the Miners’ Wellness TeleECHO Program nationally and globally.

Limitations
There are also limitations to this study. We are unable to correlate knowledge transfer to patient outcomes or changes in provider behavior. We have, however, previously published a listing of qualitative changes that our ECHO participants reported they were going to make in their practice, obtained as part of a CME survey requested at the end of each TeleECHO session [36]. Although a small sample size raises the possibility of a type I error, individual professionals and teams of professionals trained in the ECHO model can reach a large number of miners, with the potential for creating exponential change. High-risk individuals who did not volunteer to participate in this study would have not provided information in the estimation of the program effects, thus introducing an element of potential participation bias. The knowledge transfer instrument was not validated in this study. Program participants had variable competencies, with varying levels of sophistication, commitment, expertise, experience, and historic levels of collaboration within the TeleECHO Program. However, adjustment for participant experience with the TeleECHO Program or length of care for miners in the multivariable models did not change our study findings. Intergenerational, interinstitutional, and rural-urban disparities in ability to leverage technology by participating professionals may challenge empirical examinations of knowledge transfer. Finally, data limitations, including survey nonparticipation by the core program faculty and survey nonresponse among the TeleECHO Program participants, preclude the use of complex social network analysis methods (eg, exponential random graph models) commonly used to examine network selection processes in our study. Although our methods are adequate for examining associations between participating characteristics and the number, range, and heterogeneity of knowledge sources, we are unable to examine how network processes such as reciprocal knowledge transfer operate within the learning community.

Conclusions
Despite these limitations, our findings suggest participants who serve rural areas especially benefit from knowledge transfer within the TeleECHO community of practice. Additionally, the knowledge they receive comes from diverse information sources, emphasizing its multidisciplinary nature. Our results underscore the capacity of the Project ECHO model to leverage technology and workforce diversity to facilitate knowledge transfer to rural professionals and ultimately promote health equity among rural and medically underserved mining communities. Although this approach addresses a critical gap related to the emerging pneumoconiosis epidemic in rural United States, future research will evaluate whether this translates into improved patient outcomes in rural mining communities.

Acknowledgments
This study was funded by the Alpha Foundation (grant AFC719-14). The sponsors played no role in the conduct, analysis, and dissemination of this study and in the review and approval of this paper for publication.

Data Availability
The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
All authors were involved with the following: (1) substantial contributions to the conception or design of the work (BS, OM, and AS) or the acquisition (BS, OM, and AS), analysis (BS and OM), or interpretation of data for the work (BS, OM, and AS); (2) drafting the work or reviewing it critically for important intellectual content (BS, OM, and AS); (3) final approval of the version to be published (BS, OM, and AS); and (4) agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved (BS, OM, and AS).

Conflicts of Interest
None declared.

References


Abbreviations

CME: continuing medical education
ECHO: Extension for Community Health Outcomes
REDCap: Research Electronic Data Capture

©Brian Soller, Orrin Myers, Akshay Sood. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Improving Medical Photography in a Level 1 Trauma Center by Implementing a Specialized Smartphone-Based App in Comparison to the Usage of Digital Cameras: Prospective Panel Study

Jan Siad El Barbari¹, MD; Maxim Fikuart¹, MD; Nils Beisemann¹, MD; Michael Müller², PhD; Hannah Syrek², MSc; Paul Alfred Grützner¹, Prof Dr, PhD; Jochen Franke³, Prof Dr, PhD; Sven Yves Vetter¹, Prof Dr, PhD

¹Department of Orthopaedics and Traumatology, BG Klinik Ludwigshafen, Ludwigshafen am Rhein, Germany
²mbits imaging GmbH, Heidelberg, Germany
³Department of Orthopaedics and Traumatology, Tauernklinikum, Zell am See, Austria

Corresponding Author:
Sven Yves Vetter, Prof Dr, PhD
Department of Orthopaedics and Traumatology
BG Klinik Ludwigshafen
Ludwig-Guttmann-Str 13
Ludwigshafen am Rhein, 67071
Germany
Phone: 49 621 6810 2480
Email: sven.vetter@bgu-ludwigshafen.de

Abstract

Background: Medical photography plays a pivotal role in modern health care, serving multiple purposes ranging from patient care to medical documentation and education. Specifically, it aids in wound management, surgical planning, and medical training. While digital cameras have traditionally been used, smartphones equipped with specialized apps present an intriguing alternative. Smartphones offer several advantages, including increased usability and efficiency and the capability to uphold medicolegal standards more effectively and consistently.

Objective: This study aims to assess whether implementing a specialized smartphone app could lead to more frequent and efficient use of medical photography.

Methods: We carried out this study as a comprehensive single-center panel investigation at a level 1 trauma center, encompassing various settings including the emergency department, operating theaters, and surgical wards, over a 6-month period from June to November 2020. Using weekly questionnaires, health care providers were asked about their experiences and preferences with using both digital cameras and smartphones equipped with a specialized medical photography app. Parameters such as the frequency of use, time taken for image upload, and general usability were assessed.

Results: A total of 65 questionnaires were assessed for digital camera use and 68 for smartphone use. Usage increased significantly by 5.4 (SD 1.9) times per week (95% CI 1.7-9.2; P=.005) when the smartphone was used. The time it took to upload pictures to the clinical picture and archiving system was significantly shorter for the app (mean 1.8, SD 1.2 min) than for the camera (mean 14.9, SD 24.0 h; P<.001). Smartphone usage also outperformed the digital camera in terms of technical failure (4.4% vs 9.7%; P=.04) and for the technical process of archiving (P<.001) pictures to the picture archiving and communication system (PACS) and display images (P<.001) from it. No difference was found in regard to the photographer’s intent (P=.31) or reasoning (P=.94) behind the pictures. Additionally, the study highlighted that potential concerns regarding data security and patient confidentiality were also better addressed through the smartphone app, given its encryption capabilities and password protection.

Conclusions: Specialized smartphone apps provide a secure, rapid, and user-friendly platform for medical photography, showing significant advantages over traditional digital cameras. This study supports the notion that these apps not only have the potential to improve patient care, particularly in the realm of wound management, but also offer substantial medicolegal and economic benefits. Future research should focus on additional aspects such as patient comfort and preference, image resolution, and the quality of photographs, as well as seek to corroborate these findings through a larger sample size.
Introduction

Medical photography serves 3 primary purposes: documentation of diseases and procedures, education of patients and medical personnel, and publications in various forms [1-3].

The potential of medical photography lies in its ability to objectify conditions that cannot be properly illustrated by laboratory results or medical imaging. This mitigates the risk of biased descriptions or inconsistent measurements across clinicians, particularly those from different specialties [4]. Additionally, unlike written diagnoses, photos can also be proof of missed diagnoses or negative findings, as they are not limited to the perception and experience of the examiner [5].

Additionally, medical photography provides several key advantages, including supporting medical diagnoses in legal cases, enhancing diagnostic accuracy and therapeutic outcomes, improving the quality of consultations, and offering documentation for billing purposes [6-10].

The digital era and the technological revolution with digital imaging and smart devices have further lowered the threshold of medical photography [11-13]. Now, every adequately instructed person can produce medical photos anywhere at any time, repealing restrictions such as the availability of a trained medical photographer, time pressure in an emergency setting, or missing equipment. An exemption from this are specialized photographs for scientific or educational purposes, or in certain kinds of fields, that is, aesthetic surgeries, in which higher resolution and quality necessitate the use of more professional equipment.

However, data security and patient confidentiality need to be upheld. Thus, current guidelines, such as Recital 26 of the European Directive (EU) 2016/679, demand informed consent of the patient; a defined medical need for this photography; correct documentation; and safe, restricted, and password-protected storage with an access log [14].

Nonetheless, in a recent systematic review analyzing ethical aspects of medical photography, the consent process was found to be insufficient or inadequate in 95% of all cases [15].

Digital cameras are mainly used for medical photography in the clinical setting, and most patients seem to prefer these over smartphones [1,11,16]. This is because it is not clear how the data are stored and protected on either a clinically owned or private device. In both cases, people tend to estimate a higher risk of data-protection infringement in smartphones than in cameras, impairing their general acceptance as a reliable tool for medical photography [17-19]. Additionally, patients’ will to approve is influenced not only by individual consent depending on the reasoning, particularly concerning web-based publication, but also by situational preferences, such as the difference between emergency departments and aesthetic surgeries [15].

In high-paced emergency settings such as trauma units, obtaining immediate verbal consent, witnessed by another health care provider, can often be the most practical approach. This should be followed up with written consent as soon as the patient is stabilized or conscious. In contrast, nonemergency cases allow for a more thorough process where the patient can take the time to understand the implications before giving written consent. Across both scenarios, the ethical principles of autonomy, beneficence, and confidentiality remain paramount, ensuring that patient data are secure and used only for medical purposes.

Inherently, both devices bear the same risk of data infringement. Digital cameras cannot be password protected, do not encode their data, and are not usually stored as would be required: either under supervision or locked away. The last aspect is not a problem with smartphones since they are usually kept within reach all the time. Yet if the phone is not password protected or the pictures are saved in the photo app, they can be accessed by people close to the owner or may accidentally be transferred to cloud storage that is not properly protected and where access is not documented [17].

However, if the photos are taken with a password-protected app and are not stored on the device but directly in the picture archiving and communication system (PACS), data protection would be secured. Moreover, this would diminish the risk of false identification of the photo, and so all legislative demands would be met.

The use of smartphones with apps that fulfill the data-protection requirements in medical photography is being increasingly examined. Yet an extensive literature search revealed that no study has compared the use of such an app with digital cameras in terms of the quantity and efficiency of medical photography [1,3,16-18,20-22].

The following hypothesis was tested: using an app for medical photography would increase the quantity of pictures taken and the efficiency of this process.

Methods

Study Design

A prospective panel study design with 3 stages was chosen. This was realized in the period from June to November 2020 at a level 1 trauma center. No restrictions were made on where and how pictures should be taken, as the usage in clinical routine was to be evaluated. Pictures could be taken in the emergency department, as well as during surgical procedures, in the surgical ward, or in the outpatient clinic. The study focused on general usage patterns and did not collect data on the specific clinical situations in which the photographs were taken or the type of photographs captured. As a first step, the use of digital cameras was assessed using a printed questionnaire, which was handed
out to trauma and orthopedic residents of different years of training with the instruction to fill one out at the end of every week. As a second step, this process was repeated after the installation of a specialized medical imaging app on the clinically owned smartphones, using the same questions adapted to smartphone use. At the end of this second stage, a separate web-based questionnaire about the experiences with the app and its usability and interface could be filled out by all members of the medical staff, not just the residents participating in steps 1 and 2. Since the questionnaire was designed to assess what opportunities and benefits could result from the implementation of the smartphone app in comparison to the digital camera, no respective questionnaire for the usage of the camera was deemed necessary. Both questionnaires were specifically designed for this study; an example of each is included as Multimedia Appendices 1-3.

We used a Likert scale (ranging from –2 to +2) to express experiences with smartphone usage, with –2 representing “strongly nonbeneficial,” +2 as “highly beneficial,” and 0 resembling indifference or no benefit, allowing an intuitive interpretation of the results.

**Digital Camera**

Every resident who entered employment received a digital camera (Lumix DMC-FT30, 16 megapixels; Panasonic Corporation) to be used for medical photography. After taking a photo, the resident had to use 1 of 3 workstations in the clinic that offer the capability to upload the photos to the clinical PACS with certain predefined keywords, that is, preoperative or surgical site, to categorize what kind of image had been taken. After uploading, the images had to be deleted from the secure digital card. The digital camera has a 28-mm² sensor with a pixel pitch of 1.3 µm and a resolution of 16 megapixels.

**Specialized Smartphone Application**

For clinical communication, each resident received a smartphone (Galaxy xCover 4; Samsung) that only allowed phone calls and viewing of radiological images through mRay (version 6.0.3; mbits imaging GmbH), which is a certified app for medical imaging and processing. The smartphone camera has a 20-mm² sensor with a pixel pitch of 1.1 µm and a resolution of 16 megapixels.

In this study, the fully digitalized photo documentation of mRay was used. This is divided into 3 main steps (Figure 1). First, an existing wound is photographed using the smartphone camera (Figure 1A and B). More than 1 picture can be taken if necessary for the patient’s case. In the next step, the wound image can be assigned to the respective body area (Figure 1C). Using a barcode scan or direct search in the clinical PACS, the images can be keyworded and assigned directly to the associated patient’s data (Figure 1D).

**Patient Confidentiality and Data Protection**

In accordance with the hospital’s standard operating procedures (Multimedia Appendix 4), each patient is informed upon admission that, in addition to radiological images, clinical photographs necessary for their treatment may also be taken during their course of stay, and a written consent is signed. Additionally, as soon as a photo is to be taken, the patient is educated again about what kind of picture will be taken, where it will be stored, and why it is necessary, and verbal consent is obtained. In emergency situations, another staff member acts as a witness during the process of taking the photograph. Patient consent is subsequently obtained as soon as the individual regains responsiveness.

For digital cameras, data protection protocols require staff to promptly upload images to the clinical PACS, associating them with the respective patient’s file. Once uploaded, images must be deleted from the secure digital card. When not monitored, the camera should be securely stored. Regarding smartphones,
they are designated solely for clinical use and feature password protection. Additionally, photographs are exclusively taken through a specialized app, which is also password protected, ensuring direct storage of clinical photographs in the PACS.

Statistics

Primary end points were effective usage of a camera or smartphone in times per week and the time taken from capturing to uploading the taken pictures in minutes. Secondary end points were the estimated time necessary to archive and display images from the PACS, as well as the intention and reasoning behind the photographs. Additionally, it evaluated how users experienced the introduction and usage of the app, but this was not statistically analyzed. The continuous variables, usage and time to upload, were expressed using mean (SD), and time to archive photos and display them from the PACS were expressed using median (IQR). Evaluations were conducted using the Mann-Whitney U test, as these variables were considered estimations despite being interval-scaled as an International System of Units variable. The other categorical variables were analyzed using the chi-square test. The level for statistical significance was set at $P<.05$. Statistics were made using Prism (version 8.2.1; GraphPad Software).

Ethical Considerations

All procedures performed in this study involving human participants were in accordance with the ethical standards of the Ethics Committee of the State Medical Association of the Rhineland-Palatinate and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Since the actual photographs taken were acquired as part of the daily clinical routine and were not part of this study, neither informed nor written consent from the patients was necessary. Informed consent was obtained from all individual participants included in the study, and all data were deidentified. No financial compensation was provided to any of the study participants. Data collection, coding, routing, and analysis were in accordance with the legal data protection policy.

Results

A total of 65 questionnaires regarding digital camera use were collected from June to July 2020, and 68 questionnaires regarding the smartphone app were collected from September to November 2020. The questionnaires were filled out by 5 orthopedic residents. Additionally, 19 fully completed web-based questionnaires were received. A comparison of the usage of both devices revealed no significant differences. Cameras were used 16.4 (SD 7.7) times per week for taking pictures and 11.2 (SD 9.7) times per week for showing pictures for consultation, whereas for smartphones, these values were 18.8 (SD 5.9; $P=.10$) times per week and 9.8 (SD 4.4; $P=.47$) times per week, respectively. In 17.5% (SD 16.1%) of cases for taking pictures and 18.6% (SD 22.6%) for showing pictures, a missing digital camera was mentioned; however, this issue never arose with smartphones. Technical failure occurred significantly less if the smartphone was used, with a rate of 9.7% (SD 18.2%) of cases with the digital camera and 4.4% (SD 9.1%) with the smartphone ($P=.04$). If the total amount of usage (taking photos and demonstrating them) is adjusted for the cases of missing devices and technical failure, then the corrected usage for the digital camera is 20.8 (SD 11.4) times per week and for the smartphone, 26.2 (SD 10.1) times per week. This difference was statistically significant ($P=.005$; Table 1).

Table 1. Primary end points. Values are presented as mean (SD), and $P$ values were calculated using the Mann-Whitney U test.

<table>
<thead>
<tr>
<th>Primary end points</th>
<th>Camera (n=65), mean (SD)</th>
<th>Smartphone (n=68), mean (SD)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage, adjusted total (times per week)</td>
<td>20.8 (11.4)</td>
<td>26.2 (10.1)</td>
<td>.005</td>
</tr>
<tr>
<td>Time to upload (hours or minutes)</td>
<td>14.9 (24.0)$^a$</td>
<td>1.8 (1.2)$^b$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Usage (times per week)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taking images</td>
<td>16.4 (7.7)</td>
<td>18.8 (5.9)</td>
<td>.10</td>
</tr>
<tr>
<td>Displaying images</td>
<td>11.2 (9.7)</td>
<td>9.8 (4.4)</td>
<td>.47</td>
</tr>
<tr>
<td>Missing device (% of usage)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taking images</td>
<td>17.5 (16.1)</td>
<td>0 (0)</td>
<td>N/A$^c$</td>
</tr>
<tr>
<td>Displaying images</td>
<td>18.6 (22.6)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Technical failure (% of usage)</td>
<td>9.7 (18.2)</td>
<td>4.4 (9.1)</td>
<td>.04</td>
</tr>
</tbody>
</table>

$^a$Hours.
$^b$Minutes.
$^c$N/A: not applicable.

Statistical differences were also found for the time taken from taking pictures until completion of the upload, the time the technical upload took, and the amount of time needed to view pictures after request (all $P<.001$; Table 1). The time until upload presented the biggest difference, with a mean time of 14.9 (SD 24.0) hours with the digital camera compared to 1.8 (SD 1.2) minutes with the smartphone (Table 1).

A comparison of the time the technical archiving and display of pictures took revealed a significant difference in favor of the smartphone (both $P<.001$, Tables 2 and 3).
Table 2. Secondary end points.

<table>
<thead>
<tr>
<th>Secondary end points</th>
<th>Camera (n=65), n (%)</th>
<th>Smartphone (n=68), n (%)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention(^a)</td>
<td></td>
<td></td>
<td>.31</td>
</tr>
<tr>
<td>Soft tissue conditions</td>
<td>48 (74)</td>
<td>53 (78)</td>
<td></td>
</tr>
<tr>
<td>Wounds</td>
<td>53 (82)</td>
<td>56 (82)</td>
<td></td>
</tr>
<tr>
<td>Deformities</td>
<td>19 (29)</td>
<td>22 (32)</td>
<td></td>
</tr>
<tr>
<td>Range of motion</td>
<td>2 (3)</td>
<td>9 (13)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>19 (29)</td>
<td>15 (22)</td>
<td></td>
</tr>
<tr>
<td>Reasoning(^a)</td>
<td></td>
<td></td>
<td>.94</td>
</tr>
<tr>
<td>Legal requirement</td>
<td>34 (52)</td>
<td>32 (47)</td>
<td></td>
</tr>
<tr>
<td>Improving therapy</td>
<td>17 (26)</td>
<td>21 (31)</td>
<td></td>
</tr>
<tr>
<td>Preoperative planning</td>
<td>33 (51)</td>
<td>42 (62)</td>
<td></td>
</tr>
<tr>
<td>Postoperative control</td>
<td>6 (9)</td>
<td>9 (13)</td>
<td></td>
</tr>
<tr>
<td>Consultation</td>
<td>29 (45)</td>
<td>31 (46)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>14 (22)</td>
<td>15 (22)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Percentages exceed 100% because multiple selections were allowed, and the \( P \) value was calculated using the chi-square test.

Table 3. Comparison of time to upload, time to view, and reasons for delay.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Camera (n=65), n (%)</th>
<th>Smartphone (n=68), n (%)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time taken</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;10 s</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>10-30 s</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>8 (12)</td>
</tr>
<tr>
<td>30-60 s</td>
<td>4 (5)</td>
<td>11 (17)</td>
<td>27 (40)</td>
</tr>
<tr>
<td>1-5 min</td>
<td>35 (53)</td>
<td>24 (37)</td>
<td>32 (47)</td>
</tr>
<tr>
<td>&gt;5 min</td>
<td>26 (40)</td>
<td>30 (46)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Reasons for delay(^a)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Technical issues</td>
<td>4 (6)</td>
<td>8 (12)</td>
<td>11 (16)</td>
</tr>
<tr>
<td>Distance to workstation</td>
<td>49 (75)</td>
<td>37 (57)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Organizational reasons</td>
<td>42 (65)</td>
<td>29 (45)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>None</td>
<td>7 (11)</td>
<td>20 (31)</td>
<td>44 (65)</td>
</tr>
<tr>
<td>Others</td>
<td>0 (0)</td>
<td>15 (23)</td>
<td>16 (24)</td>
</tr>
</tbody>
</table>

\(^a\)Percentages exceed 100% because multiple selections were allowed, and the \( P \) value was calculated using the chi-square test.

However, both groups did differ in the reasons as to why there had been a delay (\( P<.001 \)). The main reasons mentioned with the digital camera were the distance to one of the workstations and organizational reasons, that is, being preoccupied in the operating theater. With the smartphone, there were mostly no reasons for a delay, yet in a quarter (16/68, 24% and 17/68, 25%) of cases, a time lag or app crashes were mentioned (Table 3).

No difference was found in the intention behind the photo, which was mostly documentation of soft tissue conditions (74% and 78%, respectively) and wounds (both 82%; \( P=.31 \)), nor in the reasoning why the photo had been taken (legal requirements, improving therapy, and consultation; \( P=.94 \)).

The smartphone app’s high acceptance and approval could be deduced from the web-based questionnaire, especially in terms of time savings and an easier workflow (Table 4), with a mostly positive rating on the applied Likert scale (ranging from –2 to 2; Figure 2). There were 2 indifferent evaluations regarding higher usage and improved communication. Only the responsiveness of the app was evaluated negatively with a median of –1, which concurs with the written answers about the occurring time lag and crashes of the app (Table 3 and Figure 2).
Table 4. Web-based questionnaire.

<table>
<thead>
<tr>
<th>Question</th>
<th>Value (n=19), n (%)^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where do you see the main benefit of mRay in an inpatient trauma setting?</td>
<td></td>
</tr>
<tr>
<td>Time savings</td>
<td>17 (90)</td>
</tr>
<tr>
<td>Easier communication with increased quality of treatment</td>
<td>8 (42)</td>
</tr>
<tr>
<td>Easier and more comfortable workflow</td>
<td>13 (68)</td>
</tr>
<tr>
<td>No benefit</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Where the functions of mRay sufficient for carrying out image evaluation?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18 (95)</td>
</tr>
<tr>
<td>No</td>
<td>1 (5)</td>
</tr>
<tr>
<td>What other functions would you like to see in mRay?</td>
<td></td>
</tr>
<tr>
<td>Automated wound measurement</td>
<td>9 (47)</td>
</tr>
<tr>
<td>Automated assessment of wound conditions</td>
<td>5 (26)</td>
</tr>
<tr>
<td>Entering comments on photo findings</td>
<td>11 (58)</td>
</tr>
<tr>
<td>Others</td>
<td>3 (16)</td>
</tr>
</tbody>
</table>

^aPercentages may exceed 100% because multiple selections were allowed.

Figure 2. Results of the web-based questionnaire.

Usage and usability

Discussion

Overview

This prospective study highlighted the advantages of workflow and data security for medical photography by integrating a smartphone app. One key indication is the reduced time taken from capturing the photo to its storage in the PACS: almost instantaneous with smartphones, in contrast to an average 12-hour duration using digital cameras. While viewing photos is feasible at all workstations, uploading is confined to specific stations due to network security concerns. Especially in a time-critical specialty such as traumatology and emergency medicine, such a tool could be particularly beneficial. By lowering the threshold and simplifying the cumbersome workflows of medical photography, the photography process and the number of photos taken could be increased to the benefit of the patient. This would take away the argument that there is no structured assessment or procedure for documentation of acute wounds because the required effort is considered too high and time-consuming [13,23,24].

Such reasoning might originate from studies such as Bronsard et al [23], which tried to establish a pathway for the secure handling of patient data. In their workflow, the photo was sent to a coordinator after it had been taken. This person assessed it and converted it to a DICOM file, which was then cropped by a secretary and finally uploaded. Each week, it took 3 individuals 1-2 hours to generate 1 image from 3-5 photos. It is therefore hardly surprising that they only managed to produce 300 images in 2 years.

Furthermore, previous studies showed that adequate medical photography can improve the care for and decision-making about complex injuries, especially when soft tissues are at risk [7-9,25]. In the case of traumatology, this would mainly be open
fractures. Inaccuracies in the description and extent of the concomitant soft tissue injury could affect the planning of the surgical approach and, in the worst case, necessitate the recurrent removal of the dressings to reassess the wound, facilitating a rise in infection risk [26].

The need for an easier and safer way to perform medical photography in traumatology is also enhanced by the argument of Friesen et al [27]. With the rising incidence of older patients in trauma wards, they estimated that 25% of in-house patients could exhibit chronic wounds requiring structured care and documentation [27].

The key aspect that needs to be addressed is the medicolegal aspect and risk of data infringement, which have been shown to be insufficiently addressed in most studies focusing on the adequacy of protocols for patient consent and publication in current practice [15]. First, because the workflow must uphold data security and patient confidentiality, and second, because the acceptance of photography performed with smartphones still needs to be increased, particularly among older patients [18,20].

Anonymous interviews from O’Farrell and Ferreira [1] showed that in 74% of cases, photos taken using a smartphone were not deleted. Furthermore, 58% were stored on a laptop and 26% on a flash drive, while 16% admitted that the device in question was not password protected, and in 21% of cases, third parties could have accessed the pictures. The distribution methods further raise concerns: 58% of the photos were sent for consultation through WhatsApp and 80% through email. Given European Union regulations, these findings underscore a pressing need to address security and privacy challenges in medical photography [1,14].

On the contrary, although the acceptance of clinically owned cameras is fairly high, ranging from 75% to 95%, they also pose a considerable data breach risk. They cannot be password protected, the data are not encoded, and they are mostly not stored safely when not in use [1,17,18,20]. Using an app akin to the one examined in this study, these concerns can be dismissed. Furthermore, the study of Accetta et al [20] showed that in such cases, smartphone photography, under the premise of special information, could reach a comparable acceptance rate of 88%.

Requirements for performing medical photography are an easy and fast appliance, secure storage of data, and prevention of data infringement. This can be achieved using specialized smartphone apps [28].

In addition to the medical value, efficient and extensive medical imaging can also provide economic benefits. If the pictures taken lead to a new diagnosis or therapeutic purpose, then this could be a billable service [2,21,29]. In a study examining smartphone-based medical photography, Jordan et al [21] demonstrated that in 20% of medical audit cases where photos were used, they helped confirm a diagnosis or procedure. This resulted in additional revenue of US $330 per case, amounting to a total of approximately US $70,000 annually.

Besides the possible benefits for acute fracture care and inpatient management, the third benefit could lie in the effect on medical certificates [2,30-33]. These aim to offer an objective assessment of medical outcomes after injuries and rely, therefore, on measurable findings and reliable tools to avoid bias and achieve interrater agreement. Using goniometry to clinically examine the range of motion is important in this regard, but the interrater reliability and agreement for this are not remarkably high. However, Naylor et al [32] could prove that measuring the range of motion from photos taken could achieve an agreement rate of >0.983.

Finally, a key aspect of modern medicine is the informed consent of patients and patient education, as patient compliance and outcomes could be beneficially affected by this [1,34]. Nair et al [22] showed that over two-thirds of patients stated that after being shown images of their condition, their understanding of their condition increased, they believed that this had improved their therapy, and they would therefore recommend this approach to other patients.

In the future, additional applications, such as automatic measurement and categorization of wounds, could be possible if standardized acquisition of these photos can be achieved [35-38]. This could be further simplified if technology such as light detection and ranging scanners becomes widely available on smartphones. Then maintaining specific distances or including measurement references would no longer be necessary for accurate measurements, especially in depth [38-40].

Limitations

This study has some limitations. Despite its prospective design, the sample size is quite small, and so the evidence base is limited. Additionally, the study was restricted to a single study site. As the data were acquired using questionnaires, a certain amount of bias cannot be excluded. This is especially true for the outcome parameter “time to upload or view,” which is only a subjective estimation but has been treated as a categorical variable. That is the main limitation of this study. For digital camera usage, in particular, an electronic measurement of these parameters was not feasible, and neither were such analyses incorporated in the app. However, any bias should influence both the data acquired from the digital camera and the smartphone app similarly, and we only aimed to analyze any differences found between them. Therefore, the evidence should not be relevantly impaired by these limitations. Another limitation is that smartphone photography can compete with digital cameras in regard to the standards and quality of small versions meant for small everyday tasks but not for scientific, educational, or other more challenging purposes requiring higher resolution and quality. Especially in light of this, another limitation is that usage in general was assessed, not the situations, the content, or the quality of the photographs. In this study, however, the sensor and resolution of the cameras were comparable on both devices.

Finally, no questioning or evaluation of the patients’ comfort and preference with both devices has been conducted.

Conclusions

Specialized smartphone apps offer a secure, fast, and easy way to acquire medical photos and could possibly improve patient education and care in terms of wound management, in particular,
while also offering medicolegal and economic benefits. Future studies should focus on a more objective assessment of differences and take factors such as patient comfort and preference, image resolution, and picture quality into consideration, as well as a larger sample size.

Acknowledgments
The authors would like to thank the IT department of the trauma center for supporting this study by implementing the infrastructural environment.

Data Availability
The data sets used and analyzed during this study are available from the corresponding author on request.

Authors' Contributions
SYV designed and supervised the study. The study was conducted and analyzed by JSEB, who wrote the first draft of the manuscript. MM and HS provided the software used, and HS wrote the technical part of the methods. All authors commented on previous versions of the manuscript and approved the final version.

Conflicts of Interest
The authors declare the following potential conflicts of interest concerning the research, authorship, and publication of this article: the BG Trauma Center Ludwigshafen and mbits imaging GmbH (Heidelberg, Germany) cooperate in the research field of medical digitalization without economic ties. This cooperation influenced neither the outcome of the study nor the manuscript. MM and HS are employees of mbits imaging GmbH. They provided the software "mRay".

Multimedia Appendix 1
Questionnaire on the daily usage of a digital camera for photo documentation in everyday clinical practice. [PDF File (Adobe PDF File), 100 KB - formative_v8i1e47572_app1.pdf ]

Multimedia Appendix 2
Questionnaire on the daily usage of the mobile application “mRay Foto” for photo documentation in everyday clinical practice. [PDF File (Adobe PDF File), 101 KB - formative_v8i1e47572_app2.pdf ]

Multimedia Appendix 3
Questionnaire for the qualitative survey at study completion. [PDF File (Adobe PDF File), 98 KB - formative_v8i1e47572_app3.pdf ]

Multimedia Appendix 4
Standard Operating Procedure Medical Photography. [DOCX File, 17 KB - formative_v8i1e47572_app4.docx ]

References


Abbreviations

PACS: picture archiving and communication system
Abstract

Background: Degenerative cervical myelopathy (DCM), a progressive spinal cord injury caused by spinal cord compression from degenerative pathology, often presents with neck pain, sensorimotor dysfunction in the upper or lower limbs, gait disturbance, and bladder or bowel dysfunction. Its symptomatology is very heterogeneous, making early detection as well as the measurement or understanding of the underlying factors and their consequences challenging. Increasingly, evidence suggests that DCM may consist of subgroups of the disease, which are yet to be defined.

Objective: This study aimed to explore whether machine learning can identify clinically meaningful groups of patients based solely on clinical features.

Methods: A survey was conducted wherein participants were asked to specify the clinical features they had experienced, their principal presenting complaint, and time to diagnosis as well as demographic information, including disease severity, age, and sex. K-means clustering was used to divide respondents into clusters according to their clinical features using the Euclidean distance measure and the Hartigan-Wong algorithm. The clinical significance of groups was subsequently explored by comparing their time to presentation, time with disease severity, and other demographics.

Results: After a review of both ancillary and cluster data, it was determined by consensus that the optimal number of DCM response groups was 3. In Cluster 1, there were 40 respondents, and the ratio of male to female participants was 13:21. In Cluster 2, there were 92 respondents, with a male to female participant ratio of 27:65. Cluster 3 had 57 respondents, with a male to female participant ratio of 9:48. A total of 6 people did not report biological sex in Cluster 1. The mean age in this Cluster was 56.2 (SD 10.5) years; in Cluster 2, it was 54.7 (SD 9.63) years; and in Cluster 3, it was 51.8 (SD 8.4) years. Patients across clusters significantly differed in the total number of clinical features reported, with more clinical features in Cluster 3 and the least clinical features in Cluster 1 (Kruskal-Wallis rank sum test: $\chi^2_{2}=159.46; P<.001$). There was no relationship between the pattern of clinical features and severity. There were also no differences between clusters regarding time since diagnosis and time with DCM.

Conclusions: Using machine learning and patient-reported experience, 3 groups of patients with DCM were defined, which were different in the number of clinical features but not in the severity of DCM or time with DCM. Although a clearer biological basis for the clusters may have been missed, the findings are consistent with the emerging observation that DCM is a heterogeneous disease, difficult to diagnose or stratify. There is a place for machine learning methods to efficiently assist with pattern recognition. However, the challenge lies in creating quality data sets necessary to derive benefit from such approaches.

(JMIR Form Res 2024;8:e54747) doi:10.2196/54747
cervical; myelopathy; machine learning; cluster; clusters; clustering; spine; spinal; compression; neck; degenerative; k-means; patient reported; degenerative cervical myelopathy

Introduction

Degenerative cervical myelopathy (DCM) is a progressive spinal cord injury caused by spinal cord compression from degenerative pathology and consists of various subcategories of pathology, including cervical spondylotic myelopathy, ossification of the posterior longitudinal ligament, ossification of the ligamentum flavum, and degenerative disc disease [1-4]. It is estimated to affect 2% of adults, although fewer than 10% are currently diagnosed [5,6]. Surgery is the mainstay of treatment for DCM, aiming to decompress the spinal cord [2,7-9].

DCM often presents with neck pain, sensorimotor dysfunction in the upper or lower limbs, gait disturbance, and bladder or bowel dysfunction [2,10-14]. Examination findings include upper motor neuron signs in the limbs, such as positive Babinski sign, positive Hoffman sign, hyperreflexia, and increased tone [2,10-13]. Its symptomatology is very heterogeneous, making early detection difficult. This heterogeneity makes it difficult to measure or understand what drives consequences. For instance, the heterogeneity has made it harder to understand health-related quality of life [15]. This has also hindered comparisons between studies and the development of clinical practice guidelines and recommendations for DCM [16,17]. Additionally, once detected, DCM is unpredictable due to a lack of reliable methods to determine prognosis.

Increasingly, evidence suggests that DCM may consist of subgroups of the disease, which still need to be defined [18-20]. Machine learning can help in finding them. In fact, machine learning has shown potential in predicting health-related quality of life after surgery for mild DCM and outcome after surgery, although external validation and prospective analysis are still needed [21,22]. The use of machine learning in identifying these subgroups is dependent on the data set. Munro et al [23] (2023) provided a unique and comprehensive description of the effects of DCM from the perspective of people living with DCM [24,25]. This is a data set that could lend itself to machine learning analysis due to its comprehensiveness.

The objective of this study was to explore whether machine learning can identify clinically meaningful groups of patients based on solely clinical features.

Methods

Data Set

A mixed methods cross-sectional study was conducted by a team from the University of Cambridge through Myelopathy.org [26], a global charity dedicated to DCM. A focus group session of people with DCM and their supporters was used to inform the development of a web-based survey to explore the consequences of living with DCM. The survey was advertised using the Myelopathy.org website, an international nonprofit organization dedicated to promoting understanding and awareness of DCM. Survey participants were asked to specify the clinical features they had experienced, their principal presenting complaint, and time to diagnosis as well as demographic information, including disease severity, age, and sex. The data consist of 189 yes or no responses to a list of 76 clinical features. This was published in a paper, titled “Targeting earlier diagnosis: what symptoms come first in degenerative cervical myelopathy?” [23], wherein the full methodology is detailed.

Analysis

Patients were grouped into subsets with similar characteristics using k-means clustering. K-means clustering is a method that groups data into “k” nonoverlapping, distinct subsets by finding centroids in the data representing each cluster’s center and allocating data points to each cluster by minimizing within-cluster variance around centroids. K-means clustering was used due to its efficiency for small data sets and explainability, aiming to group respondents into clusters based on their clinical features, using the Euclidean distance measure and the Hartigan-Wong algorithm [27]. The optimal number of clusters (k) was determined through the inspection of 3 ancillary methods, namely, the elbow, silhouette, and gap statistic methods [28]. The clinical significance of groups was subsequently explored by comparing their time to presentation, time with disease severity, and other demographics. DCM severity was assessed using total Modified Japanese Orthopaedic Association scores [29]. Noncomplete records were not excluded, and missing data were not imputed. All analyses were conducted in R (version 4.1.0; R Foundation for Statistical Computing) [30].

Ethical Considerations

This study was conducted with ethical approval from the University of Cambridge (HBREC.2019.14). At the start of the survey, participants were provided with an overview of the study and definition of DCM, and by continuing into the survey, participants were confirming their diagnosis of DCM and providing informed consent to participate. All data collected were anonymous. No incentives were offered for the completion of the surveys.

Results

Cohort Demographics

Of the 189 participants, 134 were female and 49 were male (6 did not report biological sex). Respondents were on average 54.1 years of age. A total of 29 of them had mild DCM, 68 had moderate DCM, and 92 had severe DCM. The majority (131/189, 69%) reported having had surgery for DCM.

Cluster Analysis

Ancillary methods suggested different optimal numbers of clusters (k). Elbow, silhouette, and gap statistic methods identified k=3, k=2, and k=5, respectively (Figure 1A). The

https://formative.jmir.org/2024/1/e54747 JMIR Form Res 2024 | vol. 8 | e54747 | p.442 https://formative.jmir.org/2024/1/e54747 (page number not for citation purposes)
data were hence clustered into multiple values of k and inspected (Figure 1B). After a review of both ancillary and cluster data, it was determined by consensus between AYT and BD that the optimal number of DCM response groups was 3. The reasoning behind this was that the ancillary curves in 2 out of 3 ancillary methods plateaued from k≥3 (Figure 1A), but clusters above k≥4 overlapped (Figure 1B).

Figure 1. (A) Determining the optimal number of clusters; (B) k-means clustering (euclidean).

Characterization Analysis

In Cluster 1, there were 40 respondents, and the ratio of male to female participants was 13:21. In Cluster 2, there were 92 respondents, with a male to female participant ratio of 27:65. Cluster 3 had 57 respondents, with a male to female participant ratio of 9:48. A total of 6 people did not report biological sex in Cluster 1. The mean age was 56.2 (SD 10.5) years in this cluster; in Cluster 2, it was 54.7 (SD 9.63) years; and in Cluster 3, it was 51.8 (SD 8.4) years.

A spider chart was subsequently generated to explore the clinical significance of the clusters, wherein the curves did not cross (Figure 2A; Multimedia Appendix 1).

It was also checked if patients in the different groups experienced differing numbers of features (Figure 2B). Patients across clusters significantly differed in the total number of clinical features reported, with more clinical features in Cluster 3 and the least clinical features in Cluster 1 (Kruskal-Wallis rank sum test: \( \chi^2 = 159.46; P < .001 \)).
To check whether patients with more clinical features had a more severe form of DCM, patterns of clinical features against severity were compared. The results showed no relationship between the pattern of clinical features and severity (Figure 2C).

Patterns of clinical features against both time since diagnosis and time with DCM were also analyzed. As shown in Figures 2D and 2E, there did not seem to be any differences between clusters in these distributions.

**Figure 2.** (A) Spider charts showing survey responses across clusters (the abbreviations along the circumference are detailed in the table in the Multimedia Appendix 1); the radius represents relative frequency, normalized to 1; (B) total number of clinical features reported across clusters; (C) proportions of degenerative cervical myelopathy (DCM) severity across clusters (based on the Modified Japanese Orthopaedic Association scores); (D) distribution of time taken to be diagnosed with DCM in each cluster; (E) distribution of time with DCM in each cluster.
Discussion

Principal Findings
Cluster analysis suggested 3 optimal subgroups based on clinical features. When exploring why these groups differed in terms of cohort demographics, only the number of reported symptoms differed significantly. The pattern of clinical features within each of the 3 groups was similar. Notably, the 3 curves in the spider chart appear to peak and trough in a similar pattern, suggesting that there was no difference in the pattern of clinical features. The concentricity of curves, however, suggested that clustering may be due to the total number of features experienced. This possibility was statistically significant (Kruskal-Wallis rank sum test). Finally, there was no link between the groups and disease severity, time with DCM, and time since diagnosis.

Limitations
This study has several limitations. The data represent a single time point cross-sectional survey of an internet-recruited cohort of patients, which could limit the generalizability of the findings. Additionally, information on disease characteristics, used for exploring the clinical significance of clusters, was limited to time with symptoms and a self-reported modified Japanese Orthopaedic Association score [31]. A more diverse data set would be more insightful, especially in DCM, wherein the nuances of symptom presentation and progression are critical. The sample size is also relatively small by machine learning standards. Finally, only 1 analysis method (k-means clustering) was performed, which may prevent us from capturing the full complexity of DCM symptomatology, especially with the increasing prominence of personalized approaches [32].

That being said, this is a unique data set, formed from the unrestricted perspectives of almost 200 patients; it was formed without any preconceptions regarding what symptoms were considered related to DCM. The result is also not unexpected. Standard analytical approaches, using more traditional data sets, have failed to stratify patients by symptoms [33]. Consequently, although a clearer biological basis for the clusters may have been missed, the findings are consistent with the emerging observation that DCM is a heterogeneous disease, difficult to diagnose or stratify [15-17]. This has been highlighted by the work of Cook et al [34] (2022) and is perhaps reflected in our inability to explain the variability in the quality of life in DCM [15].

This study shows that there is certainly a role for machine learning methods to efficiently assist with pattern recognition, but data sets must be large, valid, and comprehensive. In DCM, the challenge and priority appear to be less focused on data set size and more focused on the type of data [35]. For example, our redefinition of DCM in terms of time, mechanical stress, and vulnerability to sustain a spinal cord injury has highlighted the potential significance of various disease factors; these factors range from frailty and genetics to the type of pathology causing compression, encompassing the likely heterogeneous mechanical loading they induce [20]. Further, there are few valid and reliable outcome measures available, with most relying on face-to-face presentations to measure changes over the course of months, exhibiting low statistical power. The work of Cook et al [34] (2022) has highlighted that the experience of DCM is driven by social determinants—features such as ethnicity as well as educational, and economic status [34]. This means subjectivity in outcomes will drive current variability. Novel biomarkers, including imaging, blood, and digital biomarkers, are likely to hold value in this context, offering more disease-specific and sensitive disease indicators [36]. The need for more comprehensive and improved measurement is a firm priority in DCM [16]. Therefore, artificial intelligence undoubtedly has an important role in the future of DCM research and care. To our knowledge, such measures do not currently exist. Analysis of one of the most detailed cohorts also failed to identify biologically significant strata [22,37]. Therefore, the short-term challenge for our community lies in creating quality data sets necessary to derive benefit from these emerging analytical approaches.

Conclusions
Using machine learning and patient-reported experience, 3 groups of patients with DCM were defined. These groups differed in the number of clinical features reported but not in the severity of DCM, time since diagnosis, or time with DCM. The significance and generalization of this study remain uncertain. Overall, this study confirms the role of machine learning in DCM research, but more pressingly, it confirms the need to curate the right data sets.

Acknowledgments
BMD is supported by a National Institute for Health Research (NIHR) Clinical Doctoral Fellowship.

The views expressed in this publication are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
BMD is the director of MoveMed Ltd.
References


Abbreviations

DCM: degenerative cervical myelopathy
Testing a Behavioral Activation Gaming App for Depression During Pregnancy: Multimethod Pilot Study

Rachel C Vanderkruik, MSc, PhD; Craig Ferguson, MS; Lauren A Kobylski, MPH; Joseph J Locascio, PhD; Gabriella E Hamlett, MA; Parker C Killenberg, BS; Robert Lewis, MS; Noah Jones, MS; Ella T Rossa, BA; Hannah Dineen; Rosalind Picard, ScD; Lee S Cohen, MD

1Center for Women’s Mental Health, Massachusetts General Hospital, Boston, MA, United States
2Department of Psychiatry, Harvard Medical School, Boston, MA, United States
3MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States
4Department of Psychological & Brain Sciences, George Washington University, Washington, DC, United States
5Department of Neurology, Harvard Medical School, Boston, MA, United States
6Department of Psychology, Harvard University, Cambridge, MA, United States

Corresponding Author:
Rachel C Vanderkruik, MSc, PhD
Center for Women’s Mental Health
Massachusetts General Hospital
185 Cambridge St
Ste 2200
Boston, MA, 02114
United States
Phone: 1 781 691 9071
Email: rvanderkruik@mgh.harvard.edu

Abstract

Background: Depression during pregnancy is increasingly recognized as a worldwide public health problem. If untreated, there can be detrimental outcomes for the mother and child. Anxiety is also often comorbid with depression. Although effective treatments exist, most women do not receive treatment. Technology is a mechanism to increase access to and engagement in mental health services.

Objective: The Guardians is a mobile app, grounded in behavioral activation principles, which seeks to leverage mobile game mechanics and in-game rewards to encourage user engagement. This study seeks to assess app satisfaction and engagement and to explore changes in clinical symptoms of depression and anxiety among a sample of pregnant women with elevated depressive symptoms.

Methods: This multimethod pilot test consisted of a single-arm, proof-of-concept trial to examine the feasibility and acceptability of The Guardians among a pregnant sample with depression (N=18). Participation included two web-based study visits: (1) a baseline assessment to collect demographic and obstetric information and to assess clinical symptoms and (2) an exit interview to administer follow-up measures and explore user experience. Participants completed biweekly questionnaires (ie, Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7) during the trial to assess depression and anxiety symptom severity. App satisfaction was measured using 2 self-report scales (ie, Mobile Application Rating Scale and Player Experience of Needs Satisfaction scale). Engagement with The Guardians was captured using game interaction metric data. We used backward-eliminated mixed effects longitudinal models to examine the effects of app engagement and satisfaction and length of time in the study on symptoms of depression and anxiety. Content analysis was conducted on qualitative data from exit interviews.

Results: The 15-day and 30-day overall app retention rates were 26.6% and 15.1%, respectively. Mixed effects models found significant negative main effects of week in study ($\beta=-.35; t_{61}=-3.05; P=.003$), number of activities completed ($\beta=-.12; t_{61}=-2.05; P=.04$), days played ($\beta=-.12; t_{58}=-2.9; P=.005$), and satisfaction, according to the Mobile Application Rating Scale ($\beta=-3.05; t_{58}=-2.19; P=.03$) on depressive symptoms. We have reported about similar analyses for anxiety. There is preliminary evidence suggesting harder activities are associated with greater mood improvement than easier activities. Qualitative content analysis resulted in feedback falling under the following themes: activities, app design, engagement, fit of the app with lifestyle, perceived impact of the app on mood, and suggestions for app modifications.
Conclusions: Preliminary results from this multimethod study of The Guardians indicate feasibility and acceptability among pregnant women with depression. Retention and engagement levels were more than double those of previous public mental health apps, and use of the app was associated with significant decrease in depressive symptom scores over the 10-week trial. The Guardians shows promise as an effective and scalable digital intervention to support women experiencing depression.

KEYWORDS
perinatal depression; pregnancy; behavioral activation; mobile app; digital intervention; mobile phone

Introduction

Background

There is growing recognition of the burden of depression during the perinatal period (ie, pregnancy and postpartum periods). Prevalence estimates of perinatal depression (PD) range between 10% and 15% in high-income countries and are even higher in low- and middle-income countries [1,2]. If left untreated, there can be significant and detrimental consequences for both the mother and infant [3,4]. In response to this public health problem, the US Preventive Services Task Force recently recommended that clinicians provide or refer pregnant and postpartum individuals who are at increased risk of PD to counseling interventions [5]. Pharmacologic and nonpharmacologic treatments to manage symptoms of depression during pregnancy are available, yet most women do not receive the treatment they need [6]. Furthermore, targeting depression specifically during pregnancy is critical in addressing PD, as depression during pregnancy is a key risk factor for postpartum depression [7], and thus, reducing depression during pregnancy can reduce the risk for postpartum depression. Given the number of individuals affected by PD and significant barriers to care [8,9], strategies that provide increased access to evidence-based treatment approaches are needed.

There is evidence suggesting that women may prefer psychotherapeutic approaches over antidepressant medications during pregnancy [10]. Behavioral activation (BA) is one such psychotherapeutic approach based on the theory that, as individuals become depressed, they tend to engage in avoidance and isolation, which then maintains or worsens their depressive symptoms. BA encourages the individual to gradually increase engagement in activities that serve as “behavioral antidepressants” and decrease their avoidance and isolation [11]. Several meta-analyses support the effectiveness of BA and antidepressant medications among adults with major depression [12,13], and it has been found to be comparable in efficacy with antidepressant medications during pregnancy [14]. BA has been considered to be advantageous compared with other treatments given the simplicity of the intervention, strong retention rates, and enduring effects over a 2-year follow-up, without the concern of side effects associated with some medications [15,16]. There is emerging evidence supporting BA, specifically among pregnant individuals, with findings that BA can offer significant depression-related, anxiety-related, and stress-related benefits for this population [17].

There is support for scalability and even global dissemination of BA given how it can be delivered by nonspecialists, is cost-effective, and has demonstrated cross-cultural fit and adaptability [18]. Furthermore, BA has been identified as being suitable for “computer-based interventions that would involve no therapist input beyond an initial assessment,” which could dramatically improve the accessibility of effective treatments for depression [15]. Digital technology, including mobile health apps, has been cited as a promising strategy for increasing access to evidence-based interventions for mental health for several reasons, including constant availability, equity of availability, immediate support, low costs, lack of geographic barriers, and reduced need for direct mental health service provision (particularly given the shortage of clinicians in certain locations) [19]. A systematic review of the effectiveness of mobile apps for monitoring and managing mental health symptoms or disorders concluded that there is support for the potential of mobile apps to effectively reduce the burden of mental health symptoms; yet, further robust studies are needed to develop and test evidence-based apps [20]. Therefore, a digital BA app could be a promising way to increase access to support for depression among vulnerable or underrepresented populations.

It is, thus, not surprising that there have been recent efforts to develop BA apps, including a self-help brief BA app focused on activity scheduling for the general population called Moodivate. A preliminary randomized controlled trial (RCT) recruited adults with elevated depression from primary care practices (N=52) and randomized them to receive (1) Moodivate, (2) an active control cognitive behavioral therapy–based mobile app (MoodKit), or (3) treatment as usual (no app) [21]. Study retention was relatively high, with approximately 70% of Moodivate participants continuing to use the app 1 month after trial enrollment and 50% at the end of the 8-week study period. Compared with treatment as usual, participants in both app conditions experienced significant reductions in depressive symptoms over time, and these treatment gains were sustained throughout the trial period. These preliminary results support the potential feasibility of a BA app, such as Moodivate, as a treatment for depression. Other preliminary studies have exhibited promising findings regarding the BA apps’ abilities to provide motivation for adults with depression to plan enjoyable activities [22] and acquire insight into their own behavior and impact on mood [23]. However, these BA apps are not widely and freely available to the public, and more rigorous studies are needed regarding their impact on depressive symptoms and among certain populations.

A recent meta-analysis provides support for the efficacy and acceptability of internet-delivered interventions for pregnant women and highlights the opportunity to leverage technology for interventions targeting this population, including for mental health symptoms during pregnancy [24]. Given the challenges...
of engaging pregnant people in treatment, preferences for depression treatment in nonspecialty settings, and limited availability of services targeting depression during pregnancy [25], a widely available and engaging BA app could be a highly impactful service for individuals with PD. Although a variety of apps designed to improve health and well-being have been developed, most are not grounded in evidence-based principles, and most struggle to keep users engaged [26-28]. Many such apps used typical “gamification” techniques (ie, awarding users with badges or points for using the app), which are designed to increase adherence but often fail to give the user reasons to care about those rewards. Thus, long-term retention and engagement in the app may be affected [29]. A different approach used by a mobile game called The Guardians: Unite the Realms provides immediate rewards for using BA therapeutic techniques as part of the app and gives those rewards inherent value and meaning through their use within the game’s mechanics [30,31].

The Guardians: Unite the Realms is a novel gaming BA app that is free to be downloaded by the general public with iPhone Operating System and Android. The app was developed by an author of this paper (CF) with input from experts in digitizing therapies to ensure that it maintains the core qualities and effectiveness of BA. The Guardians was designed to increase adherence to app-delivered BA by embedding BA techniques into the unique context of a mobile game, giving intrinsic incentives for users to continue using the app [30]. Since its launch in April 2020, The Guardians has collected anonymized gameplay data from about 12,500 users and saved them to its secure backend database for research and gameplay recovery purposes, as specified in the app’s privacy policy [29]. The 15-day and 30-day overall retention rates of 10% and 6.6%, respectively, are more than double the average retention rates for mental health apps of 3.9% (IQR 10.3%) and 3.3% (IQR 6.2%), as reported by Baumel et al [27] and Ferguson et al [30]. Furthermore, the 1-day and 28-day observed retention of 37.9% and 7.3%, respectively, suggest that The Guardians has retention rates that are comparable with those of the top 15% of mobile games. Although there has not yet been a formal study of how the use of The Guardians app may influence depressive symptoms, >80% of the BA activities completed as part of the app resulted in the user feeling at least “a little bit better” [30]. The Guardians is a widely available BA gaming app with a novel approach to engage users, and preliminary data show a positive impact on mood among its users.

This Study
To the best of our knowledge, there are no gaming BA apps that have been developed specifically for a pregnant population with elevated depressive symptoms. Given the promising preliminary data about The Guardians in terms of engagement metrics and user-reported improvement in mood from app-related activities, along with the need for more novel approaches to address depressive symptoms during pregnancy, this pilot study sought to assess app engagement and to explore the potential resulting changes in mental health symptoms among a pregnant sample with elevated depressive symptoms. As anxiety is often comorbid with depression during pregnancy [32] and there is some preliminary evidence suggesting that BA may also be beneficial for anxiety symptoms, we also sought to explore the changes in anxiety symptoms among users of The Guardians app [33,34].

Furthermore, we sought to capture feedback from study participants about how a future version of The Guardians could be specifically tailored for a pregnant population, as we used the publicly available app geared for a general population in this pilot study. Incorporating a user-centered design into app development is linked to high usability, low risk of failure, reduced costs, and high overall quality [35] and can help inform app design and implementation according to feedback about the target users’ needs [36,37]. Thus, the aims of this pilot study were to explore engagement with and impact of The Guardians among pregnant individuals with elevated depressive symptoms, while also qualitatively exploring user experience and gathering suggestions for a future iteration of the app that could be tailored specifically for the context of pregnancy.

Methods
Participants and Recruitment
Consistent with a pragmatic trial, the inclusion criteria for this single-arm pilot study were minimal. Women were eligible to participate if they were pregnant, aged >18 years, and English speaking; had access to a smartphone; and had a Patient Health Questionnaire-9 (PHQ-9) score of at least 10 (indicative of a possible depressive episode). Exclusion criteria included a diagnosis of bipolar or psychotic disorder, active mania, psychosis, substance abuse, or immediate risk of self-harm based on PHQ-9 responses and clinician judgment. For this initial pilot study, we focused on pregnancy and excluded women in the postpartum period as there are other complicating factors in the postpartum period (eg, sleep disruption and demands of caring for a newborn) that can affect the ability to engage in the app and completion of app activities. As this was a completely web-based study, participants could be located anywhere in the United States. Participants were recruited through social media advertising, clinician referrals, and the Massachusetts General Hospital Center for Women’s Mental Health website [38]. Individuals interested in participating were instructed to reach out via mobile phone or email to the study’s research assistant using the contact information provided in the study advertisements.

Procedures
Following an eligibility screening via mobile phone, participants provided verbal informed consent and completed a baseline assessment with a trained research assistant. During the baseline interview, demographic variables, pregnancy characteristics, and psychiatric history (ie, diagnosis and treatment) were collected. At the conclusion of the baseline assessment, the research assistant instructed the eligible participants about how to download The Guardians app onto their personal smartphone. To approximate real-world mobile app use, there were no further app-related engagement prompts from study staff after the baseline assessment.

During the 10-week study period, participants were invited to complete web-based biweekly surveys via REDCap (Research Electronic Data Capture; Vanderbilt University) [39,40] to
assess their depressive and anxiety symptoms (refer to details in the Measures section). If participants did not complete these biweekly assessments, a research assistant contacted the participant to remind them to complete the survey. A final assessment was conducted on mobile phone at the end of the 10-week trial to administer follow-up measures and conduct a brief qualitative exit interview. In addition to assessments via REDCap surveys, app analytics (eg, days played and activities completed) were captured through The Guardians via the game’s cloud save and gameplay recovery functionality. The participant data were compared with the anonymized gameplay data gathered from the app’s public users. All data gathered by The Guardians were anonymized and collected for research purposes, as stated in the app’s privacy policy [30].

Ethical Considerations

All the study procedures were approved by the Mass General Brigham institutional review board (protocol number 2021P001400). All participants provided informed consent, and all study data were deidentified before analysis. Participants were not provided compensation for their participation in the study.

Overview of The Guardians: Unite the Realm App

As noted previously, The Guardians: Unite the Realms is a mobile game designed to increase adherence to a modified BA therapy. A detailed description of the game has been published previously [30]. In brief, The Guardians is divided into 3 realms, each of which unlocks automatically after 28 days, regardless of game progression. In each realm, the user is asked to defeat an enemy character by collecting pets and sending them on missions. The pets automatically complete each mission after 10 to 60 seconds in real time. Once the pets complete a mission, they are given experience points and other rewards that can be used to further the in-game progress. Players are given a limited amount of regenerable “stamina,” which they must spend to send pets on missions. Thus, the challenge comes from carefully managing which pets should go on which missions based on the resources and stamina currently available. Players cannot lose the game or undo any progress. Every day, the player is prompted through the app to complete activities that will reward them with more pets. Players are notified to complete their daily activity in the game and via mobile phone notifications.

The player is prompted to either pick from a list of 75 suggested real-world activities or to choose their own activity. The preselected activities are divided into 3 effort levels (low, medium, and high) and into 5 categories (basics, arts and crafts, social, fitness, and fun); in-game rewards are the same regardless of effort level and category. Once the player chooses their activity for the day, the game instructs the participant to log the activity in the app once it is complete. Players cannot log the activity until sufficient time to complete their activity has passed. After logging an activity, players are prompted to reflect about how they feel after completing it, rating their post activity mood improvement on a scale of “1: Much worse” to “5: Much better.” Reflecting about how an activity affects mood is a key ingredient from BA therapy that has been integrated into The Guardians.

Measures

Diagnostic Assessment

In addition to collecting demographic information, the Mini International Neuropsychiatric Interview [41] was administered at the baseline assessment to evaluate the diagnostic criteria for a current major depressive episode (MDE) and to assess comorbid psychiatric illnesses. It is a structured diagnostic assessment that evaluates the current existence of a variety of psychiatric disorders based on the Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) criteria. To meet the MDE criteria, participants must report having had a depressed mood most of the day or markedly diminished interest or pleasure in activities for a period of at least 2 weeks. They must also endorse at least 5 of the following symptoms: significant unintentional weight or appetite change, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue, feelings of worthlessness or inappropriate or excessive guilt, decreased ability to concentrate or make decisions, and recurrent thoughts of death or suicidal ideation.

Primary Outcome: App Satisfaction and Engagement

Overall, 2 self-report measures were used to assess satisfaction with The Guardians among this sample at the end of the 10-week trial using the Mobile Application Rating Scale (MARS) and the Player Experience of Needs Satisfaction scale (PENS). MARS is a 23-item self-report survey that contains 4 objective quality scales (engagement, functionality, esthetics, and information quality) and 1 subjective quality scale to classify and assess the quality of mobile health apps [42]. All items are rated on a scale of 1 to 5, and the participants’ score is averaged across all items. High MARS ratings indicate a more usable and high-quality app. PENS is a 6-item survey measuring participants’ play experiences: 3 items measuring competence (eg, “I feel competent at the game”) and 3 items measuring autonomy (eg, “The game lets you do interesting things”) [43,44]. This self-report scale is based on the theory that video games have the potential to satisfy the basic psychological needs for competence, autonomy, and relatedness. Items are rated on a scale of 1 (do not agree) to 5 (strongly agree), and high PENS ratings indicate that the player felt the game met more of their otherwise unmet needs and are associated with high long-term retention [43].

Metrics of engagement captured in the app included Day-N user retention, defined as the proportion of users who interact with the game or complete an activity on the Nth day since they installed the game, where day 1 is the first day after installation, and the denominator is the number of users who install the game on day 0 [30,45]. To compare retention rates among participants in this study with those of public users of The Guardians, we used data collected from public users (as per the privacy policy [30] and stored them in The Guardians database. Other engagement metrics and data collected in the app included days played (ie, the number of days a participant logged into the app) and activities completed (ie, the total number of activities that a participant completed in the game). The exit interview, described further in the following sections, additionally captured qualitative feedback about app acceptability, quality, and overall satisfaction.
Secondary Outcomes: Changes in Depression and Anxiety Symptom Severity

Depressive symptom severity was assessed using PHQ-9 on a biweekly basis during the 10-week trial using a REDCap survey. PHQ-9 is a well-validated self-report measure consisting of 9 Likert-style items assessing various depressive symptoms [46]. If a participant endorsed suicidality at any point during the trial or via the PHQ-9 suicidality item, further assessment of suicidal ideation and behaviors was performed using the Columbia Suicide Severity Rating Scale [47]. A safety protocol was triggered if the participant endorsed item 9 of PHQ-9 (thoughts of being better off dead) or if the participant endorsed suicidal intent or plan upon completion of the Columbia Suicide Severity Rating Scale. Following a safety protocol trigger, a study clinician would contact the participant as soon as possible and, according to clinical judgment, call the participant’s emergency contact. Participants were also provided with safety resources, including the National Suicide Prevention Hotline, and the safety protocol was modeled after that used in a large federally funded study (Preventing Depressive Relapse in Pregnant Women with Recurrent Depression; National Institute of Mental Health; NCT03623620), where participants (N=500) were assessed for depression across the pregnancy and postpartum periods [48].

Participants also completed the Generalized Anxiety Disorder-7 (GAD-7) scale, a 7-item scale assessing symptoms of anxiety biweekly [49]. As noted in the description of The Guardians previously, an additional assessment of post activity mood improvement was captured in the app by prompting users to rate how they felt after completing an app-based activity on a scale of “1: Much worse” to “5: Much better.”

Qualitative Inquiry: Exit Interview

Participants were invited to complete a brief, 30-minute, web-based exit interview with a research assistant at the end of the 10-week trial. During this interview, participants were asked about perceptions regarding their level of engagement with the app, approach to selecting activities, motivation levels, and aspects of the game that they liked or disliked. Participants were also invited to share their experiences with using the app, perceptions about its impact on their mood, and any recommendations regarding how to best tailor the app for pregnant individuals. Refer to Multimedia Appendix 1 for the exit interview questions.

Data Analysis

Quantitative data were analyzed using SAS software (version 9.4; SAS Institute). Demographics, user engagement, ratings of satisfaction with the app (eg, activities completed, days played, MARS, and PENS), average post activity mood improvement, and depression and anxiety symptoms were analyzed using descriptive statistics. Pearson correlation analyses were conducted to explore the correlation between metrics of app satisfaction (ie, MARS and PENS) and engagement metrics (ie, activities completed and days played). A best-fit curve for the participant app user data was created by assuming an exponential decay in app retention. The user retention must have a day-0 retention of 1 by definition and can be assumed to have an eventual plateau of long-term dedicated users, resulting in the formula, \( y = (1 - \alpha) e^{-\lambda t} + \alpha \) [50]. A series of independent group 2-tailed \( t \) tests explored the differences in baseline measures of depression (PHQ-9) and anxiety (GAD-7) among participants who completed the final 10-week assessment and those who did not complete this final study assessment. A mixed effects General Linear Model was run with a dependent variable of post activity mood improvement rating versus the fixed effect of effort level (ie, easy, medium, and hard). The random term was the participants.

We also conducted analyses to explore how app engagement metrics (ie, activities completed and days played) were associated with symptoms of depression (PHQ-9) and anxiety (GAD-7) over the 10-week trial. To examine the effects of the intervention, as predicted by number of activities completed and days played, in separate analyses, on symptoms of depression and anxiety, we used backward-eliminated mixed effects longitudinal models for the dependent variables of PHQ-9 and GAD-7, also in separate analyses. The fixed predictors in these models were the numeric variable week in study (both linear and quadratic components) and either number of activities completed or number of days played and their interactions with week in study. We included baseline age and gestational age as additional covariates in the models. The random term was participants interacting with linear week in study. For all models, model residuals were checked for conformance to model assumptions of normality.

We also used backward-eliminated mixed effects longitudinal models to examine the effect of the overall app satisfaction rating (MARS), needs met through the app (PENS), and a binary predictor indicating whether treatment (ie, psychotropic medication or psychotherapy) was started midstudy for symptoms of depression (PHQ-9) and anxiety (GAD-7), in separate analyses. The random term was participants interacting with time, week in study. For all models, we calculated the proportion of variance in the dependent variable accounted for by the fixed effects and by fixed and random variables combined.

Content analysis [51] of the transcribed exit interviews was conducted to identify key themes regarding design features that participants reported as important for the usability and acceptability of the app. A team of 2 coders (LAK and HD) first familiarized themselves with the transcripts, making notes about the themes observed in the qualitative data. Key concepts from the transcripts were used to develop a codebook; codes were developed inductively from the data and deductively from interview topic areas, in addition to exploring the study participants’ perceptions about the design and usability of the app, the app’s impact on mood, and recommendations for the app’s improvement. The coders met to review the codebook and organize codes into broad categories, with guidance from the first author of this paper, who has experience in qualitative methods (RCV). Deidentified transcripts were coded and analyzed using Dedoose (version 9.0.17), a qualitative analysis software program [52].
Results

Overview

During the study enrollment period from September 2021 to April 2022, a total of 96 women indicated interest in participating in this pilot study. Refer to Figure 1 for the flow of participants, including reasons for ineligibility. Ultimately, 24% (23/96) women screened eligible for participation, and 19% (18/96) women enrolled in the study. Of the 18 participants, 10 (56%) completed the final assessment at the end of the 10-week study; we only have data on this 56% (10/18) of the participants for certain measures captured only at the end of the study (e.g., PENS, MARS, and exit interview). Table 1 describes the baseline characteristics and demographics of the enrolled sample.

Figure 1. Study participant flow diagram. PHQ-9: Patient Health Questionnaire–9.
### Table 1. Participant characteristics at baseline (N=18).

<table>
<thead>
<tr>
<th>Categories</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>34.3 (3.0)</td>
</tr>
<tr>
<td>Gestational age (weeks), mean (SD)</td>
<td>17.1 (7.1)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>4 (22)</td>
</tr>
<tr>
<td>Hispanic or Latina</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Non-Hispanic or Latina</td>
<td>16 (89)</td>
</tr>
<tr>
<td>White</td>
<td>12 (67)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>15 (83)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Queer</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>12 (67)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Never married</td>
<td>5 (28)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>16 (89)</td>
</tr>
<tr>
<td>Student</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Disabled or unable to work</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Insurance status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Private health insurance</td>
<td>17 (94)</td>
</tr>
<tr>
<td>Medicaid</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Postgraduate training</td>
<td>12 (67)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>4 (22)</td>
</tr>
<tr>
<td>Associate degree</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Some high school</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Treatment, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Psychiatric medication in past 2 months</td>
<td>9 (50)</td>
</tr>
<tr>
<td>Psychosocial treatment in past 2 months</td>
<td>8 (44)</td>
</tr>
<tr>
<td><strong>Symptom severity (score), mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>PHQ-9&lt;sup&gt;a&lt;/sup&gt;</td>
<td>13.2 (4.7)</td>
</tr>
<tr>
<td>GAD-7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9.8 (4.2)</td>
</tr>
<tr>
<td><strong>MINI&lt;sup&gt;c&lt;/sup&gt; diagnoses at baseline, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Major depressive episode</td>
<td>18 (100)</td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>10 (56)</td>
</tr>
<tr>
<td>Agoraphobia without panic disorder</td>
<td>5 (28)</td>
</tr>
<tr>
<td>Panic disorder with agoraphobia</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Social phobia</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>1 (6)</td>
</tr>
</tbody>
</table>
## Values

<table>
<thead>
<tr>
<th>Categories</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttraumatic stress disorder</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Drug dependence</td>
<td>1 (6)</td>
</tr>
</tbody>
</table>

\(^{a}\text{PHQ-9: Patient Health Questionnaire–9.}\)

\(^{b}\text{GAD-7: Generalized Anxiety Disorder–7 item.}\)

\(^{c}\text{MINI: Mini International Neuropsychiatric Interview.}\)

## App Satisfaction and Engagement

The average number of activities completed in *The Guardians* by participants (N=18) was 11.07 (SD 13.65), and the average number of days played was 16.67 (SD 17.52). The average activity completion rate (ie, rate at which a participant, on average, would complete an activity on any given day that they opened the gaming app) was 68.2% (SD 0.33). Of all the activities that study participants completed across the study period, 43.6% (65/149) were classified as easy, 27.5% (41/149) as medium, and 13.4% (20/149) as hard in effort level (the remaining were classified as other for effort level). Among the 56% (10/18) of the participants who completed MARS and PENS ratings at the final 10-week assessment, the average MARS rating was 3.49 (SD 0.76), and the average PENS rating was 4.13 (SD 1.66). All the correlations between engagement metrics (ie, days played and activities completed) and satisfaction measures (ie, MARS and PENS) were moderately positive (r=0.44-0.50) but not significant (with P values ranging from .13 to .20). Day-N study participant retention curves mapped on to those of the public *The Guardians* users are displayed in Figure 2. The optimal fit with all study data included is \(\alpha=0.13\) and \(\lambda=0.12\) with \(R^2=0.93\).

**Figure 2.** Day-N user retention. Day-N user retention rates from day 1 to day 50 for public users of *The Guardians: Unite the Realms* and the study participants, with both raw data and its best-fit line.

## Change in Clinical Symptoms

### Overview

A series of independent 2-tailed \(t\) tests indicated no baseline differences in depression or anxiety scores between study completers (10/18, 56%) and noncompleters (8/18, 44%; \(P>0.05\)). Table 2 summarizes the change in PHQ-9 and GAD-7 scores across the 10-week study.

Regarding postactivity mood improvement ratings, participants reported feeling at least “a little better” 76% (113/149) of the time after completing an activity as part of *The Guardians* app. On average, participants reported having a greater improved mood after completing hard activities (mean 4.47, SD 0.21) compared with after completing activities identified as medium (mean 4.17, SD 0.16) or easy (mean 3.89, SD 0.14) in effort level. There was a significant (\(P=0.05\)) relation between activity effort level and post activity mood improvement, with a gradually increasing mean post activity mood improvement rating from Easy to Medium to Hard effort levels of activities, as shown in Figure 3.

### Table 2. Average scores on the 9-item Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder Scale-7 (GAD-7) across study assessment time points.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline (N=18)</th>
<th>2 weeks (n=16)</th>
<th>4 weeks (n=12)</th>
<th>6 weeks (n=13)</th>
<th>8 weeks (n=12)</th>
<th>10 weeks (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHQ-9 score, mean (SD)</td>
<td>13.17 (4.73)</td>
<td>9.88 (5.70)</td>
<td>8.67 (3.77)</td>
<td>9.92 (6.61)</td>
<td>9.58 (5.67)</td>
<td>7.2 (3.65)</td>
</tr>
<tr>
<td>GAD-7 score, mean (SD)</td>
<td>9.8 (4.18)</td>
<td>10.29 (5.62)</td>
<td>9.25 (4.03)</td>
<td>9.39 (5.46)</td>
<td>8.50 (3.50)</td>
<td>7.10 (4.12)</td>
</tr>
</tbody>
</table>
Figure 3. Means for postactivity mood improvement for various categories of activity effort level. Means are estimated via least squares (LS) regression and thus conventionally termed as “LS Means.”

Change in PHQ-9 Score According to Time (Week in Study) and Activities Completed

For all the mixed effects models in the following sections, baseline age and gestational age were removed from the model as nonsignificant. For the first mixed effects model where the dependent variable was the PHQ-9 score, there was a significant negative main effect of linear *week in study* (unstandardized partial regression coefficient $\beta = -0.35$; $t_{61} = -3.05$; $P = 0.003$; 95% CI $-0.59$ to $-0.12$) and number of *activities completed* ($\beta = -0.12$; $t_{61} = -2.05$; $P = 0.04$; 95% CI $-0.25$ to $-0.003$). Overall, these fixed effect predictors accounted for 20.8% of the variance in the PHQ-9 score.

Change in GAD-7 Score According to Time and Activities Completed

The same model with the independent variables, *activities completed* and *week in study*, was rerun with GAD-7 score as the dependent variable. There were no significant effects of *activities completed* and only a marginally significant quadratic relation for *week in study* ($\beta$ for linear term$=0.3$; $P=0.38$; $\beta$ for quadratic term$=-0.06$; $t_{45}=-1.89$; $P=0.07$; 95% CI $-0.125$ to 0.004). Overall, fixed effects accounted for only 4.2% of the variance in the GAD-7 score, whereas combined fixed and random effects accounted for 77.96% of the variance. Figure 4 displays the predicted values for this model and essentially indicates an accelerating decline in GAD-7 score across time, beginning at approximately week 4.
Figure 4. Mean values for Generalized Anxiety Disorder Scale-7 (GAD-7; with 95% CI) predicted by significant fixed effects in the fitted longitudinal model of GAD-7 versus week in study and activities completed. For fixed effects, a quadratic function for week in study was retained (marginally significant; $P=0.07$), whereby the model–predicted GAD-7 values remained essentially stable up to approximately week 4.

*Change in GAD-7 Score According to Days Played*

The same model as described previously was run with *days played* as a fixed effect substituted for *activities completed*. Although the main effect terms for linear *week in study* and for *days played* were not significant, there was a marginally significant interaction effect of *linear week in study* and *days played* ($\beta=-0.01$; $t_{55}=-1.97$; $P=0.05$; 95% CI $-0.02$ to $0.0002$), whereby GAD-7 score was predicted as essentially stable across time for participants with low numbers of *days played* but showed increasingly steep declines across time for those with increasingly high numbers of *days played* (Figure 5). Fixed effects accounted for 11.31% of the variance in the GAD-7 score.
Figure 5. Mean values for Generalized Anxiety Disorder Scale-7 (GAD-7; with 95% CI) predicted by significant fixed effects in the fitted longitudinal model of GAD-7 versus week in study and days played. For fixed effects, only an interaction of days played × linear week in study was retained as significant ($P=.05$), whereby GAD-7 values were predicted by the model to remain essentially stable across time when the number of days played was relatively low (below 10) but showed an increasingly steep linear decline across time when the numbers of days played was increasingly high. Predicted values for GAD-7 are shown at selected illustrative strata of number of days played ranging from the minimum observed value of 2 up to the maximum observed value of 57 and at equally spaced intervals in between.

Change in PHQ-9 Score Over Time According to MARS, PENS, and Treatment Started

In the model with MARS, PENS, week in study, and whether medication was started midstudy as independent variables and PHQ-9 score as the dependent variable, we found a significant negative main effect for linear week in study ($\beta=-.35; t_{45}=-2.74; P=.009; 95\% \ CI \ -5.86 \ to \ -0.24$) and MARS ($\beta=-3.05; t_{45}=-2.19; P=.03$; Figure 6). The PENS variable was nonsignificant and was removed in the backward elimination. Additive to these effects, there was a marginally significant effect of treatment initiation ($\beta=-3.25; t_{45}=-1.84; P=.07$), with the treatment started group (ie, medication or therapy started during the study period) estimated to have an adjusted mean of 11.45 (SD 1.2588), whereas the no treatment started group had a low adjusted mean of 8.20 (SD 1.2393). Overall, fixed effects accounted for 35.01% of the variance in the PHQ-9 score. There was no interaction effect between treatment started and reduction in PHQ-9 score across time; therefore, there was no significant difference in the slopes of change across time between the 2 medication groups.
Figure 6. Mean values for Patient Health Questionnaire-9 (PHQ-9; with 95% CI) predicted by significant fixed effects in the fitted longitudinal model of PHQ-9 versus week in study, Mobile Application Rating Scale (MARS), Player Experience of Needs Satisfaction Scale, and treatment initiation (ie, whether medication or therapy was started or not started during the study). For fixed effects, only negative main effects for linear week in study ($P=.009$), MARS ($P=.03$), and treatment initiation (marginal; $P=.07$) were retained. Predicted values for PHQ-9 are shown in the respective panels for those who started or did not start treatment, across time within each panel and at selected illustrative strata of MARS ranging from the minimum observed value of 2.256 up to the maximum observed value of 4.283 and at equally spaced intervals in between.

Change in GAD-7 Score Over Time According to MARS, PENS, and Treatment Started

For the backward-eliminated mixed effects model for total GAD-7 score, there was a significant interaction between MARS and weeks in study ($\beta=-.64; t_{43}=-3.27; P=.002$), reflecting the fact that as MARS ratings increased, GAD-7 score was predicted to decline more across time (Figure 7). PENS and treatment started variables were nonsignificant and removed in the backward elimination. Overall, fixed effects accounted for 18.07% of the variance in the GAD-7 score.
Figure 7. Mean values for Generalized Anxiety Disorder Scale-7 (GAD-7; with 95% CI) predicted by significant fixed effects in the fitted longitudinal model of GAD-7 versus week in study, Mobile Application Rating Scale (MARS), Player Experience of Needs Satisfaction Scale, and treatment initiation. For fixed effects, only an interaction of MARS $\times$ linear week in study was retained as significant ($P=.002$), whereby GAD-7 values were predicted by the model to vary from rising linearly across time among individuals with relatively lower MARS ratings to declining linearly across time for those with relatively higher MARS ratings. Selected illustrative strata of MARS in the figure were chosen to range from the minimum observed value of 2.256 up to the maximum observed value of 4.283 and at equally spaced intervals in between. The slight crossover of lines before approximately week 3 is likely just an artifact of the linear constraint for the lines and probably should not be interpreted with substantive meaning.

Qualitative Feedback

Overview

Themes from the exit interviews fell into the following major categories: app activities, app design, app engagement, fit of The Guardians with lifestyle, and perceived impact of The Guardians on mood. Positive and negative feedback were identified in each of these categories. Participants also provided suggestions for app modifications, some of which were general recommendations and others were specific to pregnant app users.

App Activities

Participants reported that The Guardians encouraged completion of activities outside the game, as described in the following quote:

So, I would try to do daily house tasks...Like there was one, “clean-up a junk drawer;” and another one, like, “clean a room in your house.” So, I would try to pick those activities because, let’s face it, nobody ever wants to do those kinds of things. And that, I think, was kind of helpful for me, especially because I knew we were going to be moving, and it helped me declutter some stuff before we moved, so that was really awesome.

Participants also reported appreciating the variety of activities available. A participant shared the following:

I enjoyed the activities that were offered...there’s a wide variety of activities that would suit a lot of different people.

Some participants, however, had difficulty in completing certain activities (eg, “Doing longer [activities] was harder because at the time I was having some health issues myself...”). Others felt that it was easy to select activities that they were already doing in their daily lives, as described by a participant:

I mean, most of [the activities] were things that I was kind of doing already, like taking a shower, or I think, like, doing dishes...So honestly, I kind of picked the easy ones I was automatically doing.

App Design

Overall, participants appreciated the design of The Guardians’ characters (eg, “The aesthetics were adorable; they’re really cute.”). A participant noted that the characters in the game motivated her to continue engaging with the app:

I think within the game itself, my biggest motivator was probably getting new pets. As someone who really likes arts and things like that, the designs of the creatures were really interesting. Being able to get a new creature each day was great.
Participants also appreciated the design of _The Guardians_ itself, in terms of its esthetics (eg, “I did enjoy that the different levels had different world feels.”) and mechanics (eg, “...It’s one of the only apps that never crashed on me.”).

Some participants also had more critical feedback regarding the app’s design. For instance, a person felt that the app was difficult to navigate (eg, “If I didn’t feel like I was...fumbling my way through it, I could have maybe engaged faster and more.”). Others were frustrated that they had to wait before the next level unlocked (eg, “At one point, I think I had to wait, like, thirty days or something in order for the next level to open up, and I was like, ‘Oh, my goodness—what do I do now?’”). Similarly, some participants were frustrated by the length of the game’s introduction and overall complexity, as summarized by a woman:

> I would say that the learning curve was just too much...like the mental state I’m in, plus, like, the amount that I’m trying to juggle. It was just a lot to try to add the additional learning of a game...

**App Engagement**

Participants shared that _The Guardians_ allowed them to develop a sense of routine (eg, “When I was trying to get myself out of bed and stuff like that, [the app] just gave me another reason to get out of bed...”). They also enjoyed the gamification of daily household activities that may otherwise be difficult to complete while pregnant or experiencing elevated depressive symptoms (eg, “I definitely liked the aspect of...the little challenges of doing laundry and cleaning and stuff like that, because it kind of gamified doing that type of stuff.”). Furthermore, engagement with the app even helped a participant to start and maintain a healthy coping behavior beyond the end of the trial:

> I think having that structure, having that reward to engage in a very common coping mechanism for me, really helped to motivate and get me in a pattern of reading that has continued, even if I’m not actively engaging with the app right now.

Not all participants, however, found _The Guardians_ to be motivating or engaging (eg, “I didn’t really feel like [the app] was pushing me to come back in any way.”). Some found the app to be very complex, as previously described, whereas others felt that it was very simple (eg, “I think if it was a little bit more complex, it would have been more engaging for a longer period of time.”).

**Fit of The Guardians Into Lifestyle**

Some participants found that using _The Guardians_ was an easy fit into their existing lifestyle (eg, “I would use [the app]...on the way to work.”). Furthermore, a participant felt that the game helped them return to a previous routine:

> Once I started using the app in the mornings, I was able to get out of bed and do the things that I [had previously] been doing. So, it kind of snapped me back into my regular routine...

In terms of negative feedback, some participants felt that their busy schedules and lifestyles were not compatible with using _The Guardians_ consistently. A participant stated the following:

> I already have three kids, so I needed something to kind of be easy and mindless, and this was like...you had to be focused in on what you’re doing to even learn the game. So, for me, it was just too much.

**Perceived Impact of The Guardians On Mood**

Some participants reported that _The Guardians_ positively influenced their mood (eg, “I think that the reward [was] my mood improving.”). Other participants did not find the app effective in influencing their mood (eg, when a participant was asked whether _The Guardians_ affected their mood, the participant responded, “No, not really.”).

**Suggestions for App Modifications**

Participants had various recommendations for future improvements to _The Guardians_. Broadly, participants recommended simplifying the introduction to the app (eg, “Make it a little bit easier so it’s not so much upfront...to make sense of it.”). In addition, participants suggested a way to track activities in the game to monitor progress (eg, “I think if...it tracked [the activities], that would be really helpful, too.”).

Recommendations also centered around ideas for making _The Guardians_ more relevant and engaging for pregnant individuals. Participants suggested adding more pregnancy-specific activities, as described in the following quote:

> I think maybe for pregnant women specifically...if you made one of those activities, like, do your kick counts, or something...Just a reminder to some people, especially first-time moms, because they might not know to do those kinds of things...

Furthermore, some participants recognized that potential activities could differ based on gestational age. A participant shared the following:

> I think there are elements too, where there’s pregnancy education that could be added into a gamification setting, where you are educating people at the same time as encouraging...healthy habits and things like that.

**Discussion**

**Principal Findings**

Our findings from this pilot study provide preliminary support for _The Guardians_ as a potentially desirable intervention to engage and improve mood among some pregnant women. All enrolled participants met the criteria for a current MDE and the average baseline depressive symptoms fell into the moderate...
symptom severity level according to PHQ-9 [53], suggesting that The Guardians may be of interest to and able to engage pregnant individuals with moderate depressive symptom levels. Without providing any compensation for participation in the study, we enrolled 18 pregnant women within a brief, 8-month recruitment period, indicating that a gaming app such as The Guardians may be of interest to some individuals in the target pregnant population. Furthermore, one-third of the sample (6/18, 33.3%) identified as non-White, which is encouraging in that the app may be acceptable to a diverse population. However, we recognize that this was a small and relatively highly educated and employed sample; further studies are needed to assess the ability to engage a diverse perinatal population with elevated depressive symptoms using The Guardians.

Our user engagement data aligned with previous studies suggesting that The Guardians may be effective in improving long-term engagement relative to other digital mental health interventions [30]. The 15-day and 30-day overall app retention rates of 26.6% and 15.1%, respectively, compare favorably with the median retention rates of 3.9% (IQR 10.3%) and 3.3% (IQR 6.2%) for mental health apps, as reported by Baumel et al [27]. Furthermore, our retention rates suggest that The Guardians is favorably consistent with rates observed in the top 15% of entertainment-only mobile games that include engagement rates [54]. Such retention and engagement metrics are particularly encouraging, as there were no external (ie, outside the app) reminders for participants to use The Guardians, unlike other digital mental health interventions (eg, use of a therapist or coach to check in and encourage engagement). Although not statistically significant (likely owing to the small sample size in this study), it is not surprising that there were positive, moderate correlations between ratings of satisfaction with the app (as assessed using PENS and MARS) and engagement with the app (as assessed using days played and activities completed).

This was the first study to explore the change in depressive and anxiety symptoms among users of The Guardians. Our findings demonstrated reduction in both depressive and anxiety symptoms over the course of the 10-week study on average among study participants. Although we cannot attribute the decrease in clinical symptoms to the app owing to the lack of control group and this being an underpowered pilot study focused on feasibility, our analyses indicated statistically significant relationships between depression improvement and app engagement when there were less depression symptoms (according to PHQ-9), more activities were completed, and more days played. This offers some preliminary support that there may be improvement in depression when pregnant women engage more with The Guardians. Furthermore, low depressive symptoms were associated with high MARS rating of app satisfaction, indicating that there may be great improvement in depressive symptoms when a user is more satisfied with The Guardians. Overall, of the 2 app satisfaction ratings (ie, MARS and PENS), our mixed effects models found MARS to be a better predictor of change in clinical symptoms (main effect for decrease in PHQ-9 score; interaction with time for decrease in GAD-7 score) relative to PENS, which was removed in backward elimination for all of our analyses.

Our findings also suggest that there may be great improvement in mood after completing activities that require more effort. However, the effort level of app activities was assigned by the app developers, and activities determined to be easy or hard may not align with how a user would rate the effort for these activities. Further studies are needed to explore the relationship between the types and effort levels of activities completed and impact on mood. In addition, this study focused on individuals only during pregnancy and not during the postpartum period, as it seems there are unique aspects of the postpartum period (eg, sleep deprivation and demands of a newborn) that may influence the ability to engage with the app or complete certain activities. Future studies could assess the impact of this app in the postpartum period and further explore how certain activities in the postpartum period are perceived in terms of effort level and impact on mood.

Our findings regarding change in anxiety symptoms revealed an interaction between days played and time, indicating that there were relatively stable anxiety levels on average among participants who used the app less, but there was great improvement in anxiety scores for those who used the app more (ie, great number of days played). This provides some support for the potential benefit of using The Guardians among this sample of pregnant individuals with anxiety and elevated depression, yet only for those who used the app more often. Not surprisingly, given the nonsignificant but positive correlation between app satisfaction (ie, MARS) and days played, there was great improvement in anxiety symptoms when the MARS ratings were high (ie, high acceptability). Again, further studies are needed with a powered RCT to explore these changes in clinical symptoms among users of The Guardians, including assessment of other factors that may be contributing to change in clinical symptoms (eg, specifics about treatment changes during the study and major life events). Our findings indicate that participants who started treatment (medication or psychotherapy) during the course of the trial had great depressive symptoms both at baseline and at the end of the 10-week trial; however, the rate of change in depressive symptoms between these groups over the course of the study period was similar.

The qualitative feedback provided by participants indicated that there were contrasting opinions about certain aspects of the app. Participants tended to positively assess the activities on the app if they were novel to the participant or suited their individual lifestyles and abilities. Participants tended to negatively assess the activities if they felt that they were very difficult or did not have a significant or positive effect on their mood. There were also some contrasting views about the nature of the app; some felt that it was “mindless” and fun, whereas others felt that it was very demanding and time consuming. Future studies could explore ways to tailor the app to make it more or less challenging, depending on a person’s preferences or needs. On the basis of participants’ recommendations, other features could be added in future iterations of the app, such as an activity tracking module to help monitor activity completion and impact on mood over time. Other feedback from participants will be critical in informing the adaptation of The Guardians to a perinatal population to include pregnancy-specific activities.
and educational content. It is possible that a more tailored version of the app for perinatal individuals with depression may be even more feasible, acceptable, and effective in reducing depressive symptoms for this population.

**Limitations**

This study has several limitations. As indicated previously, the study sample was small and relatively highly educated and employed. Thus, we do not know how generalizable these findings are to a broad population. As this pilot study did not have a control condition, we cannot make any causal claims regarding the impact of the app on clinical symptoms. Furthermore, future studies should assess how access to smartphones or comfort with technology may influence who would engage with an app such as *The Guardians*. A large RCT is needed to further explore how *The Guardians* may be used among pregnant individuals across socioeconomic levels and to rigorously assess the app’s effectiveness in treating PD.

**Conclusions**

*The Guardians* is a widely available and highly engaging gaming app that incorporates BA principles. In this study, we sought to explore the usability, acceptability, and preliminary effectiveness of this app among a small sample of pregnant individuals with elevated depressive symptoms. Findings from this small pilot study provide initial support for *The Guardians* as an acceptable and engaging app, and there may be some improvement in mood and anxiety among certain users in the target population. This was one of the first longitudinal pilot studies to explore the effectiveness of a BA gaming app on PD. Further studies, including a powered RCT, is needed to follow-up on these preliminary findings.

**Acknowledgments**

The study was supported by the Brain and Behavior Research Foundation, National Institute on Aging, National Institutes of Health, and SAGE Therapeutics. Advisory and consultation support was provided by JDS Therapeutics LLC.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest**


Multimedia Appendix 1

Exit interview questions.

[DOCX File 15 KB - formative_v8i1e44029_app1.docx]

**References**


25. Flynn HA, Henshaw E, O’Mahen H, Forman J. Patient perspectives on improving the depression referral processes in


38. Home page. Massachusetts General Hospital Center for Women's Mental Health. URL: https://womensmentalhealth.org/ [accessed 2023-12-12]


42. Robertson M. Can't play, won't play. Hide & Seek. 2010 Nov 10 URL: https://kotaku.com/cant-play-wont-play-5686393 [accessed 2023-12-12]


Telephone-Based Training Intervention for Using Digital Communication Technologies for Social Housing Residents During the COVID-19 Pandemic: Mixed Methods Feasibility and Acceptability Evaluation

Tim Walker¹, BA, MSc, PhD; Sarah Ann Buckingham¹, BSc, ResM, PhD; Ria Poole¹, BSc, PhD; Lewis Roland Elliott¹, BSc, MSc, PhD; Tamaryn Menneer¹, BSc, PhD; Gengyang Tu², MA, PhD; Karyn Morrissey³, BA, MA, PhD

¹European Centre for Environment and Human Health, University of Exeter Medical School, Truro, United Kingdom
²International Business School Suzhou, Xi’an Jiaotong-Liverpool University, Suzhou, China
³Department of Technology, Management and Economics, Technical University of Denmark, Lyngby, Denmark

Corresponding Author:
Tim Walker, BA, MSc, PhD
European Centre for Environment and Human Health
University of Exeter Medical School
Knowledge Spa
Truro, TR1 3HD
United Kingdom
Phone: 44 (0) 1872 258131
Email: t.walker@exeter.ac.uk

Abstract

Background: In an era in which digital communication technologies play a pivotal role in everyday life, social housing residents remain highly susceptible to digital exclusion.

Objective: This study aims to evaluate the feasibility and acceptability of a telephone-based training intervention designed to empower people to confidently use digital communication technologies (ie, video calls and web-based messaging).

Methods: Conducted in collaboration with a UK social housing association, the intervention was facilitated by a unitary authority’s Digital Inclusion Team during the COVID-19 pandemic. A mixed methods approach was used, encompassing quantitative and qualitative data collection on demand, reach, implementation, and potential outcomes. Demographic and qualitative data on the reasons for undertaking or not undertaking the training were collected via telephone interviews during the recruitment process. Digital competency and well-being data were collected via a self-reported survey before and after the intervention.

Results: Among the 4485 residents who were offered training, 67 (1.49%) expressed interest, of whom 12 (18%) of the 67 completed the training. The findings indicate a demand for basic digital training among social housing residents. The key findings revolve around the substantial dropout rate among those who were interested in undertaking the training. Barriers were strongly influenced by socioeconomic and health circumstances, reflecting the sociodigital inequalities commonly found in this group. For the training participants, the intervention was acceptable and achieved its goals, demonstrating the potential of tailored, persistent training efforts in overcoming barriers. There were no changes in self-reported well-being or digital competency outcomes (but this was limited by the small sample size).

Conclusions: Sociodigital inequalities impact the reach, implementation, and acceptability of telephone-based digital training for social housing residents. Barriers to reaching and training digitally excluded groups can be overcome through the use of trusted intermediaries, personalized recruitment approaches, the minimization of administrative barriers, and tailored and agile training programs. Recognizing the resource-intensive nature of such initiatives, this study calls for enhanced recognition of intermediary efforts in national digital inclusion policies.

(JMIR Form Res 2024;8:e45506) doi:10.2196/45506
KEYWORDS
digital training; telephone-based; social housing; feasibility; acceptability; communication technologies; sociodigital inequalities; mobile phone

Introduction

Background
Facilitated by near-ubiquitous digital connectivity in the global north, digital technologies are increasingly used across all areas of life and bestow significant benefits upon those who use them [1,2]. In response, many organizations, not least public service providers, proceed as if access to the internet and digital technology is already universal [3]. However, although connectivity has increased, access to digital technologies and the competencies to use them are not yet universal [4,5]. In the United Kingdom, the 2021 Consumer Digital Index found that approximately 10 million people (14% of the population) do not have the basic digital competencies needed for use the use of everyday digital technologies [6]. Digital exclusion can no longer be seen as a product of rural or remote living, place-based infrastructural and connectivity issues, or older age. Instead, digital exclusion is currently most likely a product of wider social and economic inequalities. Sociodigital inequalities [7] refer to the interplay between traditional (ie, social, economic, and health) and digital inequalities and have led to systematic differences in the ability and opportunity of different groups to beneficially use digital technology and participate in society. These differences, in turn, contribute to deeper social inequalities, which lead to greater digital inequalities over time [8].

A growing body of literature has begun to document the extent of digital inequalities and highlight the specific barriers to digital use in different social contexts [8]. In the United Kingdom, the Digital Inclusion Strategy sets out how the government and partners from the public, private, and voluntary sectors will increase digital inclusion [9] by targeting the following 4 interconnected barriers to digital inclusion: digital access, skills, confidence, and motivation [10]. However, although national policies are imperative to improve inclusion from an infrastructure and digital connectivity [1] perspective, localized interventions are much better placed to address digital skills, confidence, and motivation across previously excluded groups [11]. In the United Kingdom, localized digital training interventions are typically delivered by intermediaries, such as voluntary and community sector organizations and public libraries, and include formalized group programs, one-to-one digital buddy programs [12], and peer support [13]. Specifically, the United Kingdom’s Local Government Association’s Digital Inclusion Programme is funding councils to reach residents and provide personalized digital training programs for those who do not have access to or confidence in using digital communication platforms [14]. However, there is limited evidence on whether these digital training interventions are effective in increasing “digital readiness,” the competency and increased motivation to use digital technologies [15-17]. However, evidence does suggest that individuals who participate in these digital interventions have a higher socioeconomic status, higher education, higher social participation, and a greater experience with technology [18,19]. Conversely, those who are already experiencing sociodigital inequalities remain the hardest to reach [20-22] owing to reported barriers, including physical health issues, a negative attitude toward technology, caring for a sick spouse, a lack of energy, and a lack of time [23,24].

This Study
Within the context of the digital needs of hard-to-reach populations, this study focused on social housing residents in Southwest England. In the United Kingdom, social housing associations (HAs) are not-for-profit organizations that provide rental properties at 50% to 60% of market rates to those whose circumstances exclude them from the private market [25]. Overall, 3.9 million people live in social housing for various socioeconomic reasons [25]. Studies have shown that the demographic and socioeconomic profile of social housing residents means that they are significantly more likely to be digitally excluded and harder to reach than other groups in the United Kingdom [10,26-28]. This is because the factors that are known to increase digital exclusion are found at higher incidences among social housing residents than among those outside the social housing system [29,30]. These factors include lower incomes, fewer qualifications, older age, physical and mental health issues, disabilities, and living in more deprived areas [31,32].

The Smartline project [33] worked with 200 social housing households in one of the most deprived areas of England [34] to understand the potential of everyday digital technology to address health and well-being challenges. A qualitative scoping study on the feasibility and acceptability of digital technology among Smartline participants found that although the participants had positive perceptions of technology and were keen to try new technologies, digital readiness and the desired digital destination (goals) varied greatly among the community [28]. Several concerns surrounding technology use were identified, including data security and privacy concerns and the fear of “making a mistake” or “pressing the wrong button.” Many participants expressed a strong desire for further training and support.

Following this research, a training intervention was conducted to help Smartline participants get on the web and use digital communication technologies, such as web-based video calls and messaging, with confidence. The Getting Online: Staying Connected (GO:SC) intervention was originally planned as a face-to-face intervention with peer-to-peer support on how to use video calling technology. However, as with many research interventions during this time [35], the outbreak of the COVID-19 pandemic in March 2020 meant that the study had to be redesigned as a smaller-scale telephone-based training intervention for social housing residents delivered in conjunction with the Cornwall Council’s Digital Inclusion Team (DIT). The aim of this study was to evaluate the feasibility of the telephone-based training intervention. Specifically, informed by the feasibility framework of Bowen et al [36] and RE-AIM
(Reach, Effectiveness, Adoption, Implementation, and Maintenance) model of Glasgow et al [37], this study aimed to evaluate the (1) reach and demand, (2) implementation, (3) acceptability, and (4) potential efficacy of the telephone-based digital training intervention for social housing residents [36,37]. These are established frameworks for evaluating digital health interventions [38,39] and are suitable and credible for testing an unexamined intervention in a real-life setting where constraints exist over conditions [36,38,39].

**Methods**

**Overview of the Study Design**

This was a mixed methods feasibility study. The protocol and ethics application were informed by best practices for the process evaluation of public health interventions [40]. An overview of the study procedure, including recruitment and data collection, is provided in Figure 1 and Table 1 and detailed throughout the subsequent subsections.

**Figure 1.** Feasibility study process map for the digital training program covering recruitment, data collection, and intervention delivery.
Residents who expressed an interest in the digital training either via telephone or email were posted the study information sheet, the consent form, and a baseline survey assessing well-being and digital competency (Multimedia Appendices 2-4). Participants returned the completed forms via a Freepost envelope to the research team, who forwarded the participants’ phone numbers and digital training interests to the Cornwall Council’s DIT via email. The DIT then directly telephoned each participant to arrange the provision of the training.

The inclusion criteria were adult (aged >18 y) social housing residents in Cornwall who defined themselves as lacking either the competency or confidence to use digital communication apps. These included existing participants in the Smartline project and residents of the wider Coastline HA. Participants were excluded if they did not have access to an internet connection or at least 1 internet-enabled device. Training was provided free of charge to participants (in line with the DIT’s protocol).

The Intervention

The intervention is described in line with the TIDieR (Template for Intervention Description and Replication) standards for intervention reporting [42]. The training intervention was based on a 4-week, face-to-face, and basic digital skills course that was established by the DIT and delivered in libraries and community venues across Cornwall before the COVID-19 pandemic. The course structure adhered to the UK government’s Essential Digital Skills Framework [43], encompassing fundamental computer skills, problem-solving skills, communication skills, transactional skills, and skills for handling

---

**Table 1. Feasibility study data sources and time points of data capture.**

<table>
<thead>
<tr>
<th>Data source</th>
<th>Data captured</th>
<th>Time point of data capture</th>
<th>People who captured the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment log</td>
<td>• Participant ID&lt;br&gt;• Contact details, including the best time to contact and preferred method of contact&lt;br&gt;• Whether the participant has an internet-enabled device&lt;br&gt;• The date of completion of consent forms and surveys (administered at recruitment, baseline, and follow-up)&lt;br&gt;• Field notes on the recruitment process and intervention implementation</td>
<td>At recruitment&lt;br&gt;Find the research team&lt;br&gt;Participant</td>
<td>Recruitment survey (Multimedia Appendix 1; n=168)&lt;br&gt;• Demographics (age, gender, disability, racial identity, and cultural identity)&lt;br&gt;• Interest in taking part (yes or no); reasons for taking part or not taking part&lt;br&gt;• Digital training interests (free text)</td>
</tr>
</tbody>
</table>
digital information and content, as detailed in Multimedia Appendix 5.

In response to the social distancing measures in place during the COVID-19 pandemic, the DIT adapted the training to be delivered over telephone. This adaptation was informed by best practices in digital inclusion initiatives [44] and participant preferences identified in our prior qualitative scoping work [28]. The adapted training approach emphasized informality; flexibility; person-centeredness; and task-specific, on-demand delivery, diverging from traditional, predetermined, and technically focused programs imposed by training organizations [44]. This informal model aligns with research indicating that training success factors encompass participant autonomy, personalized learning, practice opportunities, and individualized support [13,45]. Informal learning, in particular, enhances self-efficacy and proves highly effective for digitally marginalized groups [46].

The overall purpose of the training intervention was to support participants in achieving active and continued use of digital devices and communication apps of their choice (eg, WhatsApp [Meta Platforms, Inc], Zoom [Zoom Video Communications, Inc], and Facebook). The training was delivered through one-on-one phone calls facilitated by a digital inclusion officer from the DIT, all of whom possessed a background in education. To provide course content guidance and ensure consistent and best-practice training delivery, the DIT used a bank of task-specific instructions (ie, step-by-step guide to installing WhatsApp), which could be printed and posted to participants on demand (Multimedia Appendix 6). Training sessions varied in duration (between 5 min and 2.5 h) according to each participant’s individual learning needs, often including follow-up calls as necessary.

Ethical Considerations

Overarching ethics approval to conduct research with Coastline Housing (a social HA in Cornwall) residents was granted by the University of Exeter Business School Research Ethics Committee as part of the Smartline project (eUEBS002996 v4.0). Specific approval for the GO:SC project was granted by the College of Life and Environmental Science Penryn Research Ethics Committee (eCORN002229). Written informed consent was obtained from all participants involved in the study (Multimedia Appendices 3 and 4). Data were anonymized using pseudonyms. Participants who completed the follow-up survey received a £10 (US $12.7) shopping voucher as compensation for their time.

Data Overview

The following 4 qualitative and quantitative data sources were used to assess the feasibility and acceptability of the intervention: a recruitment log, recruitment survey, well-being and digital competency survey (administered at baseline and 6-month follow-up), and digital training call log (Table 1). Previously piloted by the Smartline participants, the well-being and digital competency survey (Multimedia Appendix 2) used quantitative rating scales and was completed at baseline (immediately before training commencement) and 6 months following the training, with the aim of assessing potential outcomes. This survey (disseminated via post) was based on a validated survey, the “Happiness Pulse” [47], and included 4 domains of psychological well-being (general, emotional, behavioral, and social). A bespoke module on digital attitudes, behaviors, and competence was developed by the research team for the wider Smartline project. This digital module was informed by behavior change and technology acceptance theories [48-51] and included questions adapted from existing sources, including the UK government’s Digital Inclusion Evaluation Toolkit [52]. The module contained specific questions on video calls and messaging in addition to questions on technology in general. The theoretical basis of this module is provided in Multimedia Appendix 7 [47,49-63].

Data Analysis

To determine the reach of and demand for the intervention, as well as participants’ levels of engagement with the training, descriptive statistics (frequencies and percentages) from the recruitment log, recruitment survey (n=168), and training call log (n=12) were calculated. A probit model was used with the recruitment survey data to identify the socioeconomic factors associated with initial interest in participating in the digital training program. The analysis was performed in Stata (version 17; StataCorp) [64].

For the well-being and digital competency survey, the scoring protocol for the Happiness Pulse was followed [53], with means and SDs calculated from summary scores to describe each of the 4 well-being domains. As digital competency outcomes were measured on interval scales, medians and IQRs were calculated for these outcomes.

The COREQ (Consolidated Criteria for Reporting Qualitative Research) criteria [65] for reporting qualitative research were adhered to throughout the analyses. Qualitative data analysis used data from the recruitment log, recruitment survey, and digital training call log. To manage the qualitative data analysis process with transparency and traceability [66], NVivo (QSR International) [67] was used. In line with best practices [68], 3 rounds of inductive coding were conducted using a constant comparative method [69,70]. The first round of coding was open and focused on identifying and labeling discrete incidents. For example, “I have a smartphone, but I struggle to use it” contains 2 incidents: an object (smartphone) and a construct (competency). The second round of coding was axial, where open codes were compared (via contradiction, expansion, and support) and integrated into themes, and the third round of coding was selective, where connections between themes were compared and refined to build the grounded theory. The lead researcher (TW) conducted the initial coding, and the themes identified were discussed with a second researcher (SAB). To further improve rigor and reliability [71,72], a third researcher (KM) was consulted, and minor discrepancies were resolved through discussion. Pseudonyms were applied to protect participants’ identities.

The quantitative and qualitative data were integrated within the selected feasibility criteria of reach, demand, acceptability, implementation, and potential outcomes [36] to provide a complete picture of the feasibility, acceptability, and potential impact of the intervention. This use of mixed methods enabled...
triangulation to strengthen the validity of the findings and complementarity to explore different facets of a phenomenon [73].

Results

Reach and Demand of the GO:SC Digital Training Intervention

Recruitment and Reach

Digitally excluded groups within social housing are known to be difficult to reach [20,26]. In total, 4485 social housing residents were offered the training either via email (n=4365, 97.32%) or phone call (n=120, 2.68%; Table 2). The total number of responses to the recruitment survey (conducted via phone and email) was 168. A much higher proportion of phone survey respondents (37/120, 30.8%) were interested in the training, compared with the proportion of email survey respondents who were interested (30/4365, 0.69%). Although the HA actively promoted the intervention via its various web-based and printed communication channels over several months, none of the residents responded to these advertisements.

Table 2. Demand for and uptake of the digital training in and demographic characteristics of each group.

| Residents with missing age data, n (%) | Residents aged ≥75 y, n (%) | Residents aged 65 to 74 y, n (%) | Residents aged 55 to 64 y, n (%) | Residents aged 45 to 54 y, n (%) | Residents aged 35 to 44 y, n (%) | Residents aged 25 to 34 y, n (%) | Residents aged 18 to 24 y, n (%) | Residents with missing gender data, n (%) | Total (n=168), n (%)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Not interested in the training</td>
<td>101 (60.1)</td>
<td>55 (54.5)</td>
<td>35 (34.7)</td>
<td>11 (10.9)</td>
<td>4 (4)</td>
<td>10 (9.9)</td>
<td>12 (11.9)</td>
<td>14 (13.9)</td>
<td>15 (14.9)</td>
</tr>
<tr>
<td>Interested in the training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>67 (39.9)</td>
</tr>
<tr>
<td>Interested in the training but did not complete the training</td>
<td>55 (82.1)</td>
<td>31 (56.4)</td>
<td>21 (38.2)</td>
<td>3 (5.5)</td>
<td>0 (0)</td>
<td>3 (5.5)</td>
<td>5 (9.1)</td>
<td>8 (14.5)</td>
<td>12 (21.8)</td>
</tr>
<tr>
<td>Interested in the training and completed the training</td>
<td>12 (17.9)</td>
<td>9 (75)</td>
<td>3 (25)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>5 (41.7)</td>
<td>5 (41.7)</td>
</tr>
</tbody>
</table>

aThe percentage values in columns 3 to 13 were calculated using the corresponding n value in column 2.
bThe percentage value was calculated with 67 as the denominator.

Overall, 39.9% (67/168) of the recruitment survey respondents were interested in potentially undertaking the intervention. Among those initially contacted via email and phone, older women residents and those contacted by the HA were more interested in the intervention (probit model; Table 3). None of the interested participants who responded to the email survey completed the training. Our results affirm the substantial challenge of reaching and recruiting individuals interested in foundational digital training using conventional communication channels.

Phone calls were the most successful means for recruiting older and more digitally excluded groups. This success can be attributed to their familiarity as a channel for communication and the person-centered approach they enable [44]. Phone conversations also facilitated a further understanding of personal circumstances (eg, social, financial, or health needs) and desired outcomes (eg, connecting with friends or family or accessing health services), making it easier to adapt the conversations to include relevant training benefits. Furthermore, a previous study [74] found that women tend to be more receptive to recruitment contact via phone, which may explain the gender bias found.
Table 3. Probit model analysis results, that is, factors associated with the initial interest in the digital training intervention (n=168).\(^a\)

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Description of variables</th>
<th>Coefficient</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contacted by HA(^b) (dummy)</td>
<td>If contacted by HA</td>
<td>0.863</td>
<td>.001</td>
</tr>
<tr>
<td>Disable (dummy)</td>
<td>Respondent is disabled=1</td>
<td>0.228</td>
<td>.55</td>
</tr>
<tr>
<td>Age55+ (dummy)</td>
<td>Age &gt; 55 years</td>
<td>0.828</td>
<td>.003</td>
</tr>
<tr>
<td>Disable Age55+</td>
<td>_c</td>
<td>0.221</td>
<td>.66</td>
</tr>
<tr>
<td>Women (dummy)</td>
<td>Respondent is women=1</td>
<td>0.105</td>
<td>.63</td>
</tr>
<tr>
<td>Cornish (dummy)</td>
<td>Respondent’s ethnic group is Cornish=1</td>
<td>0.009</td>
<td>.98</td>
</tr>
<tr>
<td>English (dummy)</td>
<td>Respondent’s ethnic group is English=1</td>
<td>0.135</td>
<td>.72</td>
</tr>
<tr>
<td>British (dummy)</td>
<td>Respondent’s ethnic group is British=1</td>
<td>0.159</td>
<td>.64</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>−1.314</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>−101.080</td>
<td>—</td>
</tr>
<tr>
<td>Number of respondents</td>
<td></td>
<td>168</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)Dependent dummy variable: interested in participating.
\(^b\)HA: housing association.
\(^c\)Not available.

**Digital Training and Reach**

Of the 37 potential participants who indicated interest via phone, 12 (32%) completed the training. Of these 12 participants, 9 (75%) were women. Only respondents aged ≥55 years undertook the training, of whom 42% (5/12) were aged 55 to 64 years, 42% (5/12) were aged 65 to 74 years, and 17% (2/12) were aged ≥75 years (Table 2). Overall, 50% (6/12) of the participants reported a disability, 42% (5/12) of the participants reported no disability, and 8% (1/12) of the participants were missing data on disability. The majorit of the participants racially identified as White (9/12, 75%) and culturally identified as British (5/12, 42%) or British Cornish (4/12, 33%). The demographic, disability, racial identity, and cultural identity profiles of participants who completed the training were proportionally similar to those who did not complete the training.

The small sample size recruited for digital training is consistent with the small sample sizes recruited to other feasibility studies on interventions for digitally excluded populations. For example, Barbosa Neves et al [39] recruited only 12 participants in residential care to a feasibility study concerning the use of digital communication for social connectedness. Nonetheless, given their research objectives focused on uncovering feasibility, this sample size was appropriate and useful [36,75]. Similarly, we argue that our findings are useful for interpreting the feasibility of this digital intervention among social housing residents.

**Training Demand**

The most salient factors influencing the demand for the intervention were digital competency, preference for nondigital communication, and social networks. These multifaceted factors underscore the intricate dynamics surrounding digital training interventions and highlight the need for tailored strategies to address diverse participant needs and circumstances.

Among those interested in undertaking the intervention, demand was highest for training on video calling, primarily using Zoom; setting up and using devices, primarily a tablet; and improving digital skills, knowledge, and confidence in general, with most participants (37/67, 55%) who indicated interest in the recruitment survey noting multiple training needs across the 3 areas. In line with the initial scoping study conducted by Buckingham et al [28], it was a lack of competency in using digital devices, rather than device ownership and internet connectivity, that hindered digital technology use: “I am not able to use it [device] properly” (James, a man resident aged 55 to 64 years) and “I am interested in video calling, I have a smartphone, laptop, tablet, and mini-iPad, but I struggle to connect to the internet” (Mary, a woman resident aged 55 to 64 years).

The primary reason people were not interested in accessing the training was because they were already competent in using web-based video calling and messaging tools (57/101, 56.4%). However, those who were already competent were supportive of the intervention, reporting a need for digital training in general. Further reasons for the lack of demand included preferences for nondigital communication technology (“I prefer to just pick the phone up and call people” [Thomas, a man resident of unknown age]) and a feeling that the training was not personally necessary (“No, don’t think I need to learn things like this at my age, manage just fine thanks” [Deborah, a woman resident aged 65 to 74 years]).

Social networks played a key role in demand among participants who replied to the initial recruitment survey about the intervention. The lack of a digitally engaged social network was commonplace among those contacted, as was a small social network in general: “I don’t know anyone who I would call, I only have my sister and she doesn’t use internet stuff” (John, a man resident aged 55 to 64 years). Unsurprisingly, living farther away from family and friends was a key reason for engaging with digital communication technology: “My family live a distance away, so keeping in touch is important” (Sharon, a woman resident aged 86 years). However, for us, a key finding
was that people remain happy to rely on their broader support group, particularly younger family members. Having family and friends who could be relied on for help was a disincentive to undertake the training: “No help necessary, grandchildren able to help” (Rebecca, a woman resident aged 65 to 74 years).

Regarding the overall reach of the intervention, although initial recruitment calls found a moderate level of demand for this training, this interest did not translate into the same level of reach, with only 18% (12/67) of the interested participants completing the training. A key reported factor contributing to the lack of reach was health status, either personal health status (12/67, 18%) or the health status of family members (9/67, 13%); however, long-term disability (23/67, 34%) or impairments also affected the capacity to use technology and communicate, particularly among participants with visual, speech, or hearing difficulties (4/67, 6%). Caring responsibilities also acted as an impediment to uptake, particularly among older women:

I am interested [in the training], but my head is just full of decisions about my husband who will be in a nursing home for the rest of his life, I need to focus on what is important to me now. [Martha, a woman resident aged 55 to 64 years]

**Implementation of the GO:SC Digital Training Intervention**

The main implementation finding relates to the reduction in the number of participants between those who expressed initial interest in the training and those who completed the training. The study found that key facilitators included the HA’s established relationship with a digitally susceptible population and the flexible, informal approach of the DIT. Barriers involved issues with written consent, internet access, and device functionality. Here, we discuss each of these in turn.

The tangible and intangible positive roles that the HA played as an intermediary [76,77] in the implementation of the intervention cannot be overstated. In the United Kingdom, the role of HAs has evolved over recent decades to include supporting the health and well-being of their residents [78]. As a result, many HAs have built meaningful, trust-based relationships with their residents, and this factor played an important role in this project. From a practical perspective, this meant that the collaboration of Smartline with the HA enabled a wider reach in advertising the training offer. For example, the HA was able to use its customer relationship management system with phone numbers and email addresses to contact over 4000 residents. Indeed, examining the socioeconomic factors associated with initial interest in participating in the digital training program revealed that respondents who were contacted by the HA were more likely to be interested in the training program (P<.001; Table 3).

Another factor important for reach was customer liaisons by the HA with known susceptible residents who they felt would be interested in and would benefit from the training. Although recruitment remained difficult and the uptake of the intervention was low, the implementation of the recruitment survey and associated data collection would not have been possible without the help of the HA.

Participant dropout at the recruitment stage negatively affected implementation; only a small proportion of those who expressed an initial interest went on to complete the training. Among those (n=67) who were initially interested in the training, many (n=37, 55%) required multiple callbacks (up to 4) before they could be reached again; for reasons discussed next, many (55/67, 82%) dropped out. Making multiple callbacks was an administratively complex task that required many hours and email exchanges within the recruitment team. The need for written consent was a key factor for recruitment dropout, as was the need for a working digital device and an internet connection. Completing and returning the necessary consent forms was found to be a particular challenge for some participants. Participants noted that “filling in forms is a worry” (Julia, a woman resident aged 65 to 74 years) and that making time for the task was difficult: “I have the forms but have a lot on at the moment” (Jess, a woman resident aged >75 years). The possession of an internet-enabled device and internet connection was a requirement for participation in the training. Although we found high rates of digital technology ownership and internet connection possession among the participants contacted, 17.3% (29/168) of the respondents to the recruitment survey did not have a tablet, 5.9% (10/168) of the residents did not have a smartphone, and 7.1% (12/168) of the residents did not have an internet connection. Financial constraints, in particular, were a key reason for the lack of internet:

If I could get internet I would be interested [in the training], but not at the moment due to affordability issues. [Thomas, a man resident of unknown age]

I live in poverty so am scared I will incur charges using the tablet and smartphone. [James, a man resident aged 55 to 64 years]

For other interested participants, the working condition of the device was also a barrier to participation in the training:

I need to learn how to adjust settings so my old PC can cope. [David, a man resident aged 55 to 64 years]

I am not sure if my phone is smart enough. [Grace, a woman resident aged 55 to 64 years]

In addition, it is important to reflect on the implementation of the training itself. In practice, 3 different digital inclusion officers delivered the telephone-based training to 12 participants. The calls lasted between 5 minutes and 2.5 hours. To fit around participants’ caring responsibilities and day-to-day lives, all participants required multiple calls to arrange and rearrange training times. Training sessions lasted as long as required for the participant to learn to use the digital application, and follow-up support calls (up to 5) with the digital inclusion officer were arranged to ensure that the training objectives were achieved (Table 4). Successful delivery of the training required flexibility and persistence from the DIT. This indicates that there is a wide range of digital needs that are best served by informal and one-to-one support directed by individual needs [13,46].

Table 4. Overview of participants’ training objectives and associated outcomes.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Training objectives</th>
<th>Digital training needs</th>
<th>DIT training calls, n</th>
<th>Outputs</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary, a woman resident aged 55 to 64 years who reported having a disability</td>
<td>To be able to video call daughters</td>
<td>Needed help connecting a device to the internet and to learn how to video call</td>
<td>1</td>
<td>Received training to connect device to the internet and video calling</td>
<td>Uses video calls to talk to her daughters at dinnertime every week</td>
</tr>
<tr>
<td>Julia, a woman resident aged 65 to 74 years who reported having a disability</td>
<td>To order medicine for disabled son, send flowers to family members, and get inspiration for arts and craft projects</td>
<td>Needed to learn how to use emails, use web-based prescriptions, make web-based purchases, perform an internet search, and browse a website</td>
<td>5</td>
<td>Received training on web-based form-filling and purchasing, and internet searching.</td>
<td>Ordered flowers for family members and ordered medication for her son</td>
</tr>
<tr>
<td>Jade, a woman resident aged 55 to 64 years who reported having a disability</td>
<td>To be able to video call</td>
<td>Needed to learn how to unmute the computer microphone and how to video call</td>
<td>1</td>
<td>Received video call training</td>
<td>Increased competence in video calling</td>
</tr>
<tr>
<td>Susan, a woman resident aged 55 to 64 years who reported having a disability</td>
<td>To be safe and secure on the web, digitally record and mix music, and improve digital competency</td>
<td>Needed to learn about web-based safety and security and about different music recording software products, help purchasing music recording software, help setting up music recording software, and help connecting guitar mic to the computer and recording software</td>
<td>4</td>
<td>Received training to purchase music recording software, set up the music recording software, and connected guitar mic to the computer and audio recording software</td>
<td>Increased competence in recording and mixing guitar playing audio</td>
</tr>
<tr>
<td>Grace, a woman resident aged 55 to 64 years</td>
<td>To be able to video call daughter</td>
<td>Needed to purchase a device through which a video call could be made, help sorting passwords to activate internet connection, help connecting a device to the internet, and to learn how to video call</td>
<td>1</td>
<td>Received training to fix password issue, activate internet connection and video calling</td>
<td>Video calls her daughter</td>
</tr>
<tr>
<td>Rosie, a woman resident aged 65 to 74 years</td>
<td>To record and mix audio from a singing group</td>
<td>Needed to learn about different music recording software products and learn how to use software to merge several singing voices together</td>
<td>1</td>
<td>Received training on using software to mix multiple audio tracks</td>
<td>Mixes tracks together for a barbershop singing group</td>
</tr>
<tr>
<td>Tracy, a woman resident aged 55 to 64 years</td>
<td>To be digitally competent</td>
<td>Needed help completing the Learn My Way digital training course</td>
<td>3</td>
<td>Supported to complete the Learn My Way digital training course</td>
<td>Increased digital competence</td>
</tr>
<tr>
<td>Jess, a woman resident aged &gt;75 years</td>
<td>To better manage digital communication administration for a church group</td>
<td>Needed a refresher on using Zoom, to learn how to attach pictures to emails and save emails to folders, and to learn how to rearrange text and line gaps in Microsoft Word</td>
<td>3</td>
<td>Received training on attaching pictures to emails, saving emails to folders, and using Microsoft Word</td>
<td>Increased ability to manage digital communication administration for the church group</td>
</tr>
<tr>
<td>Daniel, a man resident aged 65-74 years</td>
<td>To be able to video call mother and print envelopes</td>
<td>Needed help setting up video call software and guidance on printing</td>
<td>1</td>
<td>Received video call training and guidance on printing</td>
<td>Video calls to mother and prints envelopes</td>
</tr>
<tr>
<td>Paul, a man resident aged 65 to 74 years</td>
<td>To be able to video call family and friends</td>
<td>Needed help setting up video call software</td>
<td>1</td>
<td>Received video call training</td>
<td>Ability to video call</td>
</tr>
<tr>
<td>Holly, a woman aged &gt;75 years</td>
<td>To be able to keep in touch with family</td>
<td>Needed help connecting a device to the internet, help sorting intermittent internet problems, help navigating device settings to complete software update, and to learn how to video call</td>
<td>2</td>
<td>Received training on connect a device to the internet and sort intermittent internet problems</td>
<td>Unresolved internet connection issue</td>
</tr>
</tbody>
</table>
Acceptability of the GO:SC Digital Training Intervention

Similar to the key findings on implementation, the study’s key findings on acceptability revolve around the substantial dropout rate, highlighting the challenges of translating training interest to training participation. Participants dropped out because of competing priorities, including health issues, caregiving responsibilities, and time constraints. However, all the 12 participants who started the training completed it; this suggests high acceptability of the intervention itself.

As with all research conducted during this period, the COVID-19 pandemic impacted all aspects of the intervention. We found the lockdown to have both a positive and negative impact on acceptability. For some, it was a driver for learning how to use video calls to access health services, social groups and classes, and resident groups, which had transitioned to function on the web:

I need help to connect to video appointments with the health professionals helping me. [Katy, a woman resident aged 25 to 34 years]
I would really like to join the online Coastline meetings [HA residents’ group] but don’t know how to use Zoom. It’s a priority for me to get online now. [Lily, a woman resident aged >75 years]

The lockdown also had a negative impact on the acceptability of the training intervention, with potential participants noting a need to attend to everyday tasks and self-care, rather than learning new skills: “I have been unwell with COVID for months, I just need to focus on the day-to-day things at the moment” (Judith, a woman resident aged 45 to 54 years). Importantly, we found that the pandemic compounded and increased several preexisting barriers to undertaking digital training for our participants. For example, we found that preexisting health conditions arose as a key barrier to participation:

I’m having a few health issues at the moment, call back in a few months. [Linda, woman resident aged 65 to 74 years]

Participant’s lives were generally “busy” (Diana, a woman resident aged 65 to 74 years), and they did “not have time for this at the moment” (Ruth, a woman resident aged 55 to 64 years) or needed to focus on caring responsibilities. Regarding the acceptability of the training vis-à-vis the everyday lives of our participants, 2 further themes emerged: the method used to deliver the intervention and the timetabling of the training sessions. Understanding that the Smartline participants had a strong preference for peer-based, face-to-face activities [27], participants’ concerns about the effectiveness of support being delivered over the phone were expected:

I am apprehensive about if a phone conversation would be enough to get online, would prefer one-to-one and face-to-face. I learn best by doing. [Michael, a man resident aged 65 to 74 years]
I need baby steps with the learning as I am not confident with technology, AKA a technophobe. [Gill, a woman resident aged 65 to 74 years]

The timing of the training was an important factor for participants, particularly for those who were working or had chronic health issues: “I have ME [myalgic encephalomyelitis or chronic fatigue], so afternoon is better for me” (Susan, a woman resident aged 55 to 64 years). Therefore, despite the increased need for digital communication at this time [79], the already complex social and health needs of the Smartline participants [80] meant that the recruitment and retention of interested participants was time intensive and a major challenge for the acceptability of the intervention.

Outcomes of the GO:SC Digital Training Intervention

Of the 12 people who participated in the training, 9 (75%) completed the baseline and follow-up well-being and digital competency surveys (Multimedia Appendix 2). Table 5 provides the summary well-being scores for participants who undertook the digital training intervention; there were no changes in the mean general, emotional, or social well-being between baseline and follow-up.

Counter to the overall aim of the intervention, we found a small reduction in behavioral well-being for participants who had undertaken the intervention [53]. The behavioral well-being measures included a question asking whether participants were attending courses and a further question on whether respondents were learning a new skill. As such, respondents would have indicated “yes” to these questions at baseline; however, at follow-up, participants were unlikely to be undergoing any other training given the ongoing COVID-19 restrictions. As such, this decrease could be explained by a confounding reduction in training levels from before to after the intervention, rather than an actual decrease in behavioral well-being.

Table 6 provides the summary digital competency scores of the intervention participants (n=9) at baseline and follow-up. From
Table 6, it can be inferred that there were no clear changes over time in any of these scores.

Table 5. Summary baseline and follow-up well-being scores for participants who undertook the digital training intervention (n=9).a

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Intervention, mean (SD)</th>
<th>Baseline</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>General well-being</td>
<td>5.44 (2.36)</td>
<td>5.50 (2.51)</td>
<td></td>
</tr>
<tr>
<td>Emotional well-being</td>
<td>6.43 (1.78)</td>
<td>6.38 (1.96)</td>
<td></td>
</tr>
<tr>
<td>Behavioral well-being</td>
<td>6.85 (1.83)</td>
<td>5.70 (2.26)</td>
<td></td>
</tr>
<tr>
<td>Social well-being</td>
<td>6.75 (3.68)</td>
<td>6.78 (3.85)</td>
<td></td>
</tr>
</tbody>
</table>

aHigher scores indicate higher well-being in each domain. The range was 0 to 10 for all domains.

Table 6. Summary digital competency scores for participants who undertook the digital training intervention (n=9).

<table>
<thead>
<tr>
<th>Digital module question</th>
<th>Intervention, median (IQR)</th>
<th>Baseline</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video calling and messaging questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of useb,c</td>
<td>5 (4-6)</td>
<td>5 (2-6)</td>
<td></td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>3 (3-4)</td>
<td>3 (3-4)</td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>4 (3-4)</td>
<td>3 (3-4)</td>
<td></td>
</tr>
<tr>
<td>Perceived reliability</td>
<td>3 (3-4)</td>
<td>3 (3-4)</td>
<td></td>
</tr>
<tr>
<td>Intentions to use</td>
<td>4 (4-4)</td>
<td>4 (3-5)</td>
<td></td>
</tr>
<tr>
<td>Autonomy</td>
<td>2 (2-4)</td>
<td>3 (3-3)</td>
<td></td>
</tr>
<tr>
<td>Feeling close to others</td>
<td>4 (2-4)</td>
<td>3 (3-4)</td>
<td></td>
</tr>
<tr>
<td>Desire to use technology, as friends are using it</td>
<td>3 (2-4)</td>
<td>3 (3-4)</td>
<td></td>
</tr>
<tr>
<td>Desire to use technology, as family is using it</td>
<td>4 (4-4)</td>
<td>4 (3-4)</td>
<td></td>
</tr>
<tr>
<td>Other people think that I should use it</td>
<td>4 (3-4)</td>
<td>3 (3-4)</td>
<td></td>
</tr>
<tr>
<td>General technology questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General technology self-efficacy</td>
<td>4 (2-4)</td>
<td>3 (2-3)</td>
<td></td>
</tr>
<tr>
<td>Enjoyment of using technology</td>
<td>3 (2-4)</td>
<td>3 (2-3)</td>
<td></td>
</tr>
<tr>
<td>Self-rated ability to use the internetb</td>
<td>3 (3-3)</td>
<td>4 (3-4)d</td>
<td></td>
</tr>
<tr>
<td>Perception that the internet makes life easier</td>
<td>3 (3-4)</td>
<td>4 (3-4)</td>
<td></td>
</tr>
<tr>
<td>Self-rated ability to use smartphonesb</td>
<td>3 (3-3)</td>
<td>3 (3-4)f</td>
<td></td>
</tr>
<tr>
<td>Frequency of searching online for health informationb,c</td>
<td>2 (1-2)</td>
<td>2 (1-2)</td>
<td></td>
</tr>
<tr>
<td>There are people I can talk to online if feeling lonely</td>
<td>2 (2-4)</td>
<td>2 (1-3)</td>
<td></td>
</tr>
</tbody>
</table>

aHigher scores indicate higher perceived competence, greater frequency of use, and more positive attitudes toward technology.
bReverse-coded responses were used for these questions.
cScores range from 1 to 6 for the frequency of use questions; scores range from 1 to 5 for all other questions.
dn=7 (data are missing; therefore, the number of responses is indicated).
eThe number of responses is indicated.
fThe number of responses is indicated.

Although conclusions on potential efficacy based on the survey data are limited owing to the small sample size, the qualitative data indicated that participants had achieved their original training objectives. Table 4 provides a summary of the training objectives, training needs, and the level to which they were met.

From Table 4, it can be inferred that increased competency in using video calling apps, particularly help with the installation and setting up of these apps, were key training needs. Participants’ training needs were motivated by both social and personal goals, such as contacting family members and becoming more competent with digital technology in general. Although advertised as training on video calling and messaging, participants received a diverse range of training, from training on how to order prescriptions on the web to training on more...
complex tasks such as recording and mixing music using web-based platforms. The flexibility to deliver such diverse and tailored training was not initially planned as part of the intervention yet proved a successful strategy. Overall, from the qualitative data, we found that participants achieved their training objectives and social and personal goals and have the potential to use other digital technologies in the future.

Discussion

Principal Findings

The findings indicate a demand for basic digital training among social housing residents, and the intervention was acceptable for those who received it. However, recruitment and implementation were challenging, with potential participants experiencing barriers that reflected the sociodigital inequalities commonly found in this group [10,26-28]. Barriers were strongly influenced by socioeconomic and health circumstances, which were closely related to preexisting digital readiness (eg, preexisting skills, confidence, motivation, and access) and further compounded by the COVID-19 pandemic. Our results confirm that the factors known to increase digital exclusion are particularly high among social housing populations [29,30] and highlight the interplay between traditional inequalities (ie, social, economic, and health) and digital inequalities [7]. However, social and personal goals were achieved by the participants who received the intervention. This demonstrates that tailored, flexible, and persistent training efforts can overcome barriers.

Implications for Policy and Recommendations for Practice

Regarding policy, the UK’s Digital Inclusion Strategy aims to “equip the whole country with the skills, motivation and trust to go on the internet, be digitally capable and to make the most of the internet” [81]. Although national policies are imperative to improve infrastructure, access, and digital connectivity [1,9,10], the implications of this study are that an effective policy also needs to focus on strategies for reaching digitally excluded groups [20-22]. The essential strategies and recommendations for practice derived from our findings are listed in Textbox 1.

Textbox 1. Essential strategies and recommendations for practice.

1. Partnerships with trusted intermediaries: forge partnerships with trusted local intermediaries, such as housing associations (HAs), community organizations, councils, and public libraries [76,77]. Prioritize intermediaries with established relationships and direct contact with the target group for effective reach.
2. Personalized recruitment approaches: use personalized recruitment methods, such as personal phone calls or face-to-face conversations. Understand individuals’ social, economic, health, and digital circumstances and align training benefits with their specific goals [7].
3. Minimize administrative barriers: reduce administrative burdens by minimizing form-filling processes, which negatively impact recruitment efforts. Be mindful of research monitoring procedures that may affect recruitment numbers, aiming for a streamlined approach.
4. Tailored and agile training programs: offer a flexible combination of device provision and internet access tailored to individual needs. Implement agile, person-centered training programs that adapt to participants’ personal goals and requirements.
5. Resource allocation and recognition: recognize the resource-intensive nature of initiatives targeting digitally excluded groups. Advocate for a stronger recognition of the efforts and resources required by intermediaries in national digital inclusion policies.

By implementing these strategies, policy makers, organizations, and communities can address sociodigital inequalities. However, in making these recommendations, we recognize that this places a considerable burden on individuals delivering such interventions. Future feasibility research of this nature could investigate the burden on intervention deliverers and the associated economic costs, which were not examined here.

Limitations

A strength of this study is its focus on social housing residents, an understudied population with associated socioeconomic inequalities that can present particular barriers to digital technology use. Another strength of this study is the collection of quantitative and qualitative data from various sources to enable rich insights into feasibility, acceptability, and potential impact, including the capture of data on those who initially expressed an interest in participating but did not go on to receive the intervention.

A limitation of the study is the small sample size for quantitative evaluation; however, as this is a feasibility study, quantitative outcomes (well-being and digital competency) were only intended to be indicative of the potential impact and were supplemented by qualitative findings on the achievement of training objectives. Owing to the unexpected difficulties in recruiting participants to the intervention, follow-up interviews were not possible within the time frame of the project.

We measured psychological constructs with individual items to balance theory alignment with reducing participant burden. However, we suggest that future feasibility and acceptability studies use established multitem measures to assess such constructs. The final limitations to note are those with regard to intervention delivery fidelity and economic costs. The study did not assess how the DIT delivered the intervention, other than following the “standard operating procedure.” It is possible that variations occurred in training delivery with regard to relationships with participants. Future studies should consider structured approaches to assessing intervention fidelity [82]. Finally, the intervention’s cost could not be specified, as it was provided through a county council’s DIT. Despite their personalized nature, similar personalized digital training programs are common in UK councils [14]. Therefore, the results of this study are valuable for providers facing challenges in engaging specific resident groups.
Conclusions
This study contributes to the contemporary literature, theory of “sociodigital inequalities” [7], and need to redefine digital inequalities in terms of their relation to other forms of socioeconomic and health inequalities. To address sociodigital inequalities, this study highlights that future policies need to be more proactive in reaching excluded groups, and such initiatives need to be considerate of people’s everyday lives, which will be conditioned by social and health circumstances. To achieve this, initiatives need to be appropriately resourced and include a flexible combination of digital provision with an agile person-centered approach to training based on personal needs and goals.

Acknowledgments
The authors would like to thank Karen Spooner (Volunteer Cornwall), Phil Gilbert (Coastline Housing), and Adrian Ankers (Coastline Housing) for their help and support with recruitment. They thank the University of Exeter’s Smartline team, namely Belinda Broughton, Chloe Bines, and Claire Wilcox, for support in managing survey administration. They thank the Cornwall Council’s Digital Inclusion Team, namely Dawn Stoddern (Digital Inclusion Team lead), Simon Gooding (digital inclusion officer), Kym Ley (digital inclusion officer), and Peter Finlay (digital inclusion officer), for delivering the digital training. They thank the rest of the Smartline team, particularly Emma Bland, for their feedback on the manuscript. They are grateful to all of the Smartline participants who gave their time to be involved in this study.

The Smartline project (05R16P00305) received £3,740,920 (US $4,752,989) and the Smartline Extension project (05R18P02819) received up to £3,307,703 (US $4,203,172) of funding from the England European Regional Development Fund as part of the European Structural and Investment Funds Growth Programme 2014-2020. The Ministry of Housing, Communities and Local Government (and in London, the intermediate body Greater London Authority) is the Managing Authority for the European Regional Development Fund. Established by the European Union, the European Regional Development Fund helps local areas stimulate their economic development by investing in projects that support innovation, businesses, job creation, and local community regeneration. The study received additional funding of £25,000 (US $31,768) from the Southwest Academic Health Science Network and £200,000 (US $254,162) from the Cornwall Council.

Data Availability
The anonymized quantitative data sets analyzed during this study are available from the University of Exeter Open Research Exeter repository. Qualitative data cannot be shared because they contain sensitive participant information, identifying information, and information whose dissemination would violate the agreement to which the participants consented.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Recruitment survey.
[DOCX File, 16 KB - formative_v8i1e45506_app1.docx ]

Multimedia Appendix 2
Well-being and digital competency survey.
[PDF File (Adobe PDF File), 512 KB - formative_v8i1e45506_app2.pdf ]

Multimedia Appendix 3
Information sheet.
[PDF File (Adobe PDF File), 1751 KB - formative_v8i1e45506_app3.pdf ]

Multimedia Appendix 4
Consent form.
[PDF File (Adobe PDF File), 349 KB - formative_v8i1e45506_app4.pdf ]

Multimedia Appendix 5
Course content guidance.
[DOCX File, 285 KB - formative_v8i1e45506_app5.docx ]

Multimedia Appendix 6
Task-specific leaflet example.
[DOCX File , 1353 KB - formative_v8i1e45506_app6.docx ]

Multimedia Appendix 7
Theoretical basis of digital module.
[DOCX File , 24 KB - formative_v8i1e45506_app7.docx ]

References
12. UK Digital Eagles: Helping you get the most out of being online. Barclays. URL: https://www.barclays.co.uk/digital-confidence/eagles/ [accessed 2022-04-16]


33. Smartline homepage. Smartline. URL: https://www.smartline.org.uk/ [accessed 2021-09-09]


65. Walker et al. JMIR FORMATIVE RESEARCH 2024 | vol. 8 | e45506 | p.483 https://formative.jmir.org/2024/1/e45506

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DIT: Digital Inclusion Team
GO:SC: Getting Online: Staying Connected
HA: housing association
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance
TIDieR: Template for Intervention Description and Replication

Edited by A Mavragani; submitted 04.01.23; peer-reviewed by H Bao, J Mira; comments to author 26.07.23; revised version received 01.11.23; accepted 20.11.23; published 26.01.24.

Please cite as:
Walker T, Buckingham SA, Poole R, Elliott LR, Menneer T, Tu G, Morrissey K
Telephone-Based Training Intervention for Using Digital Communication Technologies for Social Housing Residents During the COVID-19 Pandemic: Mixed Methods Feasibility and Acceptability Evaluation
JMIR Form Res 2024;8:e45506
URL: https://formative.jmir.org/2024/1/e45506
doi:10.2196/45506
PMID:38277209

©Tim Walker, Sarah Ann Buckingham, Ria Poole, Lewis Roland Elliott, Tamaryn Menneer, Gengyang Tu, Karyn Morrissey. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Acceptance of a Web-Based Intervention in Individuals Who Committed Sexual Offenses Against Children: Cross-Sectional Study

Sonja Schröder 1, MSc; Claudia Buntrock 2, PhD; Louisa Neumann 3, MSc; Jürgen L Müller 1, MD; Peter Fromberger 1, PhD

1 Clinic for Psychiatry and Psychotherapy – Forensic Psychiatry, University Medical Center Göttingen, Göttingen, Germany
2 Institute of Social Medicine and Health Systems Research, Otto von Guericke University Magdeburg, Magdeburg, Germany
3 Clinic for Forensic Psychiatry and Psychotherapy, KRH Psychiatry Wunstorf, Wunstorf, Germany

Corresponding Author:
Sonja Schröder, MSc
Clinic for Psychiatry and Psychotherapy – Forensic Psychiatry
University Medical Center Göttingen
Rosdorfer Weg 70
Göttingen, 37081
Germany
Phone: 49 5514022114
Email: sonja.schroeder@med.uni-goettingen.de

Abstract

Background: Individuals who have committed sexual offenses against children often have difficulties finding treatment, despite its potential effectiveness. Although the development of web-based interventions could enhance therapeutic supply, up to now the acceptance thereof among this target group is unknown.

Objective: For the first time, this study assesses the acceptance of a web-based intervention among individuals who committed sexual offenses against children and analyzes variables that predict acceptance. Following the Unified Theory of Acceptance and Use of Technology (UTAUT), it is assumed that acceptance of web-based interventions in individuals who have committed sexual offenses against children follows the same mechanisms as for individuals in general psychiatry.

Methods: This cross-sectional study is based on the data from an ongoing clinical trial (@myTabu) evaluating the effectiveness of a web-based intervention in individuals who committed sexual offenses against children (N=113). Acceptance level was measured using a questionnaire based on the UTAUT and modified for the target group. Furthermore, predictors of acceptance from the UTAUT (performance expectancy, effort expectancy, and social influence [SI]), attitudes toward web-based interventions, and internet anxiety were assessed at baseline.

Results: Most participants (61.1%, 69/113), reported high acceptance, while 36.3% (41/113) of them indicated moderate acceptance, and 2.7% (3/113) of them expressed low acceptance. In a linear regression model, the predictors explained 41.2% of the variance (F 11,101 = 9.055; P = .01). Attitudes toward web-based interventions (B = 0.398, 95% CI 0.16-0.64; P = .001) and SI (B = 0.183, 95% CI 0.03-0.38; P = .04) significantly predicted acceptance. Post hoc explorative analysis showed that the participants’ belief that people close to them would recommend the use of a web-based intervention is a predictor of acceptance. In contrast, the belief that their community supervisor would recommend the use thereof was not predictive in this respect.

Conclusions: For the participants of this study, we identified high acceptance of web-based interventions for the majority of participants. SI and the participants’ attitudes toward web-based interventions were important in predicting acceptance.

Trial Registration: German Clinical Trial Registration (DRKS, Deutsches Register Klinischer Studien) DRKS 00021256; https://drks.de/search/de/trial/DRKS00021256

(JMIR Form Res 2024;8:e48880) doi:10.2196/48880

KEYWORDS

mHealth; web-based intervention; acceptance; Unified Theory of Acceptance and Use of Technology; UTAUT; sexual offenses against children; child abuse; child pornography; children; sexual offense; cross-sectional study; community; anxiety; psychiatry
Introduction

Background

Sexual abuse during childhood has disruptive short and long-term effects for children who are victims of such an offense [1,2] and the treatment of individuals who committed sexual offenses against children should be a major part of efforts to reduce the risk of recidivism. Despite findings that therapy can reduce the risk of recidivism [3], many individuals who committed sexual offenses against children struggle to find a therapist. Therapists often express a low willingness to work with individuals who are convicted of a sexual offense—especially with those who have a pedophilic disorder [4]. The result, at least in Germany, is that only limited therapeutic treatment is available [5]. Web-based interventions represent a possible enhancement in the therapeutic supply [6].

Web-based interventions can be advantageous in comparison to face-to-face (f2f) therapy for the users, as they can be anonymous, flexible in time and space, and can be cost-effective [6,7]. Anonymity could be especially advantageous, as individuals who committed sexual offenses against children can feel ashamed and guilty which may hinder the willingness to find a therapist. To date, only a few web-based interventions exist for individuals who have committed sexual offenses against children and the majority of them have not yet been evaluated [8]. In a placebo-controlled trial, Lüth et al [9], showed for the first time that a guided web-based intervention for individuals who consume child exploitation material can reduce the amount of time thus spent. In addition, the study showed that, as is the case in web-based interventions in general psychiatry [10], many persons who signed up for a web-based intervention did not participate by logging in or completing the therapeutic content [9]. Also, in f2f therapy for individuals who have committed offenses, roughly one-third of individuals do not complete therapy [11,12]. Up to now the variables that predict why and for how long individuals who have committed sexual offenses against children use web-based interventions are unknown. In general, a factor that is expected to predict whether someone uses web-based interventions in general psychiatry is acceptance [13,14]. Thus, this predictor might also be important in the treatment of individuals who committed sexual offenses against children.

To study acceptance and its predictors, research on web-based interventions for general mental health often uses the Unified Theory of Acceptance and Use of Technology (UTAUT) as a theoretical framework [15,16]. The UTAUT states that the use of a technology can be predicted by acceptance. Acceptance is thereby defined as the behavioral intention to use a technology. Further, 4 core predictors are assumed, which are performance expectancy (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FC). PE is related to whether or not the person believes that the web-based intervention can help him or her; EE is related to the perceived ease of use of the web-based intervention; SI is the perception of whether people close to him or her would recommend the use of the web-based intervention; and FC is related to the belief that there is an organizational and technical infrastructure that would help him or her in case of problems with the web-based intervention. According to the UTAUT, FC together with acceptance predict the use of technology. The other 3 variables, PE, EE, and SI, predict acceptance.

Although the UTAUT was first conceptualized to explain the use of technology in organizational settings, it has been generalized to many different fields including the use of technology for treatment in general psychiatry [16]. Philippi et al [16] conducted a secondary analysis in which they integrated the data of 1588 participants from 10 UTAUT studies. The original studies analyzed the participants’ acceptance and its predictors based on the UTAUT for web-based interventions, for example, for treating depression, chronic pain, or aftercare for inpatients. In the study by Philippi et al [16], the basic structure of the UTAUT with PE, EE, and SI predicting acceptance was replicated. PE was found to be the strongest predictor, in accordance with results from prior studies [15,17,18].

Gender, age, degree of voluntary use of technology, and experience with the technology were included next to predictors in the UTAUT as moderators [15]. The authors showed that the effect of PE was stronger for younger and male individuals; the effects of EE was stronger for older, females, and less experienced individuals; and the effect of SI was stronger for older, females, and less experienced individuals as well as under conditions of mandatory use [15]. In web-based interventions in general psychiatry, however, Philippi et al [16] could not replicate a moderating effect of age, gender, or experience on any predictor. Next to moderating effects, a direct effect on acceptance of participant age was analyzed. In some studies on web-based interventions in general psychiatry, it was found that a lower age predicted higher acceptance [19-21] whereas other studies found no effect [16,22].

In the field of web-based interventions in general psychiatry, the variables attitudes toward web-based interventions and internet anxiety were also integrated into the UTAUT to predict acceptance. Attitudes refers to the evaluative judgment of a web-based intervention, which can be expressed in attributes ranging for example, from pleasant to unpleasant or likable to dislikable [20,23]. Internet anxiety is the fear, distrust, or apprehension that is experienced when using the internet [16,24]. Attitudes and computer anxiety were removed from the final UTAUT model because the explorative power of the variable was captured by EE [15]. In recent studies, however, attitudes was found to be a strong predictor for acceptance [20,25,26]. Similarly, internet anxiety studies have shown that persons with lower internet anxiety have a higher acceptance for web-based interventions in general psychiatry [16,22,26].

Objective

The goal of this study is to address the following research questions for individuals who committed sexual offenses against children, either by contact or noncontact offense (ie, child sexual exploitation material offenses): (1) how high is the acceptance of web-based interventions? (2) Which variables predict acceptance of web-based interventions?
As shown above, no data exist for the specific target group of this study. Therefore, we assume that acceptance of web-based interventions for individuals who have committed sexual offenses against children follows the same mechanisms as for individuals who use web-based interventions in general psychiatry (Figure 1). As a consequence, it is expected that higher scores in PE, EE, SI, attitudes toward web-based interventions, and lower scores in internet anxiety predict higher scores in acceptance. In addition, we will examine whether age has a moderating and direct effect on acceptance.

**Figure 1.** Conceptual study model with the UTAUT predictors [15] and additional variables as well as age as moderator. UTAUT: Unified Theory of Acceptance and Use of Technology. *Age as a moderator variable.

**Methods**

**Study Design and Data Collection**

This cross-sectional study used data collected between March 1, 2021, and March 1, 2022, of an ongoing clinical trial to evaluate the effectiveness of the web-based intervention @myTabu [27,28]. Participants were individuals convicted of child abuse, of child sexual exploitation material use, or of both under the German Penal Code and were under community supervision. Further eligibility requirements were adulthood (18 years of age or older), a community supervision period of at least 6 months at study inclusion, internet access, no severe acute psychiatric disorder, no severe cerebro-organic disorder, and no severe cognitive impairment. For the recruiting process, research staff informed community supervisors of the clinical trial and asked them to inform eligible clients. When an eligible client was interested in the clinical trial, he or she was informed about the study by research staff in a personal interview. During the recruitment period, 118 interviews were conducted and 113 individuals agreed on taking part in the study.

**Measures**

**Sociodemographic and Criminological Data**

For each participant, 1 research staff member (out of a total of 3 research staff members) collected sociodemographic and criminological data using a standardized data collection form. The written court judgment and records of the Federal Central Criminal Register were used as the primary source of information. If information was missing from these documents, corresponding information was obtained from participants. The modified Static-99, which is a version of the original Static-99 that omits victim-related variables, was assessed [29]. The Static-99 includes variables that have been found to be predictive of sexual reoffending among individuals who have previously committed a sexual offense. A higher score represents a higher risk [30]. Scores of the modified Static-99 range from 0 to 9. Information on the additional data that were coded during that process can be found in the study protocol of the @myTabu clinical trial [28].

**Acceptance and Predictors**

To measure acceptance and its predictors, the German adaptations of the UTAUT questionnaire by Baumeister et al...
and Apolinário-Hagen et al [25] were used. These adaptations were based on the well-established UTAUT questionnaire [15] and the adaptations to the field of general psychiatry [17,32-34]. For this study, the questionnaires were modified to the context of a web-based intervention for individuals who committed sexual offenses against children based on face validity (Textbox 1, see Multimedia Appendix 1 for original German questionnaire).

**Textbox 1.** Items of the questionnaire for acceptance of technology with references to original studies; the sections that have been adapted for this study are italicized.

**Questionnaire description:** Please read the following questions carefully and answer as spontaneously as possible. The following questions refer to a therapeutically guided program, which you can complete online and which supports you during your probation to avoid recidivism and to lead a crime-free life. The program consists of sessions that are unlocked weekly. In the questions, this program is called “online program.”

**Acceptance**

1. I can imagine trying an online program [31].
2. I can imagine using an online program regularly [...] [31].
3. I would recommend an online program to a friend [31].
4. I would be willing to pay for an online program [31].

**Performance expectancy**

1. Using an online program would help me not to commit a further child abuse or to consume child sexual exploitation material [31].
2. Using an online program would improve my ability to live a crime-free life [31].
3. Overall, an online program would help me during my community supervision [31].

**Effort expectancy**

1. Using an online program would be simple [31].
2. Using an online program would be an easy task for me [31].
3. An online program would be clear and easily comprehensible to me [31].

**Social influence**

1. People close to me would recommend me to use an online program [31].
2. My community supervisor would recommend me to use an online program [31].

**Attitudes toward web-based interventions**

1. Using the online program is a good idea [25].
2. Using the online program would be interesting [25].
3. Using the online program could be fun [25].
4. I would like to work with the online program [25].

**Internet anxiety**

1. The internet is something threatening to me [31].
2. I am afraid making an irrevocable mistake while using the internet [31].

According to the UTAUT, acceptance was operationalized as behavioral intention and was measured using 4 items. The UTAUT predictors PE and EE were measured using 3 items each and SI was measured using 2 items. Attitudes toward web-based interventions and internet anxiety were measured using 4 and 2 items, respectively. Responses were made on a 5-point Likert scale ranging from 1 (does not apply at all) to 5 (applies absolutely). McDonald ω total [35,36] were 0.59 for acceptance, 0.80 for PE, 0.81 for EE, and 0.83 for attitudes toward using web-based interventions, showing good reliability for PE, EE, and attitudes toward using web-based interventions and a poor reliability for acceptance [37]. For scales with 2 items, Spearman-Brown coefficient was calculated [38]. Spearman-Brown coefficient of SI was 0.21 and of internet anxiety was 0.65, showing a questionable reliability of internet anxiety and an unacceptable reliability for SI [37].

In addition to the above named scales, scales were measured on FC [31], planning behavior [39], and study compensation for hypotheses that were not part of this study.

**Statistical Analyses**

**Research Questions 1 and 2: Acceptance and its Predictors**

Data analysis was performed using the software R (version 4.2.1; R Core Team) [40]. The mean acceptance score was...
calculated and its distribution was assessed. The acceptance mean score was categorized as low (1-2.34), moderate (2.35-3.67), and high (3.68-5), in accordance with previous studies [41,42].

To test for predictors of acceptance, a multiple linear regression with acceptance as the criterion was conducted. The variables PE, EE, SI, attitudes toward web-based interventions, internet anxiety, and age were included along with a moderation of age on all variables (age × variable). The predetermined α level was .05. Variables were included simultaneously based on correlation. For meaningful interpretation of the coefficients of the first-order terms in the presence of interactions, we mean-centered the variables prior to computation [43]. For missing items responses, the mean across the available items of each scale was calculated. There were missing items for 5 participants with a maximum of 6 missing items (mean 2.2, SD 2.17). There was no missing scale, as every participant answered at least 1 item on every scale [44]. To test for outliers, Cook distance, leverage value, and studentized deleted residuals were calculated. After correcting for coding errors, there were 23 participants who were considered outliers by the above named criteria. To test the model assumptions, we looked at linear relationships between the variables and acceptance, normality of residuals [45], homoscedasticity [46], and multicollinearity [47-49].

There were signs of nonnormality of residuals; the other analyses showed no assumption violation. Because of the outliers and the nonnormality of residuals, a bootstrap procedure was used with the number of bootstrap samples of 1000. By using bootstrapping, results are less sensitive to extreme values and thus no participant had to be excluded from the analysis [49,50].

Explorative Analysis

Because of the low internal consistency of SI, a multiple linear regression was conducted with acceptance as the criterion and the items of the SI scale as factors with the lowest value as reference. In addition, the predictors PE, EE, internet anxiety, attitudes toward web-based interventions, and age were included. For the SI item (asking whether the community supervisor recommends the use of a web-based intervention), values 1 and 2 were too infrequent for a statistical analysis and were thus combined into 1 category with 3. Because of missing values on SI items, 2 participants were excluded from the analysis. There were 9 outliers according to Cook distance, leverage value, and studentized deleted residuals. There were signs of nonnormality of residuals [45]; the other analyses showed no assumption violation. Therefore, a bootstrap procedure with the number of bootstrap samples of 1000 was used [49].

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki, was approved by the medical ethical board of the Human Medical Center Göttingen, Göttingen, Germany (16/2/20), and was preregistered on AsPredicted (107090). During study enrollment, informed consent was obtained from all participants. In the informed consent, participants agreed on the study conditions and data protection and processing. Study data were saved and deidentified by using pseudonyms for each participant. During participation, identification of each individual was only possible by the respective community supervisor. After participation, identification lists were stored separately from the study data in paper form in a safe. Individuals received monetary compensation for their participation; the compensation level was dependent on the number of sessions completed in the web-based intervention. A maximum of €120 (US $131.06; €1 is approximately US $1.2 at the start of the clinical trial) could be obtained.

Results

Demographic and Criminological Data

All 113 participants were male and had a median age of 38 years with a range of 20-72 (mean 40.67, SD 12.75 years). The participants had on average 1.25 previous convictions (SD 2.47). For 57.1% (64/112; 1 missing) of the participants, the present conviction was their first. For their present conviction, 38.9% (44/113) of the participants were convicted for sexual abuse of children (German Penal Code section 176 in the version in effect before July 01, 2021), 14.3% (16/113) for aggravated sexual abuse of children (German Penal Code section 176a in the version in effect before July 01, 2021), and 74.3% (84/113) for dissemination, procurement, and possession of child pornographic content (German Penal Code section 184b). Note that 28 participants had more than 1 present conviction. The mean score for the modified Static-99 was 1.87 (SD 1.19; range 0-6).

Descriptive Data of Acceptance and Predictors

The mean (SD) acceptance level in this study was 3.78 (SD 0.66). The distribution of acceptance is negatively skewed with 2.7% (3/113) of the participants indicating low, 36.3% (41/113) moderate, and 61.1% (69/113) high acceptance. The mean score of PE was 4.08 (SD 0.77), of EE was 4.10 (SD 0.67), of SI was 3.88 (SD 0.81), of attitudes toward web-based interventions was 4.15 (SD 0.63), and of internet anxiety was 2.02 (SD 0.93).

Prediction of Acceptance

According to the F test ($F_{11,101}=9.055$), the variables in the regression model explained 41.2% of the variance of acceptance ($R^2=0.412$; $P<.001$; Table 1). With a regression coefficient of $B=0.398$ (95% CI 0.16-0.64; $P=.01$) for attitudes toward web-based interventions and $B=0.184$ (95% CI 0.03-0.38; $P=.04$) for SI, there were significant linear effects of both variables on acceptance. The other variables did not predict acceptance above the effects of attitudes toward web-based interventions and SI (all $P>.05$). There was no moderating effect of age on any variables (all $P>.05$).
Table 1. Regression results using bootstrapping with acceptance as the criterion (N=113).

<table>
<thead>
<tr>
<th>Variables</th>
<th>B (SE)</th>
<th>95% CI</th>
<th>T</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.77 (0.056)</td>
<td>3.64 to 3.87</td>
<td>67.129</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>PE</td>
<td>0.03 (0.093)</td>
<td>–0.15 to 0.23</td>
<td>0.332</td>
<td>.74</td>
</tr>
<tr>
<td>EE</td>
<td>0.09 (0.104)</td>
<td>–0.13 to 0.29</td>
<td>0.882</td>
<td>.38</td>
</tr>
<tr>
<td>SI</td>
<td>0.18 (0.087)</td>
<td>0.03 to 0.38</td>
<td>2.109</td>
<td>.04</td>
</tr>
<tr>
<td>Attitudes</td>
<td>0.40 (0.118)</td>
<td>0.16 to 0.64</td>
<td>3.372</td>
<td>.01</td>
</tr>
<tr>
<td>Anxiety</td>
<td>–0.03 (0.068)</td>
<td>–0.19 to 0.09</td>
<td>–0.462</td>
<td>.64</td>
</tr>
<tr>
<td>Age</td>
<td>0.01 (0.005)</td>
<td>0.01 to 0.01</td>
<td>0.016</td>
<td>.90</td>
</tr>
<tr>
<td>PE × age</td>
<td>–0.01 (0.008)</td>
<td>–0.03 to 0.00</td>
<td>–1.543</td>
<td>.13</td>
</tr>
<tr>
<td>EE × age</td>
<td>–0.01 (0.009)</td>
<td>–0.03 to 0.00</td>
<td>–1.435</td>
<td>.15</td>
</tr>
<tr>
<td>Attitudes × age</td>
<td>–0.003 (0.010)</td>
<td>–0.02 to 0.01</td>
<td>–0.340</td>
<td>.73</td>
</tr>
<tr>
<td>SI × age</td>
<td>0.01 (0.007)</td>
<td>–0.001 to 0.03</td>
<td>1.630</td>
<td>.11</td>
</tr>
<tr>
<td>Anxiety × age</td>
<td>–0.01 (0.006)</td>
<td>–0.02 to 0.00</td>
<td>–1.099</td>
<td>.27</td>
</tr>
</tbody>
</table>

Explorative Analysis

The mean (SD) score for the item of the SI scale asking whether people close to the participant recommend the use of a web-based intervention was 3.24 (SD 1.20). The mean (SD) score for the item of the SI scale asking whether the community supervisor recommends the use of a web-based intervention was 4.49 (SD 0.80).

The variables in the regression model explained 35.95% of the variance of acceptance (F₁₁.₉₉ = 6.273; R² = 0.3595; P < 0.001).

Attitudes toward web-based intervention (B = 0.331; 95% CI 0.09-0.55; P = 0.01) and the perceived opinion of people close to the participant significantly predicted acceptance. Participants who indicated a score of 5 (B = 0.455; 95% CI 0.07-0.96; P = 0.04) or 4 (B = 0.102; 95% CI 0.17-0.89; P = 0.01) had significantly higher acceptance than participants who indicated a score of 1. There was no significant effect for participants who indicated a score of 3 (B = 0.340; 95% CI 0.01-0.72; P = 0.07) or 2 (B = 0.182; 95% CI –0.31 to 0.60; P = 0.42) in comparison to participants who indicated a score of 1. PE (B = 0.108; 95% CI –0.08 to 0.28; P = 2.4), EE (B = 0.106; 95% CI –0.12 to 0.29; P = 0.30), internet anxiety (B = –0.06; 95% CI –0.22 to 0.09; P = 0.46), age (B = 0.006; 95% CI –0.005 to 0.01; P = 0.22), and perceived opinion of the community supervisor, when scored 4 (B = –0.05; 95% CI –0.47 to 0.25; P = 0.79) or 5 (B = 0.094; 95% CI –0.18 to 0.42; P = 0.55) in comparison to lower or equal to 3, did not significantly predict acceptance.

Discussion

Principal Results

This study examined for the first time the acceptance of web-based interventions and variables predicting it among individuals who committed sexual offenses against children. For the majority of participants, the acceptance of web-based interventions was high. Persons with higher scores in SI and attitudes toward web-based interventions showed significantly higher acceptance. In contrast to expectations, the other predictors of the UTAUT, PE and EE, as well as internet anxiety and age did not predict acceptance. An explorative analysis of the 2 items comprising the SI scale revealed that the belief that people close to the participant would recommend the use of a web-based intervention predicts acceptance but the same is not true for the belief that the community supervisors would recommend the use thereof.

Comparison With Prior Work

In comparison to prior work from general psychiatry, the average acceptance was higher in this study with a smaller variance. In the secondary analysis from Philippi et al [16], in which results from 10 original studies were included, the mean acceptance for male participants was low to moderate (mean 2.68, SD 1.12). One explanation for the high acceptance scores in this study may be that it is difficult for individuals who committed sexual offenses against children to find f2f therapy [5]. Another explanation for the divergent acceptance levels may be differences in the sample selection. A common recruiting method to contact specific target groups in the studies incorporated in the secondary analysis by Philippi et al [16] was...
to contact patients directly in clinics [31,32,41], for example by recruiting in the waiting rooms [14]. To protect the identity of clients, recruitment in this study involved collaboration with community supervisors. Thus, the research staff did not contact eligible clients directly but instead asked the community supervisors to inform eligible clients of the study. Through this approach, it is likely that some clients were never informed about the study because they were considered unsuitable by the community supervisor. In addition, clients who declined after first being informed by the community supervisor never met with research staff. Further, the participants in this study all agreed to participate in an evaluation study for a web-based intervention. The aim of most of the studies incorporated in the secondary analysis by Philippi et al [16] was to test acceptance-facilitating interventions and participants were not given access to a web-based intervention. Thus, it is likely that the results are based on a selected group of individuals in community supervision, which may not be representative of individuals who committed sexual offenses against children in general. This explanation is in line with the study by Lin et al [22], who recruited participants by sending invitations to individuals who had earlier expressed interest in participating in an evaluation study on a web-based intervention and (thus) assessed comparatively high acceptance (mean 3.44, SD 0.89; values were divided by 4 to match the scale used in this study). Despite this potential bias, the results of this study show that there exists a group of individuals who committed sexual offenses against children in community supervision that has high acceptance of web-based interventions.

Participants in this study rated web-based interventions as more helpful (PE), easier to use (EE), more enjoyable (attitudes toward web-based interventions), and perceived that their social surroundings would recommend the use of web-based interventions more (SI) than did participants in studies of web-based interventions in general psychiatry. All of these predictors are positively correlated with acceptance [16,25]. As mentioned above, this positive view of web-based interventions can be partly explained by the selection of the sample. In comparison, internet anxiety, which has a negative correlation with acceptance, was found to be slightly lower in studies from general psychiatry [16]. For individuals who were not convicted for a crime using the internet, this result could be understood when considering that internet anxiety is negatively correlated with internet use [51] and convicted individuals often lack the skills and resources to use the internet or specific technologies [52]. Although the internet anxiety levels in this study cannot be considered as high, a lack of experience with the internet could be a more important issue for individuals who committed sexual offenses against children compared to individuals who have not been convicted of a crime.

The proportion of explained variance of 41.2% in the regression model can be considered as high according to the Cohen criteria [53]. However, this proportion is lower than in other UTAUT studies, where UTAUT predictors explained for example 57% to 63% of the variance of acceptance [21,54,55]. This could mean that, for individuals who have committed a crime, further predictors are relevant that have to be investigated in order to fully understand the acceptance of web-based interventions. In this study, to test our hypotheses, only selected variables that were replicated in previous studies on web-based interventions in general psychiatry were examined for their prediction on acceptance [16,20,25]. In studies on web-based interventions in general psychiatry, further variables that have been investigated include, among others, perceived reliability [56] and perceived privacy risks [54]. Next to these variables, those that predict the treatment motivation for f2f therapy in individuals who committed sexual offenses against children, for example, antisocial personality disorder, might also be relevant for web-based interventions [57]. In addition, web-based interventions are becoming more common in general mental health care [58] and are increasingly being developed for individuals who committed sexual offenses against children [8]. Therefore, it is likely that an increasing number of individuals have some experience with web-based interventions which could have a direct or moderating effect on acceptance [15]. These and other variables could be important when explaining the variance of acceptance of web-based interventions in individuals who committed sexual offenses against children.

In previous studies, it has been repeatedly shown that the original UTAUT predictors PE, EE, and SI are predictive of acceptance [16,59]. In this study, against our expectations, the predictive effect of PE and EE could not be replicated for individuals who had committed sexual crimes against children. In contrast, SI and attitudes toward web-based interventions were significant predictors of acceptance. Attitudes toward web-based interventions was also found to be a strong predictor of acceptance in other subject groups and was equally as strong as PE [20,25]. The importance of attitudes for acceptance may be related to the fact that the participants in this study most likely had no specific knowledge or experience with web-based interventions at the time they answered the UTAUT items. In this state of indecision, positive attitudes might be more important than cognitive beliefs about the web-based intervention. That could be a reason why the hypothesis that PE and EE are predictive for acceptance was refuted in this study [25].

The significant prediction by SI of acceptance could be explained by the fact that participants are influenced by how other people, especially their community supervisors, evaluate their community supervision time. However, in the exploratory analysis, it was found that the perceived higher opinion of people close to the participant but not the perceived lower or higher opinion of the community supervisor significantly predicted higher acceptance. A reason why the perceived opinion of the community supervisor was not predictive could be that most participants rated the opinion of their community supervisor as high. This may be because community supervisors who did not support web-based interventions may have not informed their clients. Thus, the results of this explorative analysis could imply that especially in a situation where the community supervisors support a web-based intervention, the opinion of people close to the participant predicts acceptance.

In previous studies, it was found that lower age predicted higher acceptance [19-21] and that the effect of PE, EE, and SI was moderated by age [15]. In this study, however, no direct or moderating effect of age could be observed in individuals who
committed sexual offenses against children. This is in line with studies by Philippi et al [16] and Lin et al [22], who also could not replicate a direct or moderating effect of age.

**Limitations**

The first limitation arises from the sample selection. As mentioned above, participants were preselected by community supervisors and the participants were persons who already agreed to take part in a web-based intervention study. Because of that, it is not clear if and how representative the sample is of individuals who committed sexual offenses against children and who are presently in community supervision and thus how generalizable the results of this study are.

The second limitation could have resulted from the preselection. The variances in this study are low, which could be an indicator that the sample variance is lower than the actual population variance. Because of that restriction of variance, the statistical power to detect interactions is reduced [43].

The third limitation is that the scales acceptance, SI and internet anxiety show low reliability. For this study, we used the well-established UTAUT questionnaire [15] and adaptations used in the field of general psychiatry [17,32-34]. The questionnaire for this study was based as closely as possible on this format. However, some aspects of general psychiatry may not be transferable to the context of this study. For example, the acceptance scale includes an item asking whether participants would recommend a web-based intervention to a friend. For individuals convicted of a crime, shame and the need to hide the conviction from those close to them could be relevant aspects that might influence the answer to this item [60].

The fourth limitation is that the questionnaire was completed in the presence of the research staff. Therefore, the participants might have answered in a socially desirable manner, for example, to appear cooperative toward the study.

**Future Directions and Clinical Implications**

Future research should examine the predictive power of further variables that go beyond the UTAUT model. Variables that are possibly relevant are described in the previous section (eg, perceived reliability, antisocial personality disorder, and experience with web-based interventions). To increase acceptance, it should be tested whether acceptance-facilitating interventions, that highlight the positive aspects of using a web-based intervention (attitudes toward web-based interventions) and address reasons why the potential users assume that people close to them may not be in favor of them (SI) are especially effective. To that end, it should be investigated whether there are differences in acceptance depending on the characteristics of the potential users (eg, conviction type and the number of previous convictions). By doing that, acceptance-facilitating intervention could be tailored to the specific needs of the potential participants and may be more effective [31]. Further, research should look at the actual use of web-based interventions and test whether acceptance, as hypothesized by UTAUT, can predict factors like satisfaction or need fulfillment [61] and the actual use of a web-based intervention.

**Conclusions**

This study is the first to analyze the acceptance of web-based intervention in individuals who committed sexual offenses against children. In this study the acceptance levels of the majority of participants were high. The perceived opinion of the social contacts, as well as, the attitudes toward web-based interventions was important in predicting acceptance. To increase acceptance, it may be important to incorporate these predictors when designing acceptance-facilitating interventions.
Conflicts of Interest
None declared.

Multimedia Appendix 1
German questionnaire items used in this study with references to original studies.

References


27. Bauer L, Schröder S, Tozdan S, Müller JL, Fromberger P. @myTabu—konzept einer Therapeuten-gestützten online-intervention für verurteilte Personen, die kindesmissbrauch begangen oder Missbrauchsbildungen konsumiert haben [@myTabu—Concept of a guided web-based intervention for convicted individuals who have committed sexual offenses against children or who have consumed child sexual exploitation material]. Bewährungshilfe. 2021. URL: https://www.dbh-online.de/sites/default/files/bewbwi/bewi_1-2021_inhaltsverz.pdf


Abbreviations

EE: effort expectancy
FC: facilitating conditions
f2f: face-to-face
PE: performance expectancy
SI: social influence
UTAUT: Unified Theory of Acceptance and Use of Technology
Development and Implementation of Digital Diagnostic Algorithms for Neonatal Units in Zimbabwe and Malawi: Development and Usability Study

Hannah Gannon1,2, MBChB, MRCPCH; Leyla Larsson3, MSc; Simbarashe Chimhuya4, MD; Marcia Mangiza5, MBChB; Emma Wilson1, PhD; Erin Kesler6, MSN, CRNP; Gwendoline Chimhini4, MBChB, MPH; Felicity Fitzgerald2,7, MBBS, PhD; Gloria Zailani8, MBBS; Caroline Crehan1, MBChB; Nushrat Khan1, PhD; Tim Hull-Bailey1, MPhil; Yali Sassoon9, MPhil; Morris Baradza10, BA; Michelle Heys11, MBBS, MD; Msandeni Chiume8*, MBBS, MSc

1Population, Policy and Practice, Institute of Child Health, University College London, London, United Kingdom
2Biomedical Research and Training Institute, Harare, Zimbabwe
3Institute of Computational Biology, Computational Health Centre, Helmholtz, Munich, Germany
4Department of Child, Adolescent and Women’s Health, Faculty of Medicine and Health Science, University of Zimbabwe, Harare, Zimbabwe
5Sally Mugabe Central Hospital, Harare, Zimbabwe
6Children’s Hospital of Philadelphia, Philidephia, PA, United States
7Department of Infectious Disease, Imperial College London, London, United Kingdom
8Kamuzu Central Hospital, Lilongwe, Malawi
9Snowplow Analytics, London, United Kingdom
10Baobab Web Services, City of Cape Town, South Africa
* these authors contributed equally

Corresponding Author: Hannah Gannon, MBChB, MRCPCH
Population, Policy and Practice
Institute of Child Health
University College London
30 Guildford Street
London, WC1N 1EH
United Kingdom
Phone: 44 (0) 20 7905 ext 2600
Email: h.gannon@ucl.ac.uk

Abstract

Background: Despite an increase in hospital-based deliveries, neonatal mortality remains high in low-resource settings. Due to limited laboratory diagnostics, there is significant reliance on clinical findings to inform diagnoses. Accurate, evidence-based identification and management of neonatal conditions could improve outcomes by standardizing care. This could be achieved through digital clinical decision support (CDS) tools. Neotree is a digital, quality improvement platform that incorporates CDS, aiming to improve neonatal care in low-resource health care facilities. Before this study, first-phase CDS development included developing and implementing neonatal resuscitation algorithms, creating initial versions of CDS to address a range of neonatal conditions, and a Delphi study to review key algorithms.

Objective: This second-phase study aims to codevelop and implement neonatal digital CDS algorithms in Malawi and Zimbabwe.

Methods: Overall, 11 diagnosis-specific web-based workshops with Zimbabwean, Malawian, and UK neonatal experts were conducted (August 2021 to April 2022) encompassing the following: (1) review of available evidence, (2) review of country-specific guidelines (Essential Medicines List and Standard Treatment Guidelines for Zimbabwe and Care of the Infant and Newborn, Malawi), and (3) identification of uncertainties within the literature for future studies. After agreement of clinical content, the algorithms were programmed into a test script, tested with the respective hospital’s health care professionals (HCPs), and refined according to their feedback. Once finalized, the algorithms were programmed into the Neotree software and implemented at the tertiary-level implementation sites: Sally Mugabe Central Hospital in Zimbabwe and Kamuzu Central Hospital in Malawi, in
Introduction

Background

In 2021, overall, 2.3 million children died in their first month of life, and the proportion of deaths below the age of 5 years attributed to neonates has risen from 40% in 1990 to 47% in 2021 [1]. The main causes of neonatal death include intrapartum-related complications and prematurity. Accurate, evidence-based identification and management of neonatal conditions could improve outcomes by standardizing care. It has been demonstrated in the literature that poor-quality care delivered by undertrained and overstretched health care professionals (HCPs) in low-resource settings can be attributed to approximately 60% of deaths from treatable conditions [2]. Clinical decision support (CDS) tools aim to standardize the assessment process and provide evidence-based guidance at the bedside, improving HCPs' knowledge and competence. Previous studies have demonstrated that, owing to limited laboratory diagnostic capabilities, within these low-resource settings, there is significant reliance on clinical findings to make a diagnosis [2-4]. The implementation of diagnostic pathways could improve outcomes for this vulnerable neonatal population [4]. In addition, the development of management pathways adaptable and modifiable to resource availability is paramount [5]. There is potential for digital CDS tools to support HCPs to deliver improved neonatal care despite resource limitations and hence optimize outcomes.

The use of decision support tools has been well established and supported in health systems globally [6]. In this digital age, mobile phone coverage and digital uptake in low-resource settings has surged. Together with the low-dose, high-frequency education strategies [7,8] to improve newborn care, digital interventions appear to present a promising approach to improving newborn care [9,10].

Digital CDS tools are designed to aid directly in clinical decision-making and can use individualized patient data to guide management to implement evidence-based clinical guidelines at the bedside [11]. Paper-based guidelines are established resources, but these often get lost, are placed on shelves, and become easily outdated and therefore redundant. The systematic review of CDS systems by Kawamoto et al [12] identified four features for improved clinical practice:

1. Automatic provision of decision support as part of clinician workflow
2. Provision of recommendations rather than just assessments
3. Provision of decision support at the time and location of decision-making
4. Computer-based decision support

To date, most digital tools have implemented CDS as stand-alone tools to provide additional support to clinicians [9,13]. Consequently, the use of these tools can be limited as their use is voluntary and assumes clinician readiness and preparedness to adopt digital tools for which they may not have preexisting training or knowledge [14]. There is a growing evidence base for newborn digital interventions in low-resource settings ranging from patient level (NeMo tool trialed in Uganda, designed to support mothers in identifying newborns who are sick [15]) to system level (such as the World Health Organization standards-based, machine-readable, adaptive, requirements-based, and testable guidelines) for systematic pathways for the development and implementation of localized digital systems [16]. Digital interventions have the potential to both improve neonatal quality of care in low-resource settings and, in turn, improve neonatal outcomes, but there is a growing need for these interventions to be rigorously evaluated and to move from pilot or small-scale projects to scale [17]. Key elements to optimize the generalizability and scalability of such interventions are the vital importance of software sharing and the ability to have the intervention open source, which offers lower costs; greater transparency; faster iterations; and importantly, for the low-resource setting, local ownership in comparison with proprietary and, often siloed, digital interventions.
Neotree

Neotree was developed as an integrated, open-source, digital, quality improvement system for hospital-based sick and vulnerable newborns [18-20]. The evidence suggests that up to 70% of newborn deaths could be prevented by the implementation of evidence-based interventions [21]. The system guides HCPs through the admission and clinical care of newborns. It combines evidence-based clinical guidelines with real-time newborn data collection, data visualization, data export, and newborn education on a single platform [22]. This tablet-based digital system is for use at the hospital bedside by HCPs with a range of skills and competencies supporting the care and treatment of newborns. It is currently implemented in the neonatal units at Sally Mugabe Central Hospital (SMCH) and Chinhoyi Provincial Hospital in Zimbabwe and at Kamuzu Central Hospital (KCH) in Malawi. SMCH’s neonatal unit is a physician-led unit with mainly junior physicians (1-2 years after graduating) using Neotree and providing neonatal care, with consultant oversight; Chinhoyi Provincial Hospital and KCH are both nurse-led or midwife-led units with consultant oversight. Neotree offers a solution that integrates directly into everyday care for all admitted newborns. Crucially, the CDS functionality is not dependent on nurses or clinicians having any specialist knowledge about the care of the newborn. This integration and development process is consistent with the suggestions by Kawamoto et al [12]. An overview of the development and implementation experience so far can be found elsewhere [18,22]. To date, Neotree has aided in the care of approximately 35,000 newborns and the education and training of >1000 HCPs at the implementation sites in Malawi and Zimbabwe and has demonstrated high acceptability, feasibility, and usability with perceived and observed improvements in newborn care [18,22].

The CDS Function Development

The aim of the CDS function was to ultimately provide a personalized, integrated, evidence-based management plan (with embedded educational messaging around newborn care), tailored to the health care facility’s workflow and resource availability. In addition, a list of potential clinical problems was provided to facilitate pattern recognition and for education purposes. Evidence-based algorithms were precoded into Neotree management guidelines for the programmed conditions. Therefore, the entered data instantaneously provide a list of potential clinical problems and associated clinical management actions at the bedside.

The algorithms fell into four categories:

1. Simple condition-based decision trees based on good evidence (e.g., prematurity or low birth weight or resuscitation)
2. Simple conditional expressions based on weak evidence and best clinical judgment (e.g., jaundice)—these require additional refinement with evidence review and clinical consensus
3. Complex conditional expressions based on good evidence (e.g., Thompson score for neonatal encephalopathy)
4. Complex conditional expressions with weak evidence base (e.g., sepsis)—these require analytical refinement of the algorithm and testing

Phase 1

First-phase CDS development in Zomba Central Hospital, Malawi, included developing and implementing neonatal resuscitation and stabilization algorithms, conducting a Delphi study to review key algorithms (neonatal sepsis, hypoxic ischemic encephalopathy, respiratory distress of the newborn, and hypothermia) [23], and developing the initial versions of CDS to address a range of neonatal conditions [20]. The resuscitation and stabilization algorithms were configured in 2016 based on the World Health Organization guidelines and Helping Babies Breathe and have been activated throughout implementation, so that, for example, if a baby is not breathing, Neotree will advise how to resuscitate [20]. Clinical management support pages were also configured for several potential diagnoses. These appeared in Neotree at the end of the admission process according to the HCPs’ chosen diagnosis and included hypoxic ischemic encephalopathy or birth asphyxia, prematurity, neonatal sepsis, gastrochisis, and transient tachypnoea of the newborn. The development of these clinical management support pages continued through 2018 and 2019. Neotree uses a web-based editor function using “variable expressions” to describe the condition, which is coding depicted by a “$” and brackets (“’). The paper by Khan et al [22] describing the software development of Neotree describes the web editor function in detail; an example can be found in Multimedia Appendix 1.

Evaluating solutions such as CDS is important for both improvement and assessment of the effectiveness of the app [24]. For Neotree, evaluation and user feedback has been collected in three distinct ways: (1) on an ongoing basis, (2) during usability workshops, and (3) through questionnaire evaluation [20,25].

Phase 2

This study aimed to conduct the second phase of the algorithm development. In addition, we aimed to understand the user experience of the CDS functionality. As integrating such tools in routine care is not part of standard practice, we hoped to understand the strengths and weaknesses of these solutions and how we can optimize the implementation of such solutions in the future.

Methods

Overview

The second phase of CDS development was to integrate the clinical data entered during admission to trigger evidence-based diagnostic algorithms alongside personalized management guidelines, tailored to health care facility’s resource availability. On the basis of the clinical data entered by the HCP using Neotree, nested algorithms would determine whether a diagnosis should be triggered and, if so, would provide a suggested diagnosis, as shown in Figure 1.
First, it allows the HCPs to select their own diagnoses from a list of differentials. It then provides all the algorithmic diagnoses generated from clinical data (but not selected by the HCP) as “suggested diagnoses.” Phase-2 development of the CDS algorithms was planned in 3 stages (Figure 2).

Figure 1. Overview of the clinical decision support (CDS) pipeline in Neotree. HCP: health care professional.
Stage 1
Diagnosis-specific web-based workshops with Zimbabwean and Malawian neonatal experts encompassing the following:
1. Review of available evidence and refinement of existing algorithms
2. Review of country-specific guidelines
3. Identification of uncertainties within the literature for future studies

Stage 2
Algorithms programmed into the Neotree software and implemented at the tertiary-level study sites in two phases, in December 2021 and May 2022.
Study sites were the following:
1. Sally Mugabe Central Hospital, Zimbabwe
2. Kamuzu Central Hospital, Malawi

Stage 3
Usability was evaluated at Sally Mugabe Central Hospital through two usability workshops and usability questionnaires:
1. Post-Study System Usability Questionnaire
2. Systems Usability Scale
3. Variable expression (how the diagnosis is or will be coded within Neotree)
4. Clinical management advice (with reference to evidence)
5. Needs for further refinement (next steps required)

The algorithm triggers were clearly defined and agreed upon, and subsequent clinical management was based on local, national, and international clinical guidelines and adapted based on local resource availability. Textbox 1 shows an example of the discussions, and “$” denotes the variable being described. The national guidelines used were the Essential Medicines List and Standard Treatment Guidelines for Zimbabwe [26] and Care of the Infant and Newborn, Malawi.

Mock clinical test cases were created by the expert team and were used for active testing in Neotree to ensure that the algorithms were triggered in the correct circumstances and were intuitive to use, leading to further refinement. This testing was performed by Neotree team members and then a cohort of HCPs who had been trained and used Neotree in their everyday clinical practice.
Textbox 1. Example of stage-1 discussions—thermoregulation (“$” denotes the variable being described).

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>Measured or recorded</strong>—based on temperature recorded on admission</td>
</tr>
<tr>
<td>• <strong>Categorization</strong>—according to the World Health Organization guidelines, there are 5 possible categories for temperature on admission:</td>
</tr>
<tr>
<td>• Mild hypothermia: temperature 36°C-36.4°C</td>
</tr>
<tr>
<td>• Moderate hypothermia: 32°C-35.9°C</td>
</tr>
<tr>
<td>• Severe hypothermia: &lt;32°C</td>
</tr>
<tr>
<td>• Normothermia: 36.5°C-37.5°C</td>
</tr>
<tr>
<td>• Hyperthermia: &gt;37.5°C</td>
</tr>
<tr>
<td>• <strong>Variable expression</strong> ($ denotes the variable being described)</td>
</tr>
<tr>
<td>• Algorithm version 2: proposed from the workshop</td>
</tr>
<tr>
<td>• Mild hypothermia: $temperature &gt;35.9°C and $temperature &lt;36.5°C</td>
</tr>
<tr>
<td>• Moderate hypothermia: $temperature &gt;31.9°C and $temperature &lt;36°C</td>
</tr>
<tr>
<td>• Severe hypothermia: $temperature &lt;32°C</td>
</tr>
</tbody>
</table>

Stage 2
Following the finalization of the individual algorithm pathways, the CDS functionality was deployed in 2 cycles (December 2021 and May 2022). The first cycle included the first set of clinical diagnoses (update to resuscitation and stabilization, thermoregulation, convulsions, low birth weight, prematurity, hypoglycemia, HIV, and respiratory distress), and the second cycle comprised clinical diagnoses that required more extensive review and definition by the expert team (neonatal encephalopathy; sepsis; jaundice; and congenital abnormalities, which include cleft lip or palate, congenital dislocation of the hip, talipes, gastrochisis, omphalocele, and spina bifida). These were to be deployed only in the 2 central hospitals (SMCH in Zimbabwe and KCH in Malawi), as both centers have senior neonatal clinical oversight.

Stage 3
Usability evaluation was planned to include usability workshops and usability questionnaires.

The usability workshops were conducted to understand the user experience and to provide specific information about the steps and items that require improvement in CDS design and development. The workshops were conducted using a think-aloud approach [27]. The HCPs, familiar with Neotree, were provided with a specific medical scenario (designed by the expert team) and asked to complete a Neotree admission in real time. These workshops were conducted in a nonneonatal environment to allow users to focus on their experience and understanding of the app instead of having to focus on neonatal clinical care. The provided scenario simulated a typical admission. Interviews were conducted by the interviewer in English (the clinical working language of both Zimbabwe and Malawi). The interview was recorded and later transcribed by the interviewer. While the user thought out loud and provided feedback about the app, the interviewer took short notes to prompt further questioning, especially about the CDS functionality.

Usability questionnaires such as the System Usability Scale (SUS) and the Post-Study System Usability Questionnaire (PSSUQ) provide additional quantifiable insights and metrics into the user experience of CDS [28]. PSSUQ assesses 3 categories: system usefulness, information quality, and interface quality. These 3 domains cover 16 questions, rated on a Likert scale ranging from 1 (strongly agree) to 7 (strongly disagree). Participants could also indicate if the question was not applicable to them. The SUS is a 10-item Likert scale questionnaire with options ranging from 1 (strongly disagree) to 7 (strongly agree). The SUS score was compared with an internationally recognized measure of acceptable usability of 68 out of 100. Both survey questionnaires were chosen because of their popularity in assessing mobile health solutions in low-resource settings. For the quantitative measurements (SUS and PSSUQ), means and SDs were calculated using Stata (version 17.1).

Ethical Considerations
The Neotree study, including the wider implementation evaluation of Neotree and the CDS functionality development, received approval from SMCH Research ethics committee (reference HCHEC070618/58), University College London ethics committee (5019/004), Biomedical Research and Training Institute (AP148/18), Medical Research Council of Zimbabwe (MRCZ/A/2570), and Electronic Health Records Department of the Zimbabwe Ministry of Health and Child Care. The collection of qualitative data from HCPs using Neotree was performed with the approval of SMCH Research ethics committee, Biomedical Research and Training Institute, Zimbabwe Medical Research Council and the Ministry of Health and Child Care. The collection and use of data were approved by the SMCH Research ethics committee.

Written informed consent was obtained from all usability evaluation participants, and their results were deidentified and stored anonymously. Attendance at the workshops was compensated through reimbursement of travel costs and refreshments. Participants were reimbursed US $10 for their time and travel costs as per the Zimbabwean Medical Research Council policy.
Results

Algorithm Development

Overall, 11 diagnosis-specific web-based workshops with Zimbabwean and Malawian neonatal experts were conducted between August 2021 and April 2022, encompassing the following:

1. Review of available evidence
2. Review of country-specific guidelines (Essential Medicines List and Standard Treatment Guidelines for Zimbabwe and Care of the Infant and Newborn for Malawi)
3. Identification of uncertainties within the literature for future studies

Table 1 shows the diagnoses discussed and the phase they were introduced in.

Table 1. Diagnosis-specific algorithms and their introduction (cycle 1 and cycle 2).

<table>
<thead>
<tr>
<th>Data source and diagnosis</th>
<th>Cycle 1 introduction (December 2021)</th>
<th>Cycle 2 introduction (May 2022)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple conditional-based decision trees based on strong evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low birth weight</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Prematurity</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>HIV exposure</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Thermoregulation</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Hypoglycemia</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Simple conditional expressions based on weak evidence</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Jaundice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convulsions</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Complex conditional expressions based on strong evidence</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Neonatal encephalopathy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory distress</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Complex conditional expressions based on weak evidence (in low-resource settings)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sepsis</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Congenital abnormalities</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Detailed discussions and clinical triggers, documented in the A to E format (discussed previously), alongside the country-specific management plans and the coding used for Neotree can be found in Multimedia Appendix 2 [23,26,29-35]. Where national management guidelines differed, for example, birth weight thresholds for admissions to the neonatal unit, country-specific guidelines were created. Management guidelines were created based on resource and equipment availability in each setting. These international meetings also facilitated cross-site learning and quality improvement discussions. The clinical content and Neotree algorithm coding for each developed algorithm can be found in Multimedia Appendix 2.

The first-cycle algorithms developed, including low birth weight, convulsions, and hypoglycemia, had internationally defined descriptions and management guidelines, with minimal variation between nationally agreed management pathways. These diagnoses were relatively straightforward to discuss among the experts and gain unanimous agreement on the clinical content.

The more complex diagnoses proved to be more challenging and required multiple meetings with the experts to be clearly defined. As a result, neonatal encephalopathy and neonatal sepsis were assigned to ongoing international working groups for continuing research. The respiratory distress algorithm was split into term (defined as ≥37 weeks of gestation) respiratory distress and preterm (defined as <37 weeks of gestation) respiratory distress, and we identified the need for a more formal review of the literature within the global health context.

After agreement of clinical content, the algorithms were programmed into a test script, tested with the respective hospital’s HCPs, and refined according to their feedback. Figure 3 shows how the algorithm-derived diagnoses are presented within Neotree.
Usability Workshop Results

Overview

In total, 2 workshops were conducted in both Zimbabwe and Malawi after the release of each CDS—in January 2022 and in June 2022. An example of the recorded feedback is shown in Figure 4. In Zimbabwe, the workshops were attended by 5 HCPs—2 (40%) female participants and 3 (60%) male participants. In Malawi, the workshops were attended by 4 HCPs—3 (75%) female participants and 1 (25%) male participant, and feedback was recorded.

From the usability workshops that were conducted, three main themes emerged, two of which related directly to CDS, and the last one related to the Neotree tablet:

1. Education
2. Streamlining of processes
3. Technical challenges

Education

The main theme prevailing across interviews was the positive feedback toward the availability of information and clinical guidance provided through Neotree and CDS. As most of the clinicians using Neotree are junior physicians, especially at SMCH, they have limited experience in the management of neonates:

This might seem obvious but on your first day it is not. [Participant 2]

Streamlining of Processes

HCPs highlighted that CDS in Neotree was found to streamline and systematize clinical diagnosis and management at the hospital, resulting in better and more effective care. This was done by implementing diagnostic algorithms that provide...
data-driven results and by providing management for the clinical diagnoses that are based on the most up-to-date guidelines. By providing the management pages for each diagnosis (and a repertoire of all management pages) in a centralized location, Neotree provides streamlined care for newborns as clinicians do not resort to a multitude of different resources that may not agree in management:

*These are the nevirapine doses? You have outdone yourself. I like this very much because we learn about HIV guidelines but sometimes you just forget so this is nice.* [Participant 5]

Despite the ability of Neotree to provide data capture, CDS, and clinical management, this proved to extend the admission process. Clinicians highlighted the duration of admission completion to be a problem, especially in an already overstretched and understaffed hospital:

*Sometimes when you’re on call you don’t have time to go through it.* [Participant 2]

*Takes a lot of time when you have 6 babies screaming.* [Participant 4]

*Admitting a patient really does take 15-20 minutes.* [Participant 1]

### Technical Challenges

During the workshops, the tablets had issues with general degradation as they had just reached 3 years of use in the hospitals. These challenges resulted in the app crashing or not reacting to user input. The users, however, did not show frustration and mentioned that because it happens in the ward, they have grown used to it and have adapted to the challenges. Despite the clinicians adapting to these issues, this is severely affecting their experience of using the tablets and CDS, which may provide explanation for the relatively low scores in the SUS and PSSUQ questionnaires with respect to technical capability:

*Can we have better tablets.* [Participant 2]

*Just delete the ones that clicked themselves here.* [Participant 4]

*I guess I’ll wait for it.* [Participant 4]

### Usability Questionnaire Results

In this study, usability of the Neotree CDS was also assessed using the PSSUQ (8/9, 89%) and SUS (9/9, 100%) questionnaire at SMCH.

The same group of HCPs was asked to complete 2 questionnaires (SUS and PSSUQ). One person failed to complete the PSSUQ due to time constraints on the neonatal ward. The SUS results are shown in Table 2, and the PSSUQ results are shown in Table 3. SUS showed generally high usability of the CDS system in Neotree, especially relating to aspects such as desire to use the system (questions 1, 7, and 9) and ease of use (questions 3 and 7). In contrast, the CDS system scored low on questions relating to technical complexity and difficulties (questions 2, 4, 6, 8, and 10).

The PSSUQ showed results similar to those of the SUS, where the technical complexities of the system (question 8) was highlighted as an issue.

Overall, both usability scores (SUS mean score 73.8 out of 100 [higher score is better]; PSSUQ overall score 2.28 out of 7 [lower score is better]) demonstrated high usability of the CDS function (comparable with previous SUS scores assessing the usability of Neotree data capture and dashboard functionalities in Malawi [mean score 88.1 out of 100] [25]) but highlighted issues around technical complexity, which have been subsequently addressed.

### Table 2. System Usability Scale (SUS) results (N=9)

<table>
<thead>
<tr>
<th>Questions</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think I would like to use this system frequently</td>
<td>4 (1.4)</td>
</tr>
<tr>
<td>2. I found the system unnecessarily complex</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>3. I thought the system was easy to use</td>
<td>3.9 (1.4)</td>
</tr>
<tr>
<td>4. I think that I would need the support of a technical person to be able to use this system</td>
<td>2.1 (0.9)</td>
</tr>
<tr>
<td>5. I found the various functions in this system were well-integrated</td>
<td>3.7 (1.4)</td>
</tr>
<tr>
<td>6. I thought there was too much inconsistency in this system</td>
<td>1.9 (0.6)</td>
</tr>
<tr>
<td>7. I would imagine that most people would learn to use this system</td>
<td>3.9 (1.4)</td>
</tr>
<tr>
<td>8. I found the system very cumbersome to use</td>
<td>1.8 (1.1)</td>
</tr>
<tr>
<td>9. I felt very confident using the system</td>
<td>4.4 (1.1)</td>
</tr>
<tr>
<td>10. I needed to learn a lot of things before I could get going with this system</td>
<td>1.8 (0.8)</td>
</tr>
</tbody>
</table>

aQuestions were answered using a Likert scale ranging from 1-5, where 1=strongly disagree and 5=strongly agree.

bCalculated SUS score (total converted mean scores × 2.5)=73.8.
Table 3. Post-Study System Usability Questionnaire (PSSUQ) results (n=8)<sup>a,b</sup>.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. It was simple to use this system</td>
<td>1.9 (1.7)</td>
</tr>
<tr>
<td>2. I was able to complete the tasks and scenarios quickly using this system</td>
<td>1.9 (1.7)</td>
</tr>
<tr>
<td>3. I felt comfortable using this system</td>
<td>2.3 (1.6)</td>
</tr>
<tr>
<td>4. It was easy to learn to use this system</td>
<td>2.0 (1.7)</td>
</tr>
<tr>
<td>5. I believe I could become productive quickly using this system</td>
<td>1.8 (1.8)</td>
</tr>
<tr>
<td>6. The system gave error messages that clearly told me how to fix problems</td>
<td>2.1 (1.4)</td>
</tr>
<tr>
<td>7. Whenever I made a mistake using the system, I could recover easily and quickly</td>
<td>2.3 (0.9)</td>
</tr>
<tr>
<td>8. The information (such as online help, on-screen messages, and other documentation) provided with this system was clear</td>
<td>4.1 (2)</td>
</tr>
<tr>
<td>9. I needed to learn a lot of things before I could get going with this system</td>
<td>2.5 (1.6)</td>
</tr>
<tr>
<td>10. It was easy to find the information I needed</td>
<td>2.6 (1.5)</td>
</tr>
<tr>
<td>11. The information was effective in helping me complete the tasks and scenarios</td>
<td>2.1 (1.4)</td>
</tr>
<tr>
<td>12. The organisation of information on the system screens was clear</td>
<td>1.9 (1.7)</td>
</tr>
<tr>
<td>13. The interface of this system was pleasant</td>
<td>1.6 (0.5)</td>
</tr>
<tr>
<td>14. I liked using the interface of this system</td>
<td>1.6 (0.7)</td>
</tr>
<tr>
<td>15. This system has all the functions and capabilities I expect it to have</td>
<td>3.0 (1.3)</td>
</tr>
<tr>
<td>16. Overall, I am satisfied with this system</td>
<td>2.1 (1.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Questions were answered using a Likert scale ranging from 1-7, where 1=strongly agree and 7=strongly disagree.

<sup>b</sup>Calculated total PSSUQ score=2.28.

Discussion

Principal Findings

This study describes the successful development and pilot implementation of the only known neonatal CDS system incorporated within a bedside data capture system with the ability to deliver up-to-date management guidelines, tailored to local resource availability. Bucher et al [9], Mitchell et al [36], and Long et al [17], as examples, describe CDS digital tools generally designed for the pediatric population, not specifically neonatal, and without incorporated digital data capture. To date and to the best of the authors knowledge, there is no other neonatal digital intervention designed to improve neonatal care using these methods. As a significant proportion of the global mortality below the age of 5 years is attributed to neonatal deaths in low-resource settings, the development of sustainable interventions targeting this vulnerable group are vital. The incorporated real-time usability evaluation demonstrated high usability from HCPs actively using Neotree in their everyday practice, within 2 low-resource settings—Zimbabwe and Malawi. The adaptability of the Neotree system allowed for real-time changes to the app to improve the processing times and algorithm function. This enabled changing the surfacing format of the algorithm-selected diagnoses to improve the usability of the CDS function.

It has been well documented in the literature that improving adherence to evidence-based clinical guidelines can improve the quality of care and, in turn, patient outcomes. However, this has been challenging to effectively implement, particularly in low-resource settings [37,38]. A multifaceted approach is required, and a system with the ability to adapt is therefore essential, both to resource availability and guideline updates. With the plethora of digital health interventions implemented in low-resource settings, centralizing information for HCPs instead of relying on memory, wall charts, or printed guidelines has been demonstrated to improve adherence [36]. The target product profile developed by Pelle et al [37], for electronic CDS algorithms in low-resource settings, sets out specific guidance about the development of such interventions. The algorithms need to be based on evidenced clinical guidelines and be customizable to the context. This has been heavily incorporated into the design and methodology of the development of the Neotree CDS functionality.

Comparison can be made with studies within the pediatric population. Several studies have reviewed the feasibility of the implementation of electronic, user-friendly Integrated Management of Childhood Illness tools, for example, in both Tanzania and South Africa [36,39]. In the Tanzanian setting, adherence to guidelines was demonstrated to improve the use of electronic versus paper-based systems [36]. In comparison, the South African study demonstrated the challenges with technical issues and workload implications [39]. Technical challenges and workload implications were also highlighted in the usability workshops for Neotree; however, despite these challenges, the system scored well on the usability questionnaires. The HCPs stated that they had adapted to these challenges.

Detailed understanding of the context and setting within which the digital intervention has been implemented is an essential component for sustainability. Many of the comments from the
interviewed HCPs described the understaffed, demanding, and poorly resourced environment they work within, which has been described in the literature and is common across many health institutions in low-resource settings [40]. Understanding this context and the complexities of adding a digital intervention, such as power failures within degrading infrastructures or additional workload, is key to successful implementation [17]. Alongside this aspect is the importance of collaborative participatory design. Many health care interventions have been developed from the high-income—setting perspective with minimal attention to the real-life adaptation within low-resource health care settings. Collaborative participatory design and research can help to close the gap between research, policy, and practice [41]. The contextual reality that many low-resource settings face with frequent crises or shocks, such as economic crises, natural disasters, or health crises, inevitably has an impact on how well the interventions are implemented. Damschroder et al [42] have stated that “Many intervention studies have been found to be effective in health services research but fail to translate into meaningful patient care outcomes across multiple contexts.” The Neotree algorithm workshops including Zimbabwean and Malawian experts demonstrated the key importance of these international relationships for the CDS algorithms to be usable within their intended context. The customizability of the CDS algorithms to resource availability and the perceived streamlining of processes from diagnosis to management increased the usability of Neotree by incorporating the detailed knowledge about the working environment.

**Limitations**

The introduction of the algorithms had a perceived impact on the duration of the admission process. In an already overburdened neonatal unit, the increased time of use could have implications on both its acceptability and use. However, the time-use analysis and cost analysis (paper in progress) by Haghparast-Bidgoli et al [43] found that admission times were similar when comparing Neotree with paper-based systems. Implementation in low-resource settings is always going to be influenced by the environment and context it is used in. In Zimbabwe, a degrading health care system [40] poses challenges to any digital tool to be implemented in this setting. The usability questionnaires were only performed in SMCH in Zimbabwe. As SMCH is a predominantly physician-led unit, comparative Malawian and Zimbabwean CDS functionality qualitative data are needed to ensure that usability evaluation across all HCP cadres is reviewed; more comprehensive usability data are required and therefore planned to be collected.

**Next Steps**

Further usability evaluation is planned 1 year after implementation as part of the wider sustainability program of Neotree. This will again include the usability workshops and questionnaires with real-world observation, similar to previous usability testing of the data capture functionality. Approvals from the Pediatric Associations of Zimbabwe and Malawi are pending. The assigned international working groups for continuing research are ongoing. The Neonatal Sepsis Working Group is currently in the process of an in-depth review of neonatal sepsis guideline use within low-resource settings [44], conducting a scoping review [45], and developing a robust clinical prediction model with integrated machine learning technology to be implemented within Neotree. The Neonatal Encephalopathy Working Group published their study of risk factors for mortality [46], is evaluating the use of the Thompson score and its applicability in the low-resource setting, and is developing a neurodevelopmental follow-up pathway to be implemented in Zimbabwe and Malawi. Ongoing studies from the respiratory distress workshop include an in-depth scoping review of the literature about clinical features and risk factors for neonatal respiratory distress in low-resource settings to be incorporated into the respiratory distress algorithm and evaluation of the uptake of the TRY (T: Tone is good, R: Respiratory distress, and Y: Yes) continuous positive airway pressure algorithm [47] and associated outcomes in Malawi. A similar project is underway for neonatal jaundice. A mixed methods, large-scale evaluation of impact on neonatal quality of care and survival is planned. A key gap within the Neotree system is the potential to link with point-of-care diagnostic technology that is low cost, suitable and adaptable to the low-resource setting; early discussions have been made with potential collaborators.

**Conclusions**

This study describes the successful development and implementation of the only known neonatal CDS system incorporated within a bedside data capture system with the ability to deliver up-to-date management guidelines, tailored to local resource availability. The methods used in this study highlighted the importance of collaborative participatory design and South to South learning. Real-world usability testing could further enhance effective implementation. Detailed understanding of the context and setting within which the digital intervention has been implemented is an essential component for sustainability. Effective, motivated local and international partnerships are key to the success of CDS integration into routine practice. Further review of newborn CDS functionality impact and implementation evaluation is planned to guide and inform the development of health system and program strategies to support HCPs, who are overstretched and underresourced, with the ultimate goal of reducing preventable neonatal morbidity and mortality in low-resource settings.

**Acknowledgments**

The funders include the Wellcome Trust: Digital Innovation Award (215742/Z/19/Z to MH). This study, MH, and FF were further supported by the National Institute for Health Research Great Ormond Street Hospital Biomedical Research Centre. HG was further supported by the Medical Research Council Clinical Research and Training Fellowship (MR/V029126/1). The funders had no role in the design of the study; collection, analysis, and interpretation of the data; preparation of this report; or decision
to publish the results. The authors are very grateful to the staff members and families at Sally Mugabe Central Hospital’s neonatal unit and Kamuzu Central Hospital’s neonatal unit for their enthusiasm and commitment to the Neotree project, without which this study would not be possible.

Data Availability

Researchers interested in accessing the usability data will first need to send a request to the medical research council of Zimbabwe by email.

Authors’ Contributions

The study concept was designed by MH, SC, HG, THB, and MC. The workshops were designed and conducted by HG, MH, and THB, with support and attendance from GC, SC, MC, LL, EK, MM, CC, and FF. Stage 2 was supported by MB, YS, NK, and THB. User evaluation was planned by EW and MH and conducted by LL. The original draft was written by HG and LL, with review and edits for the final draft from all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of the web editor function of Neotree.

[DOCX File, 240 KB - formative_v8i1e54274_app1.docx ]

Multimedia Appendix 2

Detailed description of the diagnosis specific algorithms with the associated country-specific management guidelines.

[DOCX File, 3164 KB - formative_v8i1e54274_app2.docx ]

References


44. Fitzgerald F. Assessing the use of neonatal bloodstream infection guidelines in two sub-Saharan African countries: what is used and what is useful? In: Proceedings of the European Society for Paediatric Infectious Diseases Annual Congress. 2023 Presented at: European Society for Paediatric Infectious Diseases Annual Congress; May 8-12, 2023; Lisbon, Portugal. URL: https://cslide.ctimeetingtech.com/espid23/attendee/confcal/session/list?q=fitzgerald


SUS: System Usability Scale

Edited by A Mavragani; submitted 06.11.23; peer-reviewed by V Kain, A Carrandi; comments to author 02.12.23; revised version received 19.12.23; accepted 20.12.23; published 26.01.24.

Please cite as:
Development and Implementation of Digital Diagnostic Algorithms for Neonatal Units in Zimbabwe and Malawi: Development and Usability Study
JMIR Form Res 2024;8:e54274
URL: https://formative.jmir.org/2024/1/e54274
doi:10.2196/54274
PMID:38277198

©Hannah Gannon, Leyla Larsson, Simbarashe Chimhuya, Marcia Mangiza, Emma Wilson, Erin Kesler, Gwendoline Chimhini, Felicity Fitzgerald, Gloria Zailani, Caroline Crehan, Nushrat Khan, Tim Hull-Bailey, Yali Sassoon, Morris Baradza, Michelle Heys, Msandeni Chiume. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Patient Phenotyping for Atopic Dermatitis With Transformers and Machine Learning: Algorithm Development and Validation Study

Andrew Wang¹, BSc; Rachel Fulton², MD; Sy Hwang¹, MS; David J Margolis¹*, MD, PhD; Danielle Mowery¹*, MS, PhD

¹University of Pennsylvania, Philadelphia, PA, United States
²Lankenau Medical Center, Wynnewood, PA, United States
*these authors contributed equally

Corresponding Author:
Danielle Mowery, MS, PhD
University of Pennsylvania
A206 Richards Building
3700 Hamilton Walk
Philadelphia, PA, 19104
United States
Phone: 1 2157466677
Email: dlmowery@pennmedicine.upenn.edu

Abstract

Background: Atopic dermatitis (AD) is a chronic skin condition that millions of people around the world live with each day. Performing research into identifying the causes and treatment for this disease has great potential to provide benefits for these individuals. However, AD clinical trial recruitment is not a trivial task due to the variance in diagnostic precision and phenotypic definitions leveraged by different clinicians, as well as the time spent finding, recruiting, and enrolling patients by clinicians to become study participants. Thus, there is a need for automatic and effective patient phenotyping for cohort recruitment.

Objective: This study aims to present an approach for identifying patients whose electronic health records suggest that they may have AD.

Methods: We created a vectorized representation of each patient and trained various supervised machine learning methods to classify when a patient has AD. Each patient is represented by a vector of either probabilities or binary values, where each value indicates whether they meet a different criteria for AD diagnosis.

Results: The most accurate AD classifier performed with a class-balanced accuracy of 0.8036, a precision of 0.8400, and a recall of 0.7500 when using XGBoost (Extreme Gradient Boosting).

Conclusions: Creating an automated approach for identifying patient cohorts has the potential to accelerate, standardize, and automate the process of patient recruitment for AD studies; therefore, reducing clinician burden and informing the discovery of better treatment options for AD.

(JMIR Form Res 2024;8:e52200) doi:10.2196/52200

KEYWORDS
atopic dermatitis; classification; classifier; dermatitis; dermatology; EHR; electronic health record; health records; health; informatics; machine learning; natural language processing; NLP; patient phenotyping; phenotype; skin; transformer; transformers

Introduction

Background
Atopic dermatitis (AD) is a common skin disease with a population prevalence of approximately 30% [1]. It is often diagnosed in early childhood, but onset can occur at any age [2-5]. Symptoms of AD include inflamed, red, irritated, and itchy skin and can cause significant physical and emotional distress. AD is often associated with other allergic illnesses, including asthma, seasonal allergies, and food allergies [2,3,5-7]. AD is thought to be associated with skin barrier dysfunction and immune dysregulation [5]. AD has also been associated with genetic variation as well as environmental factors [5]. Classic treatment for AD has included the use of moisturizers, topical steroids, and other topical anti-inflammatory agents [8]. However, in the past few years, there have been significant
treatment advances, which include systemic agents that alter immune function, such as dupilumab. Therefore, due to the widespread nature of AD, the need for improved knowledge of the natural history of AD, the need to understand the efficacy of new treatments, and the need to develop new treatments, there is an urgent need to understand the clinical course of individuals with AD. However, identifying appropriate cohorts of patients for medical studies can be difficult and time-consuming. Because AD is so common as well as being diagnosed and managed by many different clinicians in varying health care settings, a potential source population would be patients from a health system’s electronic health records (EHRs) [9]. Investigators often ascertain a patient’s illness using International Classification of Disease (ICD) hospital billing codes as recorded during routine office visits. However, it has been previously demonstrated that reliance on ICD codes is not an accurate method for the ascertainment of study cohorts with AD [9,10]. Furthermore, epidemiologic studies have used different methods and algorithms, including the UK Working Party (UKWP) diagnostic criteria and the Hanifin and Rajka (HR) criteria [11,12]. Investigators attempting to conduct clinical trials and observational studies have also relied on manual, large-scale chart review, a process that is inefficient, slow, and tedious [9]. This motivates the need for a standard method to accurately, automatically, and efficiently identify potential patient cohorts from their text medical records by using natural language processing (NLP) and machine learning (ML) techniques.

Prior Work

Previously, researchers aimed to phenotype patients with AD using EHR data. In particular, Gustafson et al [10] trained a logistic regression model with lasso regularization to identify cases of AD from the Northwestern Medical Enterprise Data Warehouse, which contained both structured data (ICD Ninth and Tenth Revision codes, medication prescriptions, and laboratory results) as well as unstructured data (clinician notes from patient encounters). A gold standard diagnosis was assigned to each patient in their data set by 2 rheumatologists following a chart review when using the UKWP criteria and (alternatively) when using the HR criteria. Although similar, this study differs in the following ways: (1) we survey a wide range of supervised ML algorithms as opposed to only using lasso regularized logistic regression, (2) we use transformer embeddings of sentences to represent information in each patient’s records and aggregate these embeddings with multilayer perceptron (MLP) networks to create a patient vector representation for patient phenotyping, and (3) we performed an ablation study of processing methods to compare the impact on performance in using a probability-based versus binary label of whether each patient meets various AD diagnostic criteria when creating a vector to represent each patient for input to our final patient phenotyping algorithms.

Contributions

The primary contributions of this study are as follows:

• We introduce and validate a rules-based approach for aggregating information from patient EHR data to generate binary-valued patient vectors that are used with standard ML algorithms for patient phenotyping.

• We introduce and validate a transformer-based approach for aggregating information and patient phenotyping by using “Bidirectional Encoder Representations from Transformers” (BERT) models (ie, BERT Base Uncased and BioClinical BERT) to generate patient vectors of probabilities, which are used with standard ML algorithms for patient phenotyping.

• We compare the aforementioned approaches to (1) discern whether a transformer model pretrained on clinical text can provide performance benefits over a transformer model not pretrained on clinical text, and (2) discern whether a transformer-based approach for aggregating information could outperform a rules-based approach for aggregating information.

• We demonstrate that MLP networks can be used with BERT sentence embeddings to identify which sentences in patient records are relevant to the diagnosis of AD. These MLP networks can then be used during clinician chart review to highlight sentences that are relevant to diagnosis and therefore accelerate the process of chart review during clinical trial recruitment.

Methods

Overview

To predict whether a patient may qualify as a participant for an AD study based on their EHR, we first assigned patients in our data set to either the training or testing sets. Then, for each patient, we aggregated the text from their EHR and constructed a vector representation of clinical features indicative of AD according to the UKWP criteria. Lastly, we leveraged our vectorized patient representations to train several ML classifiers to predict whether each patient has AD. In the following sections, we detail this process.

Data Set Creation

We initially sampled 2000 patients and their clinical records from Epic Clarity, Penn Medicine’s EHR database. We selected Penn Medicine patients who were diagnosed with a subset of AD-related ICD codes [9]. As shown in Figure 1, of the 2000 sampled patients, we identified 1926 patients who had clinical notes for processing. We then deidentified these patient records according to the Safe Harbor method using the “Protected Health Information filter” (Philter) [13]. Each patient in the data set was also manually reviewed and labeled according to the UKWP diagnostic criteria for AD. According to the UKWP criteria, in order to qualify as having AD, a patient must have an itchy skin condition along with 3 or more of the following: a history of flexural involvement, a history of asthma or hay fever, a history of dry skin, an onset of rash when aged 2 years or younger, or a visible flexural dermatitis. Our data set was validated by 2 clinicians (a board-certified dermatologist [DJM] and a medical fellow [RF]), resulting in 137 patients with AD and 1789 patients without AD.
Training and Testing Split

We first created our training set. Due to the heavy class imbalance in our data set, we decided to create a balanced training set to prevent biasing the model toward either patients with AD or patients without AD. In particular, we created the training set by assigning 80% (109/137) of the 137 patients with AD to our training set and undersampling the patients without AD to match the number of patients with AD. The remaining 20% (28/137) of the 137 patients were assigned to both of our testing sets. This resulted in a training set that had 109 patients with AD and 109 patients without AD.

Next, we created 2 testing sets. The first testing set was class-balanced and was intended to show how our patient classification model can generalize to unseen samples if the class distribution is kept the same. The second testing set was class-imbalanced (28/91, 30% of patients with AD and 63/91, 70% of patients without AD) and was intended to show how our patient classification model can perform when the class-distribution of the data set matches the prevalence of AD in the United States.

We created the first (balanced) testing set by including the 20% (28/137; previously reserved for testing) of the 137 patients with AD and combining them with an equal number of patients without AD who have not been used during training. This resulted in a balanced testing set that had 28 patients with AD and 28 patients without AD.

Furthermore, we created the second (unbalanced) testing set by including the same 20% (28/137) patients with AD but instead combining them with a greater number of patients without AD to match the 30% prevalence rate of AD found in the United States [1]. This resulted in an unbalanced testing set with 28 patients who have AD and 63 patients without AD.

We chose not to create a separate hyperparameter tuning set and instead applied cross-validation for hyperparameter tuning on the training set due to the data-scarce setting of our experiments.

Vector Representation for AD Classification

Next, we created a vector representation for each patient. We performed 3 experiments to compare different methods of creating each patient’s vector representation (Figure 2).
Description of Patient Vector Representation

Each patient’s vector representation is 8 elements long, where each element of the vector is representative of whether the patient fulfills a different AD diagnosis criteria based on the UKWP criteria as well as clinician feedback (Table 1). Across all 3 experiments, each element in the patient vector corresponds to a distinct classification task; however, in experiments 1 and 2, each element is a probability, and in experiment 3, each element is a binary value.

In experiments 1 and 2, elements 1-8 of each patient’s vector represent the highest probability that any sentence in the patient’s EHR mentions (1) AD or synonyms of AD, (2) keywords that suggest hay fever allergies, (3) keywords that suggest atopic allergies, (4) keywords that suggest eczema or rashes, (5) keywords that indicate dry or itchy skin, (6) keywords denoting nonasthma medications, (7) keywords suggesting the presence of asthma, and (8) keywords indicating the use of asthma medications.

In experiment 3, instead of each element representing a probability, each element represents a binary value of whether there was at least 1 sentence in the corresponding patient record suggesting the presence of the corresponding AD indicator.

In the first 2 experiments, each patient’s vector elements represent probabilities (ranging from 0 to 1). Each probability value is derived from a distinct MLP classifier. Experiments 1 and 2 were performed to compare the use of 2 BERT models (BERT Base Uncased [14,15] in experiment 1 and BioClinical BERT [16,17] in experiment 2) for creating sentence embeddings used to train MLP networks (or alternatively, sentence classifiers). A separate MLP network is trained for each element of the patient vector. Each MLP network is trained to distinguish sentences in 1 of the 8 AD indicator categories from sentences in all other categories. Furthermore, medSpacy (Eyre et al[18]) was used to split documents into sentences and label each sentence with different categories. After each sentence classifier is trained, embeddings of all sentences in each patient’s full EHR are passed through each sentence classifier, and an aggregation function (max operator) is used to assign a value to each element of each patient’s vector. Our goal in experiments 1 and 2 was to test the hypothesis that a BERT model pretrained on clinical text (BioClinical BERT) could outperform a BERT model trained on nonclinical text (BERT Base Uncased).

In experiment 3, each patient’s vector elements are binary (either 0 or 1). Each element corresponds to a diagnostic criterion and represents whether medSpacy was able to identify at least 1 sentence in the patient’s record with a keyword and affirming context that suggests the patient meets the corresponding diagnostic criteria. Our goal was to conduct an ablation study to test the hypothesis that an AD phenotyping classifier leveraging BERT embeddings to create the patient vector representation will better discern whether a patient has AD than an AD Phenotyping Classifier without BERT embeddings.

<table>
<thead>
<tr>
<th>Element</th>
<th>AD indicator (diagnostic criteria)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EHR directly mentions patient has AD</td>
</tr>
<tr>
<td>2</td>
<td>Patient has hay fever allergies</td>
</tr>
<tr>
<td>3</td>
<td>Patient has atopic allergies</td>
</tr>
<tr>
<td>4</td>
<td>Patient has eczema or rashes</td>
</tr>
<tr>
<td>5</td>
<td>Patient has dry or itchy skin</td>
</tr>
<tr>
<td>6</td>
<td>Patient uses nonasthma medications related to treating AD</td>
</tr>
<tr>
<td>7</td>
<td>Patient has asthma</td>
</tr>
<tr>
<td>8</td>
<td>Patient uses asthma medications</td>
</tr>
</tbody>
</table>

Table 1. Meaning of each patient vector element.

aAD: atopic dermatitis.
bEHR: electronic health record.

Preprocessing for Experiments 1-3

Before each experiment, we applied the same preprocessing steps to assign 1 or more labels to each sentence in our corpus of documents in both our training and testing sets. Each sentence can be labeled as applying to 1, multiple, or none of the 8 AD indicators previously defined.

For each of the 8 diagnostic criteria, we first created a list of keywords and phrases (for each vector element) that suggested the presence of the corresponding diagnostic criteria. Next, we used medSpacy with the ConText (Harkema et al [20]) algorithm to split each document into sentences and categorize each sentence [18]. Using medSpacy allows us to obtain sentences that suggest the presence of each of the 8 diagnostic criteria due to medSpacy’s use of regex and rules-based keyword matching. Furthermore, medSpacy’s implementation of the ConText algorithm allows us to discern between sentences that affirm from negated assertions. We define negated sentences for each AD indicator as sentences where the indicator is ruled out, sentences where the indicator is experienced by someone other than the patient, and sentences where the existence of the indicator is hypothetical [19-22].

After assigning 1 or more categorical labels to each sentence with medSpacy, we then performed 3 different experiments to create a vectorized representation of each patient.

In Tables 2 and 3, we include some statistics on the data set obtained after preprocessing.
As shown in Table 2, patients with AD have approximately twice as many sentences as patients without AD. The average number of documents and sentences is the same (within patients with AD and similarly within patients without AD) between BERT Base Uncased and BioClinical BERT experiments because these values are only dependent on medSpacy’s preprocessing of documents. Furthermore, using BioClinical BERT to tokenize sentences tends to yield more tokens (on average) per patient and per document. We hypothesize this is because the BioClinical BERT tokenizer is able to recognize more clinical terms and therefore yields more tokens for the same sentence than using the tokenizer from BERT Base Uncased.

As shown in Table 3, sentences in category 5 (relating to dry or itchy skin) tend to have the most tokens, whereas sentences in category 6 (relating to the use of nonasthma medications related to treating AD) tend to have the least number of tokens. We hypothesize that this is because categories where the average number of tokens per sentence is greater tend to correspond to more general categories where many terms and sentences could apply, whereas categories where the average number of tokens per sentence is lower tend to correspond to more specific categories, thus yielding a lower average number of tokens per sentence. Additionally, similarly to before, we can see that using BioClinical BERT tends to result in a greater number of tokens per sentence than using BERT Base Uncased for the same sentence.

Table 2. Differences in the number of documents, sentences, and tokens between patients with atopic dermatitis (AD) and those without AD.

<table>
<thead>
<tr>
<th></th>
<th>Patients with AD</th>
<th>Patients without AD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BERT a Uncased</td>
<td>BioClinical BERT</td>
</tr>
<tr>
<td>Average number of documents (per patient)</td>
<td>23.44</td>
<td>23.44</td>
</tr>
<tr>
<td>Average number of sentences (per patient)</td>
<td>7.99</td>
<td>7.99</td>
</tr>
<tr>
<td>Average number of tokens (per patient)</td>
<td>392.99</td>
<td>392.99</td>
</tr>
<tr>
<td>Average number of sentences (per document)</td>
<td>16035.39</td>
<td>17054.11</td>
</tr>
<tr>
<td>Average number of tokens (per document)</td>
<td>16.77</td>
<td>16.77</td>
</tr>
<tr>
<td>Average number of tokens (per sentence)</td>
<td>684.16</td>
<td>727.63</td>
</tr>
<tr>
<td>Average number of tokens (per sentence)</td>
<td>40.80</td>
<td>43.40</td>
</tr>
</tbody>
</table>

aBERT: Bidirectional Encoder Representations from Transformers.

Table 3. Mean number of tokens for sentences identified in each category.

<table>
<thead>
<tr>
<th>Category</th>
<th>BERT a Uncased (tokens per sentence), mean</th>
<th>BioClinical BERT (tokens per sentence), mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>99.49</td>
<td>106.16</td>
</tr>
<tr>
<td>Category 2</td>
<td>81.18</td>
<td>92.41</td>
</tr>
<tr>
<td>Category 3</td>
<td>79.20</td>
<td>82.07</td>
</tr>
<tr>
<td>Category 4</td>
<td>83.74</td>
<td>92.55</td>
</tr>
<tr>
<td>Category 5</td>
<td>106.64</td>
<td>112.58</td>
</tr>
<tr>
<td>Category 6</td>
<td>74.93</td>
<td>80.17</td>
</tr>
<tr>
<td>Category 7</td>
<td>92.85</td>
<td>109.40</td>
</tr>
<tr>
<td>Category 8</td>
<td>76.13</td>
<td>83.57</td>
</tr>
</tbody>
</table>

aBERT: Bidirectional Encoder Representations from Transformers.

Experiments 1 and 2: Patient Vector Construction With BERT Embeddings

In experiments 1 and 2, we first used the sentences medSpacy identified in each category to create class-balanced training and testing sets for each MLP network classifier, as shown in Table 4. The same training and testing set was used for both experiment 1 (BioClinical BERT) and experiment 2 (BERT Base Uncased).

Next, we used pretrained BERT models to generate embeddings of the sentences in each classifier’s training and testing set. We incorporated pretrained BERT models because these models have been trained on a much larger corpus than our existing data set, and BERT provides a context-sensitive embedding of text that other techniques, such as bag of words, do not provide. Furthermore, we used BERT Base Uncased in experiment 1 and Alsentzer et al’s [16] BioClinical BERT in experiment 2 because we wanted to quantify how much of a difference in performance using a model pretrained on clinical text can provide over a model that has not been pretrained on clinical text.

Using these embeddings, we trained a MLP network to distinguish sentence embeddings in each category from sentence embeddings that are not in the corresponding category. Each of our MLPs was trained with the following architecture: a fully connected input layer of shape $768 \times 100$, followed by a Rectified Linear Unit (ReLU) activation, further followed by a fully connected output layer of shape $100 \times 2$. We trained each

https://formative.jmir.org/2024/1/e52200
of our MLPs for 10 epochs with the cross-entropy loss function, the stochastic gradient descent (SGD) optimizer, a learning rate of 0.001, and a momentum value of 0.9. The final layer of each MLP can then be used to obtain the probability that any given sentence embedding comes from the category for which the MLP is being trained by passing the logits of the final layer to the softmax function.

We used the ReLU activation function as defined below, where \( x \) is the input to the ReLU function:

\[
x = \max(0, x)
\]

We also used the softmax function as defined below, where \( e^x \) is the standard exponential function and \( [\ ] \) is element at index \( i \) of the \( K \) element long input vector \( [\ ] \).

We chose to embed our sentences once with pretrained BERT models and then feed these saved embeddings to our MLP networks as opposed to adding a classification head (a linear layer) to the end of our pretrained BERT models. Although doing so only allows us to fine-tune the weights in our MLP network (as opposed to also fine-tuning the weights BERT uses to embed the sentences), doing so allows us to iterate over different experiments more quickly and with less computational power. In particular, we are able to (1) avoid the large computational expense of gradient calculations during backpropagation for all 12 layers of transformers used by BERT when fine-tuning the model, (2) avoid the computational expense of repeatedly generating the same embeddings from BERT multiple times (if we choose to freeze the weights of BERT and only fine-tune an added classification head or linear layer), and (3) iterate more efficiently over different hyperparameter combinations across different experiments with our MLP networks.

After training a separate MLP network for each of the 8 categories, we generated a vector representation for each patient, where each of the 8 vector elements represents the highest probability that any given sentence in the patient record affirms the presence of the corresponding AD indicator (Figure 3). We accomplished this by iterating through all sentences in each patient’s full EHR and passing the sentence embedding through each of our 8 trained MLP networks to obtain 8 probabilities for each sentence corresponding to the probability that the sentence affirms each of the 8 AD indicators we previously selected. Then, for each patient and for each AD indicator, we kept the highest probability that any given sentence in the patient’s record affirms the presence of the AD indicator.

Table 4. Training and testing data set size for each classifier.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Number of training samples, n</th>
<th>Number of testing samples, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2766</td>
<td>862</td>
</tr>
<tr>
<td>2</td>
<td>1302</td>
<td>392</td>
</tr>
<tr>
<td>3</td>
<td>532</td>
<td>168</td>
</tr>
<tr>
<td>4</td>
<td>9822</td>
<td>2454</td>
</tr>
<tr>
<td>5</td>
<td>1466</td>
<td>354</td>
</tr>
<tr>
<td>6</td>
<td>9114</td>
<td>2316</td>
</tr>
<tr>
<td>7</td>
<td>1596</td>
<td>520</td>
</tr>
<tr>
<td>8</td>
<td>4764</td>
<td>1070</td>
</tr>
</tbody>
</table>

Figure 3. Patient vector representations of atopic dermatitis indicators in experiments 1 and 2. BERT: Bidirectional Encoder Representations from Transformers; MLP: multilayer perceptron.

\[
v_i = \max(p_{1,1}, p_{1,2}, \ldots, p_{1,n})
\]

\[
v_i = \max(p_{2,1}, p_{2,2}, \ldots, p_{2,n})
\]

\[
v_i = \max(p_{3,1}, p_{3,2}, \ldots, p_{3,n})
\]

\[
v_i = \max(p_{4,1}, p_{4,2}, \ldots, p_{4,n})
\]

\[
v_i = \max(p_{5,1}, p_{5,2}, \ldots, p_{5,n})
\]

\[
v_i = \max(p_{6,1}, p_{6,2}, \ldots, p_{6,n})
\]

\[
v_i = \max(p_{7,1}, p_{7,2}, \ldots, p_{7,n})
\]

\[
v_i = \max(p_{8,1}, p_{8,2}, \ldots, p_{8,n})
\]

\[
v_i = \text{Element } i \text{ of patient vector}
\]

\[
v_i = \max(p_{1,1}, p_{1,2}, \ldots, p_{1,n})
\]

\[
v_i = \max(p_{2,1}, p_{2,2}, \ldots, p_{2,n})
\]

\[
v_i = \max(p_{3,1}, p_{3,2}, \ldots, p_{3,n})
\]

\[
v_i = \max(p_{4,1}, p_{4,2}, \ldots, p_{4,n})
\]

\[
v_i = \max(p_{5,1}, p_{5,2}, \ldots, p_{5,n})
\]

\[
v_i = \max(p_{6,1}, p_{6,2}, \ldots, p_{6,n})
\]

\[
v_i = \max(p_{7,1}, p_{7,2}, \ldots, p_{7,n})
\]

\[
v_i = \max(p_{8,1}, p_{8,2}, \ldots, p_{8,n})
\]

Experiment 3: Patient Vector Construction Without BERT Embeddings

In experiment 3, we generated each patient’s vector representation by assigning a 1 to each element of the patient vector if medSpacy with the ConText algorithm identified at least 1 sentence in the patient’s record that affirms or suggests the presence of the AD indicator for which the vector element corresponds (Figure 4). Experiment 3 was conducted as an
ablation study to quantify the performance benefit (if at all) of using contextual BERT text embeddings to generate probability scores that the patient meets various AD indicators.

**Figure 4.** Patient vector representations of atopic dermatitis (AD) indicators in experiment 3.

<table>
<thead>
<tr>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
<th>$v_4$</th>
<th>$v_5$</th>
<th>$v_6$</th>
<th>$v_7$</th>
<th>$v_8$</th>
</tr>
</thead>
</table>

$v_i = \text{Element } i \text{ of patient vector}$

$$v_i = \begin{cases} 1 \text{ if } c_i > 0 \\ 0 \text{ otherwise} \end{cases}$$

$c_i = \text{Number of sentences medSpacy identified as suggesting presence of AD indicator for category } i \text{ in current patient’s records}$

**AD Phenotyping With Vector Representations**

In all 3 experiments, after generating a vector representation for each patient, we collated each patient’s vector representation with the corresponding label our clinicians assigned the patient when validating the data set. Then, we fed the vector patient representation and corresponding patient label through a variety of classification algorithms. These include logistic regression, support vector machines (SVM), decision trees, random forests, k-nearest neighbor (KNN), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost). During training for each of the previously mentioned classifiers, we used 5-fold cross validation to determine the best set of hyperparameters to use (as opposed to creating a separate validation set) due to the data-scarce setting of our experiments. We then used the selected hyperparameters to train each algorithm on the entire training set and evaluated performance on the unbalanced and balanced testing sets. In addition to using the previously mentioned classifiers, we also used the stacking algorithm provided by scikit-learn to obtain an ensemble prediction from the different classifiers [23]. To quantify performance, we calculated the accuracy, precision, recall, $F_1$-score, negative predictive value (NPV), and specificity of each algorithm on both testing sets.

We define accuracy, precision, and recall as follows, where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Additionally, we define the $F_1$-score, NPV, and specificity as follows:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{NPV} = \frac{TN}{TN + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

**Ethical Considerations**

This research protocol was reviewed and approved by the University of Pennsylvania Institute Review Board and determined to be exempt (IRB#843922).

**Results**

**Performance of MLP Networks**

In this section, we compare the performance of several MLP classifiers in distinguishing sentences relevant to the diagnosis of AD. This corresponds to the “Train separate MLP network (sentence classifier) for each of 8 AD indicators” box in Figure 2.

As part of our AD phenotyping pipeline, we trained various MLP networks to classify when a given sentence embedding indicates the presence of an AD indicator, and we compared the performance of BioClinical BERT embeddings to BERT Base Uncased embeddings when training these MLP networks. In both cases, the classifier with the highest accuracy was the classifier for category 1 (sentences with direct mentions of AD). The classifiers with the 2 lowest accuracies were either the classifier for category 5 (sentences with mentions of dry or itchy skin) or the classifier for category 7 (sentences with mentions of asthma) for both the use of BioClinical BERT embeddings and the use of BERT Base Uncased embeddings. However, the accuracy in classifier 7 was lower when using BERT Base Uncased embeddings than when using BioClinical BERT embeddings.

In experiment 1, the accuracies across AD indicator classifiers ranged from 0.7373 (classifier 5) to 0.9002 (classifier 1), as shown in Table 5.

In experiment 2, the accuracies across AD indicator classifiers ranged from 0.7269 (classifier 7) to 0.9153 (classifier 1), as shown in Table 6.
Table 5. Accuracy of different multilayer perceptron networks in discerning sentences by atopic dermatitis (AD) indicator categories using “BioClinical Bidirectional Encoder Representations from Transformers” sentence embeddings.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>AD indicator</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct mention of AD</td>
<td>0.9002</td>
</tr>
<tr>
<td>2</td>
<td>Mention of hay fever allergies</td>
<td>0.8954</td>
</tr>
<tr>
<td>3</td>
<td>Mention of atopic allergies</td>
<td>0.8214</td>
</tr>
<tr>
<td>4</td>
<td>Mention of eczema or rash</td>
<td>0.8284</td>
</tr>
<tr>
<td>5</td>
<td>Mention of dry or itchy skin</td>
<td>0.7373</td>
</tr>
<tr>
<td>6</td>
<td>Mention of nonasthma medications</td>
<td>0.8204</td>
</tr>
<tr>
<td>7</td>
<td>Mention of asthma</td>
<td>0.7712</td>
</tr>
<tr>
<td>8</td>
<td>Mention of asthma medications</td>
<td>0.8299</td>
</tr>
</tbody>
</table>

Table 6. Accuracy of different multilayer perceptron networks in discerning sentences by atopic dermatitis (AD) indicator categories using “Bidirectional Encoder Representations from Transformers Base Uncased” sentence embeddings.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>AD indicator</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct mention of AD</td>
<td>0.9153</td>
</tr>
<tr>
<td>2</td>
<td>Mention of hay fever allergies</td>
<td>0.7730</td>
</tr>
<tr>
<td>3</td>
<td>Mention of atopic allergies</td>
<td>0.7976</td>
</tr>
<tr>
<td>4</td>
<td>Mention of eczema or rash</td>
<td>0.8439</td>
</tr>
<tr>
<td>5</td>
<td>Mention of dry or itchy skin</td>
<td>0.7288</td>
</tr>
<tr>
<td>6</td>
<td>Mention of nonasthma medications</td>
<td>0.8096</td>
</tr>
<tr>
<td>7</td>
<td>Mention of asthma</td>
<td>0.7269</td>
</tr>
<tr>
<td>8</td>
<td>Mention of asthma medications</td>
<td>0.8738</td>
</tr>
</tbody>
</table>

AD Phenotyping With Patient Vector Representations

In this section, we compare performance in patient classification when using different methods for creating patient vector representations. This encompasses all 3 experiments and corresponds to the “Use vector patient representations to classify whether patient has AD” box in Figure 2.

In experiment 1, we leveraged BioClinical BERT sentence embeddings to train various MLP networks to discern sentence embeddings in different AD indicator categories. Then, we applied these trained MLP networks (sentence classifiers) along with an aggregation function (max operator) to assign values to each element of each patient’s vector representation. Lastly, we used each patient’s vector representation with their validated label to train various ML algorithms. We evaluated these on both a balanced and unbalanced testing set.

As shown in Table 7, the accuracy on the balanced testing set ranges from 0.5893 (decision tree) to 0.7321 (logistic regression and SVM).

As shown in Table 8, the range of accuracies on the unbalanced testing set is slightly lower, ranging from 0.5824 (decision tree) to 0.7253 (stacking classifier).

In experiment 2, we followed the same process as in experiment 1; however, we used BERT Base Uncased instead of BioClinical BERT. As shown in Table 9, the accuracy of our AD classifiers on the balanced testing set ranges from 0.5179 (AdaBoost) to 0.6250 (random forest).

As shown in Table 10, the range of accuracies of our AD classifiers on the unbalanced testing set is slightly higher, ranging from 0.5714 (logistic regression and SVM) to 0.6703 (random forest).

In experiment 3, we performed an ablation study and assigned binary labels to the elements of each patient’s vector based on whether medSpacy was able to identify at least 1 sentence in each of the AD indicator categories that each vector element corresponds to. As shown in Table 11, the accuracy across our AD classifiers on the balanced testing set ranges from 0.6964 (KNN) to 0.8036 (XGBoost).

As shown in Table 12, the lower bound of the range of accuracies across our AD classifiers on the unbalanced testing set is higher, and the upper bound of the accuracies is lower. The accuracies on the unbalanced testing set range from 0.7143 (Stacking Classifier) to 0.7582 (Random Forest and Stacking Classifier).
Table 7. Atopic dermatitis phenotyping performance on balanced testing set in experiment 1 (BioClinical Bidirectional Encoder Representations from Transformers).

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$-score</th>
<th>NPV $^a$</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.7321</td>
<td>0.7241</td>
<td>0.7500</td>
<td>0.7368</td>
<td>0.7407</td>
<td>0.7500</td>
</tr>
<tr>
<td>SVM $^b$</td>
<td>0.7321</td>
<td>0.7826</td>
<td>0.6429</td>
<td>0.7059</td>
<td>0.6970</td>
<td>0.7857</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.5893</td>
<td>0.6316</td>
<td>0.4286</td>
<td>0.5106</td>
<td>0.5676</td>
<td>0.7500</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.6964</td>
<td>0.7037</td>
<td>0.6786</td>
<td>0.6909</td>
<td>0.6897</td>
<td>0.8214</td>
</tr>
<tr>
<td>KNN $^c$</td>
<td>0.6786</td>
<td>0.7273</td>
<td>0.5714</td>
<td>0.6400</td>
<td>0.6471</td>
<td>0.7857</td>
</tr>
<tr>
<td>XGBoost $^d$</td>
<td>0.6071</td>
<td>0.6154</td>
<td>0.5714</td>
<td>0.5926</td>
<td>0.6000</td>
<td>0.8571</td>
</tr>
<tr>
<td>AdaBoost $^e$</td>
<td>0.6429</td>
<td>0.6538</td>
<td>0.6071</td>
<td>0.6296</td>
<td>0.6333</td>
<td>0.7857</td>
</tr>
<tr>
<td>Stacking classifier</td>
<td>0.6964</td>
<td>0.7391</td>
<td>0.6071</td>
<td>0.6667</td>
<td>0.6667</td>
<td>0.7500</td>
</tr>
</tbody>
</table>

$^a$NPV: negative predictive value.
$^b$SVM: support vector machines.
$^c$KNN: k-nearest neighbor.
$^d$XGBoost: Extreme Gradient Boosting.
$^e$AdaBoost: Adaptive Boosting.

Table 8. Atopic dermatitis phenotyping performance on unbalanced testing set in experiment 1 (BioClinical Bidirectional Encoder Representations from Transformers).

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$-score</th>
<th>NPV $^a$</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.6813</td>
<td>0.4884</td>
<td>0.7500</td>
<td>0.5915</td>
<td>0.8542</td>
<td>0.6984</td>
</tr>
<tr>
<td>SVM $^b$</td>
<td>0.6923</td>
<td>0.5000</td>
<td>0.6429</td>
<td>0.5625</td>
<td>0.8181</td>
<td>0.7302</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.5824</td>
<td>0.3438</td>
<td>0.3929</td>
<td>0.3667</td>
<td>0.7119</td>
<td>0.7143</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.7143</td>
<td>0.5313</td>
<td>0.6071</td>
<td>0.5667</td>
<td>0.6845</td>
<td>0.7619</td>
</tr>
<tr>
<td>KNN $^c$</td>
<td>0.6593</td>
<td>0.4571</td>
<td>0.5714</td>
<td>0.5079</td>
<td>0.7857</td>
<td>0.7937</td>
</tr>
<tr>
<td>XGBoost $^d$</td>
<td>0.6264</td>
<td>0.4211</td>
<td>0.5714</td>
<td>0.4848</td>
<td>0.7736</td>
<td>0.7619</td>
</tr>
<tr>
<td>AdaBoost $^e$</td>
<td>0.6044</td>
<td>0.4048</td>
<td>0.6071</td>
<td>0.4857</td>
<td>0.7755</td>
<td>0.7302</td>
</tr>
<tr>
<td>Stacking classifier</td>
<td>0.7253</td>
<td>0.5429</td>
<td>0.6786</td>
<td>0.6032</td>
<td>0.8393</td>
<td>0.6984</td>
</tr>
</tbody>
</table>

$^a$NPV: negative predictive value.
$^b$SVM: support vector machines.
$^c$KNN: k-nearest neighbor.
$^d$XGBoost: Extreme Gradient Boosting.
$^e$AdaBoost: Adaptive Boosting.
### Table 9. Atopic dermatitis phenotyping performance on balanced testing set in experiment 2 (Bidirectional Encoder Representations from Transformers Base Uncased).

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$-score</th>
<th>NPV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.5893</td>
<td>0.5758</td>
<td>0.6786</td>
<td>0.6230</td>
<td>0.6087</td>
<td>0.5000</td>
</tr>
<tr>
<td>SVM&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.6071</td>
<td>0.5938</td>
<td>0.6786</td>
<td>0.6333</td>
<td>0.6250</td>
<td>0.5357</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.6250</td>
<td>0.6522</td>
<td>0.5357</td>
<td>0.5882</td>
<td>0.6061</td>
<td>0.7143</td>
</tr>
<tr>
<td>KNN&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.5536</td>
<td>0.5714</td>
<td>0.4286</td>
<td>0.4898</td>
<td>0.5429</td>
<td>0.6786</td>
</tr>
<tr>
<td>XGBoost&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.5536</td>
<td>0.5556</td>
<td>0.5357</td>
<td>0.5455</td>
<td>0.5517</td>
<td>0.5714</td>
</tr>
<tr>
<td>AdaBoost&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.5179</td>
<td>0.5185</td>
<td>0.5000</td>
<td>0.5091</td>
<td>0.5172</td>
<td>0.5357</td>
</tr>
<tr>
<td>Stacking classifier</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
<td>0.6071</td>
</tr>
</tbody>
</table>

<sup>a</sup>NPV: negative predictive value.
<sup>b</sup>SVM: support vector machines.
<sup>c</sup>KNN: k-nearest neighbor.
<sup>d</sup>XGBoost: Extreme Gradient Boosting.
<sup>e</sup>AdaBoost: Adaptive Boosting.

### Table 10. Atopic dermatitis phenotyping performance on unbalanced testing set in experiment 2 (Bidirectional Encoder Representations from Transformers Base Uncased).

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$-score</th>
<th>NPV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.5714</td>
<td>0.3878</td>
<td>0.6786</td>
<td>0.4935</td>
<td>0.7857</td>
<td>0.5238</td>
</tr>
<tr>
<td>SVM&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.5714</td>
<td>0.3878</td>
<td>0.6786</td>
<td>0.4935</td>
<td>0.7857</td>
<td>0.5238</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.6484</td>
<td>0.4474</td>
<td>0.6071</td>
<td>0.5152</td>
<td>0.7925</td>
<td>0.6667</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.6703</td>
<td>0.4737</td>
<td>0.6429</td>
<td>0.5455</td>
<td>0.8113</td>
<td>0.6825</td>
</tr>
<tr>
<td>KNN&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.6264</td>
<td>0.4000</td>
<td>0.4286</td>
<td>0.4138</td>
<td>0.7377</td>
<td>0.7143</td>
</tr>
<tr>
<td>XGBoost&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.6374</td>
<td>0.4286</td>
<td>0.5357</td>
<td>0.4762</td>
<td>0.7679</td>
<td>0.6825</td>
</tr>
<tr>
<td>AdaBoost&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.5934</td>
<td>0.3784</td>
<td>0.5000</td>
<td>0.4308</td>
<td>0.7407</td>
<td>0.6349</td>
</tr>
<tr>
<td>Stacking classifier</td>
<td>0.6484</td>
<td>0.4474</td>
<td>0.6071</td>
<td>0.5152</td>
<td>0.7925</td>
<td>0.6667</td>
</tr>
</tbody>
</table>

<sup>a</sup>NPV: negative predictive value.
<sup>b</sup>SVM: support vector machines.
<sup>c</sup>KNN: k-nearest neighbor.
<sup>d</sup>XGBoost: Extreme Gradient Boosting.
<sup>e</sup>AdaBoost: Adaptive Boosting.
Table 11. Atopic dermatitis phenotyping performance on balanced testing set in experiment 3 (binary vector encoding).

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$-score</th>
<th>NPV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.7679</td>
<td>0.7586</td>
<td>0.7857</td>
<td>0.7719</td>
<td>0.7778</td>
<td>0.7500</td>
</tr>
<tr>
<td>SVM&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.7857</td>
<td>0.7667</td>
<td>0.8214</td>
<td>0.7931</td>
<td>0.8077</td>
<td>0.7500</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.7857</td>
<td>0.8077</td>
<td>0.7500</td>
<td>0.7778</td>
<td>0.7667</td>
<td>0.8214</td>
</tr>
<tr>
<td>KNN&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.6964</td>
<td>0.7391</td>
<td>0.6071</td>
<td>0.6667</td>
<td>0.6667</td>
<td>0.7857</td>
</tr>
<tr>
<td>XGBoost&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.8036</td>
<td>0.8400</td>
<td>0.7500</td>
<td>0.7925</td>
<td>0.7742</td>
<td>0.8571</td>
</tr>
<tr>
<td>AdaBoost&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
<td>0.7857</td>
</tr>
<tr>
<td>Stacking classifier</td>
<td>0.7500</td>
<td>0.7500</td>
<td>0.7500</td>
<td>0.7500</td>
<td>0.7500</td>
<td>0.7500</td>
</tr>
</tbody>
</table>

<sup>a</sup>NPV: negative predictive value.
<sup>b</sup>SVM: support vector machines.
<sup>c</sup>KNN: k-nearest neighbor.
<sup>d</sup>XGBoost: Extreme Gradient Boosting.
<sup>e</sup>AdaBoost: Adaptive Boosting.

Table 12. Atopic dermatitis phenotyping performance on unbalanced testing set in experiment 3 (binary vector encoding).

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$-score</th>
<th>NPV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.7253</td>
<td>0.5366</td>
<td>0.7857</td>
<td>0.6377</td>
<td>0.8800</td>
<td>0.6984</td>
</tr>
<tr>
<td>SVM&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.7473</td>
<td>0.5641</td>
<td>0.7857</td>
<td>0.6567</td>
<td>0.8846</td>
<td>0.7302</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.7473</td>
<td>0.5610</td>
<td>0.8214</td>
<td>0.6667</td>
<td>0.9000</td>
<td>0.7143</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.7582</td>
<td>0.5833</td>
<td>0.7500</td>
<td>0.6563</td>
<td>0.8727</td>
<td>0.7619</td>
</tr>
<tr>
<td>KNN&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.7363</td>
<td>0.5667</td>
<td>0.6071</td>
<td>0.5862</td>
<td>0.8197</td>
<td>0.7937</td>
</tr>
<tr>
<td>XGBoost&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.7582</td>
<td>0.5833</td>
<td>0.7500</td>
<td>0.6563</td>
<td>0.8727</td>
<td>0.7619</td>
</tr>
<tr>
<td>AdaBoost&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.7473</td>
<td>0.5641</td>
<td>0.7857</td>
<td>0.6567</td>
<td>0.8846</td>
<td>0.7302</td>
</tr>
<tr>
<td>Stacking classifier</td>
<td>0.7143</td>
<td>0.5250</td>
<td>0.7500</td>
<td>0.6176</td>
<td>0.8627</td>
<td>0.6984</td>
</tr>
</tbody>
</table>

<sup>a</sup>NPV: negative predictive value.
<sup>b</sup>SVM: support vector machines.
<sup>c</sup>KNN: k-nearest neighbor.
<sup>d</sup>XGBoost: Extreme Gradient Boosting.
<sup>e</sup>AdaBoost: Adaptive Boosting.

**Discussion**

**Sentence Classification Results**

We hypothesized that using BioClinical BERT sentence embeddings to train sentence classifiers would provide better performance than using BERT Base Uncased sentence embeddings due to the clinical setting of our data. Given the results in Tables 5 and 6, we observed that this was most often true in the context of sentence classification because we were able to achieve better performance in the majority (5 out of 8) of the sentence classification tasks when using BioClinical BERT embeddings as opposed to BERT Base Uncased embeddings.

Using BioClinical BERT sentence embeddings yielded stronger performance when distinguishing sentences in 5 of the 8 sentence categories: category 2 (mentions of hay fever allergies), category 3 (mentions of atopic allergies), category 5 (mentions of dry or itchy skin), category 6 (mentions of nonasthma medications), and category 7 (mentions of asthma). More specifically, we observed higher accuracies when using BioClinical BERT sentence embeddings for classifiers 2 (0.8954), 3 (0.8214), 5 (0.7373), 6 (0.8204), and 7 (0.7712) than their corresponding counterparts when using BERT Base Uncased embeddings for classifiers 2 (0.7730), 3 (0.7976), 5 (0.7288), 6 (0.8096), and 7 (0.7269). We observed that the differences in performance between using BioClinical BERT embeddings and BERT Base Uncased embeddings are most pronounced for classifiers 2 and 7, which correspond to mentions of hay fever allergies and asthma mentions, respectively. We hypothesize this is because hay fever allergies and asthma (and their synonyms) may be very common terms in clinical notes; therefore, models trained on clinical data (BioClinical BERT) may be able to provide stronger...
performance than models trained on nonclinical text (BERT Base Uncased), which may not have as many mentions of hay fever allergies or asthma.

Conversely, using BERT Base Uncased embeddings yielded stronger performance when distinguishing sentences in the other 3 of 8 sentence categories: category 1 (direct mentions of AD), category 4 (mentions of eczema or rashes), and category 8 (mentions of asthma medications). More specifically, we observed higher accuracies when using BERT Base Uncased sentence embeddings for classifiers 1 (0.9153), 4 (0.8439), and 8 (0.8738) than their corresponding counterparts when using BioClinical BERT embeddings for classifiers 1 (0.9002), 4 (0.8284), and 8 (0.8299). We observed differences in performance between using BERT Base Uncased embeddings and BioClinical BERT embeddings, which are most evident for classifier 8, which corresponds to mentions of asthma medications. Although this is counterintuitive at first (we would expect a classifier using embeddings generated from BioClinical BERT to be able to better recognize allergy medications), we believe that the performance benefit from using BERT Base Uncased can be attributed to the list of terms we gave to medSpacy when asking it to identify sentences in category 8. Many of the asthma medications in category 8 sentences are either monoclonal antibody medications ending in -mab (benralizumab, mepolizumab, omalizumab, etc) or hydrofluoroalkanes (hfa; atrovent hfa, flovent hfa, xopenex hfa, etc). Because monoclonal antibodies are very specialized types of medication, they may not occur as frequently as other terms in the corpus used to train BioClinical BERT, so a more general model such as BERT Base Uncased may provide more robust performance. Additionally, because the hydrofluoroalkane allergy medications in category 8 sentences are often abbreviated with “hfa,” which can have alternate medical meanings such as high-functioning autism or health facility administrator, the BioClinical BERT embeddings might not be representative of the presence of allergy medications in the sentence, so a more general model such as BERT Base Uncased may be able to provide better performance.

More broadly, looking at the results in Tables 5 and 6, we can see that the least accurate classifier has an accuracy of 0.7288, while the most accurate classifier is able to achieve an accuracy of 0.9153. Furthermore, when aggregating the most accurate classifiers from both tables we can see that we are able to achieve accuracies of 0.9153 (classifier 1) for identifying sentences that directly suggest the patient has AD, 0.8954 (classifier 2) for identifying sentences that mention hay fever allergies, 0.8214 (classifier 3) for identifying sentences that mention atopic allergies, 0.8439 (classifier 4) for identifying sentences that mention eczema or skin rashes, 0.7373 (classifier 5) for identifying sentences that mention dry or itchy skin, 0.8204 (classifier 6) for identifying sentences that mention nonasthma medications related to diagnosis of AD, 0.7712 (classifier 7) for identifying sentences that mention asthma, and 0.8738 (classifier 8) for identifying sentences that mention asthma medications. Because our training and testing sets were both class-balanced and the majority (6 of the 8) of the most accurate classifiers previously mentioned achieved accuracies between 0.8204 and 0.9153, we believe these results are promising and indicate that our sentence classifiers could potentially be used to save time in a clinical setting during chart review by identifying (and highlighting for review) sentences relevant to the diagnosis of AD when recruiting for clinical trials.

**AD Phenotyping Results**

As per Tables 7-10, our earlier hypothesis holds: using clinical embeddings (BioClinical BERT) to generate the patient vector representation does provide better performance in patient phenotyping than using nonclinical embeddings (BERT Base Uncased). Comparing evaluations on the balanced testing set in Tables 7 and 9, we observe that using BioClinical BERT embeddings provides higher accuracy in almost all models, with the exception of Decision Trees where BERT Base Uncased provides better performance (accuracy of 0.6071) as compared with BioClinical BERT (accuracy of 0.5893). Comparing evaluations on the unbalanced testing set in Tables 8 and 10, we observed that the same trend follows: using BioClinical BERT embeddings provides higher accuracy in almost all models, with the exception of Decision Trees and XGBoost, where using BERT Base Uncased embeddings provides better performance (accuracy of 0.6484 for Decision Trees and 0.6374 for XGBoost) as compared with their counterparts with BioClinical BERT embeddings (accuracy of 0.5824 for Decision Trees and 0.6264 for XGBoost).

As part of our experimental design, we included an ablation study in experiment 3 so we could compare the difference in performance during patient phenotyping when removing the use of BERT models to create each patient’s vector representations. On the class-balanced testing set, we observed that accuracies range from 0.6071 to 0.7321 when using BioClinical BERT embeddings in Table 7, accuracies range from 0.5179 to 0.6250 when using BERT Base Uncased embeddings in Table 9, and accuracies range from 0.6964 to 0.8036 when removing the use of BERT models in Table 11 (experiment 3). On the unbalanced testing set, we observed that accuracies range from 0.5824 to 0.7253 when using BioClinical BERT embeddings in Table 8, accuracies range from 0.5714 to 0.6703 when using BERT Base Uncased embeddings in Table 10, and accuracies range from 0.7143 to 0.7582 when removing the use of BERT models in Table 12 (experiment 3).

In both cases (evaluation on the balanced testing set and evaluation on the unbalanced testing set), we found that models in experiment 3 (ablation study) generally outperform (or are as good as) their corresponding counterparts in experiments 1 and 2 (BERT experiments) across all metrics (accuracy, precision, recall, F1-score, NPV, and specificity), with the exception that the stacking classifier in experiment 1 (BioClinical BERT) has marginally stronger accuracy and precision than the stacking classifier in experiment 3. This shows that traditional rules-based approaches (experiment 3) can outperform BERT-based approaches for generating a patient vector representation for downstream patient phenotyping.

We hypothesize that models in experiments 1 and 2 showed lower performance because errors from our sentence classifiers in earlier stages of the pipeline could have propagated to later stages of the pipeline during patient phenotyping. Because we...
leveraged the max operator to aggregate probabilities that any given sentence in the patient record applies to each category. More sentences in each patient record would lead to a greater chance that an erroneous prediction with a high probability would lead to a false positive error in the creation of each patient’s vector representation in experiments 1 and 2.

Although there is a wide range in performance for our patients with AD phenotyping algorithms, we believe that we have reached our goal of developing a system capable of patient with AD phenotyping for clinical trial recruitment because Tables 11 and 12 show promising results. Furthermore, our system can be used as a first step during AD clinical trial recruitment to filter out most patients who may not qualify for AD trials and therefore save valuable clinician time. We believe our pipeline is important and valuable because, unlike other diseases, such as influenza, COVID-19, and cancer, there is no gold-standard test result that can be used to determine when a patient has AD. Instead, clinicians must spend large amounts of time undergoing chart reviews to individually determine whether each patient has AD.

Limitations

One limitation of this study was the small size of our data set. Although we had a total of 1926 patients in our data set, only 137 of them were validated as having AD. During training, we leveraged 109 of the 137 patients with AD and sampled another 109 patients without AD to create a class-balanced training set. The small size of the training set could lead to overfitting and therefore result in reduced performance on the testing set. Future work could involve obtaining more data from patients with AD as well as exploring the use of an imbalanced data set but using a class-weighted loss function to counteract the class imbalance.

A second limitation of this study was the input-limit size of the large language models that were used. Both BERT Base Uncased and BioClinical BERT had an input limit of 512 tokens. This meant that any input text that was longer than 512 tokens would be ignored when training BERT. Consequently, we could not simply directly concatenate all documents from each patient’s EHR and feed the tokenized documents of each patient into BERT with an added classification head for training as well as direct prediction of whether the patient has AD. Instead, we designed a pipeline around distilling information from all documents in each patient’s EHR into a patient vector representation and then using this patient vector representation to train various classical ML algorithms for phenotyping the patient. Future work could involve exploring the use of other large language models that are suited for long inputs, such as Longformer or Doc2Vec, for predicting when a patient should be labeled as having AD.

A third limitation of this study was the list of AD indicators we selected. We did not consider additional AD indicators, and we also did not consider the use of different combinations (or subsets) of the AD indicators selected. This is particularly relevant in considering that (1) our pipeline is intended to be used for identifying patients with AD, and (2) one of our AD indicators (category 1) directly targets whether there is any given sentence in the patient’s record that mentions AD, which could be in the context of a family history of AD, a potential (but not confirmed) diagnosis of AD, as well as a confirmed diagnosis of AD, among other possibilities. If this AD indicator is removed, then 1 interesting research question could be whether our pipeline is still able to maintain performance similarly to what it is currently able to achieve. Future work could involve assessing the performance impact of removing or adding the use of various AD indicators. We could then determine if our pipeline is relying too much on or overfitting 1 or more indicators. Furthermore, we could also redesign our patient vector and separate the feature for category 1 (any sentence that mentions AD) into 3 separate indicators: whether there is (1) a family history of AD, (2) an affirmed diagnosis that the patient has AD, and (3) uncertainty of whether the patient has AD. Doing so could potentially improve precision.

Potential Applications

Given the aforementioned results, we believe our AD classifier could be operationalized to facilitate reliable and efficient EHR chart review. For example, sentence classifiers could visually indicate AD indicators inline text, therefore reducing information foraging efforts by clinicians. Additionally, AD phenotyping classifiers could indicate the strength of a patient match to UKWP criteria, exact or partial, based on AD indicator sentence classifications. Furthermore, ranking patient cases by match strength could reduce the number of cases reviewed to generate both case and matched controls.

Conclusions

In conclusion, we present and validate a promising pipeline for phenotyping patients with AD during clinical trial recruitment. To do so, we compare a rules-based and transformer-based approach for creating a vector representation of each patient and compare downstream performance in patient phenotyping with various standard ML algorithms. We find that a traditional rules-based approach outperforms using a transformer-based approach (experiment 3). We hope that our pipeline can be deployed in hospital settings during clinical trial recruitment as an initial step to automatically filter candidates before manual review. Additionally, we show that MLP networks can identify whether sentences are relevant to AD diagnosis. These MLP networks can later be deployed in clinical settings to highlight which sentences are relevant for physicians during manual chart review, therefore reducing physician burden. Future work can involve extending our patient phenotyping pipeline to other data sets and other diseases.

Acknowledgments

This study was partially funded by the National Institutes of Health (NIH) and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) P30-AR069589 as part of the Penn Skin Biology and Diseases Resource-Based Center (core: DJM and DM).

https://formative.jmir.org/2024/1/e52200 JMIR Form Res 2024 | vol. 8 | e52200 | p.524
(page number not for citation purposes)
Authors’ Contributions
AW designed the experiments, wrote the code, performed the experiments, wrote the first draft of the manuscript, and revised the manuscript. DJM conceptualized and implemented the chart abstraction study, annotated the data set, interpreted the results, and revised the manuscript. RF annotated the data set and revised the manuscript. SH queried and deidentified the data set as well as revised the manuscript. DM conceptualized the study and experiment design, interpreted results, wrote and revised the manuscript, and provided secure storage and computer resources.

Conflicts of Interest
DJM is or recently has been a consultant for Pfizer, Leo, and Sanofi with respect to studies of atopic dermatitis and served on an advisory board for the National Eczema Association.

References


Abbreviations

AD: atopic dermatitis
AdaBoost: Adaptive Boosting
BERT: Bidirectional Encoder Representations from Transformers
EHR: electronic health record
FN: false negatives
FP: false positives
Hfa: hydrofluoroalkanes
HR: Hanifin and Rajka
ICD: International Classification of Disease
KNN: k-nearest neighbor
ML: machine learning
MLP: multilayer perceptron
NLP: natural language processing
NPV: negative predictive value
Philter: Protected Health Information filter
ReLU: Rectified Linear Unit
SGD: stochastic gradient descent
SVM: support vector machines
TN: true negatives
TP: true positives
UKWP: UK Working Party
XGBoost: Extreme Gradient Boosting

©Andrew Wang, Rachel Fulton, Sy Hwang, David J Margolis, Danielle Mowery. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Examining a Remote Group-Based Type 2 Diabetes Self-Management Education Program in the COVID-19 Era Using the ORBIT Model: Small 6-Week Feasibility Study

Madison S Hiemstra¹, MSc; Sonja M Reichert², MSc, MD, CCFP; Marc S Mitchell¹, PhD

¹School of Kinesiology, Western University, London, ON, Canada
²Centre for Studies in Family Medicine, Department of Family Medicine, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

Abstract

Background: To date, most group-based diabetes self-management education (DSME) programs for type 2 diabetes (T2D) have been delivered in person. The rapid transition to remote care at the outset of the COVID-19 pandemic presented opportunities to test, evaluate, and iterate a new remote DSME program.

Objective: We aim to refine the delivery and evaluation of a multicomponent remote DSME program for adults living with T2D by examining several feasibility outcomes.

Methods: We recruited a convenience sample of patients from a London, Canada, outpatient diabetes clinic (serving high-risk, low-income adults) to participate in a 6-week, single cohort feasibility study from November 2020 to March 2021. This small ORBIT phase 1b feasibility study represents the first in a planned series guided by the ORBIT model for developing behavioral interventions for chronic diseases (phase 1: design; phase 2: preliminary testing; phase 3: efficacy; and phase 4: effectiveness). The feasibility of delivering and evaluating a remote DSME program, including (1) live video education classes, (2) individualized physical activity (PA) prescription and counseling, and (3) intermittently scanned continuous glucose and wearable PA monitoring, was assessed. Feasibility outcomes included recruitment and retention rates, program adherence, and acceptability (ie, technology issues and exit survey feedback). PA was assessed with Fitbit Inspire 2 (Fitbit Inc) and estimated glycated hemoglobin (HbA₁c) using the FreeStyle Libre (Abbot). Given the small study sample, group- and individual-level data are reported descriptively.

Results: A total of 10 adults living with T2D were recruited (female 60%; age 49.9, SD 14.3 years; estimated HbA₁c 6.2%, SD 0.5%). Recruitment and retention rates were 29% and 80%, respectively. Participants attended 83% (25/30) and 93% (37/40) of education classes and PA counseling phone calls, respectively. There were 3.2 (SD 2.6) technology issues reported per person, most of which were related to study data transfer. Exit survey responses suggest most participants (8/9, 89%) were “satisfied” with the program. Recognizing the small sample size and the fact that no inferential statistics were conducted, the mean (SD) for the weekly daily step count and estimated HbA₁c are provided for illustrative purposes. Participants accumulated 7103 (SD 2900) and 7515 (SD 3169) steps per day at baseline and week 6, respectively. The estimated HbA₁c was 6.2% (SD 0.5%) and 6.2% (SD 0.6%) at baseline and week 6, respectively.

Conclusions: This ORBIT phase 1b study served to refine the delivery (eg, automatic study data upload process recommended to reduce participant burden) and evaluation (eg, purposeful sampling of participants with baseline HbA₁c >8% recommended to address selection bias) of a remote DSME program. Preliminary proof-of-concept testing (ORBIT phase 2) incorporating some of these learnings is now warranted.

Trial Registration: ClinicalTrials.gov NCT04498819; https://clinicaltrials.gov/study/NCT04498819
KEYWORDS
activity monitor; diabetes self-management education; flash glucose monitor; glycated hemoglobin; group education; HbA1c; T2D: type 2 diabetes; virtual care; wearables

Introduction

Chronic conditions such as type 2 diabetes (T2D) are best treated when the individual living with diabetes is engaged and supported in effective self-management [1,2]. Programs, such as diabetes self-management education (DSME), that promote successful T2D self-management behaviors can dramatically lower the risk of serious complications [2]. Group DSME programs promote sustained self-care [2]; however, in response to the COVID-19 pandemic, most group DSME programs were quickly transitioned to remote, one-on-one delivery [3]. To support this rapid transition, contemporary technologies were thrust into the forefront, including videoconferencing services (eg, WebEx and Cisco), intermittently scanned continuous glucose monitors (eg, FreeStyle Libre and Abbott’s Diabetes Care Division), and wearable activity trackers (eg, Fitbit) [3]. This preparatory study examined the feasibility of delivering and evaluating remote group DSME programming for adults with T2D in the COVID-19 era. This small cohort study, aligning with phase 1b of the ORBIT model for developing behavioral treatments for chronic diseases [4], is the first in a planned series aiming to systematically develop an efficacious remote group DSME program for broad rollout in Canada (the “LIBERATE” program [5]). The central goal of this work is to refine the delivery and evaluation of a remote group DSME program (the “treatment”) to promote efficiency while producing potentially relevant changes in behavioral (eg, physical activity [PA]) and clinical risk factors (eg, glycemic control) [4].

Methods

Overview

A single-arm feasibility study (ORBIT phase 1b) was conducted between November 2020 and March 2021 in London, Ontario. Ontario COVID-19–related physical distancing policies were in place throughout the study period, with strict stay-at-home orders for 9 out of 16 weeks [6]. Despite well-documented pandemic-related recruitment challenges [7], we sought to recruit a convenience sample of 10 to 20 new patient intakes [8-12] from St Joseph’s Primary Care Diabetes Support Program, a London, Ontario outpatient clinic serving higher-risk, lower-income adults [13]. Prospective participants, physician-cleared to exercise and with access to a smartphone (ie, iPhone 7 iOS or higher or Android [operating system 5 or higher]), were invited to a study recruitment session by their physician during usual care.

Intervention

A 6-week remote group DSME program was delivered to participants between November 2020 and March 2021. A rolling intake was used, with participants completing the program over 6 consecutive weeks. The program included (1) live-video delivery (ie, WebEx) of biweekly group education classes by a multidisciplinary team (ie, physician, diabetes nurse educator, and exercise specialist) and grounded in Diabetes Canada’s Self-Management Education Guidelines [2]; (2) biweekly one-on-one PA counseling phone calls; and (3) enhanced self-monitoring using intermittently scanned continuous glucose (FreeStyle Libre; Abbot) and wrist-worn PA (Fitbit Inspire 2; Fitbit Inc) monitors. The Inspire 2 was selected as it was the most affordable Fitbit model, offering features required for the study (eg, daily step tracking and data exporting). Group education classes and PA counseling calls were designed to help participants learn from the biofeedback they were receiving (eg, draw important linkages between FreeStyle Libre-assessed glucose trends, diet, and PA behaviors) [14]. Fitbit data were used to generate individualized and adaptive daily step count goals [15] (ie, daily step count median from the past 14 days + 500 steps, equivalent to 5 more minutes of brisk walking [16]). The exercise specialist reviewed step goals with participants during biweekly PA counseling calls. Participants were instructed to submit Fitbit data (by downloading a Fitbit Excel file and uploading it to a secure file sharing website) and scan their FreeStyle Libre frequently.

Outcomes

The primary study objective was to refine the delivery and evaluation of the remote group DSME program in preparation for an ORBIT phase 2a study (preliminary testing: proof-of-concept). To do this, (1) recruitment and retention, (2) program adherence (ie, attendance, Fitbit data submission rates, and FreeStyle Libre data capture rates [active time]), and (3) acceptability (ie, number of technology issues reported and exit survey responses) data were collected. “Active time” is the mean biweekly percent of total glucose data captured by the FreeStyle Libre every 24 hours. To examine the potential impact of the program on behavioral and clinical risk factors, device-assessed (4) weekly mean daily step count and (5) estimated glycated hemoglobin (HbA1c: the average glucose level from the FreeStyle Libre readings for 14 or more days [17]) were collected.

Analysis

Given the small, single cohort and preparatory nature of this ORBIT phase 1b feasibility study, individual and group-level data are presented descriptively rather than with inferential statistics.

Ethical Considerations

This study was registered at ClinicalTrials.gov (NCT04498819) and approved by Western University’s Health Science Research Ethics Board (116071). Patients provided informed consent before participating. Deidentified study data were stored on Western University’s password-protected and encrypted OneDrive (Microsoft Corporation).
Results

Among 35 eligible new patient intakes, 10 were enrolled, meeting our a priori recruitment target (Table 1; 29% recruitment rate). Reasons for nonparticipation (n=5) included work-time conflict, a sick partner, being too busy, not wanting to wear the FreeStyle Libre, and feeling exercise and nutrition were well-managed. A total of 2 participants dropped out of the study (ie, participant 1 missed 2 consecutive biweekly group education classes, and participant 8 withdrew during week 5 citing lack of time). A total of 8 participants completed follow-up assessments (80% retention). Regarding program adherence, 83% (25/30) and 93% (37/40) of participants attended group education classes and one-on-one PA counseling phone calls, respectively. Additionally, 53% (16/30) of participants submitted Fitbit data, and the FreeStyle Libre captured 81% of blood glucose data during the intervention period (“Active time”). FreeStyle Libre use may have waned after week 4, as 4 out of 8 (50%) participants had their lowest data capture rates in weeks 5 and 6 (Table 2).

Table 1. Participants’ characteristics.

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographics</strong></td>
</tr>
<tr>
<td>Age (years), mean (SD); range</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
</tr>
<tr>
<td>Ethnicity (White), n (%)</td>
</tr>
<tr>
<td><strong>Highest education level, n (%)</strong></td>
</tr>
<tr>
<td>Less than high school</td>
</tr>
<tr>
<td>High school diploma or equivalent</td>
</tr>
<tr>
<td>College certificate, university, or higher</td>
</tr>
<tr>
<td>Employment status (employed full time), n (%)</td>
</tr>
<tr>
<td>Household income (&lt;US $39,277.30), n (%)</td>
</tr>
<tr>
<td>Car ownership, n (%)</td>
</tr>
<tr>
<td><strong>Relationship status, n (%)</strong></td>
</tr>
<tr>
<td>Single</td>
</tr>
<tr>
<td>Married or equivalent</td>
</tr>
<tr>
<td>Separated, divorced, or equivalent</td>
</tr>
<tr>
<td>Widowed</td>
</tr>
<tr>
<td><strong>Health characteristics</strong></td>
</tr>
<tr>
<td>Years since diabetes diagnosis, mean (SD)</td>
</tr>
<tr>
<td>Estimated baseline glycated hemoglobin (%), mean (SD)</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg), mean (SD)</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg), mean (SD)</td>
</tr>
<tr>
<td><strong>Comorbidities, n (%)</strong></td>
</tr>
<tr>
<td>Dyslipidemia</td>
</tr>
<tr>
<td>Hypertension</td>
</tr>
<tr>
<td>Anxiety</td>
</tr>
<tr>
<td>Depression</td>
</tr>
<tr>
<td>Other psychiatric condition</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
</tr>
<tr>
<td>Heart disease&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Baseline physical activity (steps per day), mean (SD)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Including coronary artery disease, heart failure, or arrhythmia.
Table 2. Biweekly mean estimated glycated hemoglobin (HbA1c) and “Active time” by participant and for the total sample.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline</th>
<th>Weeks 1-2</th>
<th>Weeks 3-4</th>
<th>Weeks 5-6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated HbA1c&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Active time&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Estimated HbA1c&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Active time&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A&lt;sup&gt;d&lt;/sup&gt;</td>
<td>34</td>
<td>N/A</td>
<td>35</td>
</tr>
<tr>
<td>2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.0</td>
<td>99</td>
<td>6.0</td>
<td>94</td>
</tr>
<tr>
<td>3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
<td>52</td>
<td>6.7</td>
<td>67</td>
</tr>
<tr>
<td>4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.1</td>
<td>83</td>
<td>6.0</td>
<td>85</td>
</tr>
<tr>
<td>5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>7.0</td>
<td>89</td>
<td>7.0</td>
<td>94</td>
</tr>
<tr>
<td>6&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.6</td>
<td>97</td>
<td>5.6</td>
<td>98</td>
</tr>
<tr>
<td>7&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
<td>95</td>
<td>N/A</td>
<td>100</td>
</tr>
<tr>
<td>8&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.2</td>
<td>89</td>
<td>6.2</td>
<td>87</td>
</tr>
<tr>
<td>9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.7</td>
<td>87</td>
<td>7.0</td>
<td>84</td>
</tr>
<tr>
<td>10&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.8</td>
<td>66</td>
<td>5.7</td>
<td>50</td>
</tr>
<tr>
<td>Total (%)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.2 (0.5)</td>
<td>77 (22)</td>
<td>6.3 (0.6)</td>
<td>77 (22)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The estimated HbA1c levels are percentages as measured by the FreeStyle Libre.

<sup>b</sup>Active time is the mean percentage of total blood glucose data captured in a 24-hour period.

<sup>c</sup>Participant identification number.

<sup>d</sup>N/A: not applicable.

Participants reported at least 1 technology issue, with 3.2 (SD 2.6) issues reported per person, including difficulties submitting Fitbit data (n=14), lost WebEx remote group education class link (n=6), FreeStyle Libre monitor falling off prematurely (n=7 of 32 sensors distributed in total), difficulty synchronizing Fitbit with participant smartphone (n=3), and losing a Fitbit altogether (n=2). Exit survey responses (Table S1 in Multimedia Appendix 1) suggested participants were satisfied with the program, with most (8/9, 89%) agreeing with the statement, “Overall, I was satisfied with the program.” Exit survey responses also indicated that combined biofeedback from FreeStyle Libre and Fitbit (4/9, 44%) was most helpful in optimizing self-management. Daily step counts for the total sample were 7103 (SD 2900) steps and 7515 (SD 3169) steps at baseline and week 6, respectively (Table 3). Lastly, estimated HbA1c was 6.2% (SD 0.5%) and 6.2% (SD 0.6%) at baseline and week 6, respectively (Table 2).
Table 3. Biweekly daily step count, in mean (SD), by participant and for the total sample.

<table>
<thead>
<tr>
<th>By participant</th>
<th>Baseline, mean (SD)</th>
<th>Weeks 1-2, mean (SD)</th>
<th>Weeks 3-4, mean (SD)</th>
<th>Weeks 5-6, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^a)</td>
<td>N/A(^b)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>2(^a)</td>
<td>9729 (2985)</td>
<td>11,158 (3917)</td>
<td>9521 (4002)</td>
<td>8165 (5222)</td>
</tr>
<tr>
<td>3(^a)</td>
<td>2114 (663)</td>
<td>1790 (585)</td>
<td>1779 (948)</td>
<td>2103 (746)</td>
</tr>
<tr>
<td>4(^a)</td>
<td>9191 (2794)</td>
<td>8001 (1800)</td>
<td>9187 (1654)</td>
<td>8987 (2203)</td>
</tr>
<tr>
<td>5(^a)</td>
<td>6112 (1762)</td>
<td>6154 (1590)</td>
<td>5052 (1526)</td>
<td>5487 (1823)</td>
</tr>
<tr>
<td>6(^a)</td>
<td>8954 (2569)</td>
<td>8979 (1890)</td>
<td>10,286 (1279)</td>
<td>10,111 (1744)</td>
</tr>
<tr>
<td>7(^a)</td>
<td>11,114 (3595)</td>
<td>13,114 (1736)</td>
<td>12,484 (1743)</td>
<td>12,930 (2619)</td>
</tr>
<tr>
<td>8(^a)</td>
<td>4877 (1733)</td>
<td>4355 (1212)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>9(^a)</td>
<td>4721 (2063)</td>
<td>5113 (1195)</td>
<td>5366 (2177)</td>
<td>6511 (2936)</td>
</tr>
<tr>
<td>10(^a)</td>
<td>7115 (2622)</td>
<td>7570 (2741)</td>
<td>6786 (1999)</td>
<td>6634 (4062)</td>
</tr>
<tr>
<td>Total</td>
<td>7103 (2900)</td>
<td>7359 (3485)</td>
<td>7558 (3449)</td>
<td>7515 (3169)</td>
</tr>
</tbody>
</table>

\(^a\)Participant identification number.
\(^b\)N/A: not applicable.

Discussion

Overview

The aim of this study was to refine the delivery and evaluation of a remote group DSME program in preparation for an ORBIT phase 2a study (preliminary testing: proof-of-concept). This DSME program showed promise, albeit with a small convenience sample, with most participants being satisfied with the program. Fitbit and FreeStyle Libre data capture was also high (>80% in both cases). Moreover, preliminary data suggest potential for behavioral risk factor (ie, PA) improvement over a short 6-week period. However, several opportunities to improve the protocol were identified (Table 4). These should be addressed before moving onto the ORBIT phase 2a study. For instance, reducing participant burden by offering automatic study data upload may increase evaluation and program efficiency. Additionally, purposefully sampling participants with higher baseline HbA\(_1c\) (ie, HbA\(_1c\)>8%) \[^{18}\] may help address selection bias and the “floor effect” \[^{19}\] in the future (eg, T2D was generally well-controlled among participants at baseline).

The findings should be considered in light of similar research. Our 29% recruitment rate, for example, was comparable to recruitment rates of 21% to 78% in similar studies \[^{10,11,20-22}\]. Suboptimal recruitment rates could be attributed to (1) limited physician referrals to the study recruitment session, (2) additional participant stressors amid the COVID-19 pandemic (ie, employment insecurity), (3) seasonal effects (ie, colder winter months) \[^{23}\], or (4) perceived study burden. Technology access and cost-related barriers did not appear to limit participation in this lower-income population \[^{21}\]. High program adherence and study retention (>80%) suggest remote delivery of the DSME program was generally well-received in this small convenience sample, aligning with a modest but growing number of studies in this field (eg, remote group education class attendance in similar studies has ranged from 52% to 95%) \[^{10,20}\].

This study is not without limitations. First, our sample was small, consisting of participants with well-controlled T2D (perhaps not representative of those who may benefit most from DSME). However, given the sample sizes of comparable studies \[^{10,11,24}\] and the objectives of this preparatory work, we contend it is appropriate given the conclusions being drawn. Future recruitment rates may improve with the easing of COVID-19-related restrictions \[^{25}\] and as we embark on year-round recruitment in the next study in this planned series. Second, this program was only 6 weeks long, providing little insight into sustained program adherence in a field where attrition is the norm \[^{26}\]. Third, this program was resource-intensive (eg, required significant staff time). Scaling the program (ie, having monthly classes after the first 6 weeks) to reduce resource requirements for a longer program may be necessary. Lastly, discrete exit survey responses only provided so much insight (eg, compared to conducting focus groups \[^{4,9}\]). Moving ahead to ORBIT phase 2a, focus groups will be conducted to gather deeper insights.
### Table 4. Study protocol areas for improvement and recommendations.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Areas for Improvement</th>
<th>Recommendations</th>
</tr>
</thead>
</table>
| Recruitment       | • Low recruitment rate (29%), likely due to a low referral rate to study recruitment sessions by clinicians or the colder winter season  
                    • Potentially high participant burden                                               | • Clearer recruitment instructions for clinicians  
                    • Recruit year-round for the ORBIT phase 2a trial  
                    • Streamline recruitment procedures (eg, in-person sign-up for study recruitment sessions, etc) |
| Sample            | • Possible selection bias with generally well-controlled diabetes (baseline estimated HbA$_1^c$ =6.2%) and technology as participation barriers  
                    • Both insulin and noninsulin users are included                                      | • Purposeful sampling to include baseline estimated HbA$_1^c$ >8% and individuals who do not have ready smartphone access  
                    • Conduct subanalyses of insulin versus noninsulin users                               |
| Technology        | • Issues (14 total) with study data submissions  
                    • Remote classes sound-related “feedback” issue  
                    • Early evidence of FreeStyle Libre attrition, with lower “active time”$^b$ observed for some participants in Weeks 5 and 6 | • Provide the option to email or text a screenshot of the 2-week Fitbit step count summary or implement automatic data upload  
                    • Ensure class facilitators have “mute” capabilities  
                    • Create short booster sessions to encourage exploration of personal FreeStyle Libre data and shared experiences |
| Group class       | • Limited group discussions leading to more didactic education  
                    • Difficulty “reading the room” when telephoning into remote classes  
                    • Rolling intake format                                                               | • Emphasize the importance of sharing and peer learning; provide 1-page content summaries to review before class; and provide assignments to reinforce learning  
                    • Encourage participation through a video platform with the camera on or off  
                    • Remove the rolling intake. Instead, have set start and end dates                  |
| Physical activity | • Daily step count prescriptions only. Participants had a choice on “how” to accumulate the steps  
                    • Cardiovascular fitness outcomes were not assessed  
                    • A daily step count of 500 steps or more was considered a full day worth of data (valid day) | • Offer a choice of time, type, and frequency of a preferred exercise (to compliment the daily step count goal); increase the frequency and automation of feedback  
                    • Use validated field test measures to assess cardiovascular fitness level changes (eg, a 6-minute walk test)  
                    • A “valid day” may include the time between the first and last daily step recorded |

---

$^a$HbA$_1^c$: glycated hemoglobin; an estimation provided from 2 weeks or more of glucose data collected by the FreeStyle Libre.

$^b$Active time is the mean percentage (%) of total glucose data captured in a 24-hour period every two weeks (the FreeStyle Libre requires at least 1 scan every 8 hours to collect the past 8 hours of data).

### Conclusions

This ORBIT phase 1b study has served to refine the delivery and evaluation of this remote DSME program. Proof-of-concept testing is warranted, with plans to progress to ORBIT model phases 3 (efficacy) and 4 (effectiveness). This may ultimately increase access to strong remote-group DSME programming in Canada.

### Acknowledgments

The FreeStyle Libre devices were donated in-kind by Abbott’s Diabetes Care Division. The Fitbit Inspire 2 activity monitors were purchased through funding from the St Joseph’s Health Care Foundation Innovation Grants. The authors would like to thank Betty Harvey for her contribution to the conceptualization of this program and for providing invaluable feedback throughout the intervention’s development. We would also like to thank Amanda Mikalachki for her dedication to the ongoing development and her involvement in running the intervention. Finally, we would like to thank the study participants.

### Data Availability

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

### Authors’ Contributions

MSH, SMR, and MSM were involved in conceptualizing and designing the study. MSH led the intervention delivery, data collection, and data analysis. MSH and MSM were involved in data presentation and visualization. MSH wrote the manuscript with revisions and consultations from MSM and feedback from SMR. SMR led and supervised the intervention delivery on-site and provided guidance and direction on intervention rollout and adjustments to MSH. MSM supervised the project. All authors.
Conflicts of Interest

SMR has received an honorarium for attending advisory boards for Abbott’s Diabetes Care Division and has received an investigator-initiated grant from Abbott’s Diabetes Care Division to begin a clinical trial in 2022. She also holds the Dr Brian W Gilbert Chair in Primary Health Care Research at Western University. MSH and MSM have no competing interests.

Multimedia Appendix 1
Exit survey results.

References
5. Study aims to empower patients with type 2 diabetes to take control of their health. Lawson Research. 2022. URL: https://www.lawsonresearch.ca/media-releases/study-aims-empower-patients-type-2-diabetes-take-control-their-health [accessed 2023-01-03]


Abbreviations

DSME: diabetes self-management education
HbA1c: glycated hemoglobin
PA: physical activity
T2D: type 2 diabetes

©Madison S Hiemstra, Sonja M Reichert, Marc S Mitchell. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Home Blood Pressure Telemonitoring Technology for Patients With Asymptomatic Elevated Blood Pressure Discharged From the Emergency Department: Pilot Study

Karen C Tran\textsuperscript{1,2}, MSc, MD; Meagan Mak\textsuperscript{3}, BA; Laura M Kuyper\textsuperscript{1}, MD; Jesse Bittman\textsuperscript{1}, MD; Birinder Mangat\textsuperscript{1}, MD; Heather Lindsay\textsuperscript{3}, MD; Chad Kim Sing\textsuperscript{1}, MD; Liang Xu\textsuperscript{2}, MSc; Hubert Wong\textsuperscript{2}, PhD; Martin Dawes\textsuperscript{4}, MD; Nadia Khan\textsuperscript{1,2}, MD; Kendall Ho\textsuperscript{3}, MD

\textsuperscript{1}Division of General Internal Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
\textsuperscript{2}Center for Health Evaluation and Outcome Sciences, Vancouver, BC, Canada
\textsuperscript{3}Department of Emergency Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
\textsuperscript{4}Division of Family Practice, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Karen C Tran, MSc, MD
Division of General Internal Medicine
Faculty of Medicine
University of British Columbia
2775 Laurel Street, 7th Floor
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 604 875 5181
Email: karen.tran4@vch.ca

Abstract

Background: Hypertension affects 1 in 5 Canadians and is the leading cause of morbidity and mortality globally. Hypertension control is declining due to multiple factors including lack of access to primary care. Consequently, patients with hypertension frequently visit the emergency department (ED) due to high blood pressure (BP). Telehealth for Emergency-Community Continuity of Care Connectivity via Home-Telemonitoring Blood Pressure is a pilot project that implements and evaluates a comprehensive home blood pressure telemonitoring (HBPT) and physician case management protocol designed as a postdischarge management strategy to support patients with asymptomatic elevated BP as they transition from the ED to home.

Objective: Our objective was to conduct a feasibility study of an HBPT program for patients with asymptomatic elevated BP discharged from the ED.

Methods: Patients discharged from an urban, tertiary care hospital ED with asymptomatic elevated BP were recruited in Vancouver, British Columbia, Canada, and provided with HBPT technology for 3 months of monitoring post discharge and referred to specialist hypertension clinics. Participants monitored their BP twice in the morning and evenings and tele-transmitted readings via Bluetooth Sensor each day using an app. A monitoring clinician received these data and monitored the patient’s condition daily and adjusted antihypertensive medications. Feasibility outcomes included eligibility, recruitment, adherence to monitoring, and retention rates. Secondary outcomes included proportion of those who were defined as having hypertension post-ED visits, changes in mean BP, overall BP control, medication adherence, changes to antihypertensive medications, quality of life, and end user experience at 3 months.

Results: A total of 46 multiethnic patients (mean age 63, SD 17 years, 69%, n=32 women) found to have severe hypertension (mean 191, SD 23/mean 100, SD 14 mm Hg) in the ED were recruited, initiated on HBPT with hypertension specialist physician referral and followed up for 3 months. Eligibility and recruitment rates were 40% (56/139) and 88% (49/56), respectively. The proportion of participants that completed $\geq$80% of home BP measurements at 1 and 3 months were 67% (31/46) and 41% (19/46), respectively. The proportion of individuals who achieved home systolic BP and diastolic BP control at 3 months was 71.4% (30/42) and 85.7% (36/42) respectively. Mean home systolic and diastolic BP improved by $-13/-5$ mm Hg after initiation of HBPT to the end of the study. Patients were prescribed 1 additional antihypertensive medication. No differences in medication adherence from enrollment to 3 months were noted. Most patients (76%, 25/33) were highly satisfied with the HBPT program and 76% (25/33) found digital health tools easy to use.
Conclusions: HBPT intervention is a feasible postdischarge management strategy and can be beneficial in supporting patients with asymptomatic elevated BP from the ED. A randomized trial is underway to evaluate the efficacy of this intervention on BP control.

Methods

Recruitment Procedures

The feasibility pilot study was an unblinded trial. Participants were prospectively recruited from 1 large academic urban ED in Vancouver, British Columbia, Canada, from May to December 2021. We included adults (older than 19 years of age) who presented to the ED with asymptomatic elevated BP confirmed at ED triage and average of 3 subsequent measurements performed in the ED (systolic blood pressure [SBP] ≥ 160 mm Hg or diastolic blood pressure [DBP] ≥ 100 mm Hg), who owned or had daily access to a smartphone, and agreeable to follow-up at the Vancouver General Hospital (VGH) Hypertension Clinic and tele-transmit home BP readings via the Sphygmo app. Individuals with 1 of the following conditions were excluded, patients with hypertensive emergencies with evidence of end organ involvement, stroke or acute coronary syndrome, people who are pregnant, acute intoxication, acute surgical or trauma, patients who are psychiatrically unstable, advanced cognitive impairment, patients requiring admission to hospital, inability to use or care for home BP monitor correctly, from long-term care facility, unstable housing, and are non–English-speaking with no family members who can help translate.

Participants were recruited through 2 recruitment pathways. Potential participants were first identified via referrals from hospital ED staff and were screened by a research assistant in hospital ED (recruitment pathway 1). ED staffs were asked to refer the study team any patients presenting to the ED with suspected hypertension, patients whose BP was above SBP ≥ 160 mm Hg or DBP ≥ 100 mm Hg and in stable condition. Once the research assistant was notified of the patient referrals, the research assistant reviewed their medical records and excluded any patients who met exclusion criteria from screening. Only patients who appear to meet all inclusion criteria and remain potentially eligible were approached for further screening and have subsequent BP readings measured to confirm their eligibility. During the recruitment, the study team expanded referral streams to include patients with asymptomatic elevated BP who were referred to the VGH Hypertension Clinic following discharge from the ED to increase recruitment numbers due to limits on research activities in the ED during the COVID-19 pandemic (recruitment pathway 2). Physicians at the Hypertension Clinic screened all incoming referrals based on inclusion and exclusion criteria, and selected eligible patients for further contact. Eligible patients were contacted either in person or via phone to discuss study procedures and obtain consent. Participants were enrolled if they agreed to follow-up at the Vancouver General Hospital (VGH) Hypertension Clinic and tele-transmit BP readings via the Sphygmo app. Individuals with 1 of the following conditions were excluded, patients with hypertensive emergencies with evidence of end organ involvement, stroke or acute coronary syndrome, people who are pregnant, acute intoxication, acute surgical or trauma, patients who are psychiatrically unstable, advanced cognitive impairment, patients requiring admission to hospital, inability to use or care for home BP monitor correctly, from long-term care facility, unstable housing, and are non–English-speaking with no family members who can help translate.

Participants were recruited through 2 recruitment pathways. Potential participants were first identified via referrals from hospital ED staff and were screened by a research assistant in hospital ED (recruitment pathway 1). ED staffs were asked to refer the study team any patients presenting to the ED with suspected hypertension, patients whose BP was above SBP ≥ 160 mm Hg or DBP ≥ 100 mm Hg and in stable condition. Once the research assistant was notified of the patient referrals, the research assistant reviewed their medical records and excluded any patients who met exclusion criteria from screening. Only patients who appear to meet all inclusion criteria and remain potentially eligible were approached for further screening and have subsequent BP readings measured to confirm their eligibility. During the recruitment, the study team expanded referral streams to include patients with asymptomatic elevated BP who were referred to the VGH Hypertension Clinic following discharge from the ED to increase recruitment numbers due to limits on research activities in the ED during the COVID-19 pandemic (recruitment pathway 2). Physicians at the Hypertension Clinic screened all incoming referrals based on inclusion and exclusion criteria, and selected eligible patients for further contact. Eligible patients were contacted either in person or via phone to discuss study procedures and obtain consent. Participants were enrolled if they agreed to follow-up at the Vancouver General Hospital (VGH) Hypertension Clinic and tele-transmit BP readings via the Sphygmo app. Individuals with 1 of the following conditions were excluded, patients with hypertensive emergencies with evidence of end organ involvement, stroke or acute coronary syndrome, people who are pregnant, acute intoxication, acute surgical or trauma, patients who are psychiatrically unstable, advanced cognitive impairment, patients requiring admission to hospital, inability to use or care for home BP monitor correctly, from long-term care facility, unstable housing, and are non–English-speaking with no family members who can help translate.

Participants were recruited through 2 recruitment pathways. Potential participants were first identified via referrals from hospital ED staff and were screened by a research assistant in hospital ED (recruitment pathway 1). ED staffs were asked to refer the study team any patients presenting to the ED with suspected hypertension, patients whose BP was above SBP ≥ 160 mm Hg or DBP ≥ 100 mm Hg and in stable condition. Once the research assistant was notified of the patient referrals, the research assistant reviewed their medical records and excluded any patients who met exclusion criteria from screening. Only patients who appear to meet all inclusion criteria and remain potentially eligible were approached for further screening and have subsequent BP readings measured to confirm their eligibility. During the recruitment, the study team expanded referral streams to include patients with asymptomatic elevated BP who were referred to the VGH Hypertension Clinic following discharge from the ED to increase recruitment numbers due to limits on research activities in the ED during the COVID-19 pandemic (recruitment pathway 2). Physicians at the Hypertension Clinic screened all incoming referrals based on inclusion and exclusion criteria, and selected eligible patients for further contact. Eligible patients were contacted either in person or via phone to discuss study procedures and obtain consent. Participants were enrolled if they agreed to follow-up at the Vancouver General Hospital (VGH) Hypertension Clinic and tele-transmit BP readings via the Sphygmo app. Individuals with 1 of the following conditions were excluded, patients with hypertensive emergencies with evidence of end organ involvement, stroke or acute coronary syndrome, people who are pregnant, acute intoxication, acute surgical or trauma, patients who are psychiatrically unstable, advanced cognitive impairment, patients requiring admission to hospital, inability to use or care for home BP monitor correctly, from long-term care facility, unstable housing, and are non–English-speaking with no family members who can help translate.

Participants were recruited through 2 recruitment pathways. Potential participants were first identified via referrals from hospital ED staff and were screened by a research assistant in hospital ED (recruitment pathway 1). ED staffs were asked to refer the study team any patients presenting to the ED with suspected hypertension, patients whose BP was above SBP ≥ 160 mm Hg or DBP ≥ 100 mm Hg and in stable condition. Once the research assistant was notified of the patient referrals, the research assistant reviewed their medical records and excluded any patients who met exclusion criteria from screening. Only patients who appear to meet all inclusion criteria and remain potentially eligible were approached for further screening and have subsequent BP readings measured to confirm their eligibility. During the recruitment, the study team expanded referral streams to include patients with asymptomatic elevated BP who were referred to the VGH Hypertension Clinic following discharge from the ED to increase recruitment numbers due to limits on research activities in the ED during the COVID-19 pandemic (recruitment pathway 2). Physicians at the Hypertension Clinic screened all incoming referrals based on inclusion and exclusion criteria, and selected eligible patients for further contact. Eligible patients were contacted either in person or via phone to discuss study procedures and obtain consent. Participants were enrolled if they agreed to follow-up at the Vancouver General Hospital (VGH) Hypertension Clinic and tele-transmit BP readings via the Sphygmo app. Individuals with 1 of the following conditions were excluded, patients with hypertensive emergencies with evidence of end organ involvement, stroke or acute coronary syndrome, people who are pregnant, acute intoxication, acute surgical or trauma, patients who are psychiatrically unstable, advanced cognitive impairment, patients requiring admission to hospital, inability to use or care for home BP monitor correctly, from long-term care facility, unstable housing, and are non–English-speaking with no family members who can help translate.

Hypertension is the leading cause of death and disability worldwide [1,2]. Long-term poor control of blood pressure (BP) can result in significant cardiovascular (CV) morbidity and mortality [3]. Severely asymptomatic elevated BP in the emergency department (ED) has recently been associated with undiagnosed hypertension and adverse CV outcomes [4,5], regardless of the initial reason (eg, pain or anxiety) for the ED visit [6]. In a meta-analysis of 12 studies (n=1240) of individuals with elevated BP in ED, 43.4% were diagnosed with hypertension at follow-up [7]. Among individuals discharged from ED with elevated BP, two-thirds still had uncontrolled BP at 6 months [8].

Hypertension is one of the few chronic conditions that can be monitored online, with home blood pressure telemonitoring (HBPT), which allows tele-transmission of BP in real time to central health portal, eliminating the need for in-person BP visits between patients and health care providers. Importantly, close monitoring of BP and data can be summarized by patients and providers, including calculating of BP averages, graphing temporal BP, and flagging high or low values [9,10]. This is particularly important for patients with asymptomatic elevated BP presenting to and discharged from the ED. Current ED guidelines recommends no routine ED management and outpatient follow-up without suggesting any additional monitoring or intervention with a few exceptions [11,12]. As these patients transition from acute to community settings, their BP can remain dangerously elevated or lowered with initiation of antihypertensive medications. Hypertension telemedicine studies have been shown to be highly feasible, effective, and acceptable to patients [13-17]. Although a promising intervention, to date, no studies have leveraged HBPT as a postdischarge management strategy for patients discharged from the ED with asymptomatic elevated BP.

Telehealth for Emergency-Community Continuity of Care Connectivity via Home-Telemonitoring Blood Pressure (TEC4Home-BP) is a pilot study to evaluate the feasibility of HBPT and physician case management as an integrated component of health delivery to support patients with asymptomatic elevated BP discharged from the ED to home. The primary objective of this pilot study is to report the feasibility of HBPT as a postdischarge management strategy and the secondary objective is to determine acceptability of this initiative. We hypothesize that the study procedures would be feasible and acceptable to patients.
Hypertension Clinic was received. Participants completed the consent process and were enrolled within 7 days from ED discharge.

**Ethical Considerations**

Ethics approval was obtained from the University of British Columbia Research Ethics Board (#H20-03207). The written informed consent was obtained from all participants. All data collected from the participants are de-identified and remain anonymous. A privacy impact assessment was completed to ensure the telemonitoring application is compliant with all University of British Columbia and Vancouver Coastal Health privacy policies. Participants were allowed to keep the BP telemonitor device at the end of the monitoring period as a gift for participating in the study.

**HBPT and Hypertension Clinic Intervention**

All patients were provided with a validated electronic upper arm oscillometric BP telemonitor device (A&D Ltd UA-641BLE) with wireless data transfer (Bluetooth) capabilities using a smartphone [18,19]. The research assistant assisted the patients to set up the Sphygmo app on their smartphones, and connected the BP telemonitor device to their smartphones via Bluetooth. Patients were then taught to follow the on-screen instructions in the app to complete and submit their BP readings, properly measure their BP, and view their current and previous BP readings. They were instructed to perform home BP monitoring schedules recommended by Hypertension Canada and International Society of Hypertension Virtual Management of hypertension guidelines, consisting of duplicate measurements in the morning and evening for 7 consecutive days [20,21]. For each 7-day HBPT series, the first day’s measurements were discarded and the mean of subsequent measurements calculated and used to guide medication titration. Home BP readings were transmitted and telemonitored via Sphygmo app on a smartphone, which is PIPEDA (Personal Information Protection and Electronic Documents Act)-, PIPA (Personal Information Protection Act)-, and HIPAA (Health Insurance Portability and Accountability Act)-compliant, with encryption on both ends and a medical server based in Ontario, Canada. Tele-transmitted home BP readings were monitored daily by a monitoring clinician, who would contact participants by telephone if BP was uncontrolled (SBP ≥ 180, DBP ≥ 100 mm Hg or SBP ≤ 100 mm Hg) according to BP management algorithm (Multimedia Appendix 1). Urgency of hypertension clinic follow-up was dependent on severity of telemonitored home BP readings post-ED discharge (Multimedia Appendix 1).

Patients were seen at the VGH Hypertension Clinic either in person or phone assessment. Physicians assessed patients using standardized hypertension intake forms, administered behavioral counseling, encouraged medication adherence, reviewed telemonitored BP summaries, adjusted BP medications accordingly, and arranged clinical follow-up as needed based on Hypertension Canada guidelines [20]. At minimum, patients were seen at the time of enrollment for initial consultation, and at 1 and 3 months for follow-up. Target home BP values were defined by Hypertension Canada of <135/85 mm Hg or <130/80 mm Hg for those with diabetes [20]. At the end of the study, a summary of participant’s home BP readings were sent to their primary care provider.

**Data Collection and Outcomes**

**Overview**

Baseline data including sociodemographic, ethnicity, education level, health behaviors, and relevant medical history were collected. Antihypertensive medication history was reconciled and number of antihypertensive medications, class of antihypertensive medications, and hypertensive defined daily dose (HDDD) were recorded according to patient self-report at enrollment and at 3 months (Multimedia Appendix 2). HDDD quantitatively describes the intensity of a patient’s overall antihypertensive medication regimen [22]. Medication adherences were assessed using the validated Hill Bone Medication Adherence Scale (HB-MAS) at the time of enrollment and at 3 months [23]. Health-related quality of life (QoL) was assessed using EQ-5D-5L at the end of the study [24]. Satisfaction surveys were sent to participants for completion at the end of 3 months via REDCap (Research Electronic Data Capture; Vanderbilt University; Multimedia Appendix 3). A total of 3 reminder emails were sent to participants to complete the surveys.

Feasibility outcomes were eligibility, recruitment, retention rate, and adherence rate. Eligibility rate was defined as proportion of participants who were deemed eligible to participate among all the patients that were screened. Recruitment rate was defined as the proportion of participants who are deemed eligible and who consented to participate in the study. Retention rate was defined as the proportion of the participants that completed 1 week of HBPT and attended first Hypertension Clinic visit. Home monitoring adherence was defined as the percentage of participants that completed 80% of HBPT at 1- and 3-month follow-up visits.

Secondary outcomes included the proportion of the participants meeting the definition of hypertension at 3 months (defined as mean home SBP ≥ 135 mm Hg or DBP ≥ 85 mm Hg or antihypertensive prescription), mean change in HBPM from enrollment to 1- and 3-month follow-ups, and proportion of participants meeting home BP targets (defined as mean home BP readings of <135/85 mm Hg or <130/80 mm Hg if having diabetes) at 3 months. Additional outcomes of interest were number of antihypertensive medications, HDDD, medication adherence (HB-MAS), health-related QoL (EQ-5D), and patient satisfaction questionnaires.

**Statistical Analysis**

Characteristics of included participants were described as mean, SD, median, IQR, and proportions. BP changes and changes in other study outcomes (baseline to 3 months) were summarized using means and SDs or median and IQR. For participants who had BP recordings within a window from 14 days before to 14 days after the target 90-day follow-up date, the 3-month BP was taken to be the average the BP recordings of up to 3 days closest to the target date. For participants who did not have BP recordings within this window, but who had data over at least 60 days, multiple imputation was used to assign the 3-month BP (see Multimedia Appendix 4 for details). The remaining
participants were excluded from the analysis of BP change as it was deemed not possible to assign them reliable 3-month BP values. Patients who completed either 3-month questionnaire were included in the analysis and descriptive statistics were used. Data were analyzed using Stata (StataCorp) and R (version 4.1.2; R Core Team).

**Results**

**Recruitment**
From May to December 2021, a total of 139 patients presenting to the ED with asymptomatic elevated BP were screened, including 99 identified in ED (recruitment pathway 1) and 40 identified from ED referrals to Hypertension Clinic (recruitment pathway 2). Potentially eligible patients that were referred to the research team by ED staff had an average triage BP of 192/94 and were contacted by the research assistant for additional BP measurements. Among the patients who had additional BP measured by the research assistant, those who were ineligible had an average BP of 147/81 (n=26), whereas eligible patients had an average BP of 194/99 (n=32). Including patients identified and screened through ED referrals to the Hypertension Clinic (recruitment pathway 2), 56 patients were eligible to be consented to the study (Figure 1), and 49 were enrolled. Eligibility and recruitment rates were 40% (56/139) and 88% (49/56), respectively. After enrollment, 3 participants were deemed ineligible and were excluded from the study due to requiring hospitalization (n=3). The final analyzed cohort consisted of 46 participants. Of which, 44/46 patients completed at least 1 week of HBPT and 1 clinic visit at the Hypertension Clinic, resulting in a retention rate of 98% (44/66). Home monitoring adherence was defined as the percentage of participants that completed ≥80% of the prespecified home blood pressure monitoring schedule at 1 and 3 months from enrollment. The proportion of participants that completed ≥80% of home BP measurements at 1 and 3 months were 67% (31/46) and 41% (19/46), respectively (Figure 2). A total of 4 patients did not have sufficient home BP readings at the end of 3 months to determine mean home BP and were lost to follow-up.

**Figure 1.** Flow diagram of patient participant screening and recruitment. BP: blood pressure; ED: emergency department.
Characteristics of Participants

Participants were mostly women (69%, n=32) with an average age of 63 (SD 17) years (Table 1). The ethnic diversity of participants was White (30%, n=14), Filipino (22%, n=10), and Chinese (15%, n=7). At baseline, the mean (SD) SBP and DBP in the ED were 193 (SD 23) mm Hg and 100 (SD 13) mm Hg, respectively. Of all the participants, 44 (96%) had been previously diagnosed with hypertension. Of all participants, 17 (37%) had an antihypertensive medication initiated or increased in the ED. Those who had an intensification of medications had higher mean BP in ED of 205/104 mm Hg compared to those who did not of 186/98 mm Hg.
Table 1. Baseline demographics of enrolled participants.

<table>
<thead>
<tr>
<th>Baseline</th>
<th>All participants (N=46)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographic</strong></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>63 (17)</td>
</tr>
<tr>
<td>Women, n (%)</td>
<td>32 (69)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>14 (30)</td>
</tr>
<tr>
<td>Filipino</td>
<td>10 (22)</td>
</tr>
<tr>
<td>Chinese</td>
<td>7 (15)</td>
</tr>
<tr>
<td>South Asian</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Korean</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Preferred not to answer</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Did not answer</td>
<td>2 (4)</td>
</tr>
<tr>
<td><strong>Health behavior, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>University degree above bachelor’s level</td>
<td>3 (7)</td>
</tr>
<tr>
<td>University degree at bachelor’s level</td>
<td>15 (33)</td>
</tr>
<tr>
<td>University certificate below bachelor’s level</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Trade certificate or diploma from a vocational school or apprenticeship training</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Nonuniversity certificate or diploma from a community college</td>
<td>6 (13)</td>
</tr>
<tr>
<td>High school graduation</td>
<td>11 (24)</td>
</tr>
<tr>
<td>Less than high school graduation</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Preferred not to answer</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Did not answer</td>
<td>2 (4)</td>
</tr>
<tr>
<td><strong>Medical history, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>44 (96)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>9 (20)</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>21 (46)</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Stroke</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Erectile dysfunction</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Vascular dementia</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Antihypertensive medications</strong></td>
<td></td>
</tr>
<tr>
<td>Taking antihypertensive medication, n (%)</td>
<td>10 (22)</td>
</tr>
<tr>
<td>Number of antihypertensive medication, mean (SD)</td>
<td>1.58 (1.32)</td>
</tr>
<tr>
<td><strong>Clinical measures, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Framingham risk score</td>
<td>14.8 (3.9)</td>
</tr>
</tbody>
</table>
Patient Outcomes and Experience

There were sufficient HBPT data to determine the 3-month BP end point for 42 (91%) participants, with imputation used for 1 participant. Mean home SBP and DBP decreased by 67.0 (SD 24.4) and 24.9 (SD 10.3), respectively, from ED triage or screening to study completion. Similarly, mean home SBP and DBP decreased by 13.2 (SD 17.8) mm Hg and 5.1 (SD 9.0) mm Hg, respectively, from the first 7 days to study completion. The proportion of individuals who achieved home SBP and DBP control at 3 months was 71% (30/42) and 86% (36/42), respectively (n=42). The number of adjustments in participant medications was 87 for the entire study, with most commonly having initiation of new antihypertensive medication (36/87), increase in dosage of current antihypertensive medication (23/87), and change in class of antihypertensive medication (14/87). Decreasing antihypertensive medications (9/87) and stopping antihypertensive medications (5/87) were uncommon. At the end of the study, patients were prescribed 1 additional antihypertensive medication (2 vs 1 antihypertensive medication), but no difference in HDDD was noted from enrollment to end of the study (mean 1.58, SD 1.32 vs mean 1.77, SD 1.47; P=0.39; Table 2). Most commonly patients at the beginning of the study were taking angiotensin-converting enzyme inhibitors or angiotensin receptor blockers (40%, 18/45), calcium channel blockers (40%, 18/45), and beta-blockers (24%, 11/45). At the end of the intervention, more patients were prescribed single pill combination antihypertensive medications (9%, 4/45 vs 29%, 13/45; Figure 3).

The response rate for completion of both the HB-MAS at enrollment and the end of the study was 72% (33/46). Among those who completed both questionnaires, no difference in medication adherence was noted (36 [IQR 36-33] vs 36 [IQR 36-35]) from enrollment to 3 months (Table 2). The response rate for completion of the EQ-5D validated questionnaire was 97% (44/46; Table 1). EQ-5D-5L and EQ-5D Visual Analogue Scale (EQ-5D VAS) scores were 0.77 (SD 0.23) and 69 (SD 20), respectively. Due to the pilot nature of the study, adverse events were not captured.

The response rate for the patient satisfaction surveys was 72% (33/46). Overall, patients were highly satisfied with the HBPT program (75%, 24/33) and would recommend it to others (79%, 26/33; Table 3). The majority of the participants found digital health tools easy to use (76%, 25/33) and felt that the intervention prevented the need to return to the ED with elevated BP readings (64%, 21/33). In total, 14% (7/46) of participants required family assistance to participate in the study. The most common reasons for assistance were language barrier (71%, 5/7), inability to apply the home BP monitor on their arm independently (14%, 1/7), and lack of their own smartphone (14%, 1/7). Patients who did not have their own smartphones navigated this barrier by using their family members’ phones.
Table 2. Change in blood pressure and additional secondary outcomes.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Enrollment (N=46)</th>
<th>First week of HBPT&lt;sup&gt;a&lt;/sup&gt; (N=46)</th>
<th>3-month F/U&lt;sup&gt;b&lt;/sup&gt; (N=42)</th>
<th>Change at 3-month F/U (from enrollment; N=42)</th>
<th>P value</th>
<th>Change at 3-month F/U (from first week of HBPT; N=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP&lt;sup&gt;c&lt;/sup&gt;, mean (SD)</td>
<td>193 (23)</td>
<td>140 (16)</td>
<td>127 (12)</td>
<td>-66 (24)</td>
<td>N/A&lt;sup&gt;d&lt;/sup&gt;</td>
<td>-13 (17)</td>
</tr>
<tr>
<td>DBP&lt;sup&gt;c&lt;/sup&gt;, mean (SD)</td>
<td>100 (13)</td>
<td>81 (12)</td>
<td>75 (9)</td>
<td>-25 (10)</td>
<td>N/A</td>
<td>-5 (9)</td>
</tr>
<tr>
<td>Number of antihypertensive medications, mean (SD)</td>
<td>1.58 (1.32)</td>
<td>1.58 (1.32)</td>
<td>1.77 (1.47)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>HDDD&lt;sup&gt;f&lt;/sup&gt;, mean (SD)</td>
<td>1.49 (1.22)</td>
<td>1.58 (1.32)</td>
<td>1.77 (1.47)</td>
<td>N/A</td>
<td>.39</td>
<td>N/A</td>
</tr>
<tr>
<td>HB-MAS&lt;sup&gt;g&lt;/sup&gt; (IQR), mean (SD)</td>
<td>36 (3)</td>
<td>36 (3)</td>
<td>36 (2)</td>
<td>0.66 (2)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>HBPT: home blood pressure telemonitoring.
<sup>b</sup>F/U: follow-up.
<sup>c</sup>SBP: systolic blood pressure.
<sup>d</sup>N/A: not applicable.
<sup>e</sup>DBP: diastolic blood pressure.
<sup>f</sup>HDDD: hypertensive daily defined dose.
<sup>g</sup>HB-MAS: Hill Bone Medication Adherence Scale.

Figure 3. Change in distribution of antihypertensive medication classes from study enrollment to 3-month follow-up in a group of participants with elevated blood pressure (n=46). ACEI: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; BB: beta-blocker; CCB: calcium channel blocker; SPC: single pill combination.
Table 3. The proportion of participants (n=33) that selected “highly satisfied” or “strongly agree” to questions on the participant satisfaction questionnaire following completion of the study.

<table>
<thead>
<tr>
<th>Survey question</th>
<th>Selected highly satisfied and strongly agree, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall satisfaction</td>
<td>24 (75)</td>
</tr>
<tr>
<td>Satisfaction with health care coaching</td>
<td>26 (79)</td>
</tr>
<tr>
<td>Would recommend to other patients</td>
<td>26 (79)</td>
</tr>
<tr>
<td>Easy to use</td>
<td>25 (76)</td>
</tr>
<tr>
<td>Better management of health conditions</td>
<td>24 (73)</td>
</tr>
<tr>
<td>Satisfaction with progress toward health goals</td>
<td>22 (67)</td>
</tr>
<tr>
<td>More informed about chronic conditions</td>
<td>21 (64)</td>
</tr>
<tr>
<td>Less need to visit ED&lt;sup&gt;a&lt;/sup&gt;</td>
<td>21 (64)</td>
</tr>
<tr>
<td>Improvement in QoL&lt;sup&gt;b&lt;/sup&gt;</td>
<td>17 (52)</td>
</tr>
<tr>
<td>Family satisfied with care provided</td>
<td>17 (52)</td>
</tr>
<tr>
<td>Less need to visit family doctor</td>
<td>16 (49)</td>
</tr>
</tbody>
</table>

<sup>a</sup> ED: emergency department.

<sup>b</sup> QoL: quality of life.

Discussion

Overview

TEC4Home-BP is a pilot study to evaluate the feasibility of HBPT and physician case management as an integrated component of health delivery to support patients with asymptomatic elevated BP discharged from the ED to home. We found that HBPT was feasible as a postdischarge management strategy, given our recruitment and retention rate was 88% (56/139) and 98% (44/46), respectively. Patients reported high acceptability and satisfaction with the HBPT program.

Principal Results

HBPT combined with timely physician follow-up and management is a novel and promising postdischarge management strategy to help bridge the transition from ED to home and can play a proactive role in treating asymptomatic elevated BP in the ED. Our results show that the proportion achieving home SBP and DBP control at 3 months was 71% (30/42) and 86% (36/42), respectively, which is higher than the current Canadian hypertension control rates of 65% [25]. Furthermore, home BP decreased from the first week of completing HBPT to 3 months (final BP assessment) by −13/−5 mm Hg. This may correspond to an estimated 30% and 26% risk reduction in CV disease and stroke, respectively, but likely overestimated due to the limitations of our study design with regression of the mean and lack of a comparator group [26]. Overall, patients only required 1 additional antihypertensive medication to be prescribed. Improvement in BP may be due to increased use of evidence-based medications, including single pill combination therapies or the act of self-monitoring. Patients were highly satisfied with the program and found the technology to be user-friendly.

Comparison With Prior Work

Previous studies have shown that text-messaging services to encourage measuring home BP using a wrist cuff was a feasible intervention in patients who were discharged from the ED with high BP and reduced SBP by 9.1 (95% CI 1.1-17.6) mm Hg, but patients did not use remote HBPT program where home BP readings were transmitted to physicians to review and provide management strategies directly to patients [27]. Our results showed that the adherence to home BP monitoring was 67% (31/46) and 41% (19/46) at 1 and 3 months, respectively. This is similar to meta-analysis showing that among 13 studies (n=1662 patients), the average adherence to telemedicine-based hypertension management was high (76.8%; range 48%-90%) [28]. This is not surprising that as time persists from their initial ED event, patients may not be as adherent to strict BP monitoring. It will be important for our future randomized controlled trial (RCT) to develop a more pragmatic home BP monitoring schedule that is acceptable to patients and provide accurate information to health care professionals to monitor and manage their hypertension.

Previous studies from ED and inpatient hospital settings have shown that in patients discharged with asymptomatic elevated BP, 43.4% were diagnosed with hypertension at follow-up [7]. Our study showed that among individuals presenting the ED with an average SBP ≥ 160 mm Hg and DBP ≥ 100 mm Hg in ED, most had established hypertension that was uncontrolled and required additional treatment. This reaffirms that elevated BP in the ED is not benign and a significant proportion of these individuals have undiagnosed or undertreated hypertension. Of similar importance, is the portion of patients that do not go on to have a diagnosis of hypertension and are at risk of misdiagnosis and unnecessary treatment without close follow-up. Furthermore, with worsening hypertension awareness and control rates in Canada, the ED may be a useful location to screen individuals for hypertension, as BP is measured in all individuals regardless of the presenting complaint. Despite the
guideline recommendations that close follow-up is needed for individuals with asymptomatic elevated BP discharged from ED [12], only 7%-29% of patients with elevated BP receive any hypertension assessment, treatment, or referral in the ED [29-32]. Therefore, the development of novel postdischarge management pathways to ensure that these individuals have a close follow-up for their hypertension is greatly needed. At a minimum, ED practice guidelines should consider changing to at least recommend home BP monitoring for these patients immediately postdischarge and timely follow-up with physicians should BP remain elevated.

Our intervention improved BP after 1 week of HBPT to 3 months (final BP assessment) by –13/-5 mm Hg. Meta-analyses have shown that home BP monitoring supported by co-interventions (including medication adjustments by physicians or pharmacists, education, and lifestyle counseling) results in significant BP reduction (~6.1, 95% CI –9.0 to ~3.2 mm Hg) that persists for 12 months [16]. Another meta-analysis of 23 RCTs (n=7037 patients) reported that HBPT significantly reduced BP by 5/3 mm Hg compared to the usual care (P<.001) [17]. Our results need to be verified by conducting a powered RCT to address issues with regression to the mean and lack of a comparator group.

Studies have shown that HBPT alone improved antihypertensive medication prescription and QoL [17,33], but we were not able to show this in our pilot study as it was not powered to detect these differences. Although there were no differences in HDDD prescribed, there was an increase in the number of single pill combination antihypertensive medications prescribed. We hypothesize that HDDD did not change because single pill combinations were more frequently prescribed which confers better BP lowering with combination therapy than full doses of antihypertensive medication [34]. Hypertension follow-up after an ED visit for asymptomatic elevated BP has been shown to improve evidence-based hypertension management [35]. Single pill combinations are endorsed by hypertension guidelines [20] as they have been shown to improve medication adherence and BP control while reducing medication side effects and CV events [36-39].

Our intervention was noted to be highly acceptable and usable by patients and their families, which is similar to other HBPT studies where patients reported higher satisfaction and greater convenience compared with usual hypertension care [40]. Almost 15% (7/46) of participants required assistance with our intervention, specifically due to language barriers. Given our multiethnic population, future RCT design should incorporate different languages in the technology to improve usability for everyone.

Limitations

There are a number of limitations in this study. The sample size for the TEC4Home-BP pilot study was relatively small and there may be systematic differences between those who enrolled for the study and adhered to the monitoring protocol and those who did not. Although the retention rate for participants attending the first visit was high, adherence to the monitoring schedule declined throughout the study and 4 patients were lost to follow-up. This insight on monitoring frequency has provided us with information on how to redesign on monitoring protocol to improve patient adherence. There is increasing evidence that a minimum of 12 BP readings is needed for home BP measurements to be valid [41], and more recently 3 days of BP readings appeared sufficient to prognostic home BP readings [42]. Furthermore, the exclusion criteria excluded many patients who presented to the ED with elevated BP and only stable patients who would be a good candidate for HBPT were recruited for the study. This limitation prevents our feasibility study from assessing the feasibility of HBPT in the full group of patients who present to the ED with elevated BP, and who may still benefit from and accept HBPT. The eligibility criteria may be underestimated, as we did not document all the reasons why patients were not eligible to be in the study due to resource issues. A majority of patients who had elevated triage BPs were not eligible to participate because their BP was below the inclusion criteria after the research assistant measured their BP 3 times consecutively. In addition, those who did not meet the incomplete survey questionnaires may limit the true acceptability and usability of this intervention, as participants who responded and did not respond may be very different. Finally, the lack of data from a control comparison is a true limitation of our feasibility study design and needs to be addressed. Although, these findings are promising and informative of the next steps, a more rigorous RCT design with a comparator group is required to test the true efficacy of this intervention.

Conclusions

HBPT with physician management is a feasible and acceptable postdischarge management strategy to monitor patients with asymptomatic elevated BP when they are discharged from the ED. Future multicenter RCTs are needed to evaluate the efficacy of this intervention in a large population.

Acknowledgments

We thank the TEC4Home team for answering our many questions. We also thank Vancouver Coastal Health Research Institute, University of British Columbia and Vancouver General Hospital Foundation, Vancouver General Hospital, University of British Columbia, and Sphygmo app for their support for this project.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.
Conflicts of Interest
JB has received speaker fees from Novo Nordisk, Bausch Health, and AstraZeneca; a travel bursary from Novo Nordisk; and an unrestricted grant from Bayer. JB is also on the advisory board for Novo Nordisk, Bausch Health, and Bayer.

Multimedia Appendix 1
Physician decision management tool.
[DOCX File .91 KB - formative_v8i1e49592_app1.docx]

Multimedia Appendix 2
Hypertension defined daily dose.
[DOCX File .20 KB - formative_v8i1e49592_app2.docx]

Multimedia Appendix 3
Participant survey.
[DOCX File .28 KB - formative_v8i1e49592_app3.docx]

Multimedia Appendix 4
Statistical methods.
[DOCX File .13 KB - formative_v8i1e49592_app4.docx]

References


**Abbreviations**

BP: blood pressure  
CV: cardiovascular  
DBP: diastolic blood pressure  
ED: emergency department  
EQ-5D VAS: EQ-5D Visual Analogue Scale  
HB-MAS: Hill Bone Medication Adherence Scale  
HBPT: home blood pressure telemonitoring  
HDDD: hypertensive defined daily dose  
HIPAA: Health Insurance Portability and Accountability Act  
PIPA: Personal Information Protection Act  
PIPEDA: Personal Information Protection and Electronic Documents Act  
QoL: quality of life  
RCT: randomized controlled trial  
REDCap: Research Electronic Data Capture  
SBP: systolic blood pressure  
TEC4Home-BP: Telehealth for Emergency-Community Continuity of Care Connectivity via Home-Telemonitoring Blood Pressure  
VGH: Vancouver General Hospital
Acceptability of a Self-Led Mindfulness-Based Intervention for Teens with Type 1 Diabetes: Pilot Randomized Controlled Trial

Tori Humiston1, MA; Caroline Cummings2, PhD; Stephen Suss3, MS; Laura B Cohen1, MA; Holly Hazlett-Stevens4, PhD; Amy Hughes Lansing4, PhD

1Department of Psychological Sciences, University of Vermont, Burlington, VT, United States
2Department of Psychological Sciences, Texas Tech University, Lubbock, TX, United States
3Department of Psychology, Florida International University, Miami, FL, United States
4Department of Psychology, University of Nevada, Reno, NV, United States

Corresponding Author:
Tori Humiston, MA
Department of Psychological Sciences, University of Vermont
2 Colchester Ave
Burlington, VT, 05401
United States
Phone: 1 8026562670
Email: tori.humiston@uvm.edu

Abstract

Background: Diabetes distress among adolescents with type 1 diabetes has been associated with suboptimal diabetes outcomes, including lower quality of life, increased diabetes self-management challenges, and suboptimal glycemic outcomes.

Objective: This study examined the feasibility and acceptability of a scalable self-led mindfulness-based intervention to reduce diabetes distress in adolescents with type 1 diabetes.

Methods: Adolescents (N=25) aged between 14 and 18 years diagnosed with type 1 diabetes completed a baseline assessment. Participants were randomized to receive a 10-week self-guided mindfulness-based stress reduction workbook program (e-book or paper option) immediately (n=15) or after a 10-week wait (n=10). During the intervention period, participants completed weekly assignments and feedback surveys. At 10 weeks and 20 weeks, follow-up assessments were completed.

Results: Findings indicated that participants did not find the original intervention feasible or acceptable. Adolescents reported barriers to completing the weekly material, such as that they forgot or that the material was not sufficiently related to their diabetes management. Adolescents also reported that a digital format rather than a workbook or e-book may be more acceptable. Results from weekly surveys provided the foundation for recommendations for future iterations of the mindfulness-based intervention for adolescents with type 1 diabetes.

Conclusions: Participant feedback informed recommendations for self-led mindfulness programs for youth with type 1 diabetes. Adolescents indicated that a shorter, digital mindfulness-based intervention focused on diabetes-specific behaviors may be more helpful.

Trial Registration: ClinicalTrials.gov NCT05115175; https://clinicaltrials.gov/study/NCT05115175

(JMIR Form Res 2024;8:e45659) doi:10.2196/45659

KEYWORDS
adolescents; diabetes distress; diabetes; health group intervention; intervention; mindfulness; psychosocial intervention; self-led mindfulness; type 1 diabetes

Introduction

Adolescents with type 1 diabetes, an autoimmune disease that destroys insulin-producing cells in the pancreas, must engage in numerous daily health behaviors to manage their symptoms and prevent short- and long-term health complications [1]. The persistent challenges of disease management often result in diabetes distress, characterized as the burden, worries, and frustrations associated with diabetes [2]. Adolescents with greater diabetes distress experience lower quality of life, increased diabetes self-management challenges, and suboptimal
glycemic outcomes (glycosylated hemoglobin; HbA1c) [3,4]. Heightened diabetes distress may affect type 1 diabetes outcomes both indirectly, through decreased engagement in disease self-management behaviors, and directly, through the effects of physiological arousal, hormone secretion, and insulin resistance on blood glucose levels and the microvascular system [5,6]. Current interventions aimed at fostering diabetes self-management in adolescence have targeted distress management [7] and quality of life [8], yet less research has examined mindfulness-based stress reduction (MBSR) interventions in the context of adolescents’ type 1 diabetes distress and self-management. This is despite cross-sectional data supporting the link between greater mindfulness, greater diabetes self-management, lower diabetes stress, and lower HbA1c levels in adolescents with type 1 diabetes [9].

Increased mindfulness, enhanced through mindfulness-based interventions, is linked to numerous health benefits. One example is the MBSR intervention, a mind-body public health group intervention developed for medical patients to manage the stress of chronic medical conditions. MBSR consists of eight 2.5-hour-long, weekly sessions and 1 all-day practice session [10]. In adults with type 1 or type 2 diabetes, the benefits of mindfulness-based interventions include reduced distress and cardiovascular risk [11,12], and similar outcomes emerged for adolescents with chronic diseases generally [13]. More recently, mindfulness-based interventions have been considered in the context of improving type 1 diabetes management in emerging adults [14]. For example, among emerging adults with type 1 diabetes and suboptimal glycemic levels, an MBSR intervention was found to improve psychosocial outcomes but not glycemic outcomes and was highly acceptable [15,16]. Further, an additional study within the type 1 diabetes context found that a brief self-compression intervention, which is a core component of MBSR, was acceptable and feasible among adolescents with disordered eating and improved mindfulness and coping [17]. These findings support the possible benefits of improved mindfulness in reducing diabetes distress and improving diabetes self-management in adolescents with type 1 diabetes.

The goal of this study was to begin an iterative approach to developing a self-led, scalable MBSR intervention for adolescents with type 1 diabetes. The intervention was modeled on a 10-week bibliotherapy MBSR program implemented by Hazlett-Stevens and Oren [18] that was found to be feasible and acceptable for college students and to reduce distress and anxiety for those who completed the program. A self-led MBSR program approach may be particularly helpful for disseminating MBSR to adolescents with type 1 diabetes, a population that already experiences the intensive time burden of diabetes management [19]. This study examined the acceptability, feasibility, and potential utility of a 10-week self-led workbook (e-book or paper option) MBSR intervention for adolescents with type 1 diabetes through a randomized waitlist control design. It was hypothesized that participation in a self-led MBSR intervention would be feasible and acceptable, as evidenced by low treatment attrition and positive participant feedback. Finally, we provide recommendations for future iterations of self-led, digital MBSR interventions for adolescents with type 1 diabetes based on recommendations by participants.

Methods

Ethical Considerations

All procedures and documents were approved by the institutional review board at the University of Nevada, Reno (1221205). Web-based parental consent was obtained for children aged 18 years or younger, in which case child assent was also obtained. Participants 18 years of age or older provided web-based consent. Participants earned US $10 for each assessment (up to US $30), US $10 for completing at least 6 of the 10 weekly surveys over the 10-week intervention period, and an additional US $10 for completing all 10 weekly surveys. The maximum earnings for each participant were US $50, in the form of an Amazon gift card. All measures were completed electronically through Research Electronic Data Capture (REDCap; Vanderbilt University). Participants were informed during the informed consent process that their answers and data would not be shared with individuals outside of the research team unless the research team was concerned about the participant’s safety.

Participants

Participants for this study included a sample of 25 adolescents (n=14, 56% female) aged between 14 and 18 years (mean age 16.25, SD 1.6 years) from urban and rural areas of Nevada. Most participants identified as White (n=22, 88%), 1 participant as Asian, 1 participant as Native Hawaiian or Pacific Islander, and 1 participant as biracial. Participants reported a wide range of diagnosis lengths, from less than a year to 16.15 years (mean diagnosis length 5.3, SD 4.1 years). A total of 19 (76%) participants reported using a continuous blood glucose monitor, and 17 (68%) endorsed using an insulin pump. Additionally, through self-report, 10 (40%) participants reported that they qualify for free lunch at school. Inclusion criteria included being an adolescent (aged between 13 and 19 years) with a type 1 diabetes diagnosis currently attending school or being a recent high school graduate. Participants were excluded if they were wards of the state, had severe psychiatric disturbances (eg, active psychosis), or had severe developmental delays that hindered their ability to self-report. Participants were not excluded based on the length of their type 1 diabetes diagnosis.

Procedures

Participants were recruited with flyers at a regional diabetes camp, flyers sent to community diabetes support groups, and direct recruitment by research staff in a local pediatric endocrinology clinic. All enrolled participants completed a baseline survey, including a self-report of their most recent HbA1c percentage value, and were randomized to either begin the 10-week intervention period immediately or after 10 weeks, with participants randomized to start the intervention after the 10-week waiting period acting as a control group. Randomization was computerized and stratified based on sex, duration of type 1 diabetes diagnosis (2 years or less vs more than 2 years), and most recent HbA1c (8.5% or below vs 8.6% or above). Participants also completed assessment questionnaires 10 weeks and 20 weeks after the study start date and a weekly survey for the 10 weeks of intervention participation.
Recruitment for this project began in the fall of 2018 and was completed in early spring 2020. Research staff were trained by 2 graduate students on recruitment procedures and eligibility criteria. A total of 64 participants contacted our research staff, indicating interest in participating in the study and providing permission to contact them either on the internet or in-person (Figure 1). Of the 64 participants who indicated interest in the study, 5 declined further screening, and 2 were not eligible after screening. During the study period, after 29 participants had consented and enrolled, it was determined, due to feedback from those enrolled participants as well as attrition and loss of follow-up, that the intervention required revision to increase acceptability; accordingly, recruitment was stopped at that time. An additional 28 participants expressed interest in enrolling but did not complete the consent process before the interim study end point was reached; this was also before the original goal of 60 participants was reached. Participant recruitment and enrollment are further discussed below as part of the analysis of feasibility and acceptability.
Intervention Program

The mindfulness-based intervention in this study was delivered through a teen MBSR workbook, either an e-book or paper workbook, and web-based communication across a 10-week period. If selected, the paperback workbook was mailed to participants’ home address; otherwise, access to the e-workbook was provided through email. Participants were assigned weekly readings and activities from an MBSR workbook [19]. Topics included understanding stress, an introduction to mindfulness, and mindfulness-based exercises recommended to be completed daily (e.g., mindful eating). Mindfulness-based exercises were either self-led per instructions provided in the workbook or completed using audio-recorded instructions. Participants received emails twice per week with reminders about the current week’s assignments and to practice mindfulness each day. At the end of the week, participants received a reminder to complete a brief survey on the acceptability of that week’s content. This

![ CONSORT (Consolidated Standards of Reporting Trials) table of participation.](https://formative.jmir.org/2024/1/e45659)
survey included both multiple-choice questions and open-ended questions regarding how helpful participants found each exercise and what hindered them from completing the exercises that week. Participants completed the following measures at baseline and immediately following intervention completion.

**Measures**

Multiple measures were administered in the study. Only those relevant to this study are described below.

**HbA1c Percentage**

At intake, participants self-reported their most recent HbA1c level. Higher HbA1c percentages are associated with less optimal glycemic control in the past 2-3 months.

**Intervention Acceptability and Feedback for Recommendations**

Acceptability was measured both objectively based on participant engagement as well as per participant report. First, attrition was used as a proxy variable for acceptability. Second, at the end of each week of the intervention, participants completed both multiple-choice and open-ended questions pertaining to the intervention content that week. Information from weekly surveys was used to create recommendations for future intervention development. Specifically, the following questions were asked: “How much of the suggested readings from the book did you read over the past week?” (0=”not at all” to 5=”I read all of the suggested readings”) “What percentage of exercises from the suggested readings for the previous week did you complete while reading the chapter?” (0=”not at all” to 5=”I read all of the suggested exercises”) “How often did you engage in the exercises suggested in the chapter during the previous week?” (0=”none” to 3=”almost daily”) “What got in the way of you following the workbook?” (0=”I forgot”; 1=”I didn’t have time”; 2=”I didn’t understand the material,” “the material was too difficult to follow,” or “the instructions were unclear”; or 3=”other: explain”) and “Please tell us any other thoughts you had on the workbook readings, weekly exercises, or audio tracks this week.” (open-ended).

**Analyses**

To assess the feasibility and acceptability of the self-led MBSR intervention for adolescents with type 1 diabetes, we conducted descriptive statistics of level of attrition and quantitative questions from participants’ weekly feedback surveys. Qualitative feedback from the weekly feedback surveys was brief and informally and visually analyzed and summarized. We created recommendations for a more acceptable mindfulness-based intervention based on the feedback from the weekly surveys.

**Results**

All feasibility and acceptability analyses pointed to issues with the acceptability and feasibility of the initial intervention model. First, consistent with problems with study feasibility, many participants did not complete the full study period and were lost to follow up. Among the 29 enrolled participants, 25 completed their baseline survey and were randomized to intervention now (n=15) or waitlist (n=10); groups were not balanced as enrollment was stopped early. A total of 4 of the 29 participants passively declined to complete the baseline survey and were not randomized. Overall, 2 participants withdrew from the study after intake (reasons: busy with school and study participation was stressful). Additionally, 2 participants were lost to follow-up during the program, and 7 participants were lost to follow-up at 10 weeks (3 from intervention-first and 4 from the waitlist group). Another 6 were lost to follow-up at 20 weeks, 3 of those from the intervention first group that was now completing extended follow-up. This suggested challenges with the waitlist design at a minimum, as well as with maintaining participant engagement in the study process.

Second, we examined participant feedback on both reasons for attrition and low engagement. The indicated reasons for attrition included the conflict of busy schedules and the weekly time commitment for the intervention. In the weekly surveys, participants reported, through a Likert scale, barriers to engaging in the weekly material. Options for barriers were “I forgot,” “I didn’t have time,” “I didn’t understand the material or instructions,” and “other.” Average frequencies of selected barriers indicated that the most common barrier to engaging in the material was lack of time (30/120, 41.7%). Average frequencies for the 3 other barriers included 21.7% (26/120; “I forgot”), 23.3% (28/120; “other”), and 13.3% (16/120; “I didn’t understand the material”). When participants selected “other,” they were prompted to fill in a textbox describing the barrier. For the participants who did enter a reason in the textbox, responses included that they felt they did not need a mindfulness practice and noted that diabetes-specific activities would be more helpful than general activities. Additional responses included that the workbook intervention format reflected school tasks and that some of the reading and activities felt geared toward a younger audience. Textbox responses were brief, categorized by 2 graduate students, and reviewed by the principal investigator. However, due to the brevity of responses, no formal codebook was used to code responses. Feedback on acceptability and feasibility was integrated, and recommendations for a digitally tailored MBSR intervention for teens with type 1 diabetes were generated by the study team, including experts in adolescent diabetes management and mindfulness-based interventions.

**Discussion**

**Overview**

This study evaluated the feasibility and acceptability of a self-led MBSR intervention. Although there was no support for the feasibility and acceptability of the intervention model, participants provided feedback on multiple improvements that would enhance the feasibility and acceptability of the intervention program. Participants’ weekly feedback pointed to multiple domains where an MBSR intervention might be tailored to the type 1 diabetes context. Hurdles remain in understanding the feasibility and acceptability of these programs, especially in more scalable, self-led forms. For example, in the type 1 diabetes context, there are limited studies examining the feasibility of mindfulness-based interventions for young persons.
of color with type 1 diabetes populations [15]. Additional research on mindfulness-based interventions for adolescents with type 1 diabetes indicates that brief, digital sessions may be more acceptable and feasible for this population [17]. Further research is needed to clarify the long-term benefits and acceptability of mindfulness-based interventions on diabetes distress and glycemic outcomes for adolescents with type 1 diabetes.

Our research has identified multiple potential areas for increasing the acceptability of a self-led mindfulness program for adolescents with type 1 diabetes. In this intervention, feasibility and acceptability assessments indicated that adolescents prefer a shorter, web-based intervention focused on mindfulness regarding diabetes-specific behaviors. We have since identified 5 potential avenues for future mindfulness intervention development in the adolescent type 1 diabetes context that would address the concerns raised by the adolescents in our sample while maintaining a connection to mindfulness-based therapy approaches. First, decrease the amount of content in the self-led form to emphasize the connections between everyday mindfulness, acceptance skills, and diabetes management. Some participants indicated that the original modules and instructions were difficult to understand. Therefore, it may be beneficial to focus the program on a few basic mindfulness skills that are connected to diabetes management to help build an insightful mindfulness repertoire that is meaningful to daily diabetes management. For example, the early sessions of a self-led mindfulness-based intervention might teach an introduction to mindfulness with a guided body scan aimed at increasing self-compassion around diabetes symptoms and tasks. Body scans are a foundational mindfulness practice in MBSR and can be modified to emphasize self-compassion for physiological sensations and experiences that occur during diabetes management (wearing diabetes technology or experiencing hyper- or hypoglycemia). Following this early session, adolescents might be guided to engage in a formal practice audio recording of a diabetes-oriented body scan. Follow-on sessions might continue this same training and expand the experience with a diabetes-oriented body scan with new attention toward self-monitoring and glycemic awareness (symptoms of hypo- or hyperglycemia). Increasing interoceptive awareness through guided meditation may foster adolescent awareness of hyper- and hypoglycemia. Further, acceptability and feasibility may increase if the intervention were delivered through a website or mobile app.

Limitations
This study is limited by multiple factors. The sample that consented to the study was predominantly White and therefore is not generalizable to other races and ethnicities. The study was also limited in the scope of the time period in which participants were recruited and completed the study. Some participants had already completed the study before the COVID-19 pandemic, while others completed the study during the pandemic, which may have confounded the stress and mindfulness levels of the participants. Due to the sample size, we were not able to analyze the data for potential differences across participants who were enrolled before and during the COVID-19 pandemic. Most of the participants reported an
HbA<sub>1c</sub> that was 8.5% or lower, which may limit the generalizability of the recommendations to adolescents with a higher HbA<sub>1c</sub>. Although not asked in this study, it may be helpful for future studies to ask participants about their previous history with mindfulness and meditation.

Conclusions and Future Directions

While keeping the limitations in mind, the findings of this study provide important data to contextualize the content and delivery of mindfulness-based interventions for adolescents with type 1 diabetes and provide guidance for developing an acceptable and scalable self-led mindfulness-based intervention. First, mindfulness and stress-related processes may be particularly important to understand in the context of type 1 diabetes management, given the long-term health complications that involve the vascular system. For example, the primary cause of death in middle-aged and older adults with type 1 diabetes is cardiovascular disease [24], and cardiovascular disease as well as other macro- and microvascular complications [25] contribute to an almost 17-year decrease in life span for those with an early age of type 1 diabetes onset [1]. Mindfulness-based interventions have been shown to also diminish chronic biological stress regulatory system activation that directly contributes to cardiovascular disease [26], highlighting the importance of developing mindfulness-based interventions that are acceptable and feasible for adolescents and emerging adults with type 1 diabetes. Further research is needed to examine the long-term physiological effects associated with mindfulness-based interventions for individuals with type 1 diabetes and the potential cardio-protective benefits.

Second, the data from study participants provided critical information for guiding the development of scalable self-led mindfulness-based intervention models, including emphasizing content that directly links mindfulness practice to diabetes management with the inclusion of diabetes-specific mindfulness activities. Further research examining the acceptability and effects of self-led mindfulness-based interventions that meet the recommendations provided herein and within a sample of youths that includes persons of color and with socioeconomic disadvantages is needed to better understand and support research on mindfulness interventions and type 1 diabetes management in adolescents.

Finally, mindfulness-based interventions might have a role in addressing internalizing psychopathology that is known to impact glycemic outcomes and quality of life for adolescents with type 1 diabetes. Depressive symptoms may contribute to an environment where adolescents experience decreased energy, leading to decreased engagement in diabetes-management behaviors. Decreased energy then contributes to a coercive cycle where decreased glycemic control contributes to greater depressive symptoms [27]. Adolescents may also experience anxiety symptoms, such as worry. Rechenberg and colleagues [28] found that adolescents with type 1 diabetes reported increased worry about managing their health, particularly regarding hypoglycemia and correct insulin dosing. Mindfulness-based interventions may be useful in disrupting this cycle by reducing the underlying stress associated with many psychological concerns. Further research is needed to identify if subpopulations of adolescents with type 1 diabetes, such as those with internalizing symptoms, might particularly benefit from self-led mindfulness-based interventions.

Acknowledgments

This work was supported by a grant to the AlterMed Research Foundation’s coprincipal investigators, HH-S and AHL.

Data Availability

The data sets generated and analyzed during this study are not publicly available as participants were informed during the informed consent process that their answers and data would not be shared with individuals outside of the research team unless the research team was concerned about the participant’s safety.

Conflicts of Interest

None declared.

References


Abbreviations

HbA1c: glycosylated hemoglobin
MBSR: mindfulness-based stress reduction
Original Paper

“I Just Wanted a Dentist in My Phone”—Designing Evidence-Based mHealth Prototype to Improve Preschool Children’s Oral and Dental Health: Multimethod Study of the Codevelopment of an App for Children’s Teeth

Waraf Al-yaseen1*, BDS, MSci, DipME, MDPH, PhD; Daniela Procida Raggio1, DDS, PhD; Mariana Araujo2, DDS, PhD; Nicola Innes1*, BDS, BMSc, PhD

1School of Dentistry, Cardiff University, Cardiff, United Kingdom
2FDI World Dental Federation, Geneva, Swaziland
*these authors contributed equally

Corresponding Author:
Waraf Al-yaseen, BDS, MSci, DipME, MDPH, PhD
School of Dentistry
Cardiff University
Heath Park
Cardiff, CF14 4XY
United Kingdom
Phone: 44 07413162068
Email: Al-yaseenw1@cardiff.ac.uk

Abstract

Background: Dental caries in preschool children is a global health concern. With increased access to technology and the disruption of health care during the pandemic, mobile health apps have been of interest as potential vehicles for individuals’ health maintenance. However, little is known about caring for their child’s teeth and what their preferences would be regarding the content or design of an oral health app.

Objective: This study aims to co-design the prototype of an app named App for Children’s Teeth with parents, providing a source of information for them about caring for their children’s teeth and promoting positive dental habits.

Methods: This multimethod study conducted user involvement research with a purposive sample of parents or carers of children aged ≤6 years to (1) understand their use of the internet through the eHealth Literacy Scale and interviews, (2) determine their opinions about content related to children’s oral health, and (3) collect feedback about the app’s acceptability using the Theoretical Framework of Acceptability. There were three stages: (1) interviews with parents to understand their needs, preferences, and abilities; (2) prototype design with app developers; and (3) parent feedback interviews using the think aloud method for data collection. Data were deductively analyzed using a codebook strategy, whereas data from the think aloud sessions were analyzed inductively using reflexive thematic analysis.

Results: The prototype design stage involved 10 parents who reported using the internet for health information but found it to be scattered and contradictory. Parents generally welcomed the App for Children’s Teeth but expressed concerns about screen time and practicality. They suggested guidance regarding oral hygiene practices, teething symptoms, and pain relief. Parents appreciated features such as clear fonts, categorization according to their child’s age, and “In a Nutshell” bullet points. Topics that resonated with parents included information about teething, finding a dentist, and breastfeeding. They believed that the app aligned with their goals and offered suggestions for future developments, such as outlining the process of finding a dentist and incorporating a forum for parents to communicate and exchange ideas.

Conclusions: The coproduction design approach highlighted parents’ need for solutions such as mobile health apps to access reliable information about oral health. Parents identified key design concepts for the app, including a simple and uncluttered interface, content categorization according to their child’s age, and practical guidance supported by visual aids. Despite potential challenges related to screen time restrictions, parents provided insights into how such an app could fit seamlessly into their lives.

Trial Registration: Open Science Framework; https://osf.io/uj9az
KEYWORDS
oral health promotion; mobile health; mHealth; children; oral health; behavior change; coproduction; mobile phone

Introduction

Background
Dental caries is the most common chronic disease of childhood, affecting >621 million children worldwide [1]. It negatively affects the quality of life of children and their parents [2]. The condition, dental caries, is biofilm mediated, diet modulated, multifactorial, and noncommunicable. It is dynamic in nature; it can start and stop depending on external factors (such as biofilm removal), and it can progress and regress (demineralization and remineralization), again depending on external factors such as the amount and frequency of sugar in the diet [3], removal of the biofilm, and fluoride application to the teeth during toothbrushing [4]. Evidence-based guidelines recommend that parents take responsibility for their children’s daily activities, care, and food choices to prevent dental caries [5]. These guidelines emphasize managing the external factors through personal care, including toothbrushing with fluoridated toothpaste (>1000 parts per million) and sugar reduction in food choices [5,6].

In recent years, the development and growth of mobile health (mHealth) apps have provided an opportunity to improve access to health information and promote health and well-being using mobile technologies [7]. mHealth has been defined by the World Health Organization as the “use of mobile and wireless technologies to support the achievement of health objectives” [8]. Smartphones are ubiquitous, with an estimated 88% of adults owning one in the United Kingdom in 2021 [7]. Internet use exclusively through smartphones is increasing, and in 2022, overall, 21% of users accessed the internet exclusively through their phones. As such, there is potential for high-quality and engaging health information to be easily accessible and affordable through mHealth. Although there is evidence that mHealth apps can play a role in improving health outcomes, including for oral health and dental plaque control [9], the quality and accuracy of internet-based health information is variable, with potential for harm from misinformation.

To maximize the potential of mHealth apps, it is crucial to prioritize user engagement by involving target stakeholders such as patients, caregivers, and health care providers in the co-design process. Co-design fosters collaboration and leverages the valuable insights of users and designers in the development of products or services. By incorporating user feedback and preferences into the design process, mHealth apps can better meet the needs of their target audience. Theoretical Framework of Acceptability (TFA) is a valuable tool to guide the co-design process. TFA proposes that the acceptability of any intervention, such as an mHealth app, is determined by its perceived appropriateness, effectiveness, and feasibility [10]. TFA provides an evidence-based checklist to assess these individual components in an mHealth app prototype, allowing deficiencies to be addressed and ensuring that it meets the needs and preferences of the target stakeholders, and guides the next steps in the development process.

As co-design is a user-centered process, it considers the target audience’s specific characteristics. Context, education, age, and eHealth competence of the end users, among other factors, can significantly influence the design and structure of mHealth apps. The eHealth Literacy Scale (eHLS) is a tool that measures individuals’ ability to access, comprehend, and use health information to make health care judgments. By considering the eHealth literacy of the target stakeholders, an app can be designed to be more appropriate, accessible, and user-friendly, thus improving the likelihood of it being used and, ultimately, its effectiveness [11].

Aims and Objectives
This study aimed to provide proof of concept for designing eHealth technologies in collaboration with the parents of young children, as end users, by designing a prototype of a medical health app to identify gaps in information, develop content, and collect feedback about acceptability.

The objectives were to conduct user involvement studies with parents or carers of children aged ≤6 years at the initial stage to do the following:

- Understand their use of the internet through eHLS and interview
- Determine their opinions about mHealth app content related to children’s oral health

Then, at a later stage of development of the medical health application, we aimed to do the following:

- Collect feedback about the acceptability of the medical health application according to the domains of TFA

Methods

Design
This was a single-site, multimethod, qualitative research study that evaluated parents or carers’ acceptability of a smartphone app prototype.

Ethical Considerations and Data Management
A favorable ethical opinion from Cardiff School of Dentistry Research Ethical Committee (reference number 2210a) was provided by the Dental School Research Ethics Committee on September 13, 2022. The study was conducted from October 2022 to May 2023. The project ensured participant safety and privacy by conducting interviews in safe and private locations.
implementing measures to keep both participants and researchers safe, and maintaining confidentiality of personal information. If sensitive information or harm was disclosed, the research team collectively decided about the further steps and documented the process. Participants seeking dental advice were directed to National Health Service (NHS) Wales. No identifiable information was used in reports or research papers. Study protocol was registered on the Open Science Framework.

All information collected from participants during the study was maintained strictly confidential, and any personal information they provided was managed according to Cardiff University requirements and following General Data Protection Regulation recommendations for data protection (2018) [12]. Records and study documents were stored in a password-protected folder within the university’s OneDrive (Microsoft Corp) cloud space and were accessible only to the research team. Paper-based consent forms were stored in a secure, locked drawer within a locked room on the university premises. Anonymized information and consent forms will be retained for 5 years before being destroyed. However, it could be retained indefinitely when it is likely to have continuing value for research purposes. None of the individuals were identifiable from the data in the reports provided to the funders and partners or in the published research papers.

**Participants**

**Sample Size**

A sample size was not set a priori with recruitment continuing until the research team agreed that the data set was sufficiently comprehensive and rich to address the research objectives. On the basis of similar research projects, the sample size was expected to be approximately ≤10 parents or carers [13].

**Inclusion Criteria**

Adults (aged ≥18 y) who (1) were parents or carers of children aged ≤6 years; (2) were willing to attend 2 in-person interviews; and (3) considered themselves to be fluent in English, Arabic, or Portuguese were eligible to participate.

**Exclusion Criteria**

Anyone who was working or had worked in the dental profession in any capacity was not eligible to participate.

**Procedure**

**Overview**

An overview of the project design is shown in Figure 1. In this study, a multimethod approach was used to comprehensively investigate the research question. Multiple qualitative methods were used, consisting of structured interviews conducted during the 2 stages of data collection, complemented by think aloud methods during feedback interviews. The structured interviews were designed in accordance with TFA, enabling a systematic exploration of various acceptability dimensions. Concurrently, think aloud methods provided real-time insights into participants’ interactions with the app. This integrated approach aimed to thoroughly assess parental perceptions about app acceptability and elucidate the underlying factors. The structured interviews offered a systematic framework, whereas think aloud methods provided spontaneous insights, augmenting the depth and validity of the analysis.

The study used structured interviews as the primary data collection method. These interviews followed a standardized format, with predetermined questions aligned with the research objectives and the TFA. The structured interview approach ensured consistency across all interviews and facilitated a comprehensive exploration of acceptability dimensions. The interviews were conducted in 2 stages to capture the initial impressions and feedback after app use and were set up in one-to-one format between the participant and one of the research team members (DPR or WA). They were conducted at a mutually agreed time and in either a quiet room in the university or at a community center, depending on participants’ preferences. Both were appropriate for conducting interviews—safe, accessible, and private.
Phase 1: Recruitment

Overview

A purposive recruitment strategy was adopted, in which 2 different routes were targeted (Cardiff University employees and attendees of a community center). A web-based invitation was posted on the “Taking Part in the Research” Cardiff University Yammer and on Cardiff University’s social media pages and shared within the researchers’ local community networks (Multimedia Appendix 1). Recruitment from Grangetown Community Centre was conducted by attending public engagement events, such as community fairs and local meetings, and by directly approaching the attendees. The researchers worked with the staff at the community center to help coordinate the recruitment and data collection efforts.

Participant Screening, Information Sharing, and Consent Process

Figure 1 describes the process further. The web-based invitation for both groups referred interested individuals to a web-based questionnaire for further information (Multimedia Appendix 2). This also acted as a first-line screen to guide potential participants regarding whether they met the inclusion criteria and asked them to give their contact details. A member of the research team contacted each person via phone or email within 2 weeks to perform further screening to identify those meeting the inclusion criteria and to explain the study, its aim, that it involved one-to-one interviews with the researchers, and any risks and benefits anticipated by participating in the study.

If they were eligible and still interested in participating, the participant information sheet (Multimedia Appendix 3) and a
consent form (Multimedia Appendix 4) were sent to the potential participants via email for them to review before meeting with the researchers during the face-to-face interview.

During the interview meetings, the interviewers discussed the study and checked the participants’ understanding, and a paper-based consent was signed if they were still agreeable to participate, before conducting the interview.

**Phase 2: Requirement Definition and Design Interview**

For the first data collection session (design interviews), participants were invited to complete a demographic questionnaire and the eHLS questionnaire. The interviewer explored with the participants the way they currently found and accessed information related to oral health, experience with using technology for health purposes, and potential ideas for the technical development of the app (Multimedia Appendix 5).

**Phase 3: Develop a Prototype With an App Development Team**

Participants’ inputs were analyzed and shared with the International Dental Federation (Fédération Dentaire Internationale) research team and the Swiss Tomato app development team to conduct brainstorming discussion meetings. The multidisciplinary team examined the results, proposed content and features for the app, and developed the blueprint for the app. The development team used the blueprint to further develop the prototype.

**Phase 4: Feedback Interview (Assess the App’s Acceptability)**

For the second data collection session (feedback interviews), the researchers maintained the same arrangements as in the first session and conducted the interview in person at a suitable and quiet location. Before the interview, a member of the research team contacted the participants to confirm the location, time, and date of the interview.

During the interview, the interviewer presented the prototype to the participant as a web page and explained how to use it (Multimedia Appendix 6). The data were collected using 2 strategies. First, “Think Aloud,” where the participants were asked to navigate the prototype while expressing their thoughts and perceptions out loud. In addition, the interviewer asked questions based on the TFA theory at this stage. These questions may have focused on the user’s attitudes and perceptions toward the prototype, such as its perceived usefulness, ease of use, and intention to use the system.

**Data Handling and Analysis**

Recordings were transcribed verbatim using a Cardiff University–approved service. The research team anonymized all personal data collected from or about the participants, except for signed consent forms. Personal information in transcripts that could identify the participants was masked with pseudonyms or omitted if it did not affect the transcript’s context.

Descriptive analysis was used for the participant demographic questionnaire and eHLS data. Qualitative data were analyzed separately for each data collection session using NVivo software (Lumivero) under the Cardiff University license and then interpreted together using a complementary approach. Codebook thematic analysis was used for the design interviews and TFA data [10]. Data from think aloud transcripts were analyzed inductively using reflective analysis.

Participants’ demographic and eHLS details were presented in frequencies. Qualitative data were presented according to the topic summary, with direct quotations from participants’ accounts.

**Results**

**Participants and Interviews**

**Demographics**

The study included 10 parents, with 8 (80%) from Cardiff University and 2 (20%) from Grangetown Community Centre. Table 1 presents their respective demographic characteristics.
Table 1. Participant demographics and interview summary information (n=10).

<table>
<thead>
<tr>
<th>Characteristics and groups</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Woman</td>
<td>8 (80)</td>
</tr>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>2 (20)</td>
</tr>
<tr>
<td>35-44</td>
<td>8 (80)</td>
</tr>
<tr>
<td>≥45</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Self-identified ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>White British</td>
<td>7 (70)</td>
</tr>
<tr>
<td>White non-British</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (20)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Attended college course</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Attended university course</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Other postgraduate education</td>
<td>2 (20)</td>
</tr>
</tbody>
</table>

**Interview Characteristics**

All participants (10/10, 100%) participated in interview 1. The interviews ranged from 17 to 58 (mean 29.2, SD 11.4) minutes and lasted for a total duration of 292 minutes. However, for interview 2, a participant was unable to attend any of the proposed dates or times owing to a family illness. Therefore, 90% (9/10) of the participants was interviewed during this phase. The duration of the interviews varied, ranging from 23 to 52 (mean 31.1, SD 11.4) minutes and lasting for a total of 280 minutes.

**Phase 2: Understand Parents’ Needs and Their Thoughts About and Experience With mHealth Apps**

Parents were asked to fill an eHLS survey before the first interview. The survey was completed by 90% (9/10) of the participants. Most of them (7/9, 78%) thought that the internet was useful or very useful in helping them to make health decisions, and 89% (8/9) of the parents felt that it was important or very important for them to be able to access health resources on the internet (Multimedia Appendix 7).

The remaining 7 questions (Figure 2) showed that most parents (7/10, 70%) agreed or strongly agreed with the positive statements about their abilities related to using the internet for health purposes.

**Figure 2.** Participants’ responses to the eHealth Literary Scale questions (Qs) 3 to 10 (n=9).
Thematic Analysis of Interviews

Overview

The topic guide prompted the parents to discuss several key areas. First, they shared insights about their current sources of information, highlighting both the positive and negative aspects of the different sources they used. Second, they reflected upon their previous experiences with mHealth apps. Finally, they expressed their opinions about the development of an app to assist them in caring for their children’s teeth. During the discussions, the parents freely discussed a range of topics, which could be categorized into three main groups: (1) general comments, encompassing the perceived need for such an app and its compatibility with their lifestyles; (2) desired content and features they would like to see in the app; and (3) preferences regarding the app’s design and visual elements.

Current Sources of Information

Parents stated that they consulted many sources for a broad range of oral health information, with a focus on teething, oral hygiene practices, and identifying oral conditions. These searches were mainly through the internet, especially Google and government websites such as the NHS website. Although they perceived this as an easy-to-access and readily available tool to find information and give peace of mind regarding a topic, they highlighted several issues, including the often scattered and sometimes contradictory nature of the information on the internet:

...It’s so segregated, so almost every question you have is a separate search. [Parent 8]

Another issue they felt was that the information they found was very general in terms of the management without appreciating the variety of normal presentations:

...The problem with looking on the Internet for a picture of a condition is you’re often shown the worst example and also, you’re not showing the full spectrum of possible images...various health conditions can present in various different ways. [Parent 4]

They also felt that technical details about oral hygiene advice were often lacking:

The main things around the technicalities of it [Brushing]...Everything from selection of toothbrush, the firmness of bristles, the length of brushing and whether the age range that it’s safe for a child to brush themselves. [Parent 6]

There was a perceived lack of information on what parents thought were reliable sources of information, such as NHS website, especially about common oral issues that affected them, such as teething. Web-based forums such as Mumsnet and Facebook groups such as Gentle Parents were also commonly used, providing interactive, specific tips from other parents. A participant found Facebook groups to be particularly helpful for finding green alternatives to oral products:

They are [Gentle Parents Facebook group], generally very ethical and environmentally aware and quite into natural products rather than more standardised. [Parent 9]

When it comes to credibility of other parents’ advice, parents used their experience to judge:

...If it resonates with my experience, then I would probably go ahead with that. [Parent 9]

Parent’s Previous Experience With mHealth Apps

The parents who were interviewed shared a diverse range of experiences with mHealth apps. Mental health apps, particularly the Headspace app, were the most common types of apps that parents had used. Some parents stated that they downloaded the app because their university provided free access to the “Paid” service, which they considered to be a good deal. Some parents mentioned that they still occasionally used the app as it provided easy-to-follow relaxing exercises, and they appreciated receiving regular updates and notifications. Other popular apps among the parents were the NHS COVID-19 app, which provided instant help about laboratory tests and nearest testing centers during the pandemic, and apps that focus on weight loss and tracking physical activity. None of the interviewed parents (0/10, 0%) reported using or looking for an app to help them find health information.

Parents’ Thoughts About the Development of an mHealth App for Oral Health

General Comments

Participants were positive about having access to an app to assist with children’s dental care, citing the need for a reliable and convenient source of information. They talked about potential challenges to using it. Several parents (2/10, 20%) questioned the app’s effectiveness, as people can easily search for answers to their questions on Google. They also worried about whether users would remember to use the app regularly. A possible solution was suggested by a parent: incentivizing app use by giving users “gold stars” for using it every day to encourage regular use. Another issue discussed was the challenge of integrating the app into busy morning routines. Finally, a parent noted the potential difficulty of using the app alongside other devices or apps and fitting it into their lifestyle but also tried to address these challenges with suggestions:

I still haven’t quite worked out how we could fit it in, in the morning because it’s so staggered...So, the whole idea of teeth time is this time in the morning is much more difficult to kind of work out rather than we’re all heading towards bed and therefore we’re all doing our teeth. [Parent 8]

Screen time concerns were highlighted by some parents as the app is intended to be used by both parents and their child:

I think a lot of parents in this day and age struggle already with the amount of screen time. So that would be a bit of a two-edged sword for me. [Parent 5]

Content Suggestions

Suggestions for the content of the app include accessible guidance to help identify potential oral health problems and about how to manage them. They particularly mentioned...
teething, including timings, managing symptoms, providing relief, and understanding when things were normal or not in development:

I just wanted a dentist in my phone...I can check whether I need to go and see the doctor, or this is something I shouldn’t be worried about. [Parent 3]

They also wanted guidance and more precise details about how to perform oral hygiene practices, such as when to start brushing their child’s teeth, introducing flossing, how to handle brushing when the child is using inhalers, and importance of supervised brushing for proper cleaning:

Stuff about how you do it [brushing]? at what age do you let them have more control over it? Till when am I still doing it? Is the whole 2 minutes is still a thing if they don’t have all their teeth? [Parent 7]

Design Suggestions

Parents suggested incorporating notifications, videos, engaging games, stories, and brushing timers into the app. They emphasized the need for brushing timers to be in the form of a song as children may not have a proper perception of time:

Kids don’t actually have the ability to understand time...I would say five more minutes till we got to leave the park. That doesn’t actually necessarily mean anything to a child, but following a song they would be able to know that they’re in the middle of the song. [Parent 6]

Parents stressed the importance of using concise and straightforward language in the app and requested a simple design with the ability to return to the main screen at any time. They also suggested categorizing the app according to age for more targeted guidance. Incentives to download the app were discussed, with recommendations from health care professionals and trusted individuals being the most popular suggestions:

The midwives, they give you like a list of some good apps on new-born babies or these are good apps for tracking how often they’re feeding. [Parent 2]

I go to mother and baby groups, and they always have like posters or information. I guess if I saw something in that environment, I’d kind of be like, Oh, okay, maybe this is something that I should look at. [Parent 3]

Phase 3: Design and Develop a Prototype

The design stage of the study resulted in a prototype of an app named App for Children’s Teeth (ACT) based on parents’ views and their suggestions for content and design (Figure 3) The app was aligned with the Health Belief Model [14], including features to increase parents’ knowledge and self-efficacy regarding their children’s oral health practices as follows:

1. Perceived susceptibility— informs parents about the risks associated with poor oral health practices for their child’s overall health
2. Perceived severity—highlights the consequences of dental diseases if left untreated
3. Perceived benefits—educates parents about the benefits of proper oral health practices and incentivizes children with the stamps feature to brush their teeth regularly for better oral health outcomes
4. Perceived barriers—addresses common barriers to proper oral health practices, such as lack of knowledge, time, and resources, and provides strategies to overcome them, such as finding a local dentist
5. Cues to action—includes reminders and notifications to help parents establish good oral hygiene habits for their child
6. Self-efficacy—provides guidance and resources to help parents feel confident in their ability to perform proper oral health practices and prevent dental diseases from developing
Phase 4: Assess the App’s Acceptability

Demographics
All participants, except for a woman parent (9/10, 90%), were interviewed for feedback about the app. Demographic information was similar to that in stage 1.

Main Findings
These data were gathered using the think aloud methodology by asking parents to share their thoughts about the design app prototype as they used it.

Content
The subjects that parents particularly wanted to see incorporated into the app included information about teething, how to brush children’s teeth, how much toothpaste to use, and what kind of toothpaste and toothbrush should be used. They also wanted a feature included to help them find a dentist. Although it was not possible to include everything in the prototype, they recognized when something they had asked about appeared in the app:

> When I had my daughter, one of the things that I have researched the most was teething symptoms, signs, everything like that. The fact that it’s there and signs of it, how to alleviate. Amazing, love that. [Parent 3]

Design
Parents felt that the structure fitted with their suggestions and it was simple, easy to navigate, and user-friendly. Having a unique character (the beaver) was perceived as a good way to engage the children with the app activities. A parent suggested allowing the child to name their beavers to enhance familiarity and, hence, engagement. Stamps provided upon completing brushing was also another feature that they felt would improve children’s engagement.

Other positive feedback about the design was having a “home button” on the screen all the time, which allows the users to go back to the main screen regardless of wherever they are. The font used (Sans Serif font) was found to be clear; however, a parent suggested checking the accessibility of the text and whether it is sufficiently clear for people with neurodiverse needs.

During the feedback interview, parents found several aspects of the app design to be useful, including the “In a Nutshell” bulletin points at the beginning of each subject, topics being categorized according to child’s age, and the ability to set reminders. In addition, some parents suggested sending notifications about new updates to keep parents engaged with the app for long term but with the ability for parents to control the frequency of these notifications.
You have reminders for toothbrushing. That’s really good for an older child you can get them to put their reminders on. I think they’ll like that. You could also have some notification, if new research comes out, for example, on a topic that you have in...not daily because people are likely to turn them off. [Parent 9]

**TFA Questionnaire Analysis**

As part of assessing the acceptability of the app, the TFA constructs were analyzed using the parent responses (Table 2).
<table>
<thead>
<tr>
<th>TFA constructs</th>
<th>Definition</th>
<th>Relevance to this study</th>
</tr>
</thead>
</table>
| Affective attitude | “Experienced Affective Attitude: How an individual feels about the intervention, after taking part” | • Parents thought that the app prototype satisfied parents’ needs for oral health education, but because it is only a prototype, here is scope for improvement:  
  • “It’s very intuitive and I think parents will find this really helpful and easy to use. Obviously, there’s a few things to complete.” [Parent 9] |
| Ethicality | “The extent to which the intervention has good fit with an individual’s value system” | • Parents found the app to be aligned with what they value about their children’s oral health:  
  • “I would say so, especially this day and age, everything is so busy. I rely on my phone for many things. It’s helpful to have an app that offers options for setting reminders and schedules.” [Parent 6]  
  • “The app is prioritising children, I think that’s brilliant.” [Parent 7] |
| Burden | “Experienced burden: the amount of effort that was required to participate in the intervention” | • Thoughts about using an app such as ACT on a routine basis are variable; some parents thought that they can use such an app, but it will not be realistic to use it on a routine basis:  
  • “I don’t think a parent is likely to read about oral hygiene every day or every other day. You’ll read it if there’s an issue.” [Parent 5]  
  • Others felt that it provides convenience owing to its structure and being an app:  
  • “No not at all, I think something that you can download on your phone, you can choose a convenient time when you want to look at it and you can just access it anytime and just look at the tabs that you want to, I think it’s really useful you’ve got all of the mini tabs, just to quickly click into, right at the front page, so you don’t even have to go digging for information because you can just choose what information you want to access straight away really, so yes I think it would be really useful.” [Parent 5] |
| Opportunity costs | “Experienced opportunity cost: the benefits, profits or values that were given up engaging in the intervention” | • Participants feel that using such a tool will be beneficial to learn about oral health:  
  • “I think it will be easier because usually, whenever it’s a health-related issue about my children, I just Google things, but now I know there is one place for all information.” [Parent 1]  
  • It can also be a useful option if the dentist is not accessible:  
  • “If I couldn’t get on a list to go to a dentist which is a big problem for lots of people, it would maybe feel like, well, I can’t do that but at least- Here’s an app that’s giving me information that doesn’t feel like a stretch, I’ve downloaded this app so I’m doing something.” [Parent 9] |
| Perceived effectiveness | “Experienced effectiveness: the extent to which the intervention is perceived to have achieved its intended purpose” | • ACT prototype was found to achieve its goals albeit having different priorities, but most parents emphasized the need to keep the app updated to continue achieving its objectives:  
  • “I’ve learned two things just by looking at it. But the information should be regularly updated, I suppose.” [Parent 7] |
| Self-efficacy | “The participant’s confidence that they can perform the behaviour(s) required to participate in the intervention” | • Parents have different views about whether they feel confident that using the ACT app will keep them informed. Most cautious opinions were around whether the app follows the child’s growth:  
  • “You’d have to see more. It certainly will keep me informed because I have a resource to go to, but it would depend on the full functionality. As the child ages up, does the information change to stay current? Does it prompt you that maybe you should be buying a different type of toothbrush at this age?” [Parent 3] |
Suggestions for Further Development

A wide range of suggestions was made for the app’s next stages of development, but one of the main suggestions was around finding a dentist:

*If you could outline the process of applying for a dentist, and the approximate wait times and things like that? I know that a lot of parents won’t be thinking about it until they need to do it.* [Parent 4]

Other suggestions were to enhance the app by addressing teeth shedding time, creating forum spaces for parent communication, and integrating augmented reality for guided brushing.

Discussion

Principal Findings

This study found that the interviewed parents felt there was a need for a reliable tool that they had confidence in, to assist them in managing their children’s oral health. This resonates with the worldwide trend of people using health apps to improve public health. With >55,000 health apps available globally on Google Store [15], these tools have, in some cases, been carefully evaluated to see how well they help with things such as diet [16,17], physical activity [18], or managing blood sugar [19]. mHealth apps are the most popular downloads from app stores.

They felt this would be particularly useful for young children who experience multiple landmarks during dental development such as teething, breastfeeding, and starting to brush their teeth.

An mHealth app for oral health was considered by them to be a viable possibility in filling this gap and would be easily accessible and useful for parents. Although a loose topic guide was used for the interviews, the parents were given the opportunity to talk broadly about oral health information, its sources, and the potential use of apps. Unsurprisingly, engaging the target population helped us understand their needs and preferences, and the iterative nature of the design process emphasized the need to continuously refine the prototype based on feedback. This is important in the development of digital health interventions such as mobile apps, which require a user-friendly interface and a seamless user experience to be effective.

Health Literacy and Participants’ Demographics

Health literacy is a relatively new concept. It is “linked to literacy and entails people’s knowledge, motivation and competencies to access, understand, appraise and apply health information in order to make judgments and take decisions in everyday life concerning health care, disease prevention and health promotion to maintain or improve quality of life during the life course” [20]. It is known to be linked to better health outcomes, as individuals with high levels of health literacy are more able to navigate health care systems, advocate for themselves, and make good health and prevention-based choices [20]. Approximately half of the adults in the United States have been found to have a low or marginal level of health literacy [21], with similar findings in Europe [22]. With both the risk of dental caries and low health literacy being linked to socioeconomically disadvantaged groups, there is a risk that those who are most in need of easily available evidence-based preventive advice and oral health information are the least likely to be able to find, access, and use it. Using eHLS, the population that volunteered for this study showed an overall high level of health literacy and was likely not representative of the overall Welsh and UK populations. Nevertheless, they were in a good position to help inform the development of the prototype and comment about it. We tried to recruit different populations by accessing a group through university employees and a separate population in an area of Cardiff with low socioeconomic status. We used a community hub to invite participants and it may be that those who attend the hub are more likely to have reached a higher level of education and therefore health literacy than those who do not access the hub. For the further development stages of the app, we would actively seek a more representative group, possibly by accessing patient groups.

It was interesting that some of the parents mentioned screen time for their children as perhaps being a barrier to having the child engage with the app. The 2022 Ofcom Report about media use for parents and children found that in the UK nations, parents in Wales were more likely to be very concerned about their child’s media use [7,23].

Design Process

Parents’ Current Source of Information

During the design interviews, parents reported that they primarily relied on search engines and government websites, such as the NHS website, for information about children’s oral health. However, they did not mention any specific resources dedicated to children’s oral health, indicating a need for better education and awareness about reliable sources of information. This may suggest that existing reliable resources have limited impact and reach, despite the funding and man power that are

---

<table>
<thead>
<tr>
<th>TFA constructs</th>
<th>Definition</th>
<th>Relevance to this study</th>
</tr>
</thead>
</table>
| Intervention coherence | “The extent to which the participant understands the intervention and how it works” | - Parents thought that using technology such as an interactive app to achieve engagement with the target population (parents and young children) is a suitable intervention to improve children’s oral health:  
  - “Thinking of my daughter, now, I know that she loves interacting with apps because they have iPads at school. They have Chromebooks. So, I think children over five, maybe, who are in school already, would love to have something for them in there, like the one here, like the brushing, the videos, the timers and music and all that.” [Parent 9] |

---

*aACT: App for Children’s Teeth.*
potentially invested in developing them. It will be crucial to develop a clear marketing plan during the further development stage of the app to increase its visibility. For example, some parents mentioned using apps related to pregnancy and parenting that were recommended by their midwives. This suggests that leveraging home visitors’ accessibility or dentists and general medical practitioners could be effective routes to promote the app. Parents also emphasized the value of peer experiences and word-of-mouth recommendations, as Facebook groups were frequently mentioned as a source of information and advice. In a similar way to mental health apps, leveraging social media platforms through health care providers might be a way to reach out to parents and promote the app through these channels [24].

Acknowledging the significance of trusted endorsements in mHealth apps for parents is crucial. In our study, all parents (10/10, 100%) consistently emphasized the importance of NHS endorsement as a key factor in their decision to download and trust the app. NHS holds weight and credibility as it signifies alignment with recognized health care standards. Therefore, it is essential to include a plan for obtaining such an endorsement during the design and development process. However, this journey is complex, lengthy, and multidisciplinary, involving registration as a medical device and adherence to NHS values and specifications.

Parents’ General Thoughts About the Development of an mHealth App for the Oral Health of Children

The parents were positive about the idea of an app for delivering oral health education to children. This positive reception was further evident in the recruitment process, which surpassed expectations, with parents actively providing feedback and suggestions. In addition, their willingness to receive follow-up emails showed their engagement. The enthusiastic response from parents highlights the potential impact of a well-designed and user-centric mHealth app in promoting positive health outcomes for children. By actively involving parents in the development and implementation processes, their perspectives and insights can be integrated into the app, ensuring better alignment with their expectations and health care needs.

Previous Experiences With Health Apps

Parents’ experiences with mHealth apps such as Headspace and MyFitnessPal provide valuable insights into the factors that contribute to their positive perceptions. The simplicity of these apps, characterized by user-friendly interfaces and intuitive navigation, resonated with parents seeking a hassle-free and user-friendly experience. The availability of free content in Headspace allowed parents to explore meditation and mindfulness practices without financial constraints, which was highly appreciated. Notifications related to mental health and well-being served as gentle reminders and motivators for parents to engage with the app. The wide applicability and provision of relevant information in MyFitnessPal were also highly valued by parents, highlighting the importance of addressing users’ needs and expectations. These insights were carefully considered and incorporated into the design stage of our app prototype, as demonstrated in the Results section.

Development With an App Development Team

During the development of the dental app, collaboration with the app development team brought both excitement and the natural occurrence of differing priorities. Cross-disciplinary synergy was a highlight, with the app developers effectively responding to parent feedback by creating a user-friendly interface with quick 1-click buttons and a home button for easy navigation. However, differing priorities emerged with the app developers aiming for minimization of the number of screens for cost control, whereas the dental team prioritized clinical relevance and usability. These variations in focus were resolved through dialogue and compromise, leading to a better product.

Feedback About the Prototype

The ACT app design and its content, created based on parents’ feedback and underpinned by the Health Belief Model, were felt by parents to have the potential to promote good oral health practices among children and parents. The positive findings of our project are consistent with the available evidence regarding the end user’s accessibility of using such technologies for oral health promotion [25-27]. Parents emphasized the importance of simplicity and user-friendliness of the app. These features were considered just as important as the quality of information provided in the app, highlighting the need to consider user experience in the design process and the importance of involving users in the development of health interventions. These findings are consistent with those of previous studies.

Key features of the app that were positively received included the incorporation of unique features such as a relatable character (the beaver). The character may have provided a sense of personalization and connection to the app. The “stamps” feature was designed based on gamification principles, providing a tangible and visible reward for completing oral-related tasks [9]. This type of engagement strategy has been shown in a meta-analysis to be effective in promoting positive oral health behavior change, such as regular toothbrushing and flossing, among children and parents [28]. The reminder and notification feature of the ACT app was positively received by parents, indicating its potential value in promoting long-term engagement with the app. Reminders and notifications can increase engagement and retention in digital health interventions and can lead to improved health outcomes.

Although the app has several positive features, some feedback highlighted the aspects to be considered for future development. Many parents (6/10, 60%) expressed the need for NHS endorsement as a crucial factor in engaging with the app. Securing NHS endorsement poses significant challenges owing to the stringent standards and guidelines established by NHS [29]. These complexities span clinical validation, regulatory adherence, incorporation of user feedback, engagement with stakeholders, conducting health economic assessments, and intricate management of legal and ethical considerations. Each of these facets demand meticulous attention and allocation of resources to align with NHS’s rigorous criteria and secure their indispensable support. Therefore, careful planning and consideration of these factors are necessary in further development stages.
A parent also commented about the accessibility considerations of the app for individuals with neurodiverse needs. Incorporating accessible, evidence-based recommendations into the design process can ensure that the app reaches a broad audience and promotes better oral health outcomes [25].

Finally, concerns about screen time were raised by parents. Studies have highlighted the adverse impacts of excessive screen time on various aspects of children’s well-being, including cognitive and socioemotional development [30,31].

Hence, to ensure successful adoption and sustained use of the mHealth app, it is essential to address these concerns and mitigate the potential impact of increased screen time. Strategies can be implemented within the app to promote responsible screen time practices. This may include incorporating features that allow parents to set time limits on app use, providing guidance about optimal screen time duration for different age groups, and emphasizing the importance of engaging in offline activities and interactions to complement the app’s use [32]. Including ideas for interactive, off-screen activities such as games and quizzes may also be beneficial [33].

Limitations
The study’s results may have been limited by several factors. First, most of the participants (8/10, 80%) were highly educated women, most (participants had university education; 9/10, 90%), and the age range was relatively narrow, with all participants (10/10, 100%) aged between 25 and 44 years. This was probably related to the recruitment strategy despite efforts to recruit diverse groups by targeting both a university employee cohort and local community center members.

Furthermore, the recruitment process relied on convenience sampling, which may have resulted in a sample that is not representative of the broad population. Although the researchers made efforts to recruit participants through multiple channels, all participants (10/10, 100%) were from 1 city, which may limit the applicability of the findings. This is particularly relevant because the app was developed in collaboration with the International Dental Federation (Fédération Dentaire Internationale), which aims to further refine the app with global perspective.

It is crucial to highlight that the exclusion of children as participants in our study significantly hampers our comprehension of their perspectives regarding the app. This is particularly important considering that parents expressed the belief that the app should be used collaboratively by both the parent and the child. By omitting children from the study, we miss valuable insights that could contribute to a more comprehensive understanding of the app’s use dynamics.

In addition, the development process may have been constrained by limited funding and man power, which may have influenced the extent of user testing and refinement of the app. However, despite these limitations, the study provides valuable insights into the importance of user-centered design and testing in the development of mHealth apps. Future studies can use these insights to further refine and evaluate the effectiveness of the app in improving oral health outcomes in diverse populations.

Conclusions
Digital health interventions, such as the ACT app, have the potential to promote healthy behaviors among children and have important implications for public health. The incorporation of theoretical framework, user cocreation in the design process, and emphasis on simplicity and user-friendliness have been identified as key factors contributing to the success of the intervention. These strategies have also been shown to be effective in promoting engagement and adoption of the intervention. Further studies are needed to build the app based on parents’ views and pilot test to evaluate its impact on promoting healthy behaviors among children.

The study highlights the importance of user-centered design when developing health-related tools and the value of conducting user research and iterative testing to ensure that the tool meets the needs of the target audience. The findings of this study can inform the development of similar tools in the future and ultimately help to improve oral health outcomes for young children.
Multimedia Appendix 1
Document to engage participants.
[DOCX File, 100 KB - formative_v8i1e49561_app1.docx]

Multimedia Appendix 2
Registration form.
[DOCX File, 458 KB - formative_v8i1e49561_app2.docx]

Multimedia Appendix 3
Participant information sheet.
[DOCX File, 89 KB - formative_v8i1e49561_app3.docx]

Multimedia Appendix 4
Consent form.
[DOCX File, 101 KB - formative_v8i1e49561_app4.docx]

Multimedia Appendix 5
Interview guide.
[DOCX File, 81 KB - formative_v8i1e49561_app5.docx]

Multimedia Appendix 6
Example of the mobile app’s layout.
[DOCX File, 1019 KB - formative_v8i1e49561_app6.docx]

Multimedia Appendix 7
Participants’ responses to the eHealth Literacy Scale (eHLS) questions (Qs; A) 1 and (B) 2 (n=9).
[PNG File, 82 KB - formative_v8i1e49561_app7.png]

References
8. mHealth: new horizons for health through mobile technologies. World Health Organization. URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technology [accessed 2024-01-05]

https://formative.jmir.org/2024/1/e49561

Al-yaseen et al

JMIR Form Res 2024 | vol. 8 | e49561 | p.572
(page number not for citation purposes)


Abbreviations

ACT: App for Children’s Teeth
eHLS: eHealth Literacy Scale
mHealth: mobile health
NHS: National Health Service
TFA: Theoretical Framework of Acceptability
Abstract

Background: Web-based parent training (PT) programs can strengthen parent-child relationships by equipping caregivers with knowledge and evidence-based strategies to manage behavior. Hybrid facilitation of PT includes facilitator interaction paired with self-administered and web-based PT. Web-based administrative dashboards provide users (eg, administrators, facilitators, and researchers) with an integrated platform to monitor parent progress and activities within a PT program or website. Despite the utility and prevalence of administrative dashboards for web-based behavioral interventions, to our knowledge, no research studies have explored the perspectives and insights of dashboard users to enhance user experience and program delivery.

Objective: The purpose of this study is to evaluate the usability of the administrative dashboard (ezDashboard) for the ezParent program, a 6-module web-based PT program for parents of children aged 2-5 years.

Methods: This study used a descriptive, single-group design with administrators who were overseeing the implementation of the ezParent program and trained facilitators for hybrid ezParent delivery. Participants spent at least 30 minutes reviewing and evaluating the ezDashboard and then completed a survey of their experience with the dashboard. The survey included the validated 10-item System Usability Scale and open-ended questions focusing on user performance, navigation ease, and overall usefulness of the ezDashboard.

Results: Participants (N=15) indicated high usability of the ezDashboard with System Usability Scale scoring a total mean score of 83.5 (SD 16.3). Most participants (n=13, 87%) rated the overall user-friendliness of the ezDashboard as good (n=3, 20%), excellent (n=9, 60%), or best imaginable (n=1, 7%). Open-ended questions revealed the ezDashboard is or would be useful to monitor parent progress and trends in engagement (n=8, 53%) and for reviewing topics for discussion and communicating with parents (n=5, 33%). ezParent administrators (n=4) identified that real-time data for ezParent use helps overall management of program uptake. Suggestions for features to add to the ezDashboard included the ability to track partial progress of program modules (4/14, 29%), total time spent per module (2/14, 14%), and exportable reports (4/14, 29%). Other ideas for improvement included direct messaging capabilities, videoconferencing platform integration, and being able to modify participant account and contact information.

Conclusions: Results indicate that the ezDashboard is easy to use and provides functional information to facilitators and administrators in delivering ezParent. Qualitative results indicate that integrating suggested features into the ezDashboard may help provide a smoother experience for facilitators, administrators, and ultimately the parents using the program. Providing resources for facilitators and administrators in real time to monitor intervention participants’ progress in a program can be helpful in tracking progress and providing facilitated support in tailoring program content and program completion.

(JMIR Form Res 2024;8:e53439) doi:10.2196/53439
Introduction

Parent training (PT) programs—the gold standard for prevention and treatment of child behavior problems [1-3]—aim to strengthen parent-child relationships by providing caregivers with knowledge and evidence-based strategies to effectively strengthen their parenting skills and support their child’s positive behavior. Parents who have participated in PT programs have demonstrated improvements in multiple areas, including improvements in positive parenting skills, self-efficacy, and parent-child interactions; and reductions in negative or harsh parenting and parenting stress [2,4,5]. PT participants’ children exhibit improvements in child behavior, display decreased conduct problems and aggression, improvements in academic performance, enhanced coping skills, and strengthened relationships with both caregivers and peers [2,4]. Traditionally, PT programs have been offered in-person in group or individual settings; however, web-based adaptations have emerged to mitigate the geographic, logistical, and personal barriers of face-to-face delivery [6-8].

In addition to web-based programs, web-based administrative dashboards have emerged as a promising method to improve delivery and management of web-based interventions [9-11]. These dashboards provide an integrated platform to monitor user progress and activities within a program or website. Administrative dashboards are particularly useful tools for administrators, implementers, researchers, and others who require access to detailed information regarding user usage patterns, performance and achievements, and program progress and completion [10,12,13]. For example, in a research and practice context, dashboards can be used to monitor program fidelity and provide clear metrics to understand how the program is being used in real time.

In the context of web-based programs delivering behavioral interventions, administrative dashboards prove particularly valuable. For example, administrative dashboards can include timestamps for login and logout; activity completion; and fill-in responses to in-program prompts, quiz results, page clicks, and diary entries [9,11,14,15]. These program usage metrics enable administrators to make data-driven decisions, identify areas for improvement, and provide critical support to program and research staff when offering personalized guidance to users [10]. Researchers and program facilitators working directly with participants can leverage the dashboard to prepare for discussions by reviewing the participant’s last logins and progress, using fill-in prompt responses to ask tailored questions, and providing troubleshooting recommendations to encourage program engagement [9,11]. Armed with these data points, facilitators can improve their support of participant use of a program, tailor program materials, and ultimately support improvement for child and parent social-behavioral outcomes.

Despite the utility and use of administrative dashboards, to our knowledge, there are no research studies exploring the user experience of an administrative dashboard related to web-based behavioral interventions. Thus, there is an opportunity for further research to optimize existing functionalities and address limitations to better support administrative dashboard users. Usability, which plays a crucial role in the adoption, engagement, and overall effectiveness of web-based administrative dashboards supporting behavioral health interventions, becomes paramount. The usability of a web-based innovation is commonly assessed on several domains including the efficiency, intuitiveness, ease of use, and satisfaction experienced by the user [16]. The purpose of this study was to investigate the usability of the administrative dashboard for the ezParent program, a web-based PT program. Specifically, our goal was to establish a baseline for user performance, ease of navigation, and usefulness.

Methods

Study Design

This study was a descriptive, single group survey design with participants who were overseeing the implementation of the ezParent program (administrators) or trained facilitators for hybrid ezParent delivery (facilitators).

ezParent Program

The ezParent Program is the web-based delivery of the Chicago Parent Program (CPP). The CPP has been shown to be effective in improving positive parenting skills, parenting self-efficacy, and child behavior problems in a population of low-income, urban parents of children aged 2-5 years old [17,18]. The ezParent Program teaches parents and caregivers the evidence-based strategies of CPP using 6 modules that include a video narrator and vignettes of families using the skills, reflection questions, program activities, and in-home practice assignments [19]. In a pilot randomized controlled trial (RCT) of ezParent (n=83), parents completed 82% of the 6 modules, reported high satisfaction with the program, and we found comparable effect sizes in improvements in parenting practices and reductions in parenting stress and child behavior problems to the group-based CPP [20,21]. In an RCT of self-administered ezParent in primary care (n=287), we failed to find significant main effects for parent and child behaviors [22]. Based on our findings and the extant literature suggesting web-based programs are more effective when provided along with human support [22,23] we are currently testing 2 models of hybrid delivery. The first includes 1:1 brief coaching as part of an RCT funded study (see Greene et al [24]) and community-based delivery of ezParent paired with web-based group sessions [25].

ezParent Dashboard Design and Development

The ezParent program tracks user progress as parents use the program. Custom data tables collect and store user data with...
timestamps for logins, completed modules, end-of-module surveys, badges, practice assignments, and practice reviews. The administrative dashboard (ezDashboard) was developed to include these data points for hybrid delivery facilitators to monitor parent progress in the program and to inform hybrid sessions to encourage and support parent program uptake.

ezDashboard logins are created internally by programming staff and are provided to administrators and facilitators. Users access a “Parent Lookup” form by entering the unique ID given to parents during ezParent enrollment and the parent’s last name. These form fields are used together to ensure participant data security. The parent is added to the home page or “Parent List” if the user ID and last name match an individual user in the database. The parent list displays the user ID, last name, last login date, and the last viewed module. When a parent has completed their time in the ezParent program, users can easily remove that parent from their home page.

From the parent list, ezDashboard users can access more details for an individual parent user by clicking on a “Details” button. The top of the “Details” page displays the user ID, name, phone number, and participant email for easy contact. Contact information is followed by detailed use metrics, including a scrollable list of all login dates, completed modules with a green check mark to indicate completion and the date they were completed, module survey responses, completed badges and dates earned, and responses for practice assignments and practice reviews. Finally, a downloadable, individualized completion certificate is accessible once the last (sixth) module is completed. See Figures 1 and 2 for ezDashboard screenshots.

The ezDashboard’s user interface was made in React (version 17.0.2; MetaOpenSource), using responsive web design principles to allow access across different device sizes. Microsoft SQL Server stores real time user data, which the user interface fetches and displays.

**Figure 1.** ezDashboard home page: a detailed description of parent use in the program.
Figure 2. ezDashboard badge completion to mark parent user progress through the program.

Ethical Considerations
The research protocol for this study was determined exempt by the institutional review board at The Ohio State University (study number 2023E0289). Participants completed a web-based consent form. All survey data were deidentified and participants were informed that deidentified data may be used or shared without additional informed consent. Participants who consented and completed the survey received a US $25 gift card for participation in this study.

Participants
We invited ezParent administrators (n=4; individuals overseeing implementation of ezParent) and facilitators (n=19; trained facilitators for hybrid ezParent delivery) to participate in the usability study with a goal for a sample size between 10 and 15. Administrators and facilitators were invited as they may have unique perspectives related to usability depending on their ezParent role. Macefield [26] suggests that a sample size of 3-20 is valid in a usability study to discover usability and potential problems and a sample of 10 will probably reveal a minimum of 82% of the problems and be useful in future design changes. Of the 23 invited to participate in the usability study, 16 consented to participate, and 15 completed the review of the ezDashboard and usability survey.

Procedures
Potential participants received an email inviting them to participate in the usability study. The email invitation included a description of the project and a link to the consent form in REDCap (Research Electronic Data Capture; Vanderbilt University) [27]. Once consent was obtained, participants were instructed to spend at least 30 minutes reviewing the ezDashboard and evaluating the features. Participants who were not current users of the ezDashboard were provided with login access and instructed to add 2 sample users to their Parent List. Current users of the ezDashboard were instructed to sign into their accounts and review user accounts on their existing Parent List. Both groups were instructed to evaluate their users’ activity on the main page and on the detailed page of all modules in order to evaluate the usefulness of the ezDashboard. These procedures include typical tasks that a facilitator would take in evaluating and monitoring parent use of ezParent. Then, participants were prompted to complete the survey and asked to identify their current role using the ezParent program (eg, facilitator or administrator). At the completion of the survey, participants received a US $25 gift card for their participation.

Measures
The 10-item System Usability Scale (SUS) is a tool for assessing the usability of a product (eg, websites, cell phones, and apps). The 10 items are scored on a 5-point Likert scale (range 0 strongly disagree to 4 strongly agree). A total usability score,
representing a composite measure of usability is created by reversing the score of even-numbered items, summing the items, and multiplying by 2.5 to convert the original total scores of 0-40 to a 0-100 scale [28]. A score of 70 is considered average, 80 good, and 90 or above excellent usability [29,30]. The SUS is a simple and efficient tool for assessing the usability and user-friendliness of a technological platform with demonstrated reliability across multiple studies (α=.91) and strong evidence of a single factor structure [29]. Our administration of SUS reflected the revised language and overall rating of the user-friendliness introduced in Bangor and colleagues [30]. Survey responses were collected in REDCap [31,32].

Following completion of the SUS, participants were prompted to respond to open-ended questions to elicit further information on their opinions of the e2Dashboard. Questions included the frequency of use and the helpfulness of the e2Dashboard and items that may be missing or could be changed in future iterations. Facilitators were asked specifically how access to this real time information helps them work with parents using the e2Parent program and administrators were asked how the information helps them with overall management of the program. Finally, participants were asked to self-report their race or ethnicity, age, and gender.

**Analysis Plan**

The composite SUS score was calculated by reverse scoring even-numbered items so that all items were scored in the same direction. Composite scores were then calculated by summing the item responses and multiplying them by 2.5 so that they fell on a scale of 0-100. Summary statistics were calculated for the composite scores and frequencies are reported for the overall rating of user-friendliness.

For analysis of the open-ended questions, 2 authors organized the responses by group (eg, administrators and facilitators) and conducted a thematic analysis of all responses based on the steps outlined by Braun and Clarke [33]. First, we reviewed all the responses and generated initial themes and categories. These categories were reviewed by the 2 authors and confirmed by a third author. Finally, we categorized the themes in order to provide a description and examples in this report. We quantified the comments in each category in order to provide a frequency related to participants’ ideas, suggestions, and ideas provided related to the usability of the e2Dashboard.

**Results**

**Participants**

A majority of participants (N=15) were women (n=12, 80%), White (n=11, 73%), with a mean age of 40.9 years (SD 13.9; range 20-68; Table 1). Participants were e2Parent facilitators (n=10) and administrators (n=4); 1 participant did not report their role with e2Parent. Participants (n=9) who were actively using the e2Dashboard at the time of the survey reported weekly use (n=3, 33%), once every few weeks (n=1, 11%), monthly (n=1, 11%), and less than monthly (n=4, 44%).

**Table 1.** Demographic characteristics of participants (N=15) enrolled in the e2Dashboard usability study.

<table>
<thead>
<tr>
<th>Demographic characteristic</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>12 (80)</td>
</tr>
<tr>
<td>Men</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>3 (20)</td>
</tr>
<tr>
<td>More than 1 race^a</td>
<td>1 (7)</td>
</tr>
<tr>
<td>White</td>
<td>11 (73)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Chose not to answer or missing</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>13 (87)</td>
</tr>
</tbody>
</table>

^aIn total, 1 participant reported their race as Black and other.

**About SUS**

Scoring of the SUS indicated high usability of the e2Dashboard with a total mean score of 83.5 (range 47.5-100; SD 16.3; median 90, IQR 72.5-97.5). Overall, most participants (n=10, 67%) rated the overall user-friendliness of the e2Dashboard as excellent (n=9, 60%), or best imaginable (n=1, 7%; Table 2). On average new users (n=6) ranked the e2Dashboard higher (mean 90.8, SD 8.01) than existing users (n=9; mean 78.6, SD 18.84); however, this difference was not statistically significant (P=.16) and may be the result of high variability associated with small samples.
Open-Ended Survey Responses

**Usefulness of the ezDashboard**

Participants (N=15) reported that the dashboard allowed them to keep track of parent progress and identify potential problems that may arise with parent completion of ezParent modules and serves as a helpful cue for discussion with parents. Specifically, 53% (8/15) of participants, both current users and participants who had not yet interacted with the ezDashboard, identified that the ezDashboard is or would be useful for monitoring parent progress and identifying trends in parent participation. As 1 facilitator reported, “I typically check it right before the calls to see the participant’s last login and how much of the program they have completed.” Another reported, “This makes it much easier to track program adherence and help support parents who need a little extra help in completing the program.” Finally, a participant wrote, “Overall trends of participation are helpful to learn if any systemic barriers to participation need to be addressed.” In addition to monitoring progress, participants (n=5, 33%) identified the ezDashboard as useful in reviewing topics for discussion and communicating with parents during the hybrid session. For example, 1 participant wrote “I would use the dashboard to...Refer back to specific content from modules,” and another “use it to inform the next group meeting, for example, if I noticed that everyone thought a previous module was difficult, I would give more time to that discussion.”

ezParent administrators (n=4) identified that real-time data for ezParent use helps overall management of program uptake and promotes parent motivation and accountability for module completion. For example, 1 program provides incentives for module completion and uses the ezDashboard for tracking program use for the provision of these parent incentives. Administrators also provided that this information “would also be helpful for facilitating in learning when/why/if they need to modify their approach” when facilitating hybrid ezParent and allows them to respond to questions from parents more promptly.

**Suggested ezDashboard Changes**

Participants (N=15) were asked if there was any information missing from the ezDashboard (14 responded). Overall, participants suggested potential data points in the ezDashboard to allow a more in-depth assessment of parent program use. Participants (n=4, 29%) identified that the ability to track partial progress through the modules would be useful, with suggestions of “it would be nice to see the # of logins for a week” and “it would be nice to differentiate if modules were completed or started, such as a sliding scale of how far through the module the parent has gotten so far.” In total, 2 (14%) participants suggested tracking the total time spent in each module as a method to understand meaningful engagement with the program and whether “a parent may be working a little too quickly through the program.” The administrative participants (n=4) believed “exportable, customizable reporting” would be useful for overall program management. In addition, more control in terms of modifying ezDashboard information was identified, including amending parent information (eg, name and phone number) and the ability to group parents by cohort for tracking. In total, 38% (6/14) of the participants indicated that they would change nothing to make the ezDashboard easier to use.

Beyond program metrics, participants identified changes that would integrate the hybrid delivery methods into the ezDashboard. For example, including attendance records for hybrid meetings and integrating a method to directly communicate and contact parents (eg, texting) with parents in the dashboard so they could “nudge parent to modules.” Another suggestion included integrating the videoconference system into the dashboard so all program activities could occur in 1 place.

**Discussion**

**Principal Findings: ezDashboard Usability**

The purpose of this usability study was to identify the overall performance of the ezParent administrative dashboard and understand users’ perceptions of the ease of navigation and usefulness of the ezDashboard in implementing hybrid delivery of the ezParent program. Further, we were interested in collecting users’ suggestions for changes or additions to the ezDashboard to improve the overall user experience for real-world use.

Overall, ratings on the SUS indicate good usability (mean 83.5, SD 16.3; median 90, IQR 72.5-97.5). According to Bangor and colleagues [29], a mean score of 83.5 is in the fourth quartile of scores, rated as acceptable, and falls between the good to excellent range using an adjective rating scale. In addition, overall ratings of the user-friendliness of the ezDashboard among all participants were positive (ok, good, excellent, and best imaginable). These initial ratings are promising, show an
acceptable level of usability, and the written feedback provides us with concrete methods for improving the ezDashboard.

This study’s participants reported variable use of the ezDashboard (eg, ranging from weekly to less than monthly). The differences in use may be a function of the individual’s role using ezParent (eg, administrators may only need to use the dashboard monthly for overall program management while a coach conducting weekly calls would use it weekly). We do, however, need to consider the variability of use as a potential function of the overall usefulness of the information provided in the ezDashboard. Therefore, our next steps will be to provide clear instructions and descriptions of ezDashboard use as well as integrate suggested changes. Changes to the ezDashboard to provide desired information of parent use may increase regular use and uptake.

Harrington and colleagues [10] highlight the importance of including program usage metrics in dashboards to allow interventionists to make data-driven decisions, identify areas for program improvement, and support the ability to provide personalized support to program users. Further, real-time data can support tailored approaches to increase program uptake. Overall, our participants reported that they were able to use the ezDashboard information to take an individualized approach to the hybrid delivery of the ezParent program, and administrators used the ezDashboard data to provide oversight and incentives to parent participants. In addition, there were several excellent suggestions for ezDashboard improvements. The suggested changes varied by participant role with ezParent.

Facilitator suggestions were primarily fine-tuning the data presented in the ezDashboard to provide more nuanced use metrics beyond module completion. Since most digital analytics provide summary statistics, the ability to gather use metrics at the individual level provides important information for the facilitation of program engagement [14]. The data points suggested were partial module completion, identifying the actual location within the module of last use, and average time spent on the page. While the ezDashboard currently provides time stamps for date and time, the modules were completed and allows for a rough estimate of the speed at which a parent is moving through the program; more specific analytics for time in the program could provide more meaningful estimates for engagement. In our previous work, we found that on average parents spent 37.2 (SD 22.2) minutes per module with a range from 26.4 (module 5) to 47.9 (module 2) minutes [21]. Variations across modules occurred because of variations in pages per module; however, there was a significant decrease in minutes/module over time (eg, participants spent less time on later modules) [21]. This information is important for facilitators to monitor parent engagement and to encourage active involvement in the modules, particularly in later modules when data support a decrease in overall time of engagement.

The administrative participants identified other types of metrics and functions that would be helpful for program management (eg, the ability to download individual and cohort reports of usage metrics to allow for more streamlined monitoring and administrative access to modify participant demographics). The goal of the ezDashboard is to present individual use data and adaptations to include monitoring cohorts, which could increase administrative efficiencies. A benefit of web-based interventions is the reach and accessibility of programs and the integration of administrator dashboards for monitoring and management and has the potential to increase program uptake and overall efficiencies in program delivery. Future evaluation of the ezDashboard will focus on the effects of ezDashboard use on implementation factors related to organizational and individual program uptake and delivery.

Limitations
Although deemed sufficient for usability testing to acquire feedback on user experience, our sample size was small. This study was an initial evaluation of the ezDashboard and provides valuable information for modifications. Our next steps are to integrate these findings into the ezDashboard to evaluate overall use of ezParent and dashboard use in a pragmatic ongoing trial with a larger sample of facilitators and administrators.

Additional limitations of this study include the usability testing being done virtually and at only 1 timepoint, which precluded our ability to examine participants’ actions in real time while using the ezDashboard in a pragmatic ongoing setting. Users’ feedback may change after continued use of the ezDashboard. The administrative dashboard was only tested for 1 specific intervention, ezParent; however, we believe these results could be applicable and inform the development of dashboards for different types of web-based programs.

Conclusions
The ezDashboard was initially developed to provide individual parent usage data to facilitators in real time to monitor parent progress in the program and support parent program uptake. The results of usability testing indicate that the ezDashboard is easy to use and provides functional information to facilitators in delivering the ezParent intervention. Providing resources for facilitators and administrators to aid in facilitation of the hybrid intervention may lead to improved parent uptake and outcomes [34]. Qualitative results indicate that integrating suggested features into the ezDashboard may help provide a smoother experience for facilitators, administrators, and ultimately the parents using the program. Administrative dashboards that provide real-time program usage data require an investment in the upfront cost of program development. The user facing program, in this case the ezParent, must be built to collect the user data that are to be displayed in the dashboard. For those considering integrating a dashboard for a web-based intervention, we suggest early planning during the initial development.
Acknowledgments

The authors gratefully acknowledge the coaches, facilitators, and administrators who participated in the evaluation of the eDashboard. This study is supported in part by a grant from the National Institute of Child Health and Human Development (NICHD; R01HD104072).

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request. Generative artificial intelligence was not used in any portion of this paper’s writing.

Conflicts of Interest

None declared.

References


**Abbreviations**

CPP: Chicago Parent Program  
PT: parent training  
RCT: randomized controlled trial  
REDCap: Research Electronic Data Capture  
SUS: System Usability Scale

©Susan M Breitenstein, Julia Berteletti, Shea Smoske, Charles Barger, Kyrie Tipps, Nathan P Helsabeck. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Web-Based Mindfulness Meditation as an Adjunct to Internet-Delivered Cognitive Behavioral Therapy for Public Safety Personnel: Mixed Methods Feasibility Evaluation Study

Caeleigh A Landry¹,²,³ MSc; Hugh C McCall¹,²,³ MA; Janine D Beahm¹,²,³ PhD; Nickolai Titov⁴ PhD; Blake Dear⁴ PhD; R Nicholas Carleton¹,²,³ PhD; Heather D Hadjistavropoulos¹,²,³ PhD

¹Department of Psychology, University of Regina, Regina, SK, Canada
²PSPNET, University of Regina, Regina, SK, Canada
³Canadian Institute for Public Safety Research and Treatment, Regina, SK, Canada
⁴eCentreClinic, School of Psychological Sciences, Macquarie University, Sydney, Australia

Corresponding Author:
Heather D Hadjistavropoulos, PhD
Department of Psychology
University of Regina
Administration-Humanities, AH 345
3737 Wascana Parkway
Regina, SK, S4S 0A2
Canada
Phone: 1 306 585 5133
Email: heather.hadjistavropoulos@uregina.ca

Abstract

Background: Public safety personnel (PSP) are individuals who work to ensure the safety and security of communities (eg, correctional workers, firefighters, paramedics, and police officers). PSP have a high risk of developing mental disorders and face unique barriers to traditional mental health treatments. The PSP Wellbeing Course is a transdiagnostic, internet-delivered cognitive behavioral therapy (iCBT) course tailored to assist PSP with symptoms of depression, anxiety, and posttraumatic stress disorder (PTSD). The initial course outcomes are promising, but some clients report some challenges with learning skills and recommend adding additional resources. Mindfulness meditations, which help people to experience the world and their reactions to the world in open and nonjudgmental ways, may complement the existing PSP Wellbeing Course.

Objective: This study aims to examine the feasibility of mindfulness meditations in iCBT tailored for PSP. Information was gathered to evaluate engagement and client experiences with mindfulness meditations, symptom change, and the relationship between mindfulness meditation use and symptom change.

Methods: A mixed methods study was conducted on PSP enrolled in the PSP Wellbeing Course who were offered 5 mindfulness meditations during the program (ie, 1/lesson). Clients completed questionnaires on depression, anxiety, PTSD, anger, insomnia, resilience, and mindfulness at pretreatment and at 8 weeks; an 8-week treatment satisfaction questionnaire; and brief weekly measures of mindfulness meditation engagement. We used paired sample t tests (2-tailed) to assess changes in outcomes over time and partial correlations to assess whether mindfulness meditation use predicted outcomes at posttreatment. A total of 12 clients were interviewed about their perceptions of the mindfulness meditations, and interviews were analyzed using directed content analysis.

Results: Among the 40 clients enrolled, 27 (68%) reported using the mindfulness meditations, practicing for an average of 4.8 (SD 8.1) minutes each week. Most interviewees described the mindfulness meditations as beneficial but also reported challenges, such as discomfort while sitting with their feelings. Clients provided suggestions for better integration of mindfulness into iCBT. Overall, clients who completed the PSP Wellbeing Course with mindfulness meditations experienced statistically significant improvements in symptoms of anxiety (P=.001), depression (P=.001), PTSD (P=.001), and anger (P=.001) but not insomnia (P=.02). Clients also experienced improvements in resilience (P=.01) and mindfulness (P=.001). Self-reported time spent meditating was not associated with changes in symptoms over time.

Conclusions: This study provides new insight into the integration of mindfulness meditations with iCBT for PSP. It demonstrates the partial feasibility of adding mindfulness meditations to iCBT, revealing that some, but not all, PSP engaged with the meditations...
and reported benefits. PSP reported using the mindfulness meditations inconsistently and described challenges with the meditations. Improvements can be made to better integrate mindfulness meditation into iCBT, including offering mindfulness meditation as an optional resource, providing more psychoeducation on managing challenges, and offering shorter meditations.

**KEYWORDS**

public safety personnel; PSP; internet therapy; mindfulness; meditation; internet-delivered cognitive behavioral therapy; iCBT

**Introduction**

**Background**

Public safety personnel (PSP; eg, border services officers, correctional workers, firefighters, Indigenous emergency managers, operational intelligence personnel, paramedics, police officers, public safety communicators, and search and rescue personnel) have greater exposure to potentially psychologically traumatic events than the general population and a greater risk for several psychological disorders. This heightened risk is, in part, attributable to the nature of their vocations [1-3]. In a recent pan-Canadian study, 44.5% of PSP screened positive for at least 1 mental disorder on a set of questionnaire measures [1], which is far higher than the rate of mental disorder diagnoses in the general population (ie, 10% [4]). Other studies show PSP also report high levels of anger [5] and sleep difficulties [6]. Despite the significant need for mental health treatment, PSP face several unique barriers to treatment as a function of their vocational cultures and requirements [7]. PSP report concerns about engaging in face-to-face mental health treatment options, including geographical barriers; the cost of treatment; difficulty navigating services; long wait times; and, in particular, concerns about privacy, confidentiality, and stigma [6,8-10].

Internet-delivered cognitive behavioral therapy (iCBT) appears to be well positioned to address barriers to treatment experienced by PSP, as it delivers treatment materials in a private and accessible web-based format. There is strong evidence that iCBT has comparable effectiveness with face-to-face cognitive behavioral therapy (CBT) among the general population [11]. A recent observational trial found that iCBT tailored specifically for PSP was associated with large pre-post reductions in symptoms of generalized anxiety disorder, major depressive disorder, and posttraumatic stress disorder (PTSD), although it was not compared with a control condition [9]. In written feedback, PSP who participated in iCBT reported appreciating several aspects of iCBT, including the format and content, the accessibility of the course, the additional resources and examples, and the therapist guidance [12]. Nevertheless, some PSP also reported challenges with learning skills [13] and provided suggestions for improving iCBT, including adding more resources [12].

The addition of mindfulness resources represents a promising addition to iCBT with PSP, given the growing research attention on the benefits of mindfulness [14-16] and its use in CBT [17]. Kabat-Zinn [18] provided the most commonly used definition of mindfulness: a three-component definition that calls for paying attention (1) on purpose, (2) in the present, and (3) nonjudgmentally. Mindfulness shifts awareness to focus on present-moment activities, thoughts, and sensations, which over time improves emotion regulation and, in turn, reduces symptoms [19]. Mindfulness may be cultivated through numerous activities that bring individuals into the present moment and into their experience without judgment. Mindfulness meditation, in particular, is a practice to cultivate mindfulness and refers to the practice of silently observing one’s own internal and external environment without attempting to change anything [19]. In the general population, several studies have shown that web-based mindfulness programs result in small but significant improvements in symptoms such as depression and anxiety [14-16]. In terms of PSP, mindfulness-based interventions have been associated with reductions in stress, insomnia, burnout, anger, anxiety, and depression [20,21]; however, research has been limited to pilot or feasibility studies, police populations, and in-person class delivery style [3,22].

Recent literature suggested that incorporating mindfulness may enhance the benefits derived from traditional therapies [17]. Mindfulness practice has been successfully integrated within face-to-face interventions, such as mindfulness-based stress reduction, mindfulness-based cognitive therapy, and acceptance and commitment therapy [17]. Moreover, a recent systematic review and meta-analysis reported that mindfulness-enhanced iCBT has demonstrated significant reductions in anxiety and depression in clinical populations, above and beyond comparison conditions [17], but to date, no research has focused on offering web-based mindfulness-enhanced iCBT for PSP. Although it is possible that the results of mindfulness-enhanced iCBT will generalize to the use of iCBT for PSP, this remains an empirical question. Some past research among PSP suggests that they may be more skeptical of mental health support [7]. Some studies on mindfulness among PSP have yielded mixed results, with only some participants reporting improvements [21]. Therefore, it is important to explore whether mindfulness meditations will be used and positively evaluated by PSP within the context of iCBT and whether adding mindfulness to iCBT will positively or negatively impact engagement, satisfaction, and outcomes. Adding mindfulness meditation to iCBT has the potential to increase the engagement, satisfaction, and effectiveness of iCBT. However, adding additional components could also reduce effectiveness through additional burden placed on users or if users have negative views of mindfulness meditations.

**Objectives**

As a first step in understanding the use of mindfulness meditation as part of iCBT for PSP, this study was designed to examine the feasibility of adding mindfulness meditations to an existing iCBT program for PSP, called the PSP Wellbeing Course, by (1) evaluating the level of engagement with the
mindfulness meditations (eg, practice length and frequency) as well as with the intervention; (2) assessing client experiences with mindfulness meditations during and after treatment; (3) measuring changes in measures of anxiety, depression, and PTSD to compare outcomes with previously published outcomes of the PSP Wellbeing Course [9]; and (4) assessing the relationship between mindfulness mediation use and outcome measures. We hypothesized that clients would actively participate in the mindfulness meditations, report positive experiences, and also identify areas for improvement in the mindfulness meditations. We also hypothesized that clients would report statistically significant improvements in symptoms of anxiety, depression, and PTSD and that greater mindfulness mediation use (ie, length and frequency) would predict less severe self-reported symptoms at posttreatment [23]. This feasibility study is the first step in assessing whether mindfulness meditations will be deemed usable and acceptable in iCBT for PSP.

Methods

Ethical Considerations

This study was approved by the Research Ethics Board of the University of Regina (2019-157). Clients were made aware of the details of the study and the potential risks and benefits of participating, and they provided informed consent before participation. Clients were given access to the PSP Wellbeing Course but were not otherwise offered incentives to encourage participation. Client data were stored on a secure server and deidentified before analyses.

Clients and Procedure

PSP were informed of the PSP Wellbeing Course offered by PSPNET through presentations, emails distributed through PSP organizations, social media, and word of mouth and encouraged to visit the PSPNET website [24]. PSPNET is a clinical research unit based at the University of Regina that develops, delivers, and conducts research on iCBT for PSP. In addition to offering a therapist-guided PSP Wellbeing Course, PSPNET also offers a therapist-guided PSP PTSD Course and a self-guided version of the PSP Wellbeing Course, all of which were available for PSP to select from at the time they visited the website. Prospective clients who were interested in the PSP Wellbeing Course read about the program and completed a consent form and a brief web-based screening questionnaire. Once they completed the web-based screening, they scheduled and completed a phone screening with a trained clinician. To be eligible for the intervention and thereby this study, clients needed to be current or past PSP (career or volunteer), be residing in an eligible Canadian province or territory (at the time of this study, this included Alberta, New Brunswick, Nova Scotia, Ontario, Prince Edward Island, Saskatchewan, and Nunavut), be aged at least 18 years, have access to an internet connection, and be willing to provide an emergency medical contact. Prospective clients were ineligible and referred to other services as appropriate if they reported a high suicide risk; reported a past-year suicide attempt or suicidality-related hospitalization; reported a primary problem with psychosis, mania, or substance use; or reported current involvement in another psychological treatment. The eligibility criteria were assessed during the web-based screening and the subsequent phone screening. Suicide risk was first assessed using item 9 from the Patient Health Questionnaire-9 (PHQ-9) [25], which inquires about suicidal ideation. Clinicians then conducted a clinical interview by phone to assess suicide risk, including asking about past-year attempts and hospitalizations. Severe alcohol or drug problems were assessed using validated questionnaires (ie, scored ≥20 on the Alcohol Use Disorder Identification Test [26] or ≥25 on the Drug Use Disorder Identification Test [27]). Psychosis and mania were assessed based on clinical history but were not reported by any client. Eligible clients were enrolled in the intervention and assigned to therapists who were either registered master’s-level social workers or registered psychologists. Therapists would email or call clients once or twice a week, depending on client preference. Therapist support was designed to help clients work on skills within the course, apply the skills to their lives, and troubleshoot difficulties. Clients were asked to complete regular symptom measures throughout the course on a weekly basis and at 8 weeks. At 10 weeks, the first 30 clients were invited to complete an interview assessing their perspectives on the mindfulness meditations and the course in general. There were 12 clients who agreed to participate in the interviews.

Materials

The iCBT Intervention

The PSP Wellbeing Course is a transdiagnostic iCBT course for PSP adapted from a previous Canadian iCBT course, the Wellbeing Course, which was initially developed at the Macquarie University in Australia and has been successful in treating a range of symptoms in Australia [28-31] and Canada [32-35]. More information on the origins of the intervention [25] and the adaptations of the intervention [29-32] can be found elsewhere. The course uses a theoretical, pragmatic approach to treatment that teaches clients skills that can be applied to various symptom presentations. The intervention specifically includes five core lessons: (1) introduction of CBT and identifying symptoms, (2) monitoring and challenging automatic thoughts, (3) management of physical symptoms, (4) graded exposure, and (5) goal setting and relapse prevention. Lessons are presented in a slideshow format and include text, diagrams, and case stories about PSP. Clients can download materials, homework assignments, and supplementary information on many topics (eg, panic, assertiveness, sleep, and grief). The intervention is designed to be completed in 8 weeks, but clients can have access to a therapist for up to 16 weeks and to the course materials for up to 1 year. Clients receive automated emails encouraging them to work through the materials during this period.

Mindfulness Meditations

Unlike the previous versions of the PSP Wellbeing Course, the version used in this study included a downloadable, guided, audio mindfulness meditation for each of the 5 lessons of the intervention and psychoeducational material on mindfulness meditations before the first meditation. Each lesson included a different type of mindfulness meditation designed to complement the CBT skills taught in that lesson and was based...
on evidence of successful implementation in other programs [22,36-39]:

- Lesson 1: grounding (eg, turning attention to physical sensations)
- Lesson 2: loving kindness (eg, cultivating feelings of love for self and others)
- Lesson 3: awareness of breath (eg, focusing on breathing slowly and deeply)
- Lesson 4: awareness of the 5 senses (eg, cultivating awareness of the environment through the 5 senses)
- Lesson 5: body scan (eg, focusing on the body for areas of tension).

Each mindfulness meditation was audio-recorded by a voice actor and was designed to be approximately 10 minutes long, consistent with previous recommendations [39-41]. Each mindfulness meditation was accompanied by a text script that clients could read in lieu of the audio. All clients who enrolled in the PSP Wellbeing Course were provided with access to the mindfulness meditations. Clients were encouraged to complete 10 minutes of mindfulness meditation practice per day while completing the intervention.

Measures

Overview

During eligibility screening, we administered a questionnaire to assess demographic and occupational characteristics. We also asked about pretreatment engagement in a mindfulness meditation practice. At both pretreatment and 8 weeks postenrollment, consistent with past research on the PSP Wellbeing Course, we administered the following measures: the PHQ-9 to measure symptoms of depression [25], the Generalized Anxiety Disorder-7 (GAD-7) to measure symptoms of generalized anxiety [42], and the Posttraumatic Stress Disorder Checklist for the DSM-5 (PCL-5) to measure symptoms of PTSD [43]. As anger and sleep problems are common among PSP [5,6], we also administered the Dimensions of Anger Reaction Scale-5 (DARS-5) to measure problems related to anger [44] and the Insomnia Severity Index (ISI) to measure sleep difficulties [45]. In an effort to help advance the growing research literature on resilience among PSP [46], we also administered the Brief Resilience Scale (BRS) to measure clients’ resilience to life stressors [47]. Finally, the Five Facet Mindfulness Questionnaire (FFMQ-15) was administered to measure the degree to which clients engage in 5 aspects of mindfulness (ie, observing, describing, acting with awareness, nonjudging of inner experience, and nonreactivity to inner experience) [48].

Patient Health Questionnaire-9

The PHQ-9 is a 9-item self-report measure of the frequency of symptoms of depression in the past 2 weeks. Items are rated on a 4-point Likert-type scale ranging from 0 (not at all) to 3 (nearly every day). Higher scores are indicative of greater symptoms of depression. The PHQ-9 has demonstrated good sensitivity, specificity, and convergent validity [49,50]. Reliability was good for excellent for this study (range of before and after treatment, $\alpha=.78-.85$; $\omega=0.79-0.85$).

Generalized Anxiety Disorder-7

The GAD-7 is a 7-item self-report measure of the frequency of anxiety symptoms in the past 2 weeks. Items are rated on a 4-point Likert-type scale ranging from 0 (not at all) to 3 (nearly every day). Higher scores are indicative of greater symptoms of anxiety. The GAD-7 has demonstrated good internal consistency ($\alpha=.89$) and good test-retest reliability ($r=0.83$) [42]. The reliability was good to excellent for this study ($\alpha=.87-.91$; $\omega=0.87-0.92$).

Posttraumatic Stress Disorder Checklist for the DSM-5

The PCL-5 is a 20-item measure of each of the 4 clusters of PTSD (ie, intrusive thoughts, avoidance, negative alterations in mood, and alterations in arousal and reactivity). Items are rated on a 5-point Likert-type scale ranging from 0 (not at all) to 4 (extremely). Higher scores are indicative of greater symptoms of PTSD. The PCL-5 has demonstrated strong diagnostic utility within the general population [51] as well as strong test-retest reliability ($r=0.82$) and high internal consistency ($\alpha=.94$) [43]. Reliability was good to excellent for this study ($\alpha=.85-.95$; $\omega=0.94-0.95$).

Dimensions of Anger Reaction Scale-5

The DAR-5 is a 5-item self-report measure of the dimensions of anger reactions, especially in stressful situations, over the past 4 weeks. Items are rated on a 5-point Likert-type scale ranging from 1 (none or almost none of the time) to 5 (all or almost all of the time). Higher scores are indicative of greater levels of distress. The DARS-5 has demonstrated convergent and discriminant validity [44,52] as well as high internal consistency ($\alpha=.95$) [44]. Reliability was good to excellent for this study ($\alpha=.88$; $\omega=0.89-0.90$).

Insomnia Severity Index

The ISI is a 7-item self-report measure designed to assess difficulties with sleep and insomnia [45]. Frequency items regarding how often someone experienced a problem in the last 2 weeks are rated on a 5-point Likert-type scale ranging from 0 (not at all) to 4 (very severe). Items regarding satisfaction with sleep, noticeability of sleep problems, worries about sleep, and interference with daily functioning are rated on a 5-point Likert-type scale, with higher scores indicating greater sleep difficulties. The ISI has demonstrated adequate concurrent validity and good internal consistency ($\alpha=.74$ [45]).

Brief Resilience Scale

The BRS is a 6-item self-report measure of resilience. Items are rated on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores are indicative of higher levels of resilience. The BRS has demonstrated good validity and good to excellent internal consistency ($\alpha=.80-.91$) [47]. Reliability was good to excellent for this study ($\alpha=.87-.91$; $\omega=0.87-0.91$).

Five Facet Mindfulness Questionnaire

The FFMQ-15 is a 15-item short-form self-report mindfulness measure designed to assess 5 facets of mindfulness using 5 subscales: observing, describing, acting with awareness, nonjudging of inner experience, and nonreactivity to inner
experience. Statements are rated on a 5-point Likert-type scale ranging from 1 (never or very rarely true) to 5 (very often or always true). Higher scores are indicative of higher levels of mindfulness. The FFMQ-15 has good convergent validity with the FFMQ-39 [48]. Reliability for the total scale was good for this study ($\alpha=0.79-.85; \omega=0.80$). Subscale reliability for this study was also good ($\alpha=0.77-.90; \omega=0.81-0.91$).

Treatment Use and Satisfaction
Each week during the course, clients were asked how much time they spent practicing mindfulness meditation throughout the week in minutes and how many days they practiced. At 8 weeks postenrollment, we administered a bespoke questionnaire assessing satisfaction with the mindfulness meditations and the intervention overall.

Posttreatment Semistructured Interview
After completing their 8-week measures, clients were invited to participate in semistructured telephone interviews to discuss their perspectives on the mindfulness meditations and the course in general. Invitations were extended upon completion of the course to ensure that clients had sufficient opportunities to review all course materials. Invitations continued until 12 interviews were conducted, which was deemed sufficient given the objectives and the exploratory nature of this study, and this is consistent with prior recommendations [53,54]. Interviews were conducted by a clinician and a research assistant and included questions such as “What parts of the course did you find to be the most helpful? Why?” and “Were there parts of the meditation that you did not like? Why?” Interviews ranged from 10 to 20 minutes and were recorded for later transcription and analysis. The clients were asked to provide both positive and negative constructive feedback on the mindfulness meditations. Informed consent was obtained before the commencement of each interview and confirmed verbally before each recording was initiated. Table 1 shows a summary of all measures used.

Table 1. Summary of measures.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Instrument</th>
<th>Description</th>
<th>Data collection time points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociodemographics</td>
<td>Web-based questionnaire</td>
<td>Self-report web-based questionnaires with questions about gender, province, PSP, sector, ethnicity, and age</td>
<td>Web-based screening</td>
</tr>
<tr>
<td>Depression</td>
<td>PHQ-9b</td>
<td>9-item self-report measure of symptoms of depression</td>
<td>Web-based screening; 8 weeks postenrollment</td>
</tr>
<tr>
<td>Anxiety</td>
<td>GAD-7c</td>
<td>7-item self-report measure of symptoms of anxiety</td>
<td>Web-based screening; 8 weeks postenrollment</td>
</tr>
<tr>
<td>Posttraumatic stress</td>
<td>PCL-5d</td>
<td>20-item self-report measure of symptoms of posttraumatic stress</td>
<td>Web-based screening; 8 weeks postenrollment</td>
</tr>
<tr>
<td>Anger</td>
<td>DAR-5e</td>
<td>5-item self-report measure of dimensions of anger reactions</td>
<td>Web-based screening; 8 weeks postenrollment</td>
</tr>
<tr>
<td>Insomnia</td>
<td>ISIf</td>
<td>7-item self-report measure of symptoms of insomnia</td>
<td>Web-based screening; 8 weeks postenrollment</td>
</tr>
<tr>
<td>Resilience</td>
<td>BRSg</td>
<td>6-item self-report measure of resilience</td>
<td>Web-based screening; 8 weeks postenrollment</td>
</tr>
<tr>
<td>Mindfulness</td>
<td>FFMQ-15b</td>
<td>15-item self-report measure of 5 facets of mindfulness</td>
<td>Web-based screening; 8 weeks postenrollment</td>
</tr>
<tr>
<td>Treatment use and satisfaction</td>
<td>Web-based questionnaire</td>
<td>Self-report web-based questionnaires with questions about treatment use, treatment adherence, and satisfaction with treatment</td>
<td>8 weeks postenrollment</td>
</tr>
<tr>
<td>Perspectives on the mindfulness meditations and the course</td>
<td>Posttreatment semistructured interview</td>
<td>Semistructured interview with a trained researcher inquiring about perspectives on mindfulness and the course</td>
<td>10 weeks postenrollment</td>
</tr>
</tbody>
</table>

aPSP: public safety personnel.
bPHQ-9: Patient Health Questionnaire-9.
cGAD-7: Generalized Anxiety Disorder-7.
dPCL-5: Posttraumatic Stress Disorder Checklist for the DSM-5.
eDAR-5: Dimensions of Anger Reaction Scale-5.
fISI: Insomnia Severity Index.
gBRS: Brief Resilience Scale.
hFFMQ-15: Five Facet Mindfulness Questionnaire.
Analyses

Quantitative Analyses
SPSS (version 26; IBM Corp) was used to conduct the quantitative analyses. Data were deidentified before the analysis. Descriptive statistics were used to describe the sample’s demographic and clinical characteristics, engagement in mindfulness meditation practice, and satisfaction with the mindfulness meditations and the course in general. Completer analysis was used (ie, we did not impute missing data), as is common in pilot studies [55]. Paired sample t tests (2-tailed) were conducted to measure changes from baseline to posttreatment in the total scores on the GAD-7, PHQ-9, PCL-5, DAR-5, FFMQ-15, ISI, and BRS. Partial correlations were used to test for relationships between posttreatment measures (as measured by the GAD-7, PHQ-9, DAR-5, FFMQ-15, ISI, and BRS) and minutes spent meditating, when controlling for pretreatment measures.

Qualitative Analyses
Client interview data were deidentified and analyzed using NVivo 12 (Lumivero). The data were analyzed using a directed content analysis approach [56]. An initial codebook was developed by CAL to align with the questions posed during the semistructured interview. The data were then grouped into categories by CAL using a realist approach, whereby client data were treated as their descriptions of reality [57]. New codes were created by CAL when the data did not fit into the preexisting codes. The codebook and data were independently reviewed by JDB, a research associate with extensive experience in qualitative research. CAL and JDB convened to address conflicts in the codebook and continued discussions until a consensus was reached.

Results

Client Flow and Demographics
Figure 1 shows the flow of clients through this study, with 40 clients enrolling in and initiating the course and 32 (80%) clients completing the outcome measures. The mean age of the clients was 40.57 (SD 9.75) years. Most clients identified as White (37/40, 93%) and women (26/40, 65%) and reported currently residing in Saskatchewan (27/40, 68%). Most clients (22/40, 55%) reported that they were police, although other occupational groups were also represented (Table 2). In total, 30% (11/40) of the clients had scores above the clinical cut-off on the PHQ-9, 27% (11/40) above the cut-off on the GAD-7, 15% (6/40) above the cut-off on the PCL-5, and 45% (18/40) screened positive for at least 1 mental disorder. In terms of program completion at 8 weeks, 80% (32/40) of the clients had completed all lessons. On average, clients sent 5.42 (SD 3.57) messages to their therapists and received 9.48 (SD 3.18) messages. They also had 2.39 (SD 3.13) phone calls with therapists.
Figure 1. Client flow diagram.

Completed web-based screening (n=72)

Excluded (n=13)
  • Only wanted to examine the screening process (n=1)
  • Could not be reached for telephone screening (n=12)

Completed telephone screening (n=59)

Did not enroll in the intervention (n=17)
  • Enrolled in another PSPNET course (n=15)
  • Service not required (n=2)

Enrolled in the intervention (n=42)

Did not begin the intervention (n=2)

Began the intervention (n=40)
  • Accessed lesson 1 (n=40/40, 100%), Viewed meditation (n=28/40, 70%)
  • Accessed lesson 2 (n=38/40, 95%), Viewed meditation (n=28/40, 70%)
  • Accessed lesson 3 (n=35/40, 88%), Viewed meditation (n=18/40, 45%)
  • Accessed lesson 4 (n=34/40, 85%), Viewed meditation (n=21/40, 53%)
  • Accessed lesson 5 (n=32/40, 80%), Viewed meditation (n=21/40, 53%)

Did not complete the intervention (n=8)
  • Accessed lesson 1 (n=8/8, 100%), Viewed meditation (n=3/8, 38%)
  • Accessed lesson 2 (n=6/8, 75%), Used meditation (n=2/8, 25%)
  • Accessed lesson 3 (n=3/8, 38%), Used meditation (n=1/8, 13%)
  • Accessed lesson 4 (n=2/8, 25%), Used meditation (n=1/8, 13%)
  • Accessed lesson 5 (n=0/8, 0%), Used meditation (n=0/8, 0%)

Completed primary measures at 8 weeks (n=32)

Did not complete semistructured interview (n=20)
  • Not invited to participate (n=2)
  • Declined to participate (n=9)
  • No response (n=12)

Completed the semistructured interview (n=12)
Table 2. Clients’ demographic characteristics and occupations (N=40).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Clients, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>26 (65)</td>
</tr>
<tr>
<td>Men</td>
<td>14 (35)</td>
</tr>
<tr>
<td>Province</td>
<td></td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>27 (68)</td>
</tr>
<tr>
<td>Other province (eg, Prince Edward Island, Alberta, New Brunswick, Nova Scotia, Nunavut, and Ontario)</td>
<td>13 (33)</td>
</tr>
<tr>
<td>PSP^b sector</td>
<td></td>
</tr>
<tr>
<td>Police</td>
<td>22 (55)</td>
</tr>
<tr>
<td>Fire</td>
<td>5 (13)</td>
</tr>
<tr>
<td>Corrections</td>
<td>4 (10)</td>
</tr>
<tr>
<td>Communications (eg, 911 and dispatch)</td>
<td>4 (10)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (13)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Ethnic minority (eg, First Nations, Inuit, and Metis; Asian; Middle Eastern; Black; and South Asian)^a</td>
<td>3 (8)</td>
</tr>
<tr>
<td>White</td>
<td>37 (93)</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>6 (15)</td>
</tr>
<tr>
<td>30-39</td>
<td>10 (25)</td>
</tr>
<tr>
<td>40-49</td>
<td>16 (40)</td>
</tr>
<tr>
<td>≥50</td>
<td>7 (17)</td>
</tr>
</tbody>
</table>

^aCells are merged owing to the small cell size to protect confidentiality.
^bPSP: public safety personnel.

Meditation Use

No prospective clients reported participating in regular mindfulness practice at the time of their enrollment. Figure 1 shows that of the 40 clients, 28 (70%) viewed the grounding meditation in lesson 1, 28 (70%) viewed the loving kindness meditation in lesson 2, 18 (45%) viewed the meditation on awareness of breath in lesson 3, 21 (53%) viewed the meditation on awareness of the 5 senses in lesson 4, and 21 (53%) viewed the body scan meditation in lesson 5, as tracked through automatic computer system logs. Meditations were available for download; as such, some clients may have listened to the meditations more often than indicated by the system. Clients who completed the course tended to spend more time on the meditations overall, although this difference was not statistically significant (P>.16). Table 3 shows the mean number of text views and audio listens for each meditation and suggests that clients demonstrated a preference for reviewing the text over listening to the audio. Table 3 shows that the clients who completed the course accessed the written meditations an average of 4.91 (SD 2.91) times. Table 4 shows the self-reported number of times meditating and the number of minutes spent meditating and shows that those who meditated, meditated on average 5.67 times for a total of 28.48 minutes of meditation. As shown in Table 4, there was wide variability in days meditating (0-28 days) and minutes meditating (0-290 minutes) over the course. Of note, 9 clients opened meditation materials but reported that they did not use those materials.
Table 3. Web-based meditation views and listens.

<table>
<thead>
<tr>
<th></th>
<th>All clients (N=40), mean (SD; range)</th>
<th>Course completers (n=32), mean (SD; range)</th>
<th>Course noncompleters (n=8), mean (SD; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Text views</td>
<td>Audio listens</td>
<td>Text views</td>
</tr>
<tr>
<td>Lesson 1</td>
<td>1.25 (1.19; 0-4)</td>
<td>0.73 (0.93; 0-3)</td>
<td>1.39 (1.22; 0-4)</td>
</tr>
<tr>
<td>Lesson 2</td>
<td>1.05 (0.96; 0-4)</td>
<td>0.48 (0.85; 0-4)</td>
<td>1.15 (0.97; 0-4)</td>
</tr>
<tr>
<td>Lesson 3</td>
<td>0.60 (0.84; 0-4)</td>
<td>0.43 (0.75; 0-3)</td>
<td>0.67 (0.85; 0-4)</td>
</tr>
<tr>
<td>Lesson 4</td>
<td>0.73 (0.82; 0-3)</td>
<td>0.18 (0.38; 0-1)</td>
<td>0.85 (0.83; 0-3)</td>
</tr>
<tr>
<td>Lesson 5</td>
<td>0.78 (0.92; 0-3)</td>
<td>0.25 (0.44; 0-1)</td>
<td>0.85 (0.87; 0-3)</td>
</tr>
<tr>
<td>Total</td>
<td>4.40 (3.14; 0-12)</td>
<td>2.05 (2.25; 0-9)</td>
<td>4.91 (2.91; 0-12)</td>
</tr>
</tbody>
</table>

Table 4. Days and minutes spent meditating.

<table>
<thead>
<tr>
<th></th>
<th>All clients (N=40), mean (SD; range)</th>
<th>Course completers (n=32), mean (SD; range)</th>
<th>Course noncompleters (n=8), mean (SD; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Days mediated</td>
<td>Minutes mediated</td>
<td>Days mediated</td>
</tr>
<tr>
<td>Week 2</td>
<td>0.57 (1.26; 0-6)</td>
<td>1.68 (6.90; 0-40)</td>
<td>0.70 (1.36; 0-6)</td>
</tr>
<tr>
<td>Week 3</td>
<td>0.45 (0.78; 0-3)</td>
<td>3.25 (6.94; 0-20)</td>
<td>0.55 (0.83; 0-3)</td>
</tr>
<tr>
<td>Week 4</td>
<td>0.88 (1.42; 0-6)</td>
<td>3.13 (10.66; 0-60)</td>
<td>1.06 (1.50; 0-6)</td>
</tr>
<tr>
<td>Week 5</td>
<td>0.48 (1.01; 0-3)</td>
<td>4.20 (10.75; 0-60)</td>
<td>0.58 (1.09; 0-3)</td>
</tr>
<tr>
<td>Week 6</td>
<td>0.80 (1.57; 0-6)</td>
<td>4.38 (12.26; 0-60)</td>
<td>0.97 (1.69; 0-6)</td>
</tr>
<tr>
<td>Week 7</td>
<td>0.75 (1.50; 0-6)</td>
<td>4.00 (14.99; 0-70)</td>
<td>0.91 (1.65; 0-6)</td>
</tr>
<tr>
<td>Week 8</td>
<td>0.75 (1.60; 0-6)</td>
<td>2.88 (8.00; 0-40)</td>
<td>0.91 (1.72; 0-6)</td>
</tr>
<tr>
<td>Total</td>
<td>4.68 (6.50; 0-28)</td>
<td>23.50 (49.07; 0-290)</td>
<td>5.67 (6.76; 0-28)</td>
</tr>
</tbody>
</table>

Changes in Symptoms, Resilience, and Mindfulness

The 32 clients who completed the 8-week measures reported statistically significant reductions in total scores on the GAD-7, PHQ-9, PCL-5, and DAR-5 (all \( P < .001 \)) and a statistically significant increase in BRS total scores (\( P = .01 \)). No statistically significant change in ISI total scores was observed. Clients who completed the 8-week measures reported statistically significant increases in overall FFMQ-15 scores, with examination of the subscales revealing changes in mindfulness nonjudging and nonreactivity scores but not describing, acting with awareness, or observing. Time spent meditating was not associated with posttreatment measures while controlling for pretreatment measures (all \( P = .09-.67 \)). The details are presented in Table 5.
Table 5. Changes in outcome measures.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Pretreatment score, mean (SD)</th>
<th>Posttreatment score, mean (SD)</th>
<th>t test (df)</th>
<th>P value</th>
<th>Cohen d</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAD-7\textsuperscript{a}</td>
<td>7.88 (6.13)</td>
<td>1.48 (4.07)</td>
<td>−3.74 (32)</td>
<td>&lt;.001</td>
<td>−1.23</td>
</tr>
<tr>
<td>PHQ-9\textsuperscript{b}</td>
<td>8.81 (5.41)</td>
<td>5.27 (3.51)</td>
<td>−5.10 (32)</td>
<td>&lt;.001</td>
<td>−0.78</td>
</tr>
<tr>
<td>PCL-5\textsuperscript{c}</td>
<td>20.12 (15.61)</td>
<td>10.42 (10.48)</td>
<td>−6.13 (32)</td>
<td>&lt;.001</td>
<td>−0.74</td>
</tr>
<tr>
<td>DAR-5\textsuperscript{d}</td>
<td>9.85 (4.23)</td>
<td>7.40 (2.84)</td>
<td>−3.87 (32)</td>
<td>&lt;.001</td>
<td>−0.67</td>
</tr>
<tr>
<td>ISF\textsuperscript{e}</td>
<td>10.42 (6.06)</td>
<td>9.09 (4.59)</td>
<td>−1.70 (32)</td>
<td>.02</td>
<td>−0.30</td>
</tr>
<tr>
<td>BRS\textsuperscript{f}</td>
<td>3.15 (0.74)</td>
<td>3.42 (0.71)</td>
<td>2.72 (32)</td>
<td>.01</td>
<td>−0.46</td>
</tr>
<tr>
<td>FFMQ-15\textsuperscript{g}</td>
<td>38.27 (8.25)</td>
<td>40.58 (7.40)</td>
<td>3.51 (32)</td>
<td>&lt;.001</td>
<td>0.47</td>
</tr>
<tr>
<td>Acting with awareness</td>
<td>9.52 (2.53)</td>
<td>9.48 (2.51)</td>
<td>−0.11 (32)</td>
<td>.91</td>
<td>0.03</td>
</tr>
<tr>
<td>Describing</td>
<td>9.34 (3.14)</td>
<td>9.90 (2.85)</td>
<td>1.57 (32)</td>
<td>.13</td>
<td>0.30</td>
</tr>
<tr>
<td>Nonjudgmental</td>
<td>10.00 (2.93)</td>
<td>10.85 (1.43)</td>
<td>2.68 (32)</td>
<td>.01</td>
<td>0.42</td>
</tr>
<tr>
<td>Nonreactivity</td>
<td>9.36 (2.83)</td>
<td>10.33 (2.16)</td>
<td>2.55 (32)</td>
<td>.02</td>
<td>0.44</td>
</tr>
<tr>
<td>Observing</td>
<td>8.48 (2.98)</td>
<td>8.79 (2.87)</td>
<td>2.72 (32)</td>
<td>.40</td>
<td>0.15</td>
</tr>
</tbody>
</table>

\textsuperscript{a}GAD-7: Generalized Anxiety Disorder-7.
\textsuperscript{b}PHQ-9: Patient Health Questionnaire-9.
\textsuperscript{c}PCL-5: Posttraumatic Stress Disorder Checklist for the DSM-5.
\textsuperscript{d}DAR-5: Dimensions of Anger Reaction Scale-5.
\textsuperscript{e}ISI: Insomnia Severity Index.
\textsuperscript{f}BRS: Brief Resilience Scale.
\textsuperscript{g}FFMQ-15: Five Facet Mindfulness Questionnaire.

Treatment Satisfaction
The 32 clients who completed the 8-week measures generally reported feeling satisfied with the course. The 27 clients who accessed the mindfulness meditations also generally reported that they were satisfied with the mindfulness meditations. Most clients reported that they would recommend the intervention to a friend (32/32, 100%) and that they would recommend the mindfulness meditations to a friend (25/32, 93%). Table 6 provides details.
Table 6. Treatment satisfaction (n=32).

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n=32), n (%)</th>
<th>Mindfulness meditations (n=27), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommend to a friend</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>32 (100)</td>
<td>25 (93)</td>
</tr>
<tr>
<td>No</td>
<td>0 (0)</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Satisfaction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very dissatisfied</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>1 (3)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Neutral</td>
<td>4 (13)</td>
<td>11 (41)</td>
</tr>
<tr>
<td>Satisfied</td>
<td>15 (47)</td>
<td>10 (37)</td>
</tr>
<tr>
<td>Very satisfied</td>
<td>12 (38)</td>
<td>5 (19)</td>
</tr>
<tr>
<td><strong>Worth the time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>32 (100)</td>
<td>25 (89)</td>
</tr>
<tr>
<td>No</td>
<td>0 (0)</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Increased confidence in the ability to manage symptoms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No change</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Increased</td>
<td>22 (69)</td>
<td>17 (63)</td>
</tr>
<tr>
<td>Greatly increased</td>
<td>10 (31)</td>
<td>10 (37)</td>
</tr>
</tbody>
</table>

**Qualitative Results**

Overall, clients reported that the mindfulness meditations were beneficial (Table 7). Key themes from the qualitative analysis included clients liking the simplicity of following the mindfulness meditations, incorporating other strategies into the mindfulness meditations, and the variety of mindfulness meditations that were presented. Suggestions included providing videos alongside the mindfulness meditations for individuals who consider themselves to be more “visual,” providing distinct end points in the mindfulness meditations (e.g., a bell chime to indicate when the meditation end), and providing shorter mindfulness meditations to start. Suggestions for technical or presentation changes, including creating an atmosphere for mindfulness meditations to be completed in a group setting, were also provided.
### Table 7. Results of the qualitative analyses (n=12).

<table>
<thead>
<tr>
<th>Theme</th>
<th>Endorsing theme, n (%)</th>
<th>Client quotations reflecting the themes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tried meditation before beginning the course (n=9, 75%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes, beneficial</td>
<td>7 (78)</td>
<td>“I’d used mindfulness quite a bit in the past [and I] still use it to some extent” [Client 26]</td>
</tr>
<tr>
<td>Yes, skeptical</td>
<td>2 (22)</td>
<td>“I kinda thought it would be maybe hokey-pokey...um, a little wishy-washy...yeah, I was a little skeptical shall we say” [Client 20]</td>
</tr>
<tr>
<td>Have not tried meditation before beginning the course</td>
<td>3 (25)</td>
<td>N/A*</td>
</tr>
<tr>
<td>Beneficial for PSP</td>
<td>10 (83)</td>
<td>“It’s nice to, it’s probably a really good idea for most of us, ‘cause I think anybody who is in, ah, policing and-and stuff is, ah, likely type A...go-go-go type of personality...so it’s good to take a step back, focus on what you can control, because everything we deal with is out of our control” [Client 21]</td>
</tr>
<tr>
<td>Mindfulness reduced stress and improved relaxation</td>
<td>7 (58)</td>
<td>“I think definitely with stress I found if I was stressing about something in my job or my life or whatever, once I did the meditation usually I would—during meditation I’d be able to reflect on what I was stressed about.” [Client 18]</td>
</tr>
<tr>
<td>Mindfulness helped them to slow down and regulate their bodies and emotions</td>
<td>5 (42)</td>
<td>“It was kinda nice ‘cause it just got me to slow down and actually focus on my breathing. And just doing that kinda helped regulate everything else that was going on with my body.” [Client 16]</td>
</tr>
<tr>
<td>Mindfulness can be completed on their own time</td>
<td>2 (17)</td>
<td>“I found that helpful to be able to go back and replay them basically whenever I needed.” [Client 36]</td>
</tr>
<tr>
<td>Mindfulness reminded them to be gentle with themselves</td>
<td>2 (17)</td>
<td>“It made me talk to myself a little bit differently and be a little bit kinder to myself instead of like, focusing on everything I was doing wrong.” [Client 26]</td>
</tr>
<tr>
<td><strong>Challenges with mindfulness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeling uncomfortable sitting with their feelings</td>
<td>8 (67)</td>
<td>“I think most people struggle with quieting their minds and actually taking the time to do it. It seems really painful to sit there and keep refocusing your mind as it wanders.” [Client 20]</td>
</tr>
<tr>
<td>Difficulty finding motivation, time, and quiet space</td>
<td>8 (67)</td>
<td>“It’s kind of hard sometimes to find quiet time or quiet space in my house.” [Client 18]</td>
</tr>
<tr>
<td>Technical issues</td>
<td>2 (17)</td>
<td>“I struggled with the meditations because the audio never worked for me.” [Client 32]</td>
</tr>
<tr>
<td>Suggestions for improvement</td>
<td>5 (42)</td>
<td>“For me anything more than five to ten minutes, then I’m off the rails and I can’t concentrate anymore, so shorter is better for me.” [Client 21]</td>
</tr>
</tbody>
</table>

*aN/A: not applicable.
*bPSP: public safety personnel.

**Discussion**

**Principal Findings**

In this study, we explored the feasibility of adding mindfulness meditations to the PSP Wellbeing Course, as there is a gap in the literature with regard to the use of mindfulness meditations among PSP participating in iCBT. Overall, the study suggested that there was highly variable use of mindfulness meditations among PSP in the PSP Wellbeing Course, with 70% (28/40) of the clients reviewing the first 2 meditations and then about half of clients (ie, 18/40, 45% to 21/40, 53%) viewing the last 3 meditations. Clients more often looked at the text of the meditations rather than the audio of the meditations. There was significant variability among clients in the use of the meditations, with some clients never using the meditations and others using the meditations 3 to 4 times a week. Similarly, in terms of practice, there was high variability observed, with clients on average practicing 23 minutes a week. The use data alone suggest that the incorporation of mindfulness within iCBT will not be universally used when offered. It is not fully known why some clients did not use the meditations. It is possible that the high amount of content in the PSP Wellbeing Course itself may be a factor in the lower use of mindfulness meditations by some clients (ie, some clients may not have had enough time or willingness to practice mindfulness meditation in addition to other skills taught in the course).

In terms of improvements on measures over 8 weeks, clients in this course reported statistically significant improvements in symptoms of anger, anxiety, depression, PTSD, and insomnia. These findings are consistent with previous results regarding the PSP Wellbeing Course [9]. Although this was not a randomized controlled trial comparing iCBT alone with iCBT enhanced with mindfulness meditations, the results suggest that the addition of mindfulness did not have a marked positive or negative impact on the effectiveness of the course.
Other benefits observed in this study included improved resilience and improved mindfulness scores from pre- to posttreatment. In terms of mindfulness, the results suggested that the course was specifically associated with improvements in nonjudging and nonreactivity mindfulness scores. It appears that the course reduced the frequency with which clients labeled their thoughts and feelings as “good” or “bad” and enhanced client’s ability to detach from thoughts and feelings rather than getting caught up in their thoughts and feelings. The course was not associated with changes in observing (ie, noticing internal and external experiences), describing (ie, being able to express one’s experiences in words), or acting with awareness (ie, attending to present-moment experiences) mindfulness scores. The skills of observing, describing, and acting with awareness may be more specifically associated with the practice of mindfulness, and clients may not have participated in sufficient mindfulness meditations to experience these benefits.

As this was not a randomized controlled trial, we could not determine whether the changes in mindfulness were related to the mindfulness meditations or would have resulted from the PSP Wellbeing Course alone. In general, clients who participated in the interviews on mindfulness meditations reported a number of perceived benefits of mindfulness despite variable engagement with the mindfulness meditations. Most clients reported that mindfulness meditation helped reduce stress and improved relaxation and that mindfulness meditation can be beneficial with practice. Mindfulness has previously been associated with stress reduction and improved relaxation, among other benefits [16,58,59]. Clients also reported that they liked that other skills (eg, controlled breathing and emotion labeling) could be incorporated with the mindfulness meditations.

The interviews shed light on the challenges of offering mindfulness meditations to PSP as part of iCBT. A common challenge with the meditations that clients reported was the difficulty in sitting with their emotions. Clients also commonly reported difficulties with finding time or quiet space in which to complete the mindfulness meditations and reported that the mindfulness meditations were too lengthy. Beginning with shortened meditations that slowly increase in length may be better to help PSP to learn to cultivate mindfulness gradually. Future iterations of the intervention may benefit from including information on how mindfulness meditation can be practiced throughout the day in smaller timeframes to increase the accessibility of mindfulness meditation. A suggestion for improvement included providing shorter mindfulness meditations (eg, 2-3 min) to help people learning mindfulness meditation to work their way up to longer mindfulness meditations and gain increased comfort while sitting with their feelings.

Overall, the high degree of treatment satisfaction across the previous version of the PSP Wellbeing Course and this version enhanced with mindfulness meditation suggests that mindfulness meditations, at the very least, did not markedly decrease course satisfaction or negatively impact clients’ perceptions of the course. Of note, there was also no evidence to suggest that this version of the course impacted course completion. At 8 weeks, 80% (32/40) of the clients in this trial had completed all 5 lessons. Similarly, at 8 weeks, 77% of the clients in a prior study of the original PSP Wellbeing Course had completed the course [9]. In terms of therapist engagement, clients in this course sent an average of 5.24 emails and received 9.48 emails from their therapists. In the previous version of the course, on average, clients sent 4.98 (SD 5.53) messages to their therapists and received 9.80 (SD 4.71) messages [9]. Descriptively, treatment satisfaction was lower for mindfulness meditations than for other aspects of the course (ie, lesson material, meditations, and additional resources). The lower satisfaction with the mindfulness meditations may indicate that work is needed to improve the meditations and that some PSP are not open to practicing mindfulness.

Our sample of clients, who voluntarily enrolled in a therapist-guided iCBT course, may have been more interested in CBT skills than mindfulness skills. The large variation in the quality of publicly available mindfulness meditation programs means the clients in this study may have been justifiably skeptical of mindfulness meditation or unsure about the potential benefits of mindfulness meditation [60].

Of note, in this study, no statistically significant relationships were observed between time spent practicing mindfulness meditations and reductions in symptoms, which contrasts with previous evidence that increased time spent meditating is associated with increased symptom reduction [23]. This discrepancy may be a result of the small sample size and having inadequate power to detect effects. This discrepancy may also be explained by the finding that the intervention was already associated with significant reductions in symptoms, and the addition of mindfulness meditations could not meaningfully add to the results. Furthermore, previous studies on PSP and mindfulness appear to focus on the intervention as a whole (ie, the impact of the intervention and mindfulness combined) and have not considered the independent impact of mindfulness on outcomes. It should be noted that the extant research suggests that mindfulness can have positive effects when practiced for 10 minutes daily, but shorter periods can also be associated with benefits [39-41]. Evidence also suggests that the quality of mindfulness meditation may be more important than the amount of mindfulness meditation practiced [61]. Few studies have reported the actual time their clients reported practicing mindfulness [61]. In this study, it may have been preferable to assess the quality of practice rather than the minutes practiced.

**Limitations**

This study has important limitations that will help to inform future research. First, the sample size of the study was small; however, the detailed data collected allowed for an understanding of the usability and credibility of mindfulness meditation in a sample of PSP seeking treatment. The detailed data also allow for iterative improvements to the intervention, such as providing increased psychoeducation on mindfulness meditation and introducing ultrabrief meditations for clients who want to work for longer periods. Future studies on iCBT should be conducted using larger samples of PSP from diverse communities across Canada (and internationally), as this study is not generalizable cross-culturally. It should also be noted that the clients in this study overall had variable clinical symptoms. On average, the scores of this particular subsample suggested...
that clients had milder symptoms of depression, anxiety, anger, and PTSD, although 30% (11/40) of the clients had scores above the clinical cut-off on the PHQ-9, 27% (11/40) above the cut-off on the GAD-7, 15% (6/40) above the cut-off on the PCL-5, and 45% (18/40) screened positive for at least 1 of the aforementioned mental disorders. It is possible that the lower symptom severity overall in the sample may have impacted the relationship between minutes meditating and outcomes owing to floor effects. It is also notable that in this sample, there were more women than men, which may have affected the findings. Second, this study does not contain a randomized design and involves short-term follow-up; thus, it is descriptively compared with the previously published trial on the PSP Wellbeing Course. Future studies of the program could benefit from a randomized design and a larger sample size to evaluate potential differences in experiences between versions of the course. In addition, such studies could explore how the course compares with control conditions. Third, only a subsample of the clients was interviewed, and the experiences of those who did and did not consent to an interview could have differed (eg, clients with more favorable attitudes toward the mindfulness meditations may have been more likely to consent to an interview). Fourth, our ability to determine whether time spent meditating predicted outcome measure changes may have been limited by the low engagement from clients in the mindfulness meditations. Future studies should consider means to encourage mindfulness meditation use without clients feeling forced to complete the exercises.

Clinical Implications and Future Directions

This study has several important potential implications. First, results from this study replicated previous results regarding the PSP Wellbeing Course, indicating that the intervention is associated with statistically significant reductions in symptoms of anxiety, depression, PTSD, and anger with medium to large effect sizes [62]. Contributing to the literature on iCBT for PSP, the study showed that improvements were also observed in resilience and mindfulness with this version of the course. Use data showed that a significant percentage of clients used the mindfulness meditations. The information gathered from the interviews indicated that clients enjoyed that the mindfulness meditations could be incorporated with other strategies.

Therapists may be able to help support future clients in understanding what works for them and how to incorporate their existing strategies with mindfulness meditation to achieve increased benefits. Future programs may benefit from increasing psychoeducation regarding the potential discomfort associated with mindfulness meditation and encourage clients to consider practicing more to allow them to get used to experiencing emotions. Clients also made suggestions for improvement. The mindfulness meditations may be offered after the completion of the course as an additional resource for clients who may be interested in pursuing the meditations. Additional research is needed to identify who can benefit from mindfulness meditations and to ensure that those who do not benefit are not distracted from the rest of the course content in attempting to complete mindfulness activities. One potential solution is to include mindfulness meditations as an optional component, allowing interested individuals to use mindfulness within the course while not distracting individuals who are not interested in mindfulness.

Conclusions

To the best of our knowledge, this is the first study to assess web-based mindfulness meditations as an adjunct to a preexisting iCBT intervention among PSP. The incorporation of mindfulness meditation was largely acceptable to many clients enrolled in the intervention, although our findings were limited by variable client engagement with the mindfulness meditations. The mindfulness meditations were used by approximately half of the clients, and the clients who used them reported that they enjoyed the meditations. The addition of mindfulness meditations did not appear to remarkably affect the overall engagement, satisfaction, or outcomes of the course. Clients also expressed several challenges and suggestions for improvement, which represent opportunities to improve mindfulness meditations and iCBT for PSP in general. The presentation of shortened mindfulness meditations may serve as an important adjustment to allow PSP to experience the benefits of mindfulness while maintaining their busy schedules. The results of this study suggest that mindfulness meditations, when offered along with iCBT, may be an acceptable intervention for Canadian PSP. Future research is required to further explore how best to incorporate mindfulness into iCBT for PSP and the potential benefits of doing so now that we have demonstrated partial feasibility.

Acknowledgments

The authors would like to acknowledge the contributions of all public safety personnel (PSP) who participated and the members of the PSPNET team for their support with this project. The authors would also like to acknowledge the many stakeholders and organizations that assisted them in tailoring the PSP Wellbeing Course and distributing information about PSPNET to PSP. Acknowledgments are due to the Online Therapy Unit for their support in setting up PSPNET. The authors are indebted to Information Services at the University of Regina, especially Max Ivanov, for his pivotal role in developing the Online Therapy Unit platform that is used to host PSPNET. The authors would like to thank Dr Shadi Beshai, Dr Natasha Gallant, and Dr Jenn deLugt for providing valuable feedback on this study. The authors would also like to thank Ailesh Abrams and Isabelle Dene for helping complete the interviews and Peace Dukuye for her efforts. This manuscript is based on a published thesis defended by Caelleigh Landry [63]. This research was made possible by scholarship funding provided to CAL by the Canadian Institutes for Health Research, Mental Health Research Canada, and Saskatchewan Health Research Foundation. This study was also supported by PSPNET, which is funded by the Canadian Government’s Ministry of Public Safety and Emergency Preparedness. The funders had no involvement in the design of the study or in the collection, analysis, and interpretation of the data.
Data Availability
As part of the consent form, participants were informed that their data would only be shared with members of the research team. Therefore, we are unable to share data.

Conflicts of Interest
HDH is the director of PSPNET and executive director of the Online Therapy Unit. BD and NT are the authors and developers of the Wellbeing Course but do not derive any personal or financial benefit from it. The other authors have no conflicts of interest to declare.

References


Abbreviations
- **BRS**: Brief Resilience Scale
- **CBT**: cognitive behavioral therapy
- **DAR-5**: Dimensions of Anger Reaction Scale-5
- **FFMQ-15**: Five Facet Mindfulness Questionnaire
- **GAD-7**: Generalized Anxiety Disorder-7
- **iCBT**: internet-delivered cognitive behavioral therapy
- **ISI**: Insomnia Severity Index
- **PCL-5**: Posttraumatic Stress Disorder Checklist for the DSM-5
- **PHQ-9**: Patient Health Questionnaire-9
- **PSP**: public safety personnel
- **PTSD**: posttraumatic stress disorder

Edited by A Mavragani; submitted 01.11.23; peer-reviewed by J Olthuis, R Dias; comments to author 21.11.23; revised version received 19.12.23; accepted 10.01.24; published 30.01.24.

Please cite as:
Landry CA, McCall HC, Beahm JD, Titov N, Dear B, Carleton RN, Hadjistavropoulos HD
Web-Based Mindfulness Meditation as an Adjunct to Internet-Delivered Cognitive Behavioral Therapy for Public Safety Personnel: Mixed Methods Feasibility Evaluation Study
JMIR Form Res 2024;8:e54132
URL: https://formative.jmir.org/2024/1/e54132
doi:10.2196/54132
PMID:38289655

©Caeleigh A Landry, Hugh C McCall, Janine D Beahm, Nickolai Titov, Blake Dear, R Nicholas Carleton, Heather D Hadjistavropoulos. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Formative Evaluation of a Comprehensive Self-Management Intervention for Irritable Bowel Syndrome, Comorbid Anxiety, and Depression: Mixed Methods Study

Kendra Kamp1, MS, PhD; Pei-Lin Yang2, MSN, PhD; Emily Friedman3, MID; Alejandra Lopez1, BS; Sarah Iribarren1, PhD; Pamela Barney1, MN; Sean Munson4, PhD; Margaret Heitkemper1, PhD; Rona Levy5, MSW, PhD, MPH

1Department of Biobehavioral Nursing and Health Informatics, School of Nursing, University of Washington, Seattle, WA, United States
2National Defense Medical Center, Taipei, Taiwan
3Alacrity Center, University of Washington, Seattle, WA, United States
4Human Centered Design and Engineering, University of Washington, Seattle, WA, United States
5School of Social Work, University of Washington, Seattle, WA, United States

Corresponding Author:
Kendra Kamp, MS, PhD
Department of Biobehavioral Nursing and Health Informatics, School of Nursing
University of Washington
1410 NE Campus Parkway
Seattle, WA, 98195
United States
Phone: 1 206 221 4617
Email: kamp@uw.edu

Abstract

Background: Irritable bowel syndrome (IBS) is a disorder of the gut-brain interaction that is associated with abdominal pain, altered bowel patterns, and reduced quality of life. Up to 50% of patients with IBS also report anxiety or depressive symptoms. Although effective self-management interventions exist for individuals with IBS, few have been effectively implemented, and most do not consider the unique needs of patients with comorbid IBS and anxiety or depression.

Objective: This study aimed to determine the anticipated acceptability, appropriateness, feasibility, and usability of a comprehensive self-management intervention using an implementation science and human-centered design approach among individuals with comorbid IBS and anxiety or depression and health care providers.

Methods: A convergent mixed methods design was used to elicit feedback on the comprehensive self-management intervention outline and content to identify refinement needs before testing. Patients with IBS and moderate to severe anxiety or depression and health care providers were purposefully sampled from primary care and gastroenterology settings. Participants completed semistructured interviews and surveys on anticipated acceptability, appropriateness, feasibility, and usability.

Results: Patient participants (n=12) were on average 36.8 (SD 12.2) years of age, and 42% (5/12) were currently receiving psychological therapy. Health care providers (n=14) were from primary care (n=7) and gastroenterology (n=7) settings. The mean usability scores (out of 100) were 52.5 (SD 14.5) for patients and 45.6 (SD 11.6) for providers. For patients and providers, qualitative data expanded the quantitative findings for acceptability and appropriateness. Acceptability findings were the comprehensive nature of the intervention and discussion of the gut-brain interaction. For appropriateness, participants reported that the intervention provided structure, accountability, and support. Feasibility was confirmed for patients, but there was a divergence of findings between quantitative and qualitative measures for providers. Patients focused on intervention feasibility, while providers focused on implementation feasibility in the clinic. Identified usability issues to address before implementation included the intervention delivery format, length, and lack of integration into health care settings that, if not addressed, may limit the reach of the intervention.

Conclusions: Patients and health care providers found the intervention acceptable and appropriate. Several feasibility and usability issues were identified, including intervention delivery methods, length of intervention, and the best methods to implement in the clinic setting. The next steps are to refine the intervention to address the identified issues and test in a pilot study whether addressing usability issues leads to the anticipated improvements in implementation and uptake.
Introduction

Irritable bowel syndrome (IBS) is a disorder of gut-brain interaction that affects 6%-18% of individuals worldwide [1,2]. Many evidence-based practice interventions (EBPIs) have been developed to address symptoms of IBS, such as abdominal pain, and improve quality of life [3]. However, a gap exists in translating EBPIs into clinical practice settings. Clinical practice guidelines support the use of behavioral EBPIs for IBS [4,5], yet only a small proportion of patients actually receive such interventions [6]. Implementation science is a field that focuses on translating evidence-based practice into real-world settings [7]. Attending to implementation outcomes such as acceptability, feasibility, appropriateness, and usability can assist in identifying facilitators and barriers to successful intervention implementation and adoption within clinical practice settings [7-9].

When integrating EBPIs for IBS into real-world settings, an important consideration is the common comorbidities that exist among many individuals with IBS. Most notably, 30%-50% of patients diagnosed with IBS also have a diagnosis of anxiety or depression [10]. Psychological distress is linked to the onset and exacerbation of IBS symptoms, and reducing symptom severity and burden is a key component of behavioral approaches. With the publication of the consensus-driven Rome IV criteria [11], there is a growing appreciation that IBS is a disorder of gut-brain interaction. Although many studies have examined the effectiveness of interventions for IBS, including cognitive behavioral therapy and dietary interventions [6,12,13], few have focused specifically on interventions for individuals with IBS and comorbid anxiety or depression. Given the high prevalence of IBS in the United States, along with comorbid anxiety and depression, there is a significant need to implement effective therapeutic strategies to address both IBS and psychological distress [12,14].

One EBPI, the comprehensive self-management (CSM) intervention, has been shown in multiple randomized controlled trials to reduce abdominal pain symptoms and improve quality of life [15-17]. The intervention content has been published as a book, Master Your IBS [15-17]. Although initially developed for IBS symptoms, the intervention has elements of a transdiagnostic approach, reducing other common symptoms of anxiety, depression, extraintestinal pain, fatigue, and sleep disturbances [18,19]. The intervention consists of 8 1-hour sessions, which can be provided by a psychiatric nurse practitioner or similarly trained health professional. However, there is a gap in knowledge regarding how to implement the CSM intervention into clinical practice, specifically from the perspective of key stakeholders: individuals with IBS and comorbid anxiety or depression and health care providers.

Research has argued for applying human-centered design and usability principles to address the lack of intervention implementation by redesigning interventions to improve usability while retaining the effective components [20-22]. Although usability has been most often applied in technology-based applications, usability evaluation principles can also be used to assess other products and services, including interventions and implementation strategies [9,23,24]. Human-centered design approaches focus on developing usable interventions through stakeholder input [21]. By addressing design and content issues, interventions can have increased usability and acceptability to better integrate into clinical settings.

In this research, we sought to examine the usability and acceptability of the current, paper-based CSM intervention from the perspectives of patients with IBS and comorbid anxiety or depression and health care providers in primary care and gastroenterology settings. Our formative evaluation was intended to support (1) refining and adapting the CSM to a digital format and (2) the identification of implementation strategies to facilitate adoption in clinical practice settings. The initial impressions patients and health care providers form regarding the anticipated acceptability, feasibility, appropriateness, and usability of an intervention affect their likelihood of adopting it and help characterize their needs, and these impressions can inform intervention refinement or redesign as well as the selection of intervention strategies to plan for integration into clinical practice. Our research questions were as follows: (1) What are patient and provider perspectives on the acceptability, appropriateness, feasibility, and usability of the current CSM intervention? (2) What recommendations do patients and providers have for improving this intervention?

Methods

Design

We used a convergent mixed methods design to collect both quantitative (ie, surveys) and qualitative (ie, semistructured interviews) data [25] from patients with IBS and comorbid anxiety or depression, as well as primary care and gastroenterology health care providers. In this study, participants received the paper-based intervention content to review and critique, not the intervention directly. Each patient and health care provider participated in a semistructured interview to discuss the intervention content and structure and answer survey questions. Our interview and survey questions were guided by the discover, design, build, and test framework, which combines perspectives from implementation science and human-centered design [7,8,20]. To gain a comprehensive understanding of issues and recommended strategies, we designed our qualitative interview questions to align with the concepts addressed in the surveys [25]; these included acceptability, feasibility,
appropriateness [26], and usability [9] of the CSM intervention content and structure.

Sample

Individuals with IBS and self-identified comorbid anxiety or depression symptoms (referred to as patients) were recruited on the internet through 2 methods: the University of Washington Institute of Translational Health Sciences listservs and ResearchMatch, a national health volunteer registry supported by the US National Institutes of Health as part of the Clinical Translational Science Award program.

Patients were eligible if they met the following inclusion criteria: (1) aged between 18 and 70 years; (2) ROME IV IBS criteria of recurrent abdominal pain at least 1 day per week in the past 3 months that is associated with 2 or more of defecation, onset associated with a change in frequency of stool, or change in form of the stool; (3) have a diagnosis of IBS by a health care provider; (4) report moderate to severe anxiety or depression (Generalized Anxiety Disorder-7 [GAD-7] score of >10 [27]; Patient Health Questionnaire-9 [PHQ-9] score of >10 [28]); and (5) be able to read and write in English. Participants were excluded if they had a first-degree relative with colorectal cancer before the age of 60 years or had multiple “red flag” symptoms (ie, loss of 10 or more pounds without trying, blood in stool, or anemia). Patients completed a web-based screening questionnaire to assess their eligibility. All patients who expressed thoughts of hurting themselves were immediately directed to contact the National Suicide Prevention Lifeline.

Health care providers were recruited from primary care and gastroenterology clinics in Washington State. Individuals were eligible to participate if they self-reported caring for more than 3 patients with IBS per month. Health care providers from primary care clinics were recruited through the WWAMI (Washington, Wyoming, Alaska, Montana, and Idaho) region Practice and Research Network, a practice-based research network of primary care clinics and clinical organizations. Health care providers from Seattle gastroenterology clinics were recruited through purposeful convenient sampling. Health care providers received emails regarding the study and self-identified if they met the criteria of caring for at least 3 individuals with IBS per month.

Description of the CSM Intervention at the Start of the Study

The CSM intervention was developed as a comprehensive approach to improving quality of life and reducing abdominal pain among individuals with IBS [15-17]. The intervention is delivered in 8 individual sessions, with sessions lasting 60 minutes. Participants had up to 12 weeks to complete the 8 sessions, to allow for unexpected events. Participants can elect to complete the intervention in-person, over the telephone, or through a mixture of telephone and in-person sessions since there is no difference in intervention effectiveness between in-person and telephone modalities [16]. Each participant receives a paper-bound “IBS Managing Symptoms Workbook,” which includes information, worksheets, and homework assignments. Additionally, participants received CD audio recordings of the relaxation exercises. The CSM intervention includes content such as healthy thought patterns, problem-solving, healthy eating, and relaxation. Additionally, the intervention addresses practical topics such as sleep, travel, eating out, and physical intimacy. Participants receive verbal and written instructions regarding the use of the manual.

Measures

An overview of the measures by participant group (patients and health care providers) is presented in Table S1 in Multimedia Appendix 1.

Demographics

Age, sex, and race were assessed for patients and providers. For patients, the IBS subtype of constipation, diarrhea, or mixed was reported. Anxiety was measured using the GAD-7 questionnaire, which asks how often participants have been bothered by a list of 7 symptoms over the past 2 weeks [27]. Depression was measured with the PHQ-9, which asks participants how often they are bothered by 9 problems [28]. For both anxiety and depression measures, response options are on a 4-point scale, including “not at all,” “several days,” “more than half the days,” and “nearly every day.” Health care providers answered questions on the type of provider, years of working experience, and number of patients with IBS cared for per month.

Anticipated Acceptability, Appropriateness, and Feasibility

Anticipated acceptability, appropriateness, and feasibility were each measured with 4 items [26]. Participants responded on a 5-point Likert scale from 1 (completely disagree) to 5 completely agree. A higher score indicates greater agreement. Acceptability measures anticipate responsiveness to adopting a new implementation plan. Appropriateness measures the anticipated suitability of the intervention and the perceived fit of the intervention. Feasibility measures the anticipated likelihood of implementing the intervention.

Anticipated Usability

The anticipated usability of the CSM intervention was assessed using the Intervention Usability Scale, which has 10 items and was derived from the System Usability Scale [9,29]. Participants respond from “strongly disagree” to “strongly agree.” The scale ranges from 0 to 100. A score above 68 is considered average; a score below 68 is considered below-average usability.

Procedures

Institutional review board approval was obtained from the University of Washington (IRB# 00009463) before participant recruitment. All individuals were provided with a description of the risks and benefits of participating in the research study as well as an explanation that they could stop participating at any time. Individual interviews were conducted with patients and health care providers. For this study, participants provided feedback on the intervention content and format overall without participating in individual intervention sessions. The second phase of this study (data not reported in this manuscript) focused on obtaining feedback on the intervention redesign of the first 3 intervention sessions.
Patients and health care providers were asked questions using a semistructured interview guide. For example, introductory questions were asked about symptoms (patients) and the type of practice (health care providers). Next, both patients and health care providers were shown an outline of the current CSM content and completed a card sorting activity to categorize the intervention content into 3 categories ("most helpful," "seems okay," or "least helpful"). Additionally, the current CSM intervention format was described (eg, in-person and telephone-delivered intervention that included a paper-based workbook); we asked for patients’ and health care providers’ thoughts on the intervention content outline and format and how it could best be designed to integrate into their lives and promote adherence to the intervention. At the end of the interview, patients and health care providers were sent postinterview participatory design session questionnaires regarding the anticipated acceptability, appropriateness, feasibility, and usability of the intervention.

Data Analysis and Integration
Integration in quantitative and qualitative methods occurred through merging [25]. Quantitative and qualitative data were initially analyzed separately and brought together for analysis. For qualitative data, the interview recordings were transcribed. The 2 authors (KK and PLY) coded 2 transcripts from health care providers and 2 transcripts from patients to develop the coding scheme and reach a consensus using a framework approach [30]. The coding scheme was guided by the research questions to understand anticipated barriers and facilitators to acceptability, appropriateness, feasibility, and usability. Each coder then proceeded to code half of the interviews. Any discrepancies were discussed, and consensus was reached. The codes and results were presented to the entire team for further discussion. The mixed methods findings are presented within the text and highlight how the quantitative and qualitative findings align and show a confirmation of the findings, as well as those that are disparate and demonstrate discordance in findings [25]. By integrating the qualitative and quantitative data, we expanded our insights on which intervention components were acceptable and feasible and which needed modification. Mixed methods enabled us to gain new insights into the data, particularly by assessing numeric data to further explore themes from the qualitative interviews where usability or feasibility was lowest.

Ethical Considerations
This study was reviewed by the University of Washington Institutional Review Board (IRB# 00009463). All participants provided verbal consent before the interview. Data are presented as deidentified and do not include links to participant characteristics to protect the privacy and confidentiality of the research participants. Participants received a US $50 gift card for participating in the study.

Results
Demographics
A total of 12 patients completed the qualitative interview (Table 1). Patients had a mean age of 36.8 (SD 12.2) years and were predominantly female. Overall, 42% (n=5) were currently receiving psychological therapy for anxiety or depression. A total of 14 health care providers completed the interview; half (7/14) were primary care providers, and half (7/14) specialized in gastroenterology. Professional roles included 7 physicians, 1 physician’s assistant, and 6 nurse practitioners. One provider did not complete the questionnaires. Provider experience ranged from 4 to 26 years. On average, providers cared for 26 (SD 20; range 3-80) patients with IBS per month. Interviews lasted between 27 and 43 minutes for health care providers (mean 35, SD 5 minutes) and between 32 minutes and 1 hour and 11 minutes for patients (mean 47, SD 12 minutes).

<p>| Table 1. Characteristics of patients with irritable bowel syndrome and comorbid anxiety or depression and health care providers. |</p>
<table>
<thead>
<tr>
<th>Age (years), mean (SD)</th>
<th>Patients (n=12)</th>
<th>Health care providers (n=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3 (25)</td>
<td>3 (23)</td>
</tr>
<tr>
<td>Female</td>
<td>9 (75)</td>
<td>10 (77)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1 (8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asian</td>
<td>2 (16)</td>
<td>5 (38)</td>
</tr>
<tr>
<td>White</td>
<td>9 (75)</td>
<td>7 (54)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (8)</td>
<td>1 (8)</td>
</tr>
</tbody>
</table>

Anticipated Acceptability, Appropriateness, and Feasibility

Acceptability
Mean acceptability was 4.0 (SD 0.8) out of 5 by patients and 4.4 (SD 0.5) out of 5 by providers, indicating that on average, the current CSM intervention content and format were acceptable (Table S2 in Multimedia Appendix 1 for individual acceptability items).

The qualitative results confirm the intervention was acceptable through its comprehensive nature, in which patients could try different components and see what works for them (Table 2 for
qualitative quotes). Participants reported that, given the gut-brain interaction that exists, a comprehensive approach to management of both gastrointestinal and anxiety or depressive symptoms was an important component. Patients scored acceptability lower than health care providers, particularly for the item that asks if the intervention is appealing. Patients identified that several of the intervention topics were familiar, especially those who have struggled to manage their IBS for many years. Health care providers focused on individual intervention components such as access and literacy, cost or insurance coverage, and culture or race that could be potential barriers to patients engaging in a self-management intervention like the CSM. The card sorting activity identified content that was of higher priority to participants. Patients found content related to sleep, traveling, and trigger foods most helpful, whereas providers found content related to relaxation, introduction to IBS, sleep, and trigger foods most helpful.
Table 2. Qualitative findings among 12 patients with irritable bowel syndrome and comorbid anxiety or depression and 14 primary care and gastroenterology health care providers.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Illustrative qualitative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptability</td>
<td>“I like the fact that there’s a wide variety of things; I feel confident that at least 1 or 2 of these things from the 7 kind of content weeks would be helpful.” [Patient 7]</td>
</tr>
<tr>
<td></td>
<td>“I do think something like this could fit into my day-to-day life, because... it would give me some type of structure.” [Patient 8]</td>
</tr>
<tr>
<td></td>
<td>“I really think this looks like a very comprehensive plan to address holistically what may be contributing to people who have irritable bowel syndrome.” [Provider 3]</td>
</tr>
<tr>
<td></td>
<td>“I think improving accessibility, such that it’s one that I can give to [all] sort[s] of patients, regardless of their insurance status, regardless of where they live, or sort of their profession, would be good.” [Provider 8]</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>“I feel like it would keep me more accountable, and it would give me somebody I can talk through everything with instead of just trying to figure it out on my own.” [Patient 11]</td>
</tr>
<tr>
<td></td>
<td>“I think just having that intentionality and having structure is really important if someone wants to make a change.” [Patient 4]</td>
</tr>
<tr>
<td></td>
<td>“So, I’ve done that in my psychology as a therapist, but my therapist doesn’t know very much about IBS, and my doctor that knows about IBS, I just see them like for 15 minutes every 3 months or something. So, it would be nice to have someone who is aware of the integration of those things.” [Patient 7]</td>
</tr>
<tr>
<td></td>
<td>“When you’re doing something yourself and it doesn’t rely on a medication and this gives people a little bit of power. It gives them structure.” [Provider 12]</td>
</tr>
<tr>
<td></td>
<td>“I think that it also gives some accountability in terms of, ’Did you do these exercises?’ ’Did you bring your record?’ and these kinds of things.” [Provider 8]</td>
</tr>
<tr>
<td></td>
<td>“When you’re depressed or anxious and when your body feels like it’s turning against you, which is what a lot of people with IBS I hear being said to me, it gives you a facet of control. When you’re doing something yourself and it doesn’t rely on a medication and this gives people a little bit of power. It gives them structure.” [Provider 10]</td>
</tr>
<tr>
<td>Feasibility</td>
<td>“I guess the part that might be difficult is just making sure someone actually does it, and sticking with it, which is the hard part.” [Patient 4]</td>
</tr>
<tr>
<td></td>
<td>“I just think that somebody new to [IBS] would be more apt to get into this versus somebody who’s been through all this; they’d be like, ’I’ve done all this stuff already.’” [Patient 5]</td>
</tr>
<tr>
<td></td>
<td>“I’m curious, but also, I’m skeptical. I don’t know why. Just because I feel like I’ve tried so many things and I’m like, ’Really? Fiber is going to be the thing?’ Maybe I just have more to learn.” [Patient 10]</td>
</tr>
<tr>
<td></td>
<td>“I think the hard part is having a person that’s motivated enough to actually go through and do this on their own... unless there’s some accountability.” [Provider 12]</td>
</tr>
<tr>
<td></td>
<td>“I think if they go in for counseling, they have more time to do CBT-type stuff. They have more time to talk to the patient about it. Whereas in primary care we don’t always have that kind of time, but I think if it’s something small and short that I could give them during the visit and then they can work on it.” [Provider 2]</td>
</tr>
<tr>
<td>Usability</td>
<td>“I’d like it with an app, something that’s visual on the app as well as verbal. I’d like some types of video content to actually show me certain tasks for working through planning out certain things, as well as verbalized communication.” [Patient 8]</td>
</tr>
<tr>
<td></td>
<td>“In an e-course-esque environment I think would be really helpful or an app, if that’s a possibility.” [Patient 10]</td>
</tr>
<tr>
<td></td>
<td>“If you construct your own. Build your own, I don’t know, Amazon cart, I don’t like the bundle.” [Patient 1]</td>
</tr>
<tr>
<td></td>
<td>“I’d rather do it myself and if there was somebody after the fact that wanted to check up on me for 5 minutes and say, how did it go? Do you have any questions? Did you have any concerns? Did it work?” [Patient 2]</td>
</tr>
<tr>
<td></td>
<td>“I have a very geriatric[ic] heavy panel, which certainly would not do well with an app and need kind of person kind of contacting them on a weekly basis in some shape or form phone call or something. Whereas I could definitely see my more hyper-focused, got a lot of stuff going on, needing it more as an app with an alert that pops up on their phone that says, ’Hey, it’s time to work on your skill for today. Let’s set aside 15 minutes to do this,’ or whatever.” [Provider 1]</td>
</tr>
<tr>
<td></td>
<td>“Just a very brief: Patient’s doing well. Patient does not seem to be progressing. Patient is not participating. They haven’t returned any of their journals.” [Provider 7]</td>
</tr>
<tr>
<td></td>
<td>“ Afterwards as a summary, was this overall sort of useful or which parts of it did you find use in? And, so then I know what are your residual symptoms that we can sort of work on and address because I think it’s also hard to see the clear benefit right away can sometimes take a while even with patients. Once it’s even kicked in, they have... for them to start suddenly realizing so many months down the road, ’Actually my symptoms are doing a lot better. This used to be something I would think about all the time, and now it’s kind of rare.’” [Provider 8]</td>
</tr>
</tbody>
</table>

**Appropriateness**

Mean appropriateness was rated by patients as 4.0 (SD 0.7) and by providers as 4.2 (SD 0.5) out of 5. The qualitative data confirmed the quantitative finding by indicating that the intervention was appropriate because it provided accountability, support, and structure. This was especially important, as many patients felt they had tried multiple other strategies on their own through a trial-and-error approach and viewed the addition of a support person as very helpful and important. Patients discussed accountability within the context of having someone help them navigate and talk through their experiences. Health
care providers also identified the importance of accountability, although they discussed accountability within the context of patients completing assignments and activities.

Patients and health care providers also addressed the appropriateness of the intervention in relation to patients with IBS who had comorbid anxiety or depression. Health care providers noted that structure was especially important as individuals with IBS and comorbid anxiety or depression often feel that a lot of things are out of their control. Even individuals who already received psychological therapy for anxiety or depression (n=5) saw the benefit of integrating IBS and mental health. The integration of an intervention that could address symptoms of IBS along with anxiety and depression was viewed positively by patients.

Feasibility
Mean feasibility was rated by patients as 3.9 (SD 0.7) and by providers as 3.9 (SD 0.7) out of 5. Among patients, the qualitative findings confirmed the quantitative findings. Patients identified specific situations that could influence feasibility, such as a lack of motivation to participate or “stick with” the intervention. Another factor affecting motivation is that many individuals with IBS and comorbid anxiety or depression have already tried multiple strategies to manage their symptoms. Several of the participants felt that components of the intervention were familiar and were perhaps better suited for newly diagnosed individuals. Due to having tried multiple previous strategies, some participants were skeptical that an intervention could truly help manage their symptoms. Yet, they were still interested in trying it due to experiencing symptoms. Patients primarily focused on the intervention feasibility.

Among health care providers, the quantitative and qualitative findings were discordant; the quantitative feasibility score was positive; however, the qualitative portion identified multiple barriers to implementation within a clinic setting. Similar to patients, health care providers also identified patient lack of motivation as a barrier to the intervention. Health care providers felt their clinic visits were so short that they did not have time to review the CSM self-management approaches. Health care providers’ comments regarding feasibility were focused on the feasibility of implementing the intervention in a health care clinic setting (see Perceived Usability section).

Perceived Usability

Overview
The mean score for the Intervention Usability Scale was 52.5 (SD 14.5) for patients and 45.6 (SD 11.6) for health care providers, which fell below the average usability cutoff of 68. Usability conversations provided confirmation that the intervention needed to be revised and focused on improving the delivery of the intervention, the time demands of the intervention, and integration into health care settings. Textbox 1 presents a summary of recommended changes.

Textbox 1. Summary of recommended changes.

- Patients were interested in moving through the intervention content at their own pace, but they still prefer a professional to check-in with for questions.
- Patients preferred the intervention content in a digital format.
- Reduce the face-to-face time required by providers to increase the likelihood that the intervention is adopted in clinical settings.
- Make the tracking (food, sleep, and symptoms) required by the intervention easier to do.
- Create content summarizing patient progress through the intervention to facilitate communication with their provider when they meet.

Delivery of Intervention
The original CSM was designed for in-person or telephone delivery with a paper workbook. Both patients and health care providers discussed the importance of continuing to have a person, either a health care provider or a peer, involved in the intervention. Patients, in particular, discussed the delivery of intervention content in online formats such as apps and e-courses where they could review content independently with weekly check-ins. All but 1 patient desired weekly check-ins. Some patients preferred web-based check-ins to be with an expert in IBS, whereas others noted that it may be helpful to have a peer mentor because they do not know many people with IBS. Regardless of who delivers the intervention (ie, health care provider or peer), the most important characteristic was someone who was skilled and knowledgeable in IBS. Health care providers said the delivery method depended on the age of the population. Some health care providers discussed the potential benefits of using videos to present information and demonstrate skills. However, most discussions with health care providers focused on in-person intervention delivery.

Intervention Time Demands
Patients had a variety of opinions regarding the length of the 8-week CSM intervention and the daily time commitment for practicing skills. Patients who preferred a shorter intervention typically identified content that was not applicable or of interest to them. For some, 8 weeks seemed reasonable, whereas others identified that 8 weeks may not be enough time to obtain results. Health care providers identified that the CSM, as originally designed, required more time to deliver the content than the providers were able to fit into their current practices. This creates an implementation barrier to incorporating the therapy into clinics.

Integration Into Health Care Settings
Health care providers thought the easiest integration into health care settings was for the providers to introduce the intervention to the patient and have another health care professional (eg, nurse or social worker) who was an expert in IBS deliver the intervention. Health care providers desired feedback on the patients’ progression through the intervention, such as a summary of symptoms, strategies that worked, and an overview...
of patient engagement. Most providers preferred a brief end-of-intervention update. Patients also preferred to learn about the intervention from their primary care provider or gastroenterologist.

Discussion

Overview

Overall, individuals with IBS and comorbid anxiety or depression, as well as health care providers, found the content and format of the CSM intervention acceptable and appropriate; however, challenges were identified related to anticipated feasibility and usability. The qualitative findings expanded the quantitative findings for acceptability and appropriateness. For feasibility, patient qualitative findings expanded the quantitative findings, whereas for providers, the qualitative findings indicated barriers to feasibility while the quantitative findings indicated feasibility. Overall, participants reported the intervention was comprehensive and provided structure, accountability, and support. However, participants warned that engagement in the intervention would be influenced by time, motivation, literacy, culture, and cost, in addition to a variety of usability issues (improving the delivery of the intervention, time demands of the intervention, and integration into health care settings). Addressing the anticipated acceptability, feasibility, appropriateness, and usability of the CSM intervention has the potential to influence key barriers to implementation and uptake among those with IBS and comorbid anxiety or depression.

Acceptability

Previous research has noted that patients with medically unexplained symptoms and comorbid anxiety or depression may have less favorable cognitive behavioral therapy outcomes [31]. Thus, our approach of human-centered design methods to elicit feedback from patients with IBS and comorbid anxiety or depression may serve as 1 method to create interventions to address the unique needs of this population. Specifically, patients with IBS and comorbid anxiety or depression discussed the importance of structure and support in completing the intervention. They also mentioned the importance of integrating IBS and mental health as previous health care encounters had focused independently on either IBS or mental health but did not take a holistic approach to symptoms. Recent evidence has highlighted the benefits of integrated care approaches, which include a team comprising gastroenterologists, nurses, dietitians, psychiatrists, hypnotherapists, and behavioral therapists. Individuals with IBS who were randomized to an integrative care arm had greater reductions in global symptoms as well as IBS symptom severity compared to those in the standard care group [32].

Appropriateness

Patients and health care providers both identified the importance of support from a person throughout the intervention. It is unclear if the desire for a support person is unique to patients with comorbid IBS and mental health. For instance, a meta-analysis found that computer-assisted cognitive behavioral therapy for depression in primary care is effective if clinicians offer modest support (60-194 minutes) throughout the intervention (7-16 weeks) [33]. In this study, patients preferred to review the content and practice independently and have someone available to follow up with them. Additionally, health care providers indicated that in-person sessions would be preferred but acknowledged that, with the COVID-19 pandemic, telemedicine visits could also be useful. Although a review article highlighted the effectiveness of primary care provider–delivered self-management interventions, this study highlights the time limitations of clinicians in delivering such an intervention [34]. Even if primary care providers do not deliver the intervention, there is a need for integrated care approaches so that primary care providers can receive feedback on their patients progress through the intervention. Future research should examine innovative methods to integrate comprehensive interventions into primary care, gastroenterology, and other health care settings.

Feasibility

Patients indicated that the intervention may be best suited for newly diagnosed individuals to promote self-management earlier in the disease course. Thus, additional research is needed to understand if the intervention effects differ based on time since diagnosis or time since symptom onset. Health care providers had varied opinions regarding the feasibility of the intervention, specifically highlighting barriers to implementation. Yet few studies have focused on implementation strategies within the population of patients with IBS and anxiety or depression overlap. Implementation frameworks such as the Consolidated Framework for Implementation Research [35], the Practical, Robust Implementation, and Sustainability Model [36], and the Reach Effectiveness Adoption Implementation and Maintenance framework [37] can be used to guide such research.

Usability

Both patients and health care providers identified several ways to improve the usability of the intervention. Patients emphasized the benefits of accessing intervention content on the internet and being able to track and monitor symptoms. Health care providers identified age as a factor influencing intervention delivery methods, although this theme did not arise among the small sample of patients aged between 20 and 59 years. Previous research has indicated that older adults may avoid using technology due to fear of confirming negative stereotypes [38]. Thus, considerations should be made for understanding the specific technology needs of older adults with IBS and comorbid anxiety or depression and designing accessible systems that benefit all patients.

Patients had a variety of opinions regarding the length of the intervention. Patient differences such as disease severity, previous intervention experiences, or lifestyle may influence intervention length preferences. A previous study by Lackner et al [6] identified no statistically significant differences in the proportion of patients with IBS responding to a 10-week standard cognitive behavioral therapy (87.5%) compared to a brief 4-week cognitive behavioral therapy (80%, P>.55) [6]. Thus, there is a need to identify ways to tailor the intervention length based on the content participants are familiar with and their tolerance for intervention length.
Limitations

Strengths of the study include incorporating the perspectives of both patients and health care providers into the intervention evaluation. Using implementation science and a human-centered design approach provided an established framework to elicit feedback regarding the intervention. Yet, this study has several limitations. Patients and health care providers were recruited during the COVID-19 pandemic. Patients were recruited on the internet and therefore may have different characteristics than those typically obtained in primary care and gastroenterology settings, such as greater levels of computer literacy and comfort with technology-delivered interventions. Additionally, we did not have access to clinical records for patients. It is possible that selection bias may have occurred such that health care providers with a large number of patients with IBS were more likely to participate in the study and had different barriers than health care providers seeing fewer patients with IBS. Another limitation is that patients and providers did not complete the intervention but rather provided feedback on the overall content and intervention; therefore, additional implementation barriers may be found when using the intervention.

Conclusions

Patients and health care providers reported the CSM intervention was acceptable and appropriate but identified several potential feasibility and usability challenges. Thus, before applying the intervention among a population of individuals with IBS and comorbid anxiety or depression, there is a need to modify intervention delivery methods, consider the length of the intervention, and address the best methods of implementing in the clinic setting. Considerations should be made to improve the ease of tracking, allow participants to move through the intervention at their own pace, and provide a summary of patient progress through the intervention. Future work will assess whether addressing feasibility and usability leads to the anticipated improvements in implementation and adoption.

Acknowledgments

This study was supported by the National Institute of Mental Health (grant P50 MH115837) and the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1 TR002319. KK is funded, in part, by the National Institute of Nursing Research (grant K23NR020044). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study Measures Completed by Patients with Irritable Bowel Syndrome and Comorbid Anxiety and/or Depression and Healthcare Providers.
[DOCX File, 23 KB - formative_v8i1e43286_app1.docx ]

Multimedia Appendix 2
Individual items from the acceptability, appropriateness, and feasibility measures.
[DOCX File, 21 KB - formative_v8i1e43286_app2.docx ]

References


Abbreviations

CSM: comprehensive self-management
EBPI: evidence-based practice intervention
GAD-7: General Anxiety Disorder-7
IBS: irritable bowel syndrome
PHQ-9: Patient Health Questionnaire-9

Edited by A Mavragani; submitted 13.10.22; peer-reviewed by S Schür; comments to author 12.09.23; revised version received 28.09.23; accepted 13.12.23; published 31.01.24.

Please cite as:

©Kendra Kamp, Pei-Lin Yang, Emily Friedman, Alejandra Lopez, Sarah Iribarren, Pamela Barney, Sean Munson, Margaret Heitkemper, Rona Levy. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Investigating the Potential of a Conversational Agent (Phyllis) to Support Adolescent Health and Overcome Barriers to Physical Activity: Co-Design Study

Richard Moore¹, MSc‡; Abdel-Karim Al-Tamimi², PhD; Elizabeth Freeman³, PhD

¹Sheffield Hallam University, Sport and Physical Activity Research Centre / Advanced Wellbeing Research Centre, Sheffield, United Kingdom
²Department of Computing, Sheffield Hallam University, Sheffield, United Kingdom
³Department of Psychology, Sociology & Politics, Sheffield Hallam University, Sheffield, United Kingdom
‡Conversational AI Research Cluster

Corresponding Author:
Richard Moore, MSc
Sheffield Hallam University
Sport and Physical Activity Research Centre / Advanced Wellbeing Research Centre
Sheffield Hallam University, Olympic Legacy Park, 2 Old Hall Road
Advanced Wellbeing Research Centre
Sheffield, S9 3TU
United Kingdom
Phone: 44 7751234185
Email: r.moore@shu.ac.uk

Abstract

Background: Conversational agents (CAs) are a promising solution to support people in improving physical activity (PA) behaviors. However, there is a lack of CAs targeted at adolescents that aim to provide support to overcome barriers to PA. This study reports the results of the co-design, development, and evaluation of a prototype CA called “Phyllis” to support adolescents in overcoming barriers to PA with the aim of improving PA behaviors. The study presents one of the first theory-driven CAs that use existing research, a theoretical framework, and a behavior change model.

Objective: The aim of the study is to use a mixed methods approach to investigate the potential of a CA to support adolescents in overcoming barriers to PA and enhance their confidence and motivation to engage in PA.

Methods: The methodology involved co-designing with 8 adolescents to create a relational and persuasive CA with a suitable persona and dialogue. The CA was evaluated to determine its acceptability, usability, and effectiveness, with 46 adolescents participating in the study via a web-based survey.

Results: The co-design participants were students aged 11 to 13 years, with a sex distribution of 56% (5/9) female and 44% (4/9) male, representing diverse ethnic backgrounds. Participants reported 37 specific barriers to PA, and the most common barriers included a “lack of confidence,” “fear of failure,” and a “lack of motivation.” The CA’s persona, named “Phyllis,” was co-designed with input from the students, reflecting their preferences for a friendly, understanding, and intelligent personality. Users engaged in 61 conversations with Phyllis and reported a positive user experience, and 73% of them expressed a definite intention to use the fully functional CA in the future, with a net promoter score indicating a high likelihood of recommendation. Phyllis also performed well, being able to recognize a range of different barriers to PA. The CA’s persuasive capacity was evaluated in modules focusing on confidence and motivation, with a significant increase in students’ agreement in feeling confident and motivated to engage in PA after interacting with Phyllis. Adolescents also expect to have a personalized experience and be able to personalize all aspects of the CA.

Conclusions: The results showed high acceptability and a positive user experience, indicating the CA’s potential. Promising outcomes were observed, with increasing confidence and motivation for PA. Further research and development are needed to create further interventions to address other barriers to PA and assess long-term behavior change. Addressing concerns regarding bias and privacy is crucial for achieving acceptability in the future. The CA’s potential extends to health care systems and multimodal support, providing valuable insights for designing digital health interventions including tackling global inactivity issues among adolescents.
Introduction

Background

There is indisputable evidence supporting the positive effects of regular physical activity (PA) on the physical and emotional well-being of adolescents [1], enhanced cognitive function [2], and improved educational attainment [3,4]. The UK government recommends that children and young people engage in 60 minutes of PA per day, with 30 minutes taking place at school and an additional 30 minutes outside of school hours. However, despite this guideline, only 57% of children participate in ≥30 minutes of PA outside of school, compared with 40% during school hours [5].

Inactivity is a global challenge, with a considerable number of adolescents worldwide failing to engage in sufficient PA [6]. It is crucial that adolescents maintain a consistent level of PA during adolescence as this may lead to healthy habits that extend into adulthood, lowering the risk of inactivity in later years and reducing the possibility of developing hypokinetic conditions [7-9]. There is a consensus that evidence-based approaches involving multicomponent interventions targeted at inactive adolescents are needed on a global scale [10-12].

Although schools provide equal opportunities for adolescents to participate in PA, making them valuable settings to focus resources [12], multiple barriers exist that prevent schools from allocating adequate time and resources to PA programs [13,14]. In addition, various barriers relating to physical, psychological, social, and environmental factors hinder PA participation in the adolescent population [15-17]. It is generally accepted that limited progress has been made in supporting adolescents to overcome these barriers, necessitating the development of innovative interventions [12]. This rationale underlies the launch of the Global Action Plan on PA in 2018 by the World Health Organization, aimed at promoting and supporting adolescents in achieving the recommended levels of PA [18]. As part of this plan, digital approaches were advocated, emphasizing that interventions should prioritize understanding and addressing barriers that prevent adolescents’ engagement in PA. By doing so, interventions can be effectively designed to facilitate sustainable changes in the PA habits and behaviors of adolescents [15-17].

Digital health interventions have emerged as a promising approach to promoting PA and overall well-being through digitally delivered support, providing PA education and motivational guidance [19]. However, the effectiveness of these interventions in promoting PA has yielded mixed results [20,21]. A systematic review focusing on digital interventions for PA among young people revealed positive changes in PA levels and attitudes toward PA [22]. These positive changes were attributed to numerous factors such as web-based education, goal setting, self-monitoring, parental involvement, and gamification and personalization [22]. In particular, personalization has emerged as a significant feature in engaging adolescents through digital health interventions [23]. Tailoring interventions to individual needs and preferences is crucial for promoting PA and overcoming barriers to participation. There is also some evidence of the positive engagement of young people in formal (eg, school) and informal (eg, home) settings to develop knowledge, skills, and behaviors related to PA [22]. This shows that locations, such as school and home, hold potential as promising venues for implementing such interventions.

Conversational agents (CAs) have emerged as promising tools for promoting PA, especially among adolescents, using popular digital platforms such as social media, web interfaces, and mobile apps [24-26]. These agents are accessible at any time of the day, are nonjudgmental, and can be accessed on an array of digital devices. CAs are the preferred choice for individuals seeking immediate web-based support and for those who are hesitant about seeking in-person assistance.

Systematic reviews have examined CA effectiveness in promoting PA, healthy eating, and weight loss [27,28]. Nevertheless, despite encouraging results, the efficacy of these interventions remains inconclusive owing to constraints in outcome measurement and reporting [28]. There is also a dearth of evidence regarding how such CAs have been used to help adolescents overcome barriers to PA, and there is a need for new conversational artificial intelligence (AI) approaches that draw insights from existing theories to ensure they are evidence based and effective [17,29].

To optimize CAs and maximize user engagement, previous studies have emphasized the importance of effective persona design and the integration of principles from user-centered design [30] and human-computer interaction [28,31]. However, previous research focusing on usability and feasibility has reported moderate results, highlighting challenges such as repetitive content, high attrition rates, technical difficulties, and concerns regarding safety and privacy [27]. Although rule-based approaches are commonly used, more successful outcomes have been achieved with unconstrained methods that allow for natural language input and personalized interaction. These methods offer opportunities for establishing relational and persuasive capacities [28] with adolescents, which are particularly important when delivering person-centered behavior change interventions.

Recent advancements in large language models (LLMs) as well as the advent of generative AI show promise in enhancing CAs by improving user engagement and satisfaction, potentially surpassing human performance [32]. Their ability to understand and generate natural language with complexity and accuracy allows for sophisticated conversations, including emotion detection, contextual understanding, and personalized responses [33,34]. LLMs offer automatic content generation, increasing scalability and cost-effectiveness [33,34]. All of these features...
could potentially enhance CA quality and improve adherence in the future; however, concerns about bias, misinformation, privacy, and security need to be addressed before using them safely with adolescents [33,34]. In any health-related context, it is of paramount importance that the information delivered is not only accurate and trustworthy but also firmly rooted in evidence-based practices, curated in collaboration with domain experts. Consequently, as consideration focuses on the future applications of LLMs and generative AI in this sphere, it becomes imperative to place our faith in models that are not only trustworthy but also transparent, providing a clear and comprehensive audit trail for their decision-making processes. This ensures that the highest standards of quality and reliability are maintained, thereby safeguarding the integrity of health-related information and services.

**Summary**

This study introduces the co-design and evaluation of a pilot CA called “Phyllis” to evaluate its proof of concept. Its objective is to assist adolescents in overcoming 2 barriers to PA (ie, confidence and motivation) and offer an alternative digital solution to support students and promote PA. Building upon the findings of a previous study, which identified 52 barriers to adolescents’ PA participation, appropriate intervention functions and behavior change tools were identified to provide support for each barrier [17]. A theoretical framework designed for this study and the existing behavior change model also inform the CA [29] and both aim to aid the development of CAs in promoting PA.

The hypothesis for the study is that a CA can support adolescents to overcome barriers to PA and be perceived by adolescents as being a tool to help them increase their confidence and motivation to participate in PA. The objectives of the study are as follows:

- **Co-design a CA in collaboration with adolescents,** incorporating the model by Zhang et al [29] and a theoretical framework designed for this study to assess its proof of concept.
- **Demonstrate the CA’s ability to understand user input related to one of the 52 barriers to PA identified in the previous study [17].**
- **Evaluate the usability and acceptability of the CA among adolescents.**
- **Assess the perceived effectiveness of the solutions to barriers provided by the CA.**

**Methods**

**Overview**

In this section, we present a comprehensive and systematic methodology used for the co-design and evaluation of the CA. The methodology includes 3 key phases: phase 1 focuses on co-designing the CA with adolescents to ensure the agent’s relevance and user-centeredness. In phase 2, the development of the CA and its dialogue is presented, incorporating persuasive and relational capacity elements. To facilitate intelligent and contextually relevant interactions, a natural language understanding (NLU) model and knowledge model are integrated into the cognitive system. Finally, in phase 3, we detail the mixed methods evaluation used to comprehensively assess the CA’s effectiveness and user experience with adolescents.

**Co-Design of the CA**

**Phase 1: Understanding User Background and Designing CA Characteristics and Persona**

To participate in the co-design process, all students aged 11 to 13 years were required to apply through their school and meet the following selection criteria:

- Students perceived that they did not currently meet the UK government recommendations for participating in PA for 60 minutes per day over the course of a week.
- They were interested in participating in PA more often.
- In their application, they stated barriers to participation that had prevented them from being more physically active.

All interested students were given an information sheet and provided written consent to participate. The following is an example of an application from a student who participated in the co-design, highlighting the importance that students felt in participating in the research:

_Sir, I think I would be great for this role because I have always wanted to do sports and be more fit, but I have always stopped myself because I thought I wouldn’t be good enough. I would embarrass myself, just fail altogether or I would be too scared and think that it was too hard. I think that this opportunity would bring out the best in me and give me another chance to build a better me. I also stop myself from the joy of joining in because I have always thought people would judge me and I would always get scared that I wouldn’t be perfect. Now you might be thinking why should choose this girl she has only listed things that she is bad at and why but one thing that Is that I love improving myself and with this once in a lifetime opportunity I could make a fresh start.

The primary objective of phase 1 of the study was to gain insights into the user background of the co-design participants, including sociodemographic characteristics (eg, age and sex), behavior determinants (eg, attitude toward PA), and behavior habits (eg, PA and CA use) [29].

To accomplish this, a 2-hour workshop was conducted at Wickersley School and Sports College, located in Rotherham, South Yorkshire, where the students were based. During the workshop, the following co-design activities were conducted with the students, following the principles of user-centered design [30,35]:

- Icebreaker activity: this involved a demonstration of a robot and an introduction by the research team.
- A 30-minute focus group discussion with the students, audio recorded, thematically analyzed [36], and covering the following topics:
  - Participation in PA, barriers faced, and overall experiences related to PA
The second part of phase 1 involved the completion of a co-design workbook by the participating students. This workbook was specifically designed for the project and served to gather both quantitative and qualitative data on user background and design preferences [29]. It aimed to capture information regarding various aspects of the CA’s design, including the dialogue system infrastructure (eg, buttons or open text input), media content delivered through conversation (eg, videos, documents, and text), and anthropomorphic cues (eg, identity, name, and sex) of the CA. In addition, it provided content related to building relational capacity specific to this type of CA, such as social dialogue (eg, greetings and small talk) and self-disclosure (eg, discussing the CA’s development). The following presents a summary of the data collected during this stage of the co-design process:

- Students’ age, sex, and levels of PA
- Quantifying the barriers to PA that students faced
- Detailing their emotions toward PA using the wheel of emotions by Plutchik [37]
- Using Leary’s [38] Interpersonal Complex to identify the CA’s personality along 2 dimensions: dominance (horizontal) and agreeableness (vertical)
- Identifying content that students would prefer to access via the CA (ie, conversational, videos, weblinks to other content, documents, and images)
- Identifying the preferred type of persona and key characteristics that the CA should possess

**Phase 2: Developing Relational and Persuasive Conversational Capacity**

**Persuasive Capacity**

To enhance the CA’s understanding of adolescents’ habits and behaviors and facilitate effective intervention design, it is crucial for all PA interventions to be grounded in existing evidence and theory. This phase of the study draws upon a theoretical framework that details a theory-based approach to designing CAs to support adolescents in overcoming barriers to PA. The framework used in this study is underpinned by the capability, opportunity, motivation, and behavior (COM-B) model and the Theoretical Domains Framework (TDF). Identified barriers were coded using the COM-B [39] model and TDF [40] to comprehend both the origins of adolescent behaviors and the factors influencing them. Furthermore, this approach ensured the appropriate selection of evidence-based intervention functions, policy factors, and behavior change tools for each behavior using the Behavior Change Technique Taxonomy [41]. This process involved choosing behavior change tools that could be effectively delivered conversationally through a CA [17]. The theoretical framework played a pivotal role in informing the cognitive system, including the knowledge model incorporating the study findings [17] and the conversational engine (ie, the NLU model) to accurately understand and respond to barriers input by adolescents.

The next step in the process was to evaluate the feasibility and usability of the approach and evaluate the changes resulting from the use of the CA. For this purpose, 2 CA modules were developed by the research team, specifically targeting the 2 primary barriers identified by the students: a lack of “confidence” and “motivation” to be physically active. The confidence module was designed based on the theory of planned behavior [42,43] and incorporated 3 evidence-based behavior change tools identified from a previous study [17], namely problem-solving, verbal persuasion about capability, and self-talk. These tools were developed based on 4 guiding principles:

- Setting goals: it is important to always have a goal and push oneself out of the comfort zone.
- Diverse goals: building confidence involves setting different types of goals to enhance overall confidence.
- Broad perspective: individuals with higher levels of self-efficacy can look at the bigger picture and go beyond short-term setbacks.
- Reframing setbacks: developing self-efficacy is a gradual process that may take years to reframe one’s mindset.

During the conversation, adolescents were prompted to answer the following questions. They were then given the opportunity to review their answers on the screen and receive a copy of their responses via email. This provided students with the opportunity to revisit and follow up on the actions they had stated.

- How confident are you when it comes to PA. Can you rate your answer from 1-10 with 1 being “very low,” 5 being “okay,” and 10 being “very confident.”
- When you feel confident where are you?
- Do you feel more confident being active with others or by yourself?
- And what motivates you more—being active with others or by yourself?
- When you feel most comfortable—what activity are you doing?
- If you think that activity is exercise related (eg, skipping) how could you do it more within your week? If that activity is not exercise related (eg, drawing) how could you add an element of exercise to it? (eg, standing whilst drawing, jumping on the spot every 10 min whilst drawing, etc).
- Based on what you know now, how can you bring activity into your weekly routine more?

The motivation module was developed using 2 behavior change tools: instructions on how to perform the desired behavior and behavior rehearsal or practice. Students were given a series of information on how to be more physically active and how to build small sessions (ie, 15 min) of PA into their day. These activities could be undertaken by any able-bodied person at no cost. Further information was given about how activity could...
be achieved despite educational pressure and time constraints and how PA can improve attainment. In addition, links to Sport England websites were provided to give young people further information on how to increase their PA levels in various environments, such as at home or outdoors. Useful information was provided on why PA is important and how often we should aim to engage in it.

**Relational Capacity**

To foster relational capacity, it is essential to design conversations that are evidence based, are expert led, and incorporate user feedback. However, there is a lack of empirical evidence regarding effective conversation design for CAs. To ensure that the CA is engaging and effective while adhering to the principles of conversation design, module design followed the happy and detailed conversation design process provided by the Conversation Design Institute [44]. This involved the inclusion of appropriate greetings, small talk, and module messaging that incorporated the elements of social dialogue, empathy, and humor. Sample dialogues were evaluated, and iterative refinements were performed using this process. Theoretical principles [17,29], along with the concepts of user-centered design [30] and human-computer interaction [31], were also considered.

To enable effective natural language interaction between humans and the CA, the researchers developed an NLU model [45]. They assigned at least one intent to each identified barrier from a previous study [17] to train the NLU model. For each intent, a minimum of 15 utterances were manually generated by the researchers and uploaded to Google Dialogflow’s NLU engine to train the data [46]. The training process involved providing the system with a diverse range of conversational data, allowing it to learn and comprehend various patterns, intents, and entities relevant to the specific domain. Through iterative training and refinement, the Dialogflow NLU model was optimized to accurately understand and interpret user input, forming a solid foundation for the subsequent phases of CA development and evaluation [47].

The NLU model represents a hybrid approach, falling between the spectrum of fully open-ended models, such as ChatGPT, and strictly closed branching–dialogue systems. Instead of strictly adhering to predetermined dialogue paths or being entirely unrestricted, the NLU model uses a nuanced methodology. It focuses on the interactions between the CA and students through several natural language processing techniques. Specifically, the NLU model is adept at comprehending and interpreting user inputs in natural language. It identifies the intent behind user queries and recognizes key entities mentioned in the text. Unlike rigid closed systems, the model does not follow a fixed script. Instead, it uses this understanding of user intent and entities to dynamically route the conversation. Although predefined dialogues are in place to guide the conversation, these paths are adaptable based on the context of the conversation and the user’s input. This flexibility ensures a more engaging and personalized interaction while still maintaining a degree of structure within the conversation flow.

**Figure 1** provides an example of an interaction between a student and the CA.

![Figure 1. An example of an interaction between a student and the conversational agent.](image-url)
To assess the accuracy of the model, the researchers used MindBehind, a software program that enabled the research team to input statements pertaining to each barrier and evaluate and refine the utterances used during the NLU model training. The performance of the CA was evaluated using an $F_1$-score, yielding a score of 80% [48]. MindBehind was then used to create a knowledge model and use its conversational engine, leveraging the NLU model and logic functions to route conversations based on the input of barriers and corresponding solutions. Following this, the researchers visually represented the conversational dialogue of the CA (ie, barriers and solutions) on a canvas and tested its accuracy to ensure its readiness to be deployed for testing.

**Phase 3: Evaluating Mechanisms and Outcomes**

During the final phase of the study, the CA was distributed to the students for testing via a test link. The students interacted with the CA through text messaging. The platform hosting the CA enabled the students to engage with it through various internet-enabled devices such as laptops, tablets, and smartphones. In addition, the students had the option to select appropriate responses from a predefined list of choices, which included options such as requesting help, ending the conversation, or seeking more support. This setup provided a user-friendly and versatile interface, accommodating different communication preferences and ensuring a seamless interaction experience for the participants. Initially, the adolescents were asked to engage in a conversation with the CA as users and to identify any barriers they faced that led them to seek assistance from the CA. Subsequently, they were requested to test the CA by inputting various barriers and assessing how the conversation was routed. Following this, participants were invited by each organization using convenience sampling to evaluate the CA from Wickersley School and Sports College (7 participants) and a community enterprise called Zest (31 participants), totaling 46 adolescents.

To evaluate the CA's acceptability, usability, and perceived outcomes of using the modules, the users were invited to participate in a web-based survey using Key Survey, a specialized web-based survey platform. The survey also sought to gauge their perception of the CA's effectiveness and user experience, encompassing 7 recommended themes for evaluating CAs, as outlined in a previous study [29]. Users were specifically encouraged to share both positive and negative aspects of their experience, offer suggestions for improvement, and report any observed malfunctions during their engagement. In addition, the survey included an adapted version of the AttrakDiff questionnaire, tailored specifically for young individuals, to measure usability [49].

**Analysis**

During phases 1 and 2, an analysis was conducted on both quantitative and qualitative data gathered from the participants. For the quantitative analysis, basic frequencies were used to examine data related to students’ age, sex, and levels of PA in phase 1 and to evaluate the user experience and efficacy of the CA in phase 2. This analysis provided an overview of the demographic characteristics and the distribution of PA levels among the participants, as well as an objective assessment of their experience and the outcomes of using the CA.

The qualitative data were subjected to thematic analysis [36] to identify and analyze recurring themes and patterns across several key aspects. These aspects included the persona and key characteristics that the CA should possess in phase 1, as well as user feedback on how the CA could be improved in phase 2.

**Ethical Considerations**

This study adheres to the established human subject research ethics guidelines and was granted institutional ethics approval (ER37229351) by the Sheffield Hallam University Research Ethics Committee.

Informed consent was obtained from participants’ legal guardians following a comprehensive explanation of the study’s purpose, procedures, risks, benefits, and the voluntary nature of their participation. The participants were explicitly informed that they had the right to withdraw from the study at any time without repercussions.

The privacy and confidentiality of all human subjects were rigorously safeguarded throughout the study. All the data collected were either anonymized or deidentified to protect the identities of the participants. The data were stored securely, and the data used in this study were not linked to any external databases or sources that could compromise the privacy of the participants. The participants were not offered an incentive to participate in the study.

**Results**

**Co-Design**

The following section presents the results of the co-design with adolescents and the evaluation of the performance of the CA and its efficacy.

**User Background**

The co-design of the CA involved 9 students aged between 11 and 13 years, in school years 7 to 9. The sex distribution among the students was 56% (5/9) female and 44% (4/9) male. In terms of ethnic background, 6 (67%) of the 9 students identified as White, 2 (22%) identified as Asian or Asian British, and 1 (11%) identified as Black.

All 9 students met the inclusion criteria and expressed their interest in exploring the use of a CA to support themselves and their peers to help overcome barriers to PA. Initially, the students were provided with a list of 52 barriers to PA and were asked to identify the specific barriers that they experienced. They highlighted 37 barriers, which accounted for 66% of the total barriers listed in the previous study [17].

Among the barriers identified, the most commonly highlighted included a “lack of confidence” (mentioned by 8 students), “fear of failure” (mentioned by 7 students), and a “lack of motivation” (mentioned by 5 students). When the adolescents were asked to elaborate on the barriers in their own words, they provided more specific insights, particularly concerning psychological barriers. For instance, 2 adolescents reported experiencing “anxiety” both before and during PA, especially at competitive

https://formative.jmir.org/2024/1/e51571

JMIR Form Res 2024 | vol. 8 | e51571 | p.620

(page number not for citation purposes)
expressed a preference for a gender-neutral CA, whereas only 1 (11%) person advocated for a gender-specific CA. The remaining students (3/9, 33%) suggested that the CA’s gender should be customizable, allowing users to select their preferred option. In terms of appearance, most students felt that if the CA had an avatar, it should be customizable, with suggestions ranging from human-like features to animal-like or robot-like attributes.

Regarding the tone of the CA, approximately two-thirds of the group agreed that it should have a human-like and engaging tone, regardless of whether it was in the text or voice form. Of the 9 students, 2 (22%) students even proposed that the CA could mimic the tone of a coach or celebrity. Conversely, 1 (11%) person expressed a preference for the CA to have a robotic voice. However, 3 (33%) individuals indicated a preference for customizable voices, although it remained unclear whether these voices should be exclusively human-like.

During the design phase, the concept of customization and personalization emerged as a prominent theme, reflecting the students’ desire to have control over various aspects of the CA to meet their individual needs. This preference for customization poses several design challenges, as it necessitates the creation of multiple personas and the development of appropriate messages for each persona. However, it is a crucial consideration for the future development of CAs, as it allows for genuine personalization and enhances the overall user experience.

As part of the design process, students were also invited to suggest names for the CA. The name suggestions varied, highlighting different preferences and creative ideas. Some students proposed generic brand names with a playful twist related to PA, such as “Sporta” or “Phylo.” Others suggested human names such as Robert or Patricia, whereas some students preferred names that explicitly incorporated the terms “CA” or “robot,” such as “Charlie the CA” or “Sporty Bot.”

For the prototype, only 1 persona was created, and it was named “Phyllis.” This name remained unchanged throughout the final testing phase, as some students expressed a liking for the name, whereas others did not find the other suggested names suitable. The decision to choose Phyllis as the chosen name for the CA reflects the consideration of participant preferences and the desire for consistency during the testing phase.

During the co-design process, the students were asked to indicate their preferred personality characteristics for the CA using Leary’s Interpersonal Complex (Figure 2), which measures personality along 2 dimensions: dominance (horizontal) and agreeableness (vertical). As shown in Figure 2, the results revealed that students sought a balanced combination of dominance and agreeableness in the CA’s personality. They desired a friendly CA that exhibited both submissive and dominant behaviors and even opposing viewpoints at certain points in the conversation, potentially to challenge the user’s beliefs. These findings align with the broader characteristics that the students identified during the co-design process, emphasizing their need for personalization.

**Relational and Persuasive Capacity**

During the co-design process, the students were actively involved in expressing their preferences for the persona and characteristics of the CA. All co-design group members unanimously agreed that the CA should fall within the age range of 11 to 18 years, with the majority believing that a teenage persona would be most relevant and engaging for the target audience. When it came to gender, of the 9 students, 4 (44%) expressed a preference for a gender-neutral CA, whereas only 1 (11%) person advocated for a gender-specific CA. The remaining students (3/9, 33%) suggested that the CA’s gender should be customizable, allowing users to select their preferred option. In terms of appearance, most students felt that if the CA had an avatar, it should be customizable, with suggestions ranging from human-like features to animal-like or robot-like attributes.

Regarding the tone of the CA, approximately two-thirds of the group agreed that it should have a human-like and engaging tone, regardless of whether it was in the text or voice form. Of the 9 students, 2 (22%) students even proposed that the CA could mimic the tone of a coach or celebrity. Conversely, 1 (11%) person expressed a preference for the CA to have a robotic voice. However, 3 (33%) individuals indicated a preference for customizable voices, although it remained unclear whether these voices should be exclusively human-like.

During the design phase, the concept of customization and personalization emerged as a prominent theme, reflecting the students’ desire to have control over various aspects of the CA to meet their individual needs. This preference for customization poses several design challenges, as it necessitates the creation of multiple personas and the development of appropriate messages for each persona. However, it is a crucial consideration for the future development of CAs, as it allows for genuine personalization and enhances the overall user experience.

As part of the design process, students were also invited to suggest names for the CA. The name suggestions varied, highlighting different preferences and creative ideas. Some students proposed generic brand names with a playful twist related to PA, such as “Sporta” or “Phylo.” Others suggested human names such as Robert or Patricia, whereas some students preferred names that explicitly incorporated the terms “CA” or “robot,” such as “Charlie the CA” or “Sporty Bot.”

For the prototype, only 1 persona was created, and it was named “Phyllis.” This name remained unchanged throughout the final testing phase, as some students expressed a liking for the name, whereas others did not find the other suggested names suitable. The decision to choose Phyllis as the chosen name for the CA reflects the consideration of participant preferences and the desire for consistency during the testing phase.

During the co-design process, the students were asked to indicate their preferred personality characteristics for the CA using Leary’s Interpersonal Complex (Figure 2), which measures personality along 2 dimensions: dominance (horizontal) and agreeableness (vertical). As shown in Figure 2, the results revealed that students sought a balanced combination of dominance and agreeableness in the CA’s personality. They desired a friendly CA that exhibited both submissive and dominant behaviors and even opposing viewpoints at certain points in the conversation, potentially to challenge the user’s beliefs. These findings align with the broader characteristics that the students identified during the co-design process, emphasizing their need for personalization.
In total, the students selected 30 distinct characteristics that they believed the CA should possess, underscoring their desire for a personalized and tailored experience. Among the co-design participants, only 3 characteristics were mentioned more than once: understanding (mentioned 2 times), intelligence (mentioned 3 times), and kindness (mentioned 3 times). These repeated characteristics indicate the significance of empathy, intelligence, and kindness in the preferred personality traits of the CA.

Figure 2 illustrates students’ preferred personality characteristics using Leary’s Interpersonal Complex. The graph measures personality along 2 dimensions: dominance (horizontal) and agreeableness (vertical).

Social Dialogue and Self-Disclosure

Students were asked to provide dialogue examples that represented the personas they had chosen for the CA. All personas were of peer age for the students. The dialogue examples covered various scenarios, including greetings, providing assistance, encouraging positive actions, offering reassurance, giving praise, acknowledging mistakes, and saying goodbye.

The welcome messages displayed personalization, often addressing users by their first names and minimizing small talk. The style of the messages was similar across the personas. The “help” messages were typically direct and straightforward, presenting suggestions (eg, “I have a suggestion”). The “take positive action” messages were motivational, encouraging users to take specific actions (eg, “You can do it! How about you try this?”) Reassuring messages often began with an “okay” to acknowledge the user’s concerns before offering support. “Praise” messages consistently started with “well done” and then reiterated the user’s accomplishments. The “goodbye” messages usually included a supportive or hopeful closing message to conclude the conversation.

This approach to social dialogue and self-disclosure, characterized by personalization, motivational language, acknowledgment of concerns, and positive reinforcement, was preferred by the co-design participants. Their preferences were considered and integrated into the dialogue design of the CA persona.

Content

As part of the co-design process, students expressed their preferences regarding the content they wanted to be delivered through the CA, focusing on the barriers to PA they had previously identified. Most students preferred content that would help boost their confidence, overcome insecurities, address problems, and manage negative emotions related to PA. A student (1/9, 11%) suggested that the CA should provide advice on supporting others, whereas another (1/9, 11%) preferred a more proactive approach in which the CA would listen, understand their problems, and provide relevant answers.

Different preferences emerged regarding the approach the CA should adopt to engaging with adolescents and delivering content. A student (1/9, 11%) preferred a guiding approach, allowing the adolescents to find their own solutions while also offering assistance if they struggled with that approach. Practical solutions were also highlighted, such as suggestions for various sports activities, strategies to stay physically active, ways to improve performance in sports, and tips for eating healthier. Furthermore, the students recommended including motivating quotes and information on accessing support through other organizations or websites.

Media

The students were provided with diverse options to choose from regarding the type of media they would like to use during the CA conversation. According to their preferences, the most favored options were pictures (7/9, 78%) and videos (7/9, 78%), followed by audio (6/9, 67%) and links to websites. Conversely, the least popular options were links to apps, social media (2/9, 22%) and documents (1/9, 11%). The CA being piloted incorporates pictures, animated pictures (GIFs), and links to websites and apps, whereas videos were not included owing to...
the lack of appropriate content for adolescents, corresponding to the barriers identified. In the future, filming bespoke videos would be necessary to provide engaging content to adolescents via the CA.

**Dialogue Infrastructure**

Owing to the students’ age and unwillingness to disturb their schooling, the researchers were unable to spend extended periods with them and had to rely on the students’ prior experience with CAs. In the future, it is recommended that more time be spent with the students to fully understand all the key features of CA design and improve the co-design process.

**Implementation of Co-Design Phase Recommendations**

In this section, we summarize the implementation of the findings from the co-design process in the development of the prototype CA.

Textbox 1 presents the summarized findings from the co-design research with 8 adolescents to develop the prototype chatbot. It details the findings of user requirements and context, the content of the chatbot, and the chatbot’s persona.
Textbox 1. Summary of co-design findings.

**Conversational agent (CA)**
- **Name:** Phyllis
- **Audience:** Students aged 11-16
- **Channel:** A school website or social media

**Use case**
- **Service and expertise**
  - Identify barrier to physical activity (PA) from user input.
  - Provide solution to barriers including recommending activities they could participate in.
  - Prototype provides 2 modules on motivation and confidence.
  - Future version will have solutions to all 52 barriers as well as monitoring of PA.

- **Purpose**
  - Support for students in schools to overcome barriers to PA and to understand more about PA and gain information and guidance of how to be physically active.

- **Media**
  - CA mainly conversational but provides links to other website apps and resources.
  - Future iterations will include voice and video media.

**User requirements and context**
- **User persona**
  - Students aged 11-16 that want to be more active or support others to be active

- **Motivations**
  - Want to be more physically active but aware that there are barriers which prevent them from participating in PA and they want to overcome these

- **Anxieties**
  - That the CA can understand them
  - That the CA can help them
  - Data protection

**Content**
- **Behavioural modules**
  - Improving confidence (5 min)
  - Improving fitness (4 min): General fitness
    - Balance with educational constraints
    - Links to Sport England content
    - Fitness tracking

- **General content**
  - Why it is important to be physically active
  - How long should adolescents be active for
  - Respond to input around 52 barriers to PA. Currently acknowledges and asks to confirm barrier before saying content is not yet available.

**Bot persona**
- **Gender and age**
Age: 16
Gender: Neutral
Future designs may include a choice of personas

Backstory
An understanding intelligent and kind young person with knowledge and experience of PA

Role and style
Helpful peer who is supportive to other students of a similar age

Personality
Humanoid
Helpful
Challenging
Understanding
Intelligent
Kind

Can do
Have a conversation and identify barriers to PA
Suggest solutions to barrier and ways to be more active
Provide general information about PA

Can’t do
Have a wider conversation than the stated use case
Advise on how to help others
Monitor PA

Standard vocabulary
Typical things to say
Here are some ideas.
I have a suggestion.

Introductions
Hi, my name is Phyllis how are you?
Hey, my name is Phyllis.
See you soon.

Acknowledgement
Sure, I will do that.
Okay, I have a suggestion.

Confirmations
Here you go ‘name’.
You can do it! How about you try this.

Apologies
I am sorry I cannot help you with that.
Okay, I didn’t understand that. Can you try and tell me what stops you from being more physically active?
Following the co-design process, the CA was constructed based on the gathered findings and was internally evaluated with the research team. Subsequently, the CA was prepared for deployment to be tested with adolescents.

**Evaluation**

This section presents the findings of evaluating the prototype CA (referred to as “Phyllis”) with 46 students.

**Use Patterns**

During the evaluation phase, 62 conversations were conducted with the CA. Among these conversations, 15 students concluded one conversation and then initiated a new conversation with Phyllis. Overall, the users sent 1512 messages to Phyllis, resulting in an average of 33 messages per person. Phyllis responded by sending 2806 messages to the users, averaging 61 messages per conversation. Only a small portion (n=77, 5.09%) of the messages sent per person could not be successfully matched with a corresponding intent. This was because the students sent unintelligible messages or misspelled words while discussing barriers to PA participation. The most frequently matched intents included “lack of confidence,” “lack of time,” and “lack of motivation,” which aligns with the barriers identified during the co-design focus groups.

**User Experiences**

The users generally had positive experiences with the CA, rating its conversational capacity as good, clear, simple, and straightforward to use on a comparative scale (Figure 3). Qualitative feedback indicated that the users found the CA clear and easy to use, with good options for obtaining information and interesting websites to access from the CA.

Concerning acceptability, a sizable portion of the students (34/46, 73%) expressed their definite intention to use a fully functional CA in the future. The remaining students stated that they might consider using it in the future. When asked to rate Phyllis, 65% (30/46) of respondents considered it “good,” whereas 20% (9/46) rated it as “very good.” The remaining 15% (7/46) rated Phyllis as “average.” The CA received a positive net promoter score of 13, indicating a favorable likelihood of recommendation. This suggests that not only would most students use Phyllis themselves but also they would recommend it to their friends.

On the basis of user feedback, it was suggested that Phyllis should reduce the amount of text in each message to enhance readability and recommended slowing down the speed of message delivery to assist with user recognition and comprehension. Furthermore, 2 users specifically requested that the CA offer direct answers to various questions about fitness and exercise without necessitating the initial mention of barriers. They also suggested incorporating more videos, diagrams, and real-life scenarios depicting PA. As anticipated, users expressed an interest in customizing the CA further by having the option to choose a different avatar, name, and personality.

**Figure 3.** User experiences and relational capacity.

Concerning acceptability, a sizable portion of the students (34/46, 73%) expressed their definite intention to use a fully functional CA in the future. The remaining students stated that they might consider using it in the future. When asked to rate Phyllis, 65% (30/46) of respondents considered it “good,” whereas 20% (9/46) rated it as “very good.” The remaining 15% (7/46) rated Phyllis as “average.” The CA received a positive net promoter score of 13, indicating a favorable likelihood of recommendation. This suggests that not only would most students use Phyllis themselves but also they would recommend it to their friends.

On the basis of user feedback, it was suggested that Phyllis should reduce the amount of text in each message to enhance readability and recommended slowing down the speed of message delivery to assist with user recognition and comprehension. Furthermore, 2 users specifically requested that the CA offer direct answers to various questions about fitness and exercise without necessitating the initial mention of barriers. They also suggested incorporating more videos, diagrams, and real-life scenarios depicting PA. As anticipated, users expressed an interest in customizing the CA further by having the option to choose a different avatar, name, and personality.

**Figure 3.** User experiences and relational capacity.

Concerning acceptability, a sizable portion of the students (34/46, 73%) expressed their definite intention to use a fully functional CA in the future. The remaining students stated that they might consider using it in the future. When asked to rate Phyllis, 65% (30/46) of respondents considered it “good,” whereas 20% (9/46) rated it as “very good.” The remaining 15% (7/46) rated Phyllis as “average.” The CA received a positive net promoter score of 13, indicating a favorable likelihood of recommendation. This suggests that not only would most students use Phyllis themselves but also they would recommend it to their friends.

On the basis of user feedback, it was suggested that Phyllis should reduce the amount of text in each message to enhance readability and recommended slowing down the speed of message delivery to assist with user recognition and comprehension. Furthermore, 2 users specifically requested that the CA offer direct answers to various questions about fitness and exercise without necessitating the initial mention of barriers. They also suggested incorporating more videos, diagrams, and real-life scenarios depicting PA. As anticipated, users expressed an interest in customizing the CA further by having the option to choose a different avatar, name, and personality.

**Figure 3.** User experiences and relational capacity.

Concerning acceptability, a sizable portion of the students (34/46, 73%) expressed their definite intention to use a fully functional CA in the future. The remaining students stated that they might consider using it in the future. When asked to rate Phyllis, 65% (30/46) of respondents considered it “good,” whereas 20% (9/46) rated it as “very good.” The remaining 15% (7/46) rated Phyllis as “average.” The CA received a positive net promoter score of 13, indicating a favorable likelihood of recommendation. This suggests that not only would most students use Phyllis themselves but also they would recommend it to their friends.

On the basis of user feedback, it was suggested that Phyllis should reduce the amount of text in each message to enhance readability and recommended slowing down the speed of message delivery to assist with user recognition and comprehension. Furthermore, 2 users specifically requested that the CA offer direct answers to various questions about fitness and exercise without necessitating the initial mention of barriers. They also suggested incorporating more videos, diagrams, and real-life scenarios depicting PA. As anticipated, users expressed an interest in customizing the CA further by having the option to choose a different avatar, name, and personality.

**Figure 3.** User experiences and relational capacity.

Concerning acceptability, a sizable portion of the students (34/46, 73%) expressed their definite intention to use a fully functional CA in the future. The remaining students stated that they might consider using it in the future. When asked to rate Phyllis, 65% (30/46) of respondents considered it “good,” whereas 20% (9/46) rated it as “very good.” The remaining 15% (7/46) rated Phyllis as “average.” The CA received a positive net promoter score of 13, indicating a favorable likelihood of recommendation. This suggests that not only would most students use Phyllis themselves but also they would recommend it to their friends.

On the basis of user feedback, it was suggested that Phyllis should reduce the amount of text in each message to enhance readability and recommended slowing down the speed of message delivery to assist with user recognition and comprehension. Furthermore, 2 users specifically requested that the CA offer direct answers to various questions about fitness and exercise without necessitating the initial mention of barriers. They also suggested incorporating more videos, diagrams, and real-life scenarios depicting PA. As anticipated, users expressed an interest in customizing the CA further by having the option to choose a different avatar, name, and personality.

**Figure 3.** User experiences and relational capacity.

Concerning acceptability, a sizable portion of the students (34/46, 73%) expressed their definite intention to use a fully functional CA in the future. The remaining students stated that they might consider using it in the future. When asked to rate Phyllis, 65% (30/46) of respondents considered it “good,” whereas 20% (9/46) rated it as “very good.” The remaining 15% (7/46) rated Phyllis as “average.” The CA received a positive net promoter score of 13, indicating a favorable likelihood of recommendation. This suggests that not only would most students use Phyllis themselves but also they would recommend it to their friends.

On the basis of user feedback, it was suggested that Phyllis should reduce the amount of text in each message to enhance readability and recommended slowing down the speed of message delivery to assist with user recognition and comprehension. Furthermore, 2 users specifically requested that the CA offer direct answers to various questions about fitness and exercise without necessitating the initial mention of barriers. They also suggested incorporating more videos, diagrams, and real-life scenarios depicting PA. As anticipated, users expressed an interest in customizing the CA further by having the option to choose a different avatar, name, and personality.

**Figure 3.** User experiences and relational capacity.
level with 3 statements, both before (pre) and after (post) the use of Phyllis. A total of 80% (37/46) of students reported being “confident in their ability to engage in PA or play sports,” indicating a 50% increase from the initial preresor response rate of 53%. The findings for motivation were even more notable, with 73% (34/46) of users reporting an enhanced motivation to engage in PA following the intervention. This represents a significant 120% increase in agreement with the statement from 33% (15/46) pre-to 73% (34/46) postintervention. Furthermore, students were asked about their ability to overcome challenges related to PA or sports. Over half of the respondents (24/46, 52%) agreed that they could overcome these challenges, which increased to 67% (31/46) after engaging with Phyllis. However, it is worth noting that this outcome may have been less substantial owing to the current limitations of the CA, as it is unable to provide solutions or modules for all 52 identified barriers. Overall, 67% (31/46) of respondents felt more motivated to participate in PA or sports since using Phyllis, and an overwhelming 93% (43/46) of them expressed a desire to be more active throughout the day. In addition, 73% (34/46) of respondents indicated that Phyllis helped them contemplate ways to increase their PA levels during the day.

Overall, there was no significant difference reported between the adolescents who were part of the co-design and those who were not. This was the same for users when comparing users from Wickersley School and Sports College and Zest.

Discussion

Principal Findings

The study involved co-designing and evaluating a pilot CA named Phyllis to test its proof of concept. The primary objective was to support adolescents in overcoming barriers to PA [15-17], to realize the associated benefits [1,2,4,50], and to provide an alternative digital solution to combat global inactivity [6]. Furthermore, it aims to promote the healthy behaviors during adolescence [7,8,51], thereby fostering continued participation in adulthood. The hypothesis guiding the study was that a CA could assist adolescents and that they would perceive Phyllis as a valuable tool for boosting their confidence and motivation to engage in PA. The study’s objectives were to co-design the CA in collaboration with adolescents, demonstrating the CA’s capacity to understand user input related to the identified barriers. It also aimed to evaluate the CA’s usability and acceptability among adolescents and assess the perceived effectiveness of the solutions offered by the CA to address these barriers. Overall, adolescents reported high acceptability and positive user experiences when using Phyllis, whereas the modules and personas designed to increase motivation and confidence achieved positive outcomes. This evidence positions Phyllis as a promising digital tool to alleviate global inactivity and aligns with the goals of the World Health Organization’s Global Action Plan on PA [18].

The research builds upon prior studies of PA interventions [24,25,27,28,52,53] and adolescent barriers to PA [15,16,51] and is informed by a behavior change model and theoretical framework [29]. This approach is unique as it draws upon relevant research, engages in co-design with adolescents and applies relevant theories [30,35,38,54] to design a CA targeted at adolescents. It is a novel approach in this domain targeted at this population, as indicated by prior research [27,28], which has identified a lack of CAs tailored specifically for adolescents. This innovation in design, using the COM-B model and TDF for intervention development, promises to pave the way for more impactful interventions in the future.

The co-design process engaged 8 adolescents, which represented the broader population’s sex distribution and a higher representation of “Asian” and “Black” ethnic groups who are more likely to be inactive [17]. The process yielded significant findings that support the rationale for the study. First, it highlighted the inadequate access that adolescents have to support systems for addressing barriers to PA and provided further evidence of the array of barriers faced by adolescents [15-17]. Co-design participants were able to identify 37 barriers to engaging in PA, reporting an average of 9 barriers per person. This accounts for approximately 66% of the barriers that the CA will aim to support, based on a study that served as the foundation for the current research [17]. These findings highlight the significance of developing multicomponent interventions that can effectively address multiple barriers, as advocated in the existing literature [12]. Notably, a lack of confidence and motivation emerged as significant barriers to this cohort, which is consistent with previous studies [15,16,51]. Moreover, adolescents reported experiencing 4 times as many negative feelings as positive feelings when describing the emotional impact of PA, highlighting potential reasons for limited engagement in PA. An unintended outcome of the study was the revelation of a paradox, as some adolescents reported that PA may exacerbate psychological effects such as anxiety, while perceiving it as a means of improving these conditions. Subsequent iterations of Phyllis will consider these limiting beliefs, aiming to help individuals become more self-aware of such conflicting thoughts.

For the CA to effectively address barriers to PA, it must be perceived as an acceptable tool for support by adolescents. The co-design participants expressed high levels of acceptability toward the CA and demonstrated a willingness to share personal information with it. However, some students voiced concerns regarding privacy and data protection, aligning with the findings of other studies [27]. Previous studies have highlighted the significance of personalization [23], and this study further underscores adolescents’ specific requirements. Although a consensus favored a gender-neutral teenage persona with a human-like tone, the ideal scenario would involve fully customizable features such as persona, avatar, and tone of voice. The adolescents’ diverse preferences were also reflected in their choice of the CA’s personality and characteristics, representing a range of personalities. Although challenging for designers, this evidence is valuable for guiding the future design of CAs and for advancing research in this domain. Overall, the feedback emphasized the importance of empathy, intelligence, and kindness in the CA’s approach, characterized by motivational language, acknowledgment of concerns, and positive reinforcement through both submissive and dominant behaviors at specific points of interaction. For example, some adolescents appear willing and would even encourage the CA to challenge
their beliefs and offer alternative views and solutions. Similarly, there was variation in the preferred approach to delivering content using different media; yet, it was clear that adolescents desired more complex interactions with content delivered and explained conversationally rather than being signposted to external websites or apps. This perception strengthens the case for CAs to be used for this purpose, with interventions delivered conversationally. All this knowledge provides much-needed insight into the design of effective CAs targeted at this population.

The second objective was to build a CA that, although informed by the co-design, had a specific purpose of recognizing input from adolescents around the 52 barriers [17] and providing a relevant solution to overcome each barrier. This is a novel approach that has not been applied in the design of previous CAs. In terms of recognition, an $F_1$-score was used to assess the performance of the model, which scored 80%. All relevant and negligible utterances were matched accurately during the pilot, and adolescents who typed utterances related to “a lack of confidence” or “fitness” were accurately matched to the solution. Further testing will be required when all solutions are developed, including training data, to help improve recognition further. Importantly, the model can recognize a range of input from adolescents, thereby improving the potential viability of this approach to support adolescents.

Once the prototype CA was designed and developed, the objective was to validate it as a proof of concept by assessing its acceptability, usability, and effectiveness in supporting adolescents to overcome barriers to PA. Most adolescents expressed positive acceptability toward the CA, with 73% (34/46) of students indicating they would “definitely use” a fully operational Phyllis and 85% (39/46) rating the CA as either “good” or “very good.” In addition, Phyllis received a “good” net promoter score, indicating that adolescents would also recommend it to their friends [55]. Quantitative data analysis revealed positive scores in terms of user experience, whereas qualitative feedback highlighted students’ appreciation for the CA’s personality; humor; and the use of emojis, icons, and GIFs.

The evaluation also provided valuable feedback to enhance Phyllis in the short term (eg, delivering fitness activities conversationally) and long term (eg, enabling deeper personalization and reducing message text). Students found Phyllis helpful, informative, likable, and motivating and appreciated its ability to ask open-ended questions, demonstrating the importance of unconstrained conversation [28] and the ability to provide reassurance. These results indicate that adolescents had a positive user experience and would be willing to use a CA such as Phyllis in the future, further strengthening the evidence for these tools to be used to support adolescents. Such features developed from the co-design process will help alleviate concerns expressed in the literature around usability challenges [27] faced by users and further reinforce the importance of integrating the principles of user-centered design [30] and human-computer interaction [31] with the aim of establishing relational and persuasive capacity [28] with users. These findings could also be used to inform the development of other CAs in this and other domains, serving as useful insights for designers and developers of digital health interventions.

The primary goal of the CA is to augment PA behaviors, and the prototype included 2 modules developed for this purpose. Following an interaction with the CA, 80% (37/46) of the students reported improved confidence in their ability to participate in PA or sports. This represents a 50% increase compared with the initial response. In addition, there was a significant 120% increase in the proportion of students who agreed with the statement expressing their motivation to engage in PA after engaging with the CA. This serves as a positive indicator of the CA’s efficacy. Most students (34/46, 73%) also indicated that the CA would assist them in exploring ways to incorporate more PA into their daily routines. These findings demonstrate the potential of the CA as an effective tool in helping students overcome barriers to PA, further validating the CA’s potential. However, further research is necessary to determine if these improvements translate into increased levels of PA and to identify an appropriate measurement method so that PA can be monitored via the CA.

It is important to consider the implications of these findings in the context of recent developments in LLMs and generative AI, as they have the potential to enhance the effectiveness of the CA as they can comprehend and generate natural language with heightened complexity and accuracy [33,34]. These advancements can also expedite the development process, elevate the quality of interaction, and expand their knowledge base by providing a more comprehensive conversational experience that meets the specific needs of users. In health-related fields, it is imperative that CAs provide accurate, trustworthy, and evidence-based information and address concerns related to bias, misinformation, privacy, and security that may arise from the use of LLMs.

The approach advocated in this study has the potential to bridge the gap between sophisticated LLMs and trusted, evidence-based content. To illustrate this further, interventions should continue to be co-designed and expert led, grounded in evidence-based content. This content can then serve as the foundation for the development of domain-specific models, which in turn are used to provide data to inform intervention delivery. The critical aspect of intervention delivery depends on the use of prompt engineering or the application of a cognitive system that guides the model on how and when to deliver an intervention. The process of delivering interventions may necessitate extremely precise prompting to ensure the accuracy of behavior change tools, intervention content, language, and persona. This precision may require the use of an LLM working in combination with a traditional intent-based approach to allow for predefined responses to guarantee the accuracy and quality of the expert-designed intervention content. Adhering to these principles can effectively mitigate concerns and establish an audit trail for the content provided, ultimately promoting greater transparency. This transparent and comprehensive audit trail serves not only to foster trust but also to strengthen accountability and facilitate the identification of potential biases or inaccuracies.
The limitations of the study findings lie in their reliance on a single interaction with the CA, which does not provide insight into long-term adherence or sustained behavior change resulting from CA use. Nonetheless, the results show promise in terms of children’s willingness to engage and to self-report positive attitude changes after a brief period of interacting with the CA. The evidence also demonstrates why this population perceives CAs as a potentially valuable solution. Further research is necessary to address each of the 52 identified barriers, identify the commonalities among them, enhance the algorithm, and ensure greater efficacy.

Overall, the results demonstrate that Phyllis has the potential to be a cost-effective, resource-efficient solution that organizations can offer to support adolescents and address the multifaceted barriers to PA [15-17]. With further development, the tool could serve as a self-help resource for students, with teachers administering it to inactive students to enhance their levels of PA. Positive evidence of digital interventions being used in both formal and informal settings exists to support their potential use in this setting [22]. As this research highlights, schools lack the time, resources, and capacity to adequately support adolescents [14], making the CA an effective tool to augment existing support. Data from CA interactions could also be shared with schools to identify more serious barriers such as mental and physical health issues or to highlight prevalent barriers within the student population, thereby improving support within the CA and informing policy or intervention approaches within schools. Moreover, the data could offer greater insight into adolescent needs and barriers, informing policy in this field.

The implementation of digital technology in schools, which ensures transparency, safety, and trust, can be achieved with minimal demands on school staff and teachers while also being cost-effective. The findings reinforce the importance of evidence-based self-help tools, which can be accessed by adolescents in schools at a low cost. Moreover, these tools can be supervised by trusted adults, ensuring personalized support for adolescents without overwhelming the capacity of teachers or the school. The CA and its design also have the potential for broader deployment within health care systems as part of the PA referral pathway to enhance adherence to programs or social prescribing services that aim to enhance PA levels among adolescents with long-term conditions or hypokinetic diseases. The design could also be enhanced to provide multimodal support by using robotics and be adapted and used within other populations, such as with older adults.

Conclusions

The study aimed to co-design, develop, and evaluate a prototype CA as a proof of concept to assist adolescents in overcoming barriers to PA. It presents one of the first theory-driven approaches to designing a CA. Drawing from prior studies, theoretical approaches, and insights into adolescent barriers, the research focused on achieving relational and persuasive capacity with adolescents using the CA. Highlighting inadequate access to support systems, 37 barriers were identified, with 66% aligning with previous research. The emphasis was placed on the acceptability and personalization of the CA to address privacy and data protection concerns. The results demonstrated high acceptability and positive user experiences, highlighting the potential of the CA. The modules designed for the CA showed promising outcomes, fostering increased confidence and motivation for PA. Phyllis also performed well, recognizing a range of different barriers to PA. However, further research is needed to develop other modules to overcome other barriers and explore long-term adherence and the effectiveness of interventions for sustaining behavior change. Adolescents also expect to have a personalized experience and be able to personalize all aspects of the chatbot. Phyllis holds potential as a cost-effective solution for schools to support adolescents and tackle barriers to PA. The integration of LLMs can significantly enhance the CA’s capabilities, facilitating sophisticated conversations and automated content generation with this study providing knowledge of how they can be designed to incorporate evidence-based approaches to ensure trust and transparency. The CA’s potential extends to health care systems, social prescribing services, and multimodal support, including robotics. This study provides valuable insights for designing and developing digital health interventions in other domains as well as contributing to the improvement of PA levels among adolescents and addressing global inactivity concerns.

Acknowledgments

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any author-accepted manuscript version arising from this submission. Thank you to Richard Gill (teacher) and the students at Wickersley School and Sports College who took part in this study.

Data Availability

The data sets used and, or analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions

RM contributed by conceptualizing and writing the original draft of the manuscript. AKA-T and EF contributed by reviewing and editing the manuscript.

Conflicts of Interest

None declared.
References


Abbreviations

AI: artificial intelligence
CA: conversational agent
COM-B: capability, opportunity, motivation, and behavior
LLM: large language model
NLU: natural language understanding
PA: physical activity
TDF: Theoretical Domains Framework

Please cite as:
Moore R, Al-Tamimi AK, Freeman E
Investigating the Potential of a Conversational Agent (Phyllis) to Support Adolescent Health and Overcome Barriers to Physical Activity: Co-Design Study
JMIR Form Res 2024;8:e51571
URL: https://formative.jmir.org/2024/1/e51571
doi:10.2196/51571
PMID:38294857

©Richard Moore, Abdel-Karim Al-Tamimi, Elizabeth Freeman. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete
bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Improvements in Adolescents’ Disordered Eating Behaviors in a Collaborative Care Digital Mental Health Intervention: Retrospective Observational Study

Landry Goodgame Huffman\textsuperscript{1}, PhD; Darian Lawrence-Sidebottom\textsuperscript{1*}, PhD; Aislinn Brenna Beam\textsuperscript{1}, PhD; Amit Parikh\textsuperscript{1}, MD; Rachael Guerra\textsuperscript{1}, PhD; Monika Roots\textsuperscript{1}, MD; Jennifer Huberty\textsuperscript{1,2}, PhD

\textsuperscript{1}Bend Health Inc, Madison, WI, United States
\textsuperscript{2}FitMinded Inc LLC, Phoenix, AZ, United States
\textsuperscript{*}these authors contributed equally

Corresponding Author:
Darian Lawrence-Sidebottom, PhD
Bend Health Inc
2810 Marshall Ct
Madison, WI, 53705
United States
Phone: 1 8005160975
Email: darian.lawrence@bendhealth.com

Abstract

Background: Young people today are exhibiting increasing rates of disordered eating behaviors, as well as eating disorders (EDs), alongside other mental and behavioral problems such as anxiety and depression. However, limited access to mental health care means that EDs, disordered eating behaviors, and comorbid mental health problems are often underdiagnosed and undertreated. Digital mental health interventions (DMHIs) offer accessible and scalable alternatives to traditional treatment modalities, but their effectiveness has not been well established among adolescents with EDs and disordered eating behaviors.

Objective: This study uses data from a collaborative care pediatric DMHI to determine whether participation in a DMHI is associated with a reduction in adolescents’ disordered eating behaviors.

Methods: Adolescent members in care with Bend Health Inc completed the SCOFF questionnaire at baseline (before the start of care) and approximately every month during care to assess disordered eating behaviors. They also completed assessments of mental health symptoms at baseline. Member characteristics, mental health symptoms, and disordered eating behaviors of adolescents with elevated SCOFF scores at baseline (before the start of care) were compared to those of adolescents with nonelevated SCOFF scores at baseline. Members participated in web-based coaching or therapy sessions throughout the duration of mental health care.

Results: Compared to adolescents with nonelevated SCOFF scores (n=520), adolescents with elevated SCOFF scores (n=169) were predominantly female and exhibited higher rates of elevated anxiety and depressive symptoms. SCOFF scores decreased over time in care with the DMHI for 61.4% (n=70) of adolescents with elevated SCOFF scores, and each additional month of participation was associated with greater improvements in disordered eating behaviors ($F_{1,233}=72.82; P<.001$).

Conclusions: Our findings offer promising preliminary evidence that participation in mental health care with a collaborative care DMHI may be beneficial in the reduction of disordered eating symptoms in adolescents, including those who are experiencing comorbid anxiety and depressive symptoms.

(JMIR Form Res 2024;8:e54253) doi:10.2196/54253

KEYWORDS
behavioral care; mental health; web-based coaching; web-based therapy; eating disorders; eating; anorexia; coach; coaching; pediatric; pediatrics; adolescent; adolescents; teen; teens; teenager; teenagers; digital mental health intervention; DMHI; collaborative; digital health
Introduction

Eating disorders (EDs) and associated disordered eating behaviors impact approximately 10% of the US population [1]. The peak age of onset of EDs such as anorexia nervosa, bulimia nervosa, avoidant or restrictive food intake disorder, and binge ED is between 13 and 18 years, making them particularly relevant during adolescence. Estimates of EDs in adolescence range from 1.2% among males to 5.7% in females and 9% among sexual and gender nonconforming youths [2-5]. Disordered eating behaviors (eg, obsessing over food intake and excessive worry about weight) are even more common, impacting nearly 1 in 4 children and adolescents [6]. EDs and disordered eating behaviors disrupt critical periods of physical and socioemotional development that occur during childhood and adolescence [7]. Youth with EDs experience compromised physical functioning such as malnutrition, disrupted pubertal development, and delayed menarche, as well as worsened psychosocial functioning. EDs are highly comorbid with, and often preceded by, other mental health problems such as anxiety or anxiety disorders, depression, and obsessive-compulsive disorder [8].

During the COVID-19 pandemic, the prevalence of EDs more than doubled among adolescents, exacerbating an already pressing public health issue [9,10]. Indeed, EDs confer a significant economic burden for families and hospitals, with an estimated annual disease burden of US $70 billion [11,12]. However, more than 75% of those with EDs or risk for EDs do not receive the necessary treatment [13]. Several issues underlie this gap between ED diagnosis and treatment. ED treatments are often difficult to access, especially for young people, who often experience financial, geographic, and transportation constraints as well as increased stigma [14]. Additionally, in-person services, such as inpatient and outpatient care and face-to-face therapy, are severely limited in their accessibility, largely due to shortages of trained personnel and long waitlists [14].

In recent years, digital mental health interventions (DMHIs) such as self-guided applications and web-based therapy have emerged as accessible and scalable alternatives to traditional mental health treatments. Several systematic reviews and meta-analyses suggest the potential for DMHIs in ED treatment, although results remain heterogeneous and inconclusive [15]. Moreover, few studies of DMHIs for EDs have been conducted among adolescents, despite the pressing need for accessible child and adolescent ED treatments. As argued by Loucas et al [16], most pediatric DMHIs for EDs are more similar to web-based self-help programs than therapeutic interventions, given their lack of personalized and interactive components [17-19]. However, recent advances in pediatric DMHIs using family-based therapy for ED treatment have yielded promising results, suggesting that digital ED treatments for youths are most effective when administered in the context of a holistic care team [20-22].

The collaborative care model is a team-based framework for mental health care that has been used by DMHIs with promising results. In the collaborative care model, primary care providers collaborate with behavioral care managers (BCM) and other providers to implement measurement-based mental health care. Collaboration between providers reduces the burden on primary care providers while ensuring optimal, evidence-based care via regular symptom assessments. As a result, interventions that use collaborative care models are particularly effective for the treatment of mental health problems in both youths and adults [23-26]. However, no studies to date have evaluated whether participation in collaborative care DMHIs may improve disordered eating behaviors among adolescents. As such, the purpose of this study was to use retrospective analyses of data collected from adolescents participating in care with a collaborative care DMHI to determine whether participation in a pediatric collaborative care DMHI is associated with a reduction in disordered eating behaviors. Considering established demographic and clinical correlates of disordered eating behaviors we also explored associations between potential confounds (age, sex, and anxiety and depressive symptoms) and disordered eating behaviors among those receiving care with the DMHI.

Methods

Design and Participants

All adolescents (ages 13-17 years) who met the following inclusion criteria were eligible for inclusion in the study (N=689): (1) started mental health care (first synchronous event) with Bend Health Inc between January 1, 2023, and October 1, 2023 (9 months); (2) had at least 1 synchronous session with a Bend Health Inc coach or therapist during the study time frame; and (3) completed the assessment of eating behaviors before the start of care (baseline).

Treatment

Bend Health Inc is a DMHI that provides behavioral care for adolescents (aged 13-17 years), using a whole-family approach (ie, caregivers are closely involved in care), via a web-based platform. Bend Health Inc’s behavioral care has been described elsewhere [24,25]. Members enroll via referral from a health care provider, or they use insurance, employer benefits, or self-pay. Members are assigned a BCM, who conducts an initial evaluation of the member’s mental health concerns and circumstances, and they continue to monitor the member’s care while they are enrolled in the program. BCMs assign members a coach, and also a therapist in some cases, based on their mental health symptoms (eg, type and acuity), goals for treatment, and insurance coverage; therapists tend to be assigned only to members with more severe symptoms and conditions, whereas nearly all members are assigned a coach. Members with higher symptom acuity or a psychiatric referral may also be assigned a prescribing psychiatric practitioner (eg, a medical doctor or psychiatric nurse practitioner). In synchronous video-based coaching and therapy sessions, Bend Health Inc’s practitioners guide members and their caregivers through structured care programs. These care programs are designed to deliver evidence-based tools and techniques that target common mental and behavioral health issues such as anxiety, depression, and body image. Most care programs are designed to be completed in approximately 3 months, and some care programs (eg, the

https://formative.jmir.org/2024/1/e54253

JMIR Form Res 2024 | vol. 8 | e54253 | p.635

(page number not for citation purposes)
body image program) are designed to be completed in a shorter amount of time. Once a month, caregivers and adolescent members are asked to complete validated web-based assessments of the adolescent’s mental health symptoms (see Measures section). Caregivers are required to be in the same general location for their adolescents’ synchronous sessions with a coach or therapist for safety purposes.

**Measures**

Demographic and health information of adolescent participants is gathered during enrollment with Bend Health Inc. Caregivers respond to basic demographic questions, providing their adolescent member’s date of birth, sex at birth (male, female, or others), gender identity (male, female, transgender, nonbinary, or others), and race or ethnicity. Details on the race or ethnicity response options are included in Multimedia Appendix 1.

In addition, at enrollment, caregivers and adolescent members complete a series of symptom screening questions and validated assessments to identify common mental and behavioral health concerns. To assess eating behaviors, all adolescent members complete the SCOFF (Cronbach $\alpha=0.48$) [27,28]. The SCOFF is a valid questionnaire, in which the member responds “yes” or “no” to 5 items about disordered eating behaviors that they might have. The 5 items, which each correspond with a letter of the SCOFF name, are as follows: (1) Do you make yourself sick because you feel uncomfortably full? (2) Do you worry that you have lost control over how much you eat? (3) Have you recently lost more than one stone (14 lb) in a 3-month period? (4) Do you believe yourself to be fat when others say you are too thin? and (5) Would you say that food dominates your life?

To identify members with elevated anxiety and depressive symptoms, caregivers of adolescent members respond to 5 screening questions drawn from the Diagnostic and Statistical Manual of Mental Disorder, Fifth Edition, Text Revision (DSM-5-TR) Cross-Cutting Symptom Measure [29]. For anxiety, the caregivers and adolescents together respond to the following questions: Over the last 2 weeks, how often have you been bothered by any of the following problems? (1) Feeling nervous, anxious, or on edge? and (2) Not being able to stop or control worrying? For depressive symptoms, caregivers and teens responded to the items: Over the last 2 weeks, how often have you been bothered by any of the following problems? (1) Had less fun doing things than you used to? and (2) Felt sad or depressed for several hours? Best-fit responses are selected from a 4-item Likert-type scale, with responses ranging from “not at all” (item score=0) to “nearly every day” (item score=3). Screener scores are calculated by aggregating all item scores for each screener.

Members who have an anxiety screener score of 2 or more are prompted to complete the Generalized Anxiety Disorder-7 (GAD-7) assessment (Cronbach $\alpha=0.91$) [30,31], and members who have a depressive screener score of 2 or more are prompted to take the Patient Health Questionnaire-9 for Adolescents (PHQ-9A; Cronbach $\alpha=0.85$) [32,33]. The GAD-7 has 7 questions regarding symptoms of anxiety in the prior 2 weeks with the same Likert-type scale as used in the screener questions (“not at all” to “nearly every day”). The PHQ-9A is a modified version of the PHQ-9 for adolescents aged 11-17 years. The original measure has 9 questions, but we omit a question regarding suicide and self-harm (ie, the PHQ-9A here includes 8 questions). The PHQ-9A asks adolescents about depressive symptoms in the prior week using the same Likert-type scale as in the GAD-7 and screener questions. Adolescents are asked to report their own symptoms for both GAD-7 and PHQ-9A.

**Statistical Analysis**

**Outcome Calculations**

The last assessment before the start of care (baseline) and all assessments after the start of care were considered for analysis. SCOFF scores were calculated by aggregating the number of “yes” responses, with scores ranging from 0 to 5. As has been used previously [27], members with a SCOFF score of 2 or more on their baseline assessment (ie, the last assessment before the start of care) were included in the “elevated SCOFF score” group, and members with a score of less than 2 were included in the “non-elevated SCOFF score group.” GAD-7 scores were calculated by aggregating the individual item scores. PHQ-9A scores were calculated by aggregating the individual item scores, and then, dividing by 8 and multiplying by 9 (to account for the omitted item). Members with moderate or greater severity anxiety or depressive symptoms, as determined by established criteria for the GAD-7 and PHQ-9A [29,30], were flagged as having elevated anxiety or depressive symptoms, respectively.

**Baseline Characteristics: Eating Behaviors and Member Characteristics**

SCOFF scores at baseline were characterized for all members, including total score and responses to individual SCOFF items. Then, member characteristics, anxiety and depressive symptom severity (elevated or nonelevated), and care participation characteristics were reported for each group and compared between groups to identify any differences. Member characteristics included age at baseline, sex (female, male, and nonbinary), gender-sex conformity (conforming and nonconforming), race or ethnicity (Asian, Black or African American, Hispanic or Latino, White, and other or multiracial), mental health condition (anxiety disorder diagnosis and depressive disorder diagnosis), and elevated mental health symptoms (anxiety and depression). The care participation metrics included the number of months in care (time between first session and last session), rates of members in coaching and therapy, and rates of members participating in the anxiety, depression, and body image care programs. Member demographics were reported by caregivers (described earlier), mental health conditions were identified from electronic health records, mental health symptoms were characterized based on symptom severity at baseline, and care participation characteristics were assessed using data from electronic health records. Age in years and months in care were compared between groups using Wilcoxon signed rank tests. All other between-group comparisons for member characteristics were performed using chi-square tests.

**Change in Eating Behaviors**

The number of total SCOFF assessments (ie, with baseline as the first assessment) was quantified for members in both groups.
Only data from members with at least 2 assessments (baseline and at least 1 assessment after care) were included in the analyses of change in SCOFF scores (n=233 members excluded). For members in both groups, SCOFF scores at baseline and the last assessment, as well as the change in score from baseline to last assessment, were quantified. Change scores were compared to 0 using Wilcoxon signed rank test to determine whether SCOFF scores changed significantly over the course of the care. The rates of members with a decrease and an increase in SCOFF score were reported for both groups to quantify rates of symptom improvement and symptom worsening, respectively. To identify which items contributed to a change in SCOFF score, “yes” responses to each item at the last assessment were reported for members who responded “yes” to the item at baseline.

Finally, a linear mixed effects model was used to test whether SCOFF scores decreased over months in care and to test whether mental health symptom severity (at baseline) and demographic factors predicted SCOFF scores. Only members with a baseline assessment within 1 month or less of the start of care were included in the linear mixed effects model (n=10 members excluded). The basic model included a fixed effect of months in care (at the time of SCOFF assessment) and a random effect of subject (member ID) on the intercept. Alternative models including one of the following predictors were tested for fit against the basic model using likelihood ratio tests: elevated anxiety symptoms at baseline (yes or no), elevated depressive symptoms at baseline (yes or no), sex at birth (female or nonfemale), age at baseline (in years), and participation in therapy (yes or no). When a likelihood ratio test was statistically significant, the predictor variable was included in the final model as a fixed effect.

Throughout, standard descriptive statistics (eg, percentages, mean and SD, and median and IQR) were used to describe the data, as appropriate. IQR values are reported as the range: 25th-75th percentile. An α level of .05 was used as the threshold of statistical significance for all analyses. Tests of normality (Kolmogorov-Smirnov and Jarque-Bera) were performed to determine appropriate statistical tests and descriptive statistics. P values were corrected for multiple tests using the Bonferroni correction statistical tests performed on baseline characteristics and change in SCOFF scores (2 sets of corrections).

**Ethical Considerations**

At enrollment with Bend Health Inc, all study participants provided informed consent for primary data collection (required for participation in care) and use of their data in further analyses.

### Table 1. Distribution of SCOFF scores at baseline (N=689).

<table>
<thead>
<tr>
<th>SCOFF score</th>
<th>Members, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>368 (53.4)</td>
</tr>
<tr>
<td>1</td>
<td>152 (22.1)</td>
</tr>
<tr>
<td>2</td>
<td>100 (14.5)</td>
</tr>
<tr>
<td>3</td>
<td>50 (7.3)</td>
</tr>
<tr>
<td>4</td>
<td>13 (1.9)</td>
</tr>
<tr>
<td>5</td>
<td>6 (0.9)</td>
</tr>
</tbody>
</table>

Given the retrospective observational nature of the study, participants were not compensated for their participation in the study. Procedures for this study were approved by the Biomedical Research Alliance of New York (Study 23-12-034-1374; approved on June 5, 2023). To ensure the privacy and confidentiality of the human participants in this study, all data (eg, from electronic health records) were deidentified prior to analysis.

### Results

#### Baseline Characteristics: Eating Behaviors and Member Characteristics

The distribution of baseline SCOFF scores is reported in Table 1. Ultimately, 75.5% (n=520) had nonelevated SCOFF scores, and 24.5% (n=169) had elevated SCOFF scores. For members with nonelevated SCOFF scores, 13.7% (n=71) of members responded “yes” to the item about control (item 2), and 9.2% (n=48) responded “yes” to the item about believing you are fat (item 4; Table 2). For members with elevated SCOFF scores, the most commonly reported items were loss of control (item 2: n=151, 89.3%), believing you are fat (item 4: n=113, 66.9%), and food dominating life (item 5: n=92, 54.4%).

Compared to the nonelevated SCOFF score group, the elevated SCOFF score group was more predominantly female ($\chi^2_{1}=24.2; P<.001$) and also had a higher rate of diagnoses with depressive disorders ($\chi^2_{1}=9.5; P=.005$; Table 3). Rates of elevated anxiety and depressive symptoms were higher for members with elevated SCOFF scores than members with nonelevated SCOFF scores (anxiety: $\chi^2_{1}=31.9; P<.001$ and depression: $\chi^2_{1}=63.2; P<.001$). Age, gender-sex conformity, race or ethnicity, and rates of anxiety disorder diagnoses did not differ between groups.

Members were in care for a median of 2.60 (IQR 1.27-4.23) months. In terms of participation in behavioral care, 98.8% (n=681) of all members were in coaching, and 39.8% (n=274) of all members were in therapy. Rates of participation in coaching and therapy, as well as months in care, did not differ between groups. Approximately 1 in 2 members participated in the anxiety care program, and this was similar between groups. However, participation in the depression and body image care programs was higher for members with elevated SCOFF scores versus members with nonelevated SCOFF scores (depression: $\chi^2_{1}=15.0; P<.001$).
### Table 2. Responses to each SCOFF item at baseline by group.

<table>
<thead>
<tr>
<th>SCOFF item</th>
<th>Elevated SCOFF score (n=169, n (%))</th>
<th>Nonelevated SCOFF score (n=520, n (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you make yourself <em>sick</em> because you feel uncomfortably full?</td>
<td>13 (2.5)</td>
<td>43 (25.4)</td>
</tr>
<tr>
<td>Do you worry that you have lost control over how much you eat?</td>
<td>71 (13.7)</td>
<td>151 (89.3)</td>
</tr>
<tr>
<td>Have you recently lost more than one stone (14 lb) in a 3-month period?</td>
<td>11 (2.1)</td>
<td>33 (19.5)</td>
</tr>
<tr>
<td>Do you believe yourself to be fat when others say you are too thin?</td>
<td>48 (9.2)</td>
<td>113 (66.9)</td>
</tr>
<tr>
<td>Would you say that <em>food</em> dominates your life?</td>
<td>9 (1.7)</td>
<td>92 (54.4)</td>
</tr>
</tbody>
</table>

### Table 3. Member characteristics for each group.

<table>
<thead>
<tr>
<th>Member characteristics</th>
<th>Elevated SCOFF score (n=169, 24.5%)</th>
<th>Nonelevated SCOFF score (n=520, 75.5%)</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), median (IQR)(^a)</td>
<td>15 (14-16)</td>
<td>15 (14-16)</td>
<td>Chi-square (df=1)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td>N/A(^b)</td>
</tr>
<tr>
<td>Female</td>
<td>301 (57.9)</td>
<td>134 (79.3)</td>
<td>24.2</td>
</tr>
<tr>
<td>Male</td>
<td>212 (40.8)</td>
<td>33 (19.5)</td>
<td>N/A</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>7 (1.3)</td>
<td>2 (1.2)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Gender-sex conformity, n (%)</strong></td>
<td></td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td>Conforming</td>
<td>481 (92.5)</td>
<td>155 (91.7)</td>
<td></td>
</tr>
<tr>
<td>Nonconforming</td>
<td>39 (7.5)</td>
<td>14 (8.3)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Race or ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Asian</td>
<td>30 (5.8)</td>
<td>4 (2.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Black or African American</td>
<td>35 (6.7)</td>
<td>15 (8.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>25 (4.8)</td>
<td>17 (10.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>White</td>
<td>252 (48.5)</td>
<td>87 (51.5)</td>
<td>0.4</td>
</tr>
<tr>
<td>Other or multiracial</td>
<td>178 (34.2)</td>
<td>46 (27.2)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Mental health condition, n (%)</strong></td>
<td></td>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>147 (28.3)</td>
<td>56 (33.1)</td>
<td></td>
</tr>
<tr>
<td>Depressive disorder</td>
<td>35 (6.7)</td>
<td>25 (14.8)</td>
<td>9.4</td>
</tr>
<tr>
<td><strong>Elevated mental health symptoms, n (%)</strong></td>
<td></td>
<td></td>
<td>31.9</td>
</tr>
<tr>
<td>Anxiety</td>
<td>230 (44.2)</td>
<td>118 (69.8)</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>211 (40.6)</td>
<td>129 (76.3)</td>
<td>63.2</td>
</tr>
<tr>
<td><strong>Duration in care (months), median (IQR)(^c)</strong></td>
<td>2.51 (1.23-4.20)</td>
<td>2.80 (1.50-4.37)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Behavioral care participation, n (%)</strong></td>
<td></td>
<td></td>
<td>0.92</td>
</tr>
<tr>
<td>Coaching</td>
<td>513 (98.7)</td>
<td>168 (99.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Therapy</td>
<td>201 (38.7)</td>
<td>73 (43.2)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Care program, n (%)</strong></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>Anxiety</td>
<td>243 (46.7)</td>
<td>84 (49.7)</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>77 (14.8)</td>
<td>48 (28.4)</td>
<td>15.0</td>
</tr>
<tr>
<td>Body image</td>
<td>6 (1.2)</td>
<td>7 (4.1)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(a\) Between-group comparison was performed using a Wilcoxon signed rank test (z=–0.23).

\(b\) N/A: not applicable.

\(c\) Between-group comparison was performed using a Wilcoxon signed rank test (z=–1.46).
Change in Eating Behaviors

SCOFF assessment counts during the study time frame are reported for each group in Table 4. For all members with a baseline and postcare assessment (ie, 2 or more total assessments), baseline assessments were completed at a median of 0.27 (IQR 0.5-0.13) months before the start of care, and the last assessments were completed at a median of 2.23 (IQR 1.1-3.6) months after the start of care. The timing of baseline and last assessments did not differ between groups (baseline: $z=0.70; P=.70$ and last assessment: $z=-0.38; P=.70$).

For members in the nonelevated SCOFF score group, 16.4% (n=56) had a decrease in SCOFF score from baseline to their last assessment, 13.7% (n=47) had an increase in SCOFF score, and 69.9% (n=239) had no change (Table 5). For members in the elevated SCOFF score group, on the other hand, 61.4% (n=70) had a decrease in SCOFF score from baseline to their last assessment, 16.7% (n=19) had an increase, and 21.1% (n=25) had no change. While SCOFF scores remained stable for the nonelevated SCOFF score group ($z=-0.47; P=.70$), median SCOFF scores decreased from 2 (IQR 2-3) at baseline to 1 (IQR 0-3) at the last assessment for the elevated SCOFF score group ($z=-6.39; P<.001$). Individual item responses at the last assessment for members who responded “yes” to the item at baseline are reported in Table 6.

In the linear mixed effects model of SCOFF score over time in care for members with elevated SCOFF scores, the following predictors were included in the final model: elevated anxiety symptoms at baseline ($\chi^2=7.6; P=.006$), elevated depressive symptoms at baseline ($\chi^2=8.6; P=.003$), and participation in therapy ($\chi^2=4.4; P=.03$). SCOFF score decreased by 0.26 points for each month in care ($F_{1,233}=72.82; P<.001$; Figure 1). Members with elevated anxiety symptoms at baseline had SCOFF scores 0.40 points higher than those with nonelevated anxiety symptoms ($F_{1,101}=9.87; P=.005$). Similarly, members with elevated depressive symptoms at baseline had SCOFF scores 0.43 points higher than those with nonelevated depressive symptoms ($F_{1,101}=5.40; P=.044$). Participation in therapy did not relate to SCOFF score ($F_{1,101}=2.87; P=.16$).

### Table 4. Total assessments completed for each group.

<table>
<thead>
<tr>
<th>Total assessments</th>
<th>Nonelevated SCOFF score (n=520), n (%)</th>
<th>Elevated SCOFF score (n=169), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>178 (34.2)</td>
<td>55 (32.5)</td>
</tr>
<tr>
<td>2</td>
<td>141 (27.1)</td>
<td>44 (26)</td>
</tr>
<tr>
<td>3</td>
<td>96 (18.5)</td>
<td>31 (18.3)</td>
</tr>
<tr>
<td>4</td>
<td>57 (11)</td>
<td>22 (13)</td>
</tr>
<tr>
<td>5</td>
<td>25 (4.8)</td>
<td>9 (5.3)</td>
</tr>
<tr>
<td>6</td>
<td>15 (2.9)</td>
<td>5 (3)</td>
</tr>
<tr>
<td>7+</td>
<td>8 (1.5)</td>
<td>3 (1.8)</td>
</tr>
</tbody>
</table>

### Table 5. Change in SCOFF score from baseline to last assessment for each group.

<table>
<thead>
<tr>
<th>SCOFF score, median (IQR)</th>
<th>Nonelevated SCOFF score (n=342)</th>
<th>Elevated SCOFF score (n=114)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0 (0 to 1)</td>
<td>2 (2 to 3)</td>
</tr>
<tr>
<td>Last</td>
<td>0 (0 to 0)</td>
<td>1 (0 to 3)</td>
</tr>
<tr>
<td>Change</td>
<td>0 (0 to 0)</td>
<td>-1 (~2 to 0)</td>
</tr>
<tr>
<td>Change in score, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decrease</td>
<td>56 (16.4)</td>
<td>70 (61.4)</td>
</tr>
<tr>
<td>Increase</td>
<td>47 (13.7)</td>
<td>19 (16.7)</td>
</tr>
</tbody>
</table>
Table 6. Responses to each SCOFF item at the last assessment for members who answered “yes” to the item at baseline. Results are reported for each group.

<table>
<thead>
<tr>
<th>Item</th>
<th>Nonelevated SCOFF score (n=342), n/N (%)</th>
<th>Elevated SCOFF score (n=114), n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you make yourself sick because you feel uncomfortably full?</td>
<td>1/8 (12.5)</td>
<td>7/8 (87.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9/23 (39.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14/23 (60.9)</td>
</tr>
<tr>
<td>Do you worry that you have lost control over how much you eat?</td>
<td>21/44 (47.7)</td>
<td>23/44 (52.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49/102 (48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53/102 (52)</td>
</tr>
<tr>
<td>Have you recently lost more than one stone (14 lb) in a 3-month period?</td>
<td>2/9 (22.2)</td>
<td>7/9 (77.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4/22 (18.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18/22 (81.8)</td>
</tr>
<tr>
<td>Do you believe yourself to be fat when others say you are too thin?</td>
<td>14/35 (40)</td>
<td>21/35 (60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50/77 (64.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27/77 (35.1)</td>
</tr>
<tr>
<td>Would you say that food dominates your life?</td>
<td>1/6 (16.7)</td>
<td>5/6 (83.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34/60 (56.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26/34 (43.3)</td>
</tr>
</tbody>
</table>

Figure 1. SCOFF score over months in care for members with elevated SCOFF scores. Individual scores are indicated by dark blue markers, and the linear model fit is indicated by the purple line.

**Discussion**

**Principal Results**

Using retrospective analyses of data collected from adolescents participating in care with Bend Health Inc, the purpose of this study was to determine whether mental health care with a collaborative care DMHI is associated with a reduction in disordered eating behaviors while also accounting for demographic and mental health symptom covariates. Disordered eating scores decreased throughout care with the DMHI for 61.4% (n=70) of adolescents with elevated disordered eating scores at baseline, and longer participation was associated with greater reductions in scores. Elevated disordered eating behaviors at baseline were associated with female sex and elevated mental health symptoms. This study provides preliminary evidence that mental health care with a DMHI may positively affect disordered eating behaviors.

Over the course of care with the DMHI, 61.4% (n=70) of adolescents with disordered eating behaviors exhibited fewer problematic eating behaviors throughout participation. Further, each additional month in care was associated with larger improvements in disordered eating behaviors, even while controlling for elevated mental health symptoms. These findings, which suggest the effectiveness of a pediatric DMHI in mitigating symptoms of disordered eating, are especially timely, as many youths with EDs and disordered eating symptoms are unable to access adequate care via in-person modalities—a lingering effect of the COVID-19 pandemic and current mental health crisis in youths [34,35]. Moreover, these results suggest that collaborative care DMHIs may mitigate disordered eating behaviors.
behaviors in a relatively short time frame, with the change in disordered eating behaviors assessed within this study sample after just 2.23 (IQR 1.1-3.6) months in care. Given the complex and chronic nature of disordered eating behaviors and EDs, it is understandable that health care professionals prefer team-based, integrated behavioral care (eg, the collaborative care model used by the DMHI in this study) for the identification and treatment of EDs [36-39].

Compared to those without disordered eating behaviors, adolescent members with disordered eating behaviors were more likely to be female than male. Extant literature suggests a similar trend, namely, that females tend to report more disordered eating behaviors and higher rates of ED diagnosis than their male peers [40,41]. However, this does not mean that disordered eating is not a problem among males. In this study, 19.5% (n=33) of adolescents flagged with problematic eating behaviors were male. Disordered eating behaviors may go underreported and unrecognized in males due to stigma [42], and ED diagnostic criteria may not accurately capture disordered eating behaviors that are more prevalent in males than females, such as preoccupation with gaining muscle mass and fear of losing weight [43,44]. Mental health providers and practitioners in both traditional and web-based modalities should continue to screen for disordered eating behaviors among all clients, regardless of sex and gender, while paying particular attention to muscle dysmorphia and excessive exercise [41].

We also found that adolescents with disordered eating behaviors also had higher rates of elevated anxiety and depressive symptoms than their peers who were not flagged with disordered eating behaviors. Moreover, having elevated anxiety and depressive symptoms at baseline was a significant predictor of more severe disordered eating behaviors among youths flagged with disordered eating behaviors. In children, adolescents, and adults, internalizing disorders such as anxiety and depression frequently co-occur with disordered eating behaviors and EDs [45-47]. This study sample exhibited higher rates of anxiety and depressive symptom comorbidity than previously reported. This is to be expected, given that all children and adolescents in this study were treatment-seeking, whereas previous estimates have included both treatment- and nontreatment-seeking individuals [45,46]. Notably, 95.9% (n=162) of members flagged with disordered eating behaviors participated in care programs other than the body image care program (namely, the anxiety and depression programs), and this group exhibited improvements in disordered eating behaviors, nonetheless. This finding adds to the body of literature suggesting that the overlap between disordered eating behaviors and internalizing problems such as anxiety and depression may have similar underlying constructs and thus may warrant similar treatment [45,48]. Further research is necessary to study the effectiveness of DMHIs in directly addressing disordered eating behaviors as well as treating comorbid EDs and internalizing problems among young people.

Limitations and Future Directions

While our findings provide compelling evidence that participation in a collaborative care DMHI may be associated with improvements in disordered eating behaviors, there are several limitations. The SCOFF assessment has attracted criticism for lack of sensitivity and accuracy and other questionnaires. The Eating Attitudes Test [49], for instance, may indeed allow for more fine-grained assessment and diagnosis of eating behaviors. However, the SCOFF has been used and validated extensively as a screener and assessment of common EDs among adolescents [27,50,51]. Moreover, the SCOFF has a high correlation with other validated disordered eating surveys [52] and has also shown acceptable convergent validity when compared to clinical interviews [53].

It should be noted that the SCOFF questionnaire includes an item that may not be appropriate for adolescents. Adolescence is a time of puberty and rapid growth, and therefore, losing 14 pounds may indicate extreme weight loss and a significant health issue (more so than the other SCOFF items). Further, if an adolescent responds “yes” to this item at one time point, it is unlikely that they will respond “yes” in the next several months. Therefore, we investigated whether our findings were driven or altered by the inclusion of this potentially problematic item. We found that few participants in the study sample responded “yes” to this item (n=44, 7.3% at baseline), and follow-up analyses revealed that our longitudinal results did not substantively change when this item was removed from the calculation of SCOFF scores (Multimedia Appendix 2). Thus, the inclusion of this item did not confound the primary findings reported here.

Although we evaluated cross-sectional associations between improvements in disordered eating behaviors and internalizing problems at baseline, future research should explore longitudinal and bidirectional associations between disordered eating symptoms and changes in mental health symptoms among youths involved in collaborative care DMHIs. Extant research regarding causal associations between disordered eating behaviors and other psychiatric symptoms among those engaged in treatment has yielded heterogeneous results [54-56], but no relevant studies to date have been conducted among youths involved in a collaborative care DMHI.

Concluding Remarks

Young people today are exhibiting increasing rates of disordered eating behaviors and EDs alongside other mental and behavioral problems such as anxiety and depression. Collaborative care DMHIs have the capacity to mitigate the growing mental health crisis by providing holistic and evidence-based care that is more accessible and scalable than traditional modalities. The findings from this study suggest that participation with a collaborative care DMHI such as Bend Health Inc may be beneficial in the reduction of disordered eating symptoms in adolescents. Future studies, particularly those bolstered by improved measurement of eating behaviors over time and with a larger and more diverse cohort of youths, are paramount to establishing the effectiveness of collaborative care DMHIs as an evidence-based provider of care for those with problematic eating behaviors and EDs.
Data Availability
The data sets generated and analyzed during this study are not publicly available because this would violate Bend Health Inc’s privacy. However, aggregated and anonymized data are available from the corresponding author on reasonable request.

Authors’ Contributions
LGH wrote the first draft of the paper and led further editing and paper refinement. DL-S performed formal data analysis and generated tables and figures for the paper. ABB wrote the first draft of the paper. JH supervised all paper writing and data analysis. All authors were involved in study conceptualization, development of methodology, and reviewing and editing the paper.

Conflicts of Interest
All authors are employed by, contracted with, or volunteering for Bend Health Inc, which delivered the treatment used in this retrospective study. However, authors’ employment status, salary, and any associated compensation are not dependent upon the results of their research.

Multimedia Appendix 1
Information on the race and ethnicity demographic response options, and also how these data were analyzed.
[DOCX File, 13 KB - formative_v8i1e54253_app1.docx ]

Multimedia Appendix 2
Additional analyses, and accompanying results, to determine SCOFF response patterns of participants. Follow-up analyses were performed with the response to item 3 excluded from calculations of total SCOFF score.
[DOCX File, 16 KB - formative_v8i1e54253_app2.docx ]

References


Abbreviations

BCM: behavioral care manager
DMHI: digital mental health intervention
DSM-5-TR: Diagnostic and Statistical Manual of Mental Disorder, Fifth Edition, Text Revision
ED: eating disorder
GAD-7: Generalized Anxiety Disorder-7
PHQ-9A: Patient Health Questionnaire-9 for Adolescents

©Landry Goodgame Huffman, Darian Lawrence-Sidebottom, Aislinn Brenna Beam, Amit Parikh, Rachael Guerra, Monika Roots, Jennifer Huberty. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Data Representation Structure to Support Clinical Decision-Making in the Pediatric Intensive Care Unit: Interview Study and Preliminary Decision Support Interface Design

Najia Yakob¹, MSc; Sandrine Laliberté², MSc; Philippe Doyon-Poulin², PhD; Philippe Jouvet³, PhD; Rita Noumeir¹, PhD

¹École de technologie supérieure, Montreal, QC, Canada
²Polytechnique, Montreal, QC, Canada
³Pediatric Intensive Care Unit, CHU Sainte-Justine, Montreal, QC, Canada

Corresponding Author:
Philippe Jouvet, PhD
Pediatric Intensive Care Unit, CHU Sainte-Justine
3175 Côte-Sainte-Catherine
Montreal, QC, H3T 1C5
Canada
Phone: 1 514 345 4927
Email: philippe.jouvet@umontreal.ca

Abstract

Background: Clinical decision-making is a complex cognitive process that relies on the interpretation of a large variety of data from different sources and involves the use of knowledge bases and scientific recommendations. The representation of clinical data plays a key role in the speed and efficiency of its interpretation. In addition, the increasing use of clinical decision support systems (CDSSs) provides assistance to clinicians in their practice, allowing them to improve patient outcomes. In the pediatric intensive care unit (PICU), clinicians must process high volumes of data and deal with ever-growing workloads. As they use multiple systems daily to assess patients’ status and to adjust the health care plan, including electronic health records (EHR), clinical systems (eg, laboratory, imaging and pharmacy), and connected devices (eg, bedside monitors, mechanical ventilators, intravenous pumps, and syringes), clinicians rely mostly on their judgment and ability to trace relevant data for decision-making. In these circumstances, the lack of optimal data structure and adapted visual representation hinder clinician’s cognitive processes and clinical decision-making skills.

Objective: In this study, we designed a prototype to optimize the representation of clinical data collected from existing sources (eg, EHR, clinical systems, and devices) via a structure that supports the integration of a home-developed CDSS in the PICU. This study was based on analyzing end user needs and their clinical workflow.

Methods: First, we observed clinical activities in a PICU to secure a better understanding of the workflow in terms of staff tasks and their use of EHR on a typical work shift. Second, we conducted interviews with 11 clinicians from different staff categories (eg, intensivists, fellows, nurses, and nurse practitioners) to compile their needs for decision support. Third, we structured the data to design a prototype that illustrates the proposed representation. We used a brain injury care scenario to validate the relevance of integrated data and the utility of main functionalities in a clinical context. Fourth, we held design meetings with 5 clinicians to present, revise, and adapt the prototype to meet their needs.

Results: We created a structure with 3 levels of abstraction—unit level, patient level, and system level—to optimize clinical data representation and display for efficient patient assessment and to provide a flexible platform to host the internally developed CDSS. Subsequently, we designed a preliminary prototype based on this structure.

Conclusions: The data representation structure allows prioritizing patients via criticality indicators, assessing their conditions using a personalized dashboard, and monitoring their courses based on the evolution of clinical values. Further research is required to define and model the concepts of criticality, problem recognition, and evolution. Furthermore, feasibility tests will be conducted to ensure user satisfaction.

(JMIR Form Res 2024;8:e49497) doi:10.2196/49497
KEYWORDS
data representation; decision support; critical care; clinical workflow; clinical decision-making; prototype; design; intensive care unit

Introduction

Background

In the pediatric intensive care unit (PICU), clinicians are required to make clinical decisions daily regarding inpatients’ health conditions. In critical care, data access accuracy and speed are crucial for optimizing the decision-making process. However, the following are some factors that can limit the effectiveness of this decision-making process: (1) clinicians must deal with a high volume of clinical data from several sources, such as physiological monitors, laboratory systems, and caregiver notes on the electronic health record (EHR) [1], which can lead to delays in processing data and reaching a decision; (2) decisions made in intensive care units rely on clinical judgment based on the clinician’s knowledge and experience, which are variable [2]; (3) clinical uncertainty in critical care and the variability of cases may lead to inconclusive decisions [3]; and (4) clinician’s stress, lack of sleep, and multiple stimuli can interfere with decision-making.

To gather relevant information, clinicians have to go through several systems, browsing through different sections of an EHR, laboratory systems, and imaging systems. They must sort and analyze all this information based on their personal expertise and available scientific evidence before making any decision regarding patient care.

For the last few decades, the emergence of clinical decision support systems (CDSSs) has assisted clinicians in their cognitive process by combining scientific knowledge bases with patient data for personalized and adapted management [4]. The design of the CDSS can take different forms depending on the clinical needs [5]. These systems focus on specific problems in the environment where they are to be implemented and often rely on existing systems and organizational contexts [6]. Therefore, understanding the workflow of existing systems is essential for supporting the adoption and optimal use of a new system. This understanding of the existing systems helps to determine when and how CDSS will be used [7].

To develop decision support mechanisms, participative approaches have been used to optimize the representation of clinical data. Faiola et al [8] adopted a human-centered approach to design a decision support tool, which has been shown to be effective in reducing the cognitive overload experienced by users [8]. More design and integration approaches have been developed based on domain-specific characteristics and matching users’ cognitive processes [9,10]. Clinical data display optimization, including EHRs focused on patient-centered care [11] and dedicated clinical decision support tools that depend on specialized knowledge bases [12], has attracted the interest of other researchers.

End user involvement is essential to ensure optimal data representation, which can be achieved through observation of clinical activities, individual interviews, and focus groups [13]. Data visualization must be validated by clinicians to ensure that it is understandable, relevant, useful, and readily available [14,15].

In this study, we aimed to adapt the data representation structure to the clinical processes in the PICU and allow its application in various clinical care scenarios. Although most decision support research in the literature focuses on specific clinical needs, the objective of our study is to facilitate the integration of multiple CDSSs developed for specific problems and the use of such systems for specific patients while ensuring harmonized monitoring and adequate evaluation for all hospitalized patients.

Our approach involved the end users throughout the data representation implementation process, including needs identification and prototype design, to illustrate the targeted structure.

Literature Review

Clinical Decision-Making

The clinical decision-making process is based on 2 main approaches: an intuitive heuristic approach, which is triggered in uncertain or critical situations requiring rapid intervention, and an analytical approach that involves gathering and processing information before reaching a conclusive decision. In clinical practice, the decision-making process varies with the clinician’s experience, their developed cognitive model, and processed information [16]. Furthermore, to refine clinical decisions and reduce the risk of errors, clinicians rely on knowledge bases and scientific evidence to process patient-specific data [17]. This adds complexity to the cognitive process in terms of time and effort invested.

In critical care, clinical teams typically discuss patients’ status and care during handoff meetings and medical rounds. Decision-making at these times depends on the relevance and accuracy of the data presented [18]. Decision support mechanisms are increasingly integrated into clinical processes to reduce the information gap by making relevant knowledge and data readily available through computerized systems.

Use of CDSSs

CDSSs are computer-based solutions that support clinicians and health care professionals in making clinical decisions [19] by providing them with person- or population-specific knowledge and information. This information is filtered and presented in a convenient timeframe to improve the health care of individuals and promote better population health [20]. Historically, CDSSs have been used for preventive, diagnostic, and therapeutic purposes, with the primary goal of improving the quality, safety, and efficiency of patient care [21]. Depending on the context of use, these systems may include best-practice guidelines for specific conditions or suggestions based on patient clinical data [22].
A CDSS is usually supported by an inference engine that incorporates clinical practice guidelines with patient-specific data to generate tailored suggestions [4]. However, other models are increasingly used, and artificial intelligence is used to predict condition changes or deterioration [23]. Computerized systems encompass 5 common types of decision support methods for knowledge sharing to reduce the risk of error among clinicians: order sets, information buttons, data documentation forms and templates, alerts and reminders, and relevant data representation [24].

The synthetic representation of patient data is a major challenge, mainly because of variability in data sources and format, along with the integration of medical knowledge in data processing. To implement such a representation, researchers have developed integration and structure design approaches that rely on the specificities of the work domain and adapt to users’ cognitive processes [9,10].

Improved visual representation facilitates timely information access, which has a positive impact on clinicians’ performance and cognitive processes [25,26]. Therefore, selecting adequate, reliable, and relevant content and using simple and understandable messages is highly recommended. Furthermore, clinicians’ time must be optimized by providing accurate and timely information and avoiding double entries by ensuring interoperability with EHR [27]. Finally, incorporating these systems into the users’ workflow is key to optimizing their implementation [28].

Wright et al [29] developed a taxonomy of clinical decision support tools to help categorize and compare their capabilities (eg, guidelines, notification, and order edition) in both commercially available and internally developed systems [28,29]. They found that a home-developed CDSS is more likely to achieve its goals as it focuses on local needs. Few studies have been conducted on implementing CDS components in commercial EHRs [30]. A review of 9 commercial systems found variability in decision support capabilities, which shows a significant gap between vendors [31]. Most EHRs focus on patient care and limit the scope of integrated CDS components to medication safety and managing lists of patients with common characteristics. However, standalone CDSSs are continuously evolving [32]. Recent work has demonstrated the feasibility of developing a flexible platform for hosting CDSS outside a specific EHR. The authors estimated that an EHR-agnostic approach facilitated the modification and development of new features because it implies fewer technical challenges [33].

The implementation of new technologies in the PICU settings should be performed carefully. This applies to CDSSs as their potential to improve clinical outcomes depends on how they are implemented in terms of integration into clinical workflow; process fluidity; interoperability and communication with the existing clinical systems; and data collection, analysis, and display. A previous study reported that commercial EHRs lacked features required in pediatric settings and that CDSSs were mostly integrated using home-developed tools in the unit [34].

At Sainte-Justine Hospital, several research initiatives have been undertaken in the PICU to develop CDSSs for specific needs, such as assistance in the automated diagnosis of acute respiratory distress syndrome in children based on various physiological and radiological criteria [35,36], assessment of the quality of head injury care in adherence to clinical practice guidelines [37], early detection of ventilator-associated pneumonia [38], and hypoxemia diagnosis and management [39]. Unlike the commercially available CDSSs, these tools developed at the Sainte-Justine Hospital were based on local clinical needs, adapted to patient characteristics in the PICU, and developed in harmony with the existing infrastructure, including devices, data availability, and access. Using additional tools to address individual problems can cause an excessive burden on clinicians. The integration of these initiatives into a unified structure will benefit both the clinical workflow through centralized information and the patient’s overall care, as each CDSS improves accuracy by targeting specific criteria.

**Research Objective**

The purpose of this study was to collect and analyze clinicians’ needs in an academic hospital PICU in support of clinical decision-making and establish a data representation structure for easy and quick access to relevant information required for clinical care, depending on patients’ care trajectory. Our goal was to provide a customizable visual tool allowing an overview of the patient’s data depending on their health condition (eg, diagnosis, current problem, and deterioration of the human body system) to reduce information processing time and mental overload for clinicians. This tool serves as a platform for integrating CDSSs in the PICU in response to patients’ specific needs while ensuring that the clinical flow is respected. This was the first step in implementing multimodal real-time CDSSs.

This study was not intended to replace existing clinical tools (eg, EHR, laboratory systems, bedside monitors, and ventilators) because these tools remain essential sources for acquired data and form integral parts of the intensive care unit environment. The EHR represents the core of this technological ecosystem, as it covers the patient’s trajectory from admission until being transferred or discharged. During this time, clinicians (eg, intensivists, nurses, external specialists, pharmacists, and health professionals) use the EHR’s functionalities for different purposes (eg, notes, prescriptions, reports, consultations, patient assessment, and monitoring) and have access to some decision support features, such as alerts for abnormal clinical values, task reminders, prescription aid, and events notification. Although these features help clinicians in their daily work, they do not provide further assistance in specific situations or for variable diagnoses.

In addition, we believe that an independent decision support tool allows for continuous improvement and adjustment while considering local needs. To this end, we encouraged the clinicians’ involvement throughout the study.

**Methods**

**Overview**

In our approach to implementing the new CDSS structure, we opted for the standard process of implementing computerized systems, which starts with identifying end users’ needs before beginning the modeling and prototyping phase and then...
continues with performing tests to finally allow its integration into the clinical flow [40].

Our work focuses specifically on the first 2 phases of the process: identifying requirements through observation activities and interviews, followed by modeling and prototyping using design meetings. User testing will be covered in future work.

Needs Identification
To identify clinicians’ needs in terms of decision support, we first participated in a day of routine clinical activities at the Sainte-Justine Hospital PICU to understand the general workflow by observing interactions between team members and how they used clinical systems.

Following our observations, we planned interviews with PICU clinicians to understand their workflow and collect data on their needs. We approached the main categories of clinical staff in the unit, including intensivists, fellows, residents, nurses, and nurse practitioners. To achieve the target sample level (>10 participants), we used different communication channels for recruitment, namely email invitations, announcements in weekly journals, and direct contact in the unit.

We enrolled 11 clinicians, including 5 intensivists, 1 fellow, 4 nurses, and 1 nurse practitioner. Semistructured interviews lasting between 30 and 60 minutes were conducted face-to-face or remotely via a videoconferencing platform based on the participants’ preferences and availability. The interviews were recorded and transcribed by the research team. The interview guide was designed to provide an understanding of the use of existing work systems, evaluate participants’ knowledge and familiarity with decision support systems, and identify their needs and expectations regarding CDSS implementation.

Modeling and Prototyping
On the basis of data collected from the observation activities and interviews, we defined a 3-level data representation structure (ie, unit, patient, and system). This provided us with a basis for designing the first prototype. To validate the understanding of this first prototype and the relevance of the integrated functions, we held design meetings with the enrolled participants via videoconferencing. Before these meetings, the participants received a short video explanation with an evaluation survey to introduce the general functioning of the prototype and obtain their initial feedback. Our goal was to engage in interactive discussions with participants during the design meetings. To this end, we used a clinical scenario involving a patient with a severe head injury, and then we asked the participants to perform some tasks, such as sorting the patient list and assessing the patient’s health condition based on the presented data, to use the functionalities available on the prototype and to describe their understanding. Simultaneously, the participants were given the opportunity to suggest improvements for adding, removing, or correcting the represented data. A total of 5 intensivists participated in these design meetings. Depending on their availability, 3 physicians were met individually, and 2 were brought together in the same meeting.

Ethical Considerations
The Centre Hospitalier Universitaire Sainte-Justine Ethics Review Board approved this study (CER-2022-4083), and all participants signed an informed consent form before participating in the study. Consent was obtained in person, either on the first contact or the day of the interview, after receiving a positive response to the mail invitation. All original consent forms were archived at the Sainte-Justine Research Center.

Participants’ personal information (eg, name and email) was saved separately from the study data in a password-protected Excel (Microsoft Corp) file. Personal information was linked to study data using a code for each participant. The data will be kept in a secure directory in the hospital server for 7 years, after which it will be destroyed.

No personal information was used during interviews. Only the participant codes were mentioned at the beginning of the interviews. The recordings were immediately deposited in the secure directory at the end of each interview. Once listened to and transcribed, this file was saved in another folder in the same secure directory.

A CAD $10 (US $7.5) gift card was offered to participants as a gesture of appreciation for their participation.

Results
This section presents the findings from the data collection and analysis as well as the prototype designed to illustrate the proposed structure for clinical data representation.

Description of the Existing Process
The observation activities allowed us to understand the clinical workflow related to team members’ interactions and how they used the existing clinical systems.

Clinical Workflow
Overview
Figure 1 presents a typical day at the PICU. The day usually began with a handoff meeting (1) between the last medical team and the team taking over during the day, followed by a bedside visit (2) to discuss and validate the patient’s treatment plan. Subsequently, team members performed clinical interventions (3) related to their specific roles and responsibilities before handing over patient information to the next team.
Figure 1. Clinicians workflow during a typical daily shift in the pediatric intensive care unit at Sainte-Justine hospital.

We took time to observe some clinical activities, such as patient information transfer meetings and morning medical rounds.

**Handoff Meeting**

This meeting brought together pediatric intensivists or patrons, and fellows from 3 specialties: general acute pediatrics, called Pediatrics A; chronic pediatrics, Pediatrics B; and cardiac surgery, Pediatrics C. The goal was to assess the medical conditions and illness evolution of inpatients and new admissions to establish a treatment plan for the next 24 hours. Generally, patients were presented, starting with discharged patients, followed by critical or extremely ill patients, and then stable patients. For each patient, a predefined plan covered the body’s systems, including respiratory, cardiovascular, neurological, gastrointestinal, hematologic, immunologic, renal, and metabolic systems, as well as the infectious process, tegument, and musculoskeletal system. Patients were also assessed psychosocially before the medical team concluded the global assessment by proposing a treatment plan.

**Clinical Activities**

After the handoff meeting, the team began a collaborative round at the patient’s bedside to discuss the patient’s current condition with the nurse in charge. Parents could participate in discussions to complete the information and ask about their children’s condition. Once the discussion was completed, a patient status summary was presented with a proposed treatment plan, including new laboratory or imaging orders, medication adjustments, outpatient referrals, and other diagnostic or therapeutic interventions as needed. Once the plan was approved, a medical team member recorded the assessment summary by creating a new medical progress note in the patient’s record. This note included important laboratory results, vital signs, ventilation, the patient’s global evolution in the unit, and their evolution within the human body systems. For example, the neurological level included sedation and comfort assessment data, whereas the respiratory level included ventilatory parameters assessment and likely respiratory distress signs.

**Medical Round**

After the handoff meeting, the team began a collaborative round at the patient’s bedside to discuss the patient’s current condition with the nurse in charge. Parents could participate in discussions to complete the information and ask about their children’s condition. Once the discussion was completed, a patient status summary was presented with a proposed treatment plan, including new laboratory or imaging orders, medication adjustments, outpatient referrals, and other diagnostic or therapeutic interventions as needed. Once the plan was approved, a medical team member recorded the assessment summary by creating a new medical progress note in the patient’s record. This note included important laboratory results, vital signs, ventilation, the patient’s global evolution in the unit, and their evolution within the human body systems. For example, the neurological level included sedation and comfort assessment data, whereas the respiratory level included ventilatory parameters assessment and likely respiratory distress signs.

Clinicians must document all interventions in their clinical notes on the EHR. Clinical notes were entered in free text, which meant that the information structure and volume and the terminologies and expressions used differed among clinicians. Textbox 1 illustrates the variability in the medical progress notes taken while assessing patients with respiratory problems.
Examples of respiratory assessment in medical progress notes from electronic health records in the pediatric intensive care unit at Sainte-Justine Hospital illustrating the formatting variability among clinicians.

**Note 1**
- #Decadran BID (0.6 mg/kg/day) last dose for the day
- Extubation 28/04 AM
- AA
- Minimal desaturation, spontaneous resolution overnight
- Bilateral GAE, no added noise, eupneic
- Venous gas 7.37/47/25
- Last RPL 28/04 improvement

**Note 2**
- # HFNC 20 LPM FiO2 40%
- Sat 90-92% More obstruction than usual
- RR 25-30 no drawing
- Secretory + physio in progress during passage
- GAE bilaterally
- Noise transmitted bilaterally
- No labs

**Note 3**
- # Ventolin IV 3 mcg/kg/min * 7h45 this morning
- # Solumedrol 1 mg/kg q6h
- # Ketamine infusion 0.5
- BiPAP Ai 5 / Peep 8 / FiO2 21%
- Reduced EA, but improving, absence of wheezing
- Indrawing
- 7.4/34/20.8

**Use of Existing Systems**
While observing the clinical activities in the PICU, we learned about the main working tools in the unit. We mainly targeted the TVL (tableau de visualisation de lits [beds visualization table]) unit dashboard and EHR.

**TVL Unit Dashboard**
TVL is a digital display tool developed in the PICU to evaluate the unit’s capacity to receive patients and the nurses’ workloads. Besides allowing all professionals and families to easily locate a patient, as it is displayed on a large screen at the unit’s entrance, the TVL allows PICU staff to view the patient distribution, depending on the team in charge (Pediatrics A, B, and C), and identify discharged patients and new admissions [41,42]. The tool is based on an architectural representation of the units (Figure 2). It mostly contains unit management information, including (1) the patient room or bed, (2) bedside nurse allocation, (3) the team in charge, (4) room index, (5) waiting patients, (6) PICU patients summary, and (7) the legend indicating the meanings of icons (eg, room, workload, and equipment indications). The TVL contains limited information about the patient’s condition, such as ventilatory mode and circulatory support equipment, which limits its use in clinical care.
Figure 2. Tableau de visualisation de lits (beds visualization table; TVL) unit dashboard displayed at the pediatric intensive care unit (PICU) entry (names are hidden), which presents (1) patient room and information, (2) bedside nurse allocation, (3) the PICU team in charge, (4) room index, (5) waiting patients, (6) patients summary, (7) and the legends.

EHR Tool

Patient records in the PICU were managed using a dedicated critical care system known as IntelliSpace Critical Care and Anesthesia (Philips Healthcare). This system is connected to administrative modules to manage patient admissions, transfers, and discharges. It is also connected to physiological monitors for vital signs, mechanical ventilators for respiratory parameters, intravenous pumps and syringes for drug perfusion and feeding data, the pharmacy for medication prescription management, and laboratory modules for physiological monitors, intravenous pumps, ventilators, laboratory, and pharmacy systems [43]. System interoperability consolidates all clinical data from the connected systems (eg, physiological monitors, intravenous pumps, ventilator, laboratory, and pharmacy systems) into the EHR along with free text clinical notes typed by the clinicians. However, clinicians must search several sections, gather information, and analyze it to assess the patient’s condition and adjust the treatment plan. Therefore, a synthetic representation of patient data is required to guide clinicians, limit cognitive overload, and optimize the time spent collecting information relevant to decision-making.

The EHR is an integral tool and reference for clinicians in the PICU, which is used as a source of clinical data collected continuously from the patient’s environment (eg, bedside monitors and ventilators) and data collected from punctual or recurrent interventions (eg, laboratory examinations). The EHR is also used for data entry purposes to document the patient’s assessment and to add some measured values (eg, Glasgow Coma Scale and Comfort Scale scoring). Although the EHR played an essential role in the clinical workflow, the real challenge remained in the clinician’s ability to trace the required data and process it in due course [44]. Therefore, the development of a visual tool provided a targeted view of the EHR’s content without additional entry tasks for the clinicians.

Data Analysis

Interviews with participants provided insights into the information-seeking process through existing systems and allowed discussions about decision support systems in terms of familiarity with and clinicians’ expectations of such systems.

Information-Seeking

Table 1 presents the use patterns of clinical systems among participants. To gather information for decision-making, clinicians browsed through different sections in the EHR.
Table 1. Data sources within the existing systems and their use by PICU\textsuperscript{a} clinicians.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Use by participant category</th>
<th>Physician</th>
<th>Fellow</th>
<th>Nurse practitioner</th>
<th>Nurse</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHR\textsuperscript{b}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical data</td>
<td>High\textsuperscript{c}</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Vital signs trends</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Prescriptions and medication</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Scores</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Admission notes</td>
<td>Low\textsuperscript{d}</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Medical progress notes</td>
<td>Low\textsuperscript{e}</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Brief notes</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Consultants’ notes</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Laboratory system</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Imaging system</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Clinical practice guideline</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>TVL\textsuperscript{f} unit dashboard</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low\textsuperscript{g}</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}PICU: pediatric intensive care unit.
\textsuperscript{b}EHR: electronic health record.
\textsuperscript{c}High: >3 times per shift.
\textsuperscript{d}Low: 0 to 1 time per shift.
\textsuperscript{e}Medium: 2 to 3 times per shift.
\textsuperscript{f}TVL: tableau de visualisation de lits (beds visualization table).
\textsuperscript{g}The frequency becomes high when the nurse is assigned a team leader.

Clinical data from physiological monitors, laboratory systems, and intravenous pumps were categorized by body systems. Data were displayed in a set of detailed tables containing values for each category and the results of nursing observations. Clinicians must navigate all the tables to find relevant information for patient assessment. Data were collected in the same way regardless of the patient’s problem, which made it difficult to analyze and process. Considering a case of brain injury, the clinician examined the clinical indicators, including biological examinations obtained from the laboratory system and physiological parameters collected from connected devices such as ventilators and feeding pumps. These indicators were associated with the patient’s condition by going through data categories (eg, neurological, respiratory, and cardiovascular) and then refined information to obtain a synthesis to support their decision, which took time. The EHR also displayed vital signs trends for a certain period. The vital signs were fed directly from the bedside monitors which were connected to the patient. Notably, some clinicians believed that trends could be improved by facilitating access to the graphs when analyzing patient data and by ensuring that abnormal values were quickly detected.

Regarding prescriptions, a dedicated section allowed the display of detailed drug information, such as doses and administration modalities, and tracked current prescriptions or added new ones. Furthermore, the ongoing perfusion and the drug boluses could be tracked in the EHR. The nurses collected the scores and measurements important for patient assessment and entered them in the EHR (eg, the Comfort-Behavior Scale score for intubated patient assessment and delirium and Richmond Agitation Sedation Scale score for neurological assessment).

Regarding clinical notes, their use varied based on need. Admission notes describing a patient’s illness and past medical history were generally viewed when the patient was newly admitted but continued to be important as a reference point for patient outcomes during their stay. Medical progress notes were completed daily by the medical team in charge. Data were entered in free text to describe the patient’s evolution before concluding with a treatment plan. Information entry was redundant and unstructured, which complicated its processing. To monitor patient progress in these notes, clinicians often relied on the conclusion and might also rely on brief notes to learn about reassessments made during the day. External consultant notes entered by other medical specialists and health professionals were displayed in chronological order, allowing clinicians to track them by date. However, clinicians were not notified when notes were added or modified and could not use filters to facilitate searches. This meant that PICU clinicians must repeatedly check the external consultants’ sections for new updates. In addition, they must scroll through the chronological list and search through involved specialties to locate the required note.

Although the EHR gathered the necessary data for patient care, clinicians commonly used laboratory and imaging systems for a complete examination of the test results.
Regarding the TVL dashboard, clinicians used it mostly when starting their work shift to track patients and verify who was in charge (e.g., medical team and bedside nurse). Some physicians used a printed version to organize their daily schedule by taking notes directly on paper, whereas a nurse would use it, especially when assigned as a team leader, to manage the workload and resource allocation. Most of the information integrated into the TVL was not helpful in the clinical care context because it was dedicated to bed management. However, clinicians used it to help plan medical rounds.

### Decision Support: Expectations and Needs

#### Overview

The interviews conducted enabled us to assess participants’ familiarity with the CDSS and determine their expectations with respect to these systems. In Table 2, which presents the main results, it is notable that most clinicians interviewed (8/11, 73%) reported being unfamiliar with the CDSS. Clinicians’ practice experience had no impact on their level of familiarity with the CDSS. An experienced clinician does not necessarily have specific knowledge about the CDSS or its potential use in clinical practice. Among physician intensivists, those with strong knowledge related it to their involvement in research to develop clinical decision support tools. However, we found that even clinicians with little knowledge about CDSS operations could express their expectations and needs, both for their professional development and for the benefit of their patients. For clinicians, using the CDSS would optimize their cognitive decision-making process, facilitate daily work planning and managing information flow during the busiest periods, improve clinical tools efficiency, and reduce the risk of errors and oversights by providing timely and easy access to relevant data. In addition, the CDSS could promote coaching for medical and nursing interns and support newly hired staff members. Regarding patients, the CDSS helped to improve clinical care by personalizing data processing based on the patient’s physiological and pathological characteristics while adhering to scientific recommendations and clinical practice guidelines.

#### Table 2. Participants’ experience and expectations from a decision support system to be used in PICUa (N=11).

<table>
<thead>
<tr>
<th>Category</th>
<th>Physician (n=5)</th>
<th>Fellow (n=1)</th>
<th>Nurse practitioner (n=1)</th>
<th>Nurse (n=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of practice (years)</td>
<td>7-32</td>
<td>6</td>
<td>1.5</td>
<td>7-25</td>
</tr>
<tr>
<td>Familiarity with the CDSSb</td>
<td></td>
<td>100% low</td>
<td>100% low</td>
<td>100% low</td>
</tr>
<tr>
<td></td>
<td>40% strong knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20% medium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40% low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Guide cognitive process for decision-making</td>
<td>Optimize daily work planning</td>
<td>Reduce mental overload during busy workdays and agitated nights</td>
<td>Coach newly hired staff members</td>
</tr>
<tr>
<td></td>
<td>Optimize daily work planning</td>
<td>Support students in their learning process</td>
<td>Reduce the risk of error and omission</td>
<td>Optimize use of work tools and systems</td>
</tr>
<tr>
<td></td>
<td>Support students in their learning process</td>
<td>Personalize patient care management</td>
<td>Harmonize access to knowledge and data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Personalize patient care management</td>
<td>Optimize rare disease management</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optimize rare disease management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Considerations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Have a user-friendly design.</td>
<td>Use for guided decisions.</td>
<td>Avoid intrusive alerts</td>
<td>Avoid double data entry</td>
</tr>
<tr>
<td></td>
<td>Use for guided decisions.</td>
<td>Respect clinical workflow.</td>
<td>Opt for simplicity and ease of use</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Respect clinical workflow.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aPICU: pediatric intensive care unit.
bCDSS: clinical decision support system.

To ensure the efficient and successful implementation of a CDSS in their workflow, clinicians insisted on the usability and simplicity of design features while avoiding irritating factors, such as duplication of existing data entry and disruption with highly intrusive alerts. The CDSS must also fit into the users’ workflow and contribute to decisions guided and supported by clinical judgment. This meant that a clinician might find that the guidance or recommendations generated by the CDSS did not align with their conclusions based on prior knowledge and experience. In this case, if the clinician chooses to ignore the CDSS guidance, they must justify the final decision.

On the basis of the collected data, we identified 5 main themes related to clinical decision support needs (Table 3). Furthermore, we highlighted the main objectives and the means to respond to them.
### Table 3. Clinician needs for decision support capabilities in the PICU

<table>
<thead>
<tr>
<th>Themes</th>
<th>Needs expressed by participant category</th>
<th>Physicians</th>
<th>Fellows</th>
<th>Nurse practitioners</th>
<th>Nurses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient prioritization</td>
<td>Provide stability indexes</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Categorize patients according to their condition severity</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient assessment and problem tracking</td>
<td>Provide a synthetic presentation of the patient</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provide guidance on the reasoning behind patient assessment</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assist in problem recognition</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Select indicators based on the patient’s problem</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical indicator monitoring, and notification and alert optimization</td>
<td>Combine relevant data from different sources (e.g., laboratory results, physiological parameters, and monitoring)</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provide reminders of target values for the clinical indicators, and alert when abnormal values are reached</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Integrate measurements and data collected by nurses</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access and adherence to clinical practice guidelines</td>
<td>Monitor guideline adherence Alert when actions do not align with the best recommendations.</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration of decision support algorithms</td>
<td>Diagnostic aid Ventilatory weaning Vasopressor weaning</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prescription aid Transfusion</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prediction of patient deterioration</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Automating standard prescriptions (e.g., change of route and medication)</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**a**PICU: pediatric intensive care unit.

<table>
<thead>
<tr>
<th>Needs expressed by participant category</th>
<th>Physicians</th>
<th>Fellows</th>
<th>Nurse practitioners</th>
<th>Nurses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify critically ill patients</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notify changes in patient’s condition</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quickly detect abnormal changes in patient’s status</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improve access to vital signs trends and displayed graphs</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distinguish chronic and acute problems</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimize access and display of relevant information based on patient condition</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide an overview with targeted information based on the patient’s problem</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highlight important information (e.g., abnormal values, reminders, and new results notifications)</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapt notifications for quick and easy interpretation</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harmonize and facilitate access to evidence-based references</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incorporate practice support procedures</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prediction of patient deterioration</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automating standard prescriptions (e.g., change of route and medication)</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Patient Prioritization**

Clinicians expressed a need to prioritize patients according to the severity of their condition and to quickly detect any changes. A participant mentioned the value of rapid assessment of the patient status as follows:

> ...give me a quick view, actually, of whether a patient is stable vs. not stable, or critical, or an alert for a change in situation.

To help clinicians prioritize patients during handoff meetings or medical rounds, we aimed to optimize patients’ visualization with a user-friendly, interactive, and customizable display while adding stability indexes according to the patients’ conditions.

**Patient Assessment and Problem Tracking**

- **Patient portrait:** Clinicians were looking for a synthetic presentation of patient’s data to optimize clinical assessment, as expressed by 1 participant as follows:

  > If we would be able to make a patient Dashboard with a synthetic presentation of the different elements, passing some of the elements that the electronic record should do to us, but that doesn’t do too much and that considers the temporality, that considers these important clinical elements and that are in real time, or at least close.

- **Problem monitoring:** Clinicians must recognize patient problems to guide and facilitate data analysis and monitor patient outcomes. This was highlighted by a participant as follows:

  > ...help me more finely, when I’m on rounds or when I’m assessing a patient, help me in my reasoning or in my diagnosis or in my assessment of the patient. At that time, to have a more accurate view of the patient’s condition, for conditions that are complex.

Participants highlighted the importance of monitoring patients according to their condition and early detection of problems, as follows:

> (...)Depending on the pathology of the patient, it would be nice if it was the thing that detects it on its own, if the patient has respiratory distress.
Our goal was to optimize the evaluation of the patient’s current state through a synthetic presentation by selecting relevant clinical information for the patient’s assessment and improving the patient’s progress monitoring in intensive care. The complete clinical data remained accessible in the EHRs for detailed analysis.

Clinical Indicators Monitoring and Notification or Alert Optimization

Although EHRs enable monitoring of clinical indicators across several sections, including medical and nursing progress notes, clinicians believed that notes could be optimized with automatic data extraction and updates. One participant stated as follows:

> Notes are worth what they are worth, there are people who write good notes, there are people who don’t write good notes. I think there’s a lot of copy-paste. So sometimes you go back into the notes and you’re going to see the exact same thing at 5 days online in different sections. It's the conclusion that changes a little bit.

Alerts were also crucial for early detection of changes in the patient’s disease course. One participant said:

> It helps, the little logos come out quicker. It helps identify what's more abnormal, more quickly.

The goal was to monitor indicators according to the body system or problems associated with the system, maintaining the same structure used in presenting and assessing patients during clinical activities and within the EHR. Customizing notifications facilitated data processing and intervention planning. This could be achieved by opting for simple color codes, avoiding intrusive alerts, and duplicating information.

Access to Evidence-Based Recommendations

Clinicians must incorporate evidence-based recommendations and guidelines into their decision-making processes. Easy and standardized access to scientific databases is essential. One participant explained the need for clinical practice guidance as follows:

> I think we could do better; to see the number of variations for a problem is that there’s not a lot of scientific rigor. Could this compensate for things that are done by repetition, by reflex without foundation, and that it would be more supervised or with a better scientific basis? Probably.

The aim was to optimize access to clinical practice guidelines either by integrating the recommendations directly into data analysis through reminders and suggestions or providing direct access to scientific databases.

Integration of Decision Support Algorithms

This involved development work to associate knowledge bases with patient data using inference engines or artificial intelligence algorithms. Prediction of patient condition deterioration, diagnosis support, prescription support, and sedation weaning were among the main expressed needs.

These algorithms aimed to guide clinical decision-making. Their integration could be performed at diagnostic, therapeutic, and preventive levels. This study aimed to provide a structure susceptible to accommodating new algorithms and decision support features. Therefore, according to the first reflections, we intended to associate these algorithms with body systems or the patient’s problems.

Data Structure

To structure emerging elements while adapting data representation to the clinical workflow, we opted for a 3-level system structure (Figure 3). These levels mainly reflect the first 3 identified themes, whereas the themes related to clinical practice guidelines (#4) and decision support algorithms (#5) can be attached to different levels, depending on the context of use and the problem being addressed.

Figure 3. The resulting structure for data representation in the pediatric intensive care unit (PICU) with 3 levels of abstraction. CDSS: clinical decision support system.

![Diagram showing Unit: customized view of PICU patients, Patient: synthetic dashboard, System: clinical data classification, Prioritize PICU patients, Assess patient’s condition, Monitor clinical indicators]
1. **Unit level (patient prioritization):** The first level provided a global view of patients in the PICU, with various display modes allowing efficient management of patient lists and easy identification of those who were unstable and helping to plan bedside activities. Adding stability indices helped to categorize and prioritize patients based on their condition severity.

2. **Patient level (patient assessment):** For each patient, clinicians could access this second level to obtain a quick overview of the patient’s condition and better understand the underlying cause of their instability. The patient’s synthetic presentation allowed clinicians to assess the patient’s status based on current problems and probable complications and to track important events. Eventually, incorporating guidance and evidence-based recommendations would be pertinent at the second and third levels.

3. **System level (indicator monitoring):** The third level was intended to align with the clinical flow by assessing patients according to their body systems (e.g., neurological, respiratory, cardiovascular, renal, gastrointestinal, hematological, immunologic, infectious, and musculoskeletal tissue systems). This allowed monitoring of the degree of alteration of the system based on associated clinical indicators. Moreover, this level aimed to integrate clinical decision support algorithms developed in response to specific problems (e.g., evaluation of head trauma management associated with the neurological system).

### Prototype Design

#### Overview

Using the defined structure for clinical data representation, we designed interfaces corresponding to the 3 levels of the structure. Subsequently, design meetings allowed us to adapt the design and integrate, early in the process, the necessary adjustments to meet the end users’ needs. This section presents the design and adjustments of the prototype. Each interface included a targeted functionality in response to the objectives of the associated level.

We used a brain injury care scenario to illustrate the functioning model of the preliminary prototype, knowing that a patient with a severe traumatic brain injury requires attentive monitoring that involves clinical data from different devices (e.g., mean arterial pressure, oxygen saturation, temperature, end-tidal carbon dioxide, and brain tissue oxygenation), laboratory results, ongoing sedation, and medication treatment, along with the interpretation of imaging examinations. To assess the patient’s status and, therefore, adjust their care plan, clinicians should be able to categorize the patient according to a severity scale, identify whether there is a risk of deterioration (e.g., ischemia and hyperemia), define optimal mean arterial pressure in the context of brain injury, and quickly recognize abnormal values depending on the patient’s profile. We expected that this would help clinicians to focus on pertinent details and support the prediction of changes in the patient’s condition.

**Unit Level**

The objective of this level was to visualize all inpatients in 2 display modes and introduce the concepts of stability and system alteration.

The list-mode display (Figure 4) allowed clinicians to select patients by service (Pediatric A, B, or C) or to create their personalized list (My Patients) by adding patients under their responsibility. Every patient on the list was identified with a bed number, name, age, weight, length of stay, and diagnosis. It also included scheduled tests or procedures that required off-unit transportation to assist the clinician in scheduling bedside interventions. Stability indices were added to help clinicians prioritize patients and plan their interventions for the day. These indices included a list of altered body systems and the status to categorize patients according to their conditions: critical, watcher, stable, or discharged. We used red alerts to indicate a severe alteration or criticality level and orange alerts to indicate a moderate level.
**Figure 4.** Level 1 interface in the preliminary prototype: patients list. This figure includes lists management (1.1.1), patient identification (1.1.2), and stability indices (1.1.3).

Furthermore, level 1 provided an architectural view of the unit (Figure 5) inspired by the TVL, which was adapted to assist clinicians in planning medical rounds and bedside interventions. Clinicians could customize the display to view their patients or all inpatients.
Figure 5. Level 1 interface in the preliminary prototype: unit architectural view. This figure includes color-coded boxes for patients according to their stability (1.2.1), the team in charge (1.2.2), medical team location (1.2.3).

The stability indicators display observed the same color code for the boxes: red for the critically ill and orange for less acuity. This view allowed us to see the list of available caregivers with their contact numbers and the nurses responsible for the patient’s bedside. Notably, geolocation of the team’s location during medical rounds could help plan clinical interventions. For example, a clinician who needed to join the medical round for a specific patient could check this interface to plan his or her next tasks to match the team’s arrival at the patient’s bedside.

To assess a specific patient, clinicians could select the patient from the patient list or switch to the TVL view and access synthetic data presentation at the second level.

**Patient Level**

Continuous monitoring of inpatient progress was central to clinical activities in the PICU. Therefore, the patient level (Figure 6) was incorporated into the prototype design to facilitate the evaluation of the patient status and progress during their stay in the unit. The second level provided an overview of the patient’s active problems and likely risks based on monitoring relevant indicators.
Figure 6. Level 2 interface in the preliminary prototype. In this figure, we focused level 2 on a brain injury case scenario. The features include access to personal list (2.1), patient identification and vital signs (2.2), navigation menu (2.3), primary and secondary diagnosis (2.4 and 2.5), patient’s active problems (2.6), problems under surveillance (2.7), important events (2.8), and access to clinical systems (2.9).

Clinicians could easily return to their personalized lists and search for patients. Color-coded notifications indicated the number of critical patients (red) and watcher patients (orange). The patient identification zone included demographic data, initial diagnosis, and length of stay in the unit. The same zone displayed vital signs in real values, with possible access to trends observed in the last few hours. The left menu allowed quick navigation between levels 1 and 2 and through the body systems at level 3. This interface provided information about the patient’s primary diagnosis, with the last revision date, and allowed clinicians to access a direct link to the UpToDate (Wolters Kluwer) knowledge base [45], which was widely used for medical decision support. Secondary pathologies and patient history were also listed and allowed sorting by body systems. Furthermore, active problems were displayed and an interpretation of abnormal indicator values were provided, with reminders for target values, to facilitate the recognition of the patient’s problems. Finally, problems under surveillance were shown to guide clinicians in patient care by targeting probable complications. Decision support systems could be incorporated into this level and connected to a patient’s problem. For example, a patient with respiratory failure (a medical problem) could have a CDSS for the early diagnosis of acute respiratory distress syndrome [36] and another for the management of mechanical ventilation if diagnosed with acute respiratory distress syndrome [46]. This level allowed the tracking of significant events, such as procedures performed in the operating room, specific investigations, and consultant visits. When needed, clinicians could search for additional information by directly accessing clinical applications and systems, which could be related to prescription history, treatment plans, or recent imaging or laboratory tests.

Clinicians could visualize clinical indicators on the third level to closely monitor these indicators related to patient problems (refer to system level section).

System Level

The third level (Figure 7) was designed to display groups of indicators related to human body systems and to access decision support tools developed for specific problems involving these systems. Our goal was to prioritize indicators to be monitored based on body system alterations while retaining the ability to add indicators from other systems to refine clinical decisions. In Figure 7, we included indicators of the brain injury care scenario and added a visualization tool to assess clinician adherence to the clinical practice guidelines. The development of a visualization tool will be subject to further research.
The personalized patient list could also be accessed at this level. The patient identification zone included demographic data, initial diagnosis, length of stay in the unit, and data related to patient progress in the care trajectory. For a head injury, clinicians could assess the global adherence to clinical practice guidelines and follow, through trends, changes in the patient's neurological status and Glasgow score. This zone also provided access to the last computerized tomography scan performed.

As in the previous level, the navigation menu allowed users to browse between different levels and different systems at the third level. Altered systems were easily identified using simple color-coded signs (e.g., red for highly severe indicators and orange for less severe indicators). The first group contained specific indicators related to the neurological system; this area allowed clinicians to evaluate adherence to guidelines for brain injury indicator monitoring and management. Abnormal values were systematically displayed, with access provided to trends observed in the last few hours. Furthermore, clinicians could view trends in normal indicators or add other neurological indicators not directly related to head injury care. The interface also allowed users to view indicators belonging to other systems but related to the patient's problem. For example, surveillance of a patient with a head injury is not limited to neurological indicators but covers variable indicators, such as cardiovascular and respiratory indicators. Other groups were included for bedside monitoring. In addition, the interface enabled data display by date or time range to optimize clinical indicator monitoring.

**Discussion**

**Principal Findings**

In this study, we took the first steps to develop a decision support structure that responds to clinician needs in the PICU. We analyzed the existing situation to evaluate current needs, which led us to develop a 3-level data representation structure, with the first level aimed at prioritizing inpatients based on the severity of their conditions, the second level providing an overview of the patient's condition and evolution, and the third level allowing close monitoring of clinical indicators related to a specific problem or human body system. From this perspective, the third level was intended to support CDSS integration as developed in response to specific care management needs related to the patient’s condition. In subsequent steps, there will be a testing process involving end users to validate the usability and performance of the designed prototype.
Our goal was to create a system based on the proposed representation and eventual CDSS integration. It is important to note that this system is not intended to replace EHRs designed for documenting patient care or any other existing systems. However, its use should help clinicians prioritize their interventions according to patient’s needs, which could be applied to the handoff meetings while discussing inpatient conditions and planning next-shift interventions. Furthermore, the tool could optimize clinicians’ cognitive processes by readily accessing relevant information when needed, such as for patient presentation during medical rounds, for fast checks on patient status and detection of any changes. In addition, the display of the clinical indicators could be personalized to suit the user’s preference and optimize clinical monitoring by allowing an adequate and efficient classification of indicators either by the human body system or by patient problem, which helps to contextualize data evolution.

Limitations

Although the features presented in the Principal Findings Section are generally appreciated by the clinicians, they, nonetheless, remain prudent regarding the following concepts. The first concept is patient criticality assessment, knowing that criticality could be linked to variable factors, such as a combination of a patient physiological profile, care required, and intensity of that care [47], perception related to patient prognosis, illness progression and response to treatment [48], and severity scores used to measure deviations observed in certain groups of physiological variables [49]. The second concept is problem progression, which could be difficult to track because information at the start and end of a problem is not always accurate. Although a change is usually identified by a deviation from normal or expected values, it ultimately depends on the patient’s progress in their care trajectory [50]. The third concept is that some problems affect multiple body systems and certain specific indicators related to such problems [51]. This requires classifying the indicators by problem and defining abnormal variations for each indicator according to the patient’s physiological and pathological profiles. For example, the mean arterial pressure indicator is related to the cardiovascular system, but for a patient with a head injury, this indicator directly affects cerebral perfusion; therefore, its monitoring is also linked to the neurological system. Furthermore, the thresholds for this indicator may vary with the patient’s age and illness history.

Clinical judgment is crucial for patient assessment and decision-making in critical care. This judgment varies among clinicians and relies on each clinician’s ability to synthesize relevant clinical data, which is not easy to model.

Analyses of these factors will eventually help us to optimize our data representation model in terms of the connections between problems and human body systems. In addition, identifying the factors that influence the progression of problems will help in predicting the deterioration of a patient’s condition and preparing an appropriate intervention.

In our study, we initially envisaged a sample of 30 participants (6 physicians, 4 fellows, 3 residents, 2 specialized nurse practitioners, and 15 nurses) to have a better representation of the targeted population. However, the desired sample was not achieved owing to the limited availability of PICU staff and their high workload during the project period, which was during the Covid-19 pandemic. A total of 11 participants could participate in the interviews. Only 5 (45%) physicians participated in the design meetings during the second phase of the project. Through the design meetings, we could improve the prototype design. However, we were unable to test the final version with the clinicians. Therefore, we intend to conduct usability tests afterward to identify potential issues and ensure that end users are satisfied with the resulting prototype. Future work should also investigate the integration of the prototype into the clinician’s workflow. Although the prototype intended to synthesize relevant clinical data from other sources into a consistent view, it could increase the clinician’s workload by adding another technological tool to consult the patient’s condition. This requires careful consideration of the tools’ interoperability to follow the clinician’s role.

Conclusions

This study provided a clinical data representation structure to support PICU clinicians in their decision-making process and to assist them in optimizing inpatient care management. An observation of clinical activities and interviews with participants allowed us to identify the current needs for decision support. Through an analysis of collected data, we created a structure with 3 levels of abstraction to facilitate patient prioritization, assessment, and monitoring. A prototype was designed based on the main structure and then presented to the participants to obtain feedback for improvement. Notably, the functionalities integrated into the prototype mainly met the clinicians’ expectations regarding information relevance and classification. Adjustments were made to the data representation following the design meetings with the participants. However, further tests will be conducted to ensure the tool’s usability.

To enable the deployment of the proposed decision support structure and its integration into the clinical workflow in the PICU, further analysis and development are needed to establish patient stability indices, automate problem recognition, and define the indicators associated with each body system and the respective alteration thresholds.

Acknowledgments

The authors would like to express their gratitude to the Centre Hospitalier Universitaire Sainte-Justine and the Centre Hospitalier Universitaire Sainte-Justine Research Center for their valuable support and to the caregivers for their participation. Furthermore, the authors would like to thank Cecilia Buratti and Nadia Roumeliotis for their assistance.
No generative artificial intelligence tool, including ChatGPT (OpenAI Inc), was used to draft questions for the interview guide or to write the manuscript.

This study was funded by a chair from Sainte-Justine Hospital and the Ministry of Health of Quebec (PJ) and a cochair from Fonds de Recherche du Québec Santé (RN and PJ). PJ earns a salary from Fonds de Recherche du Québec Santé.

Data Availability
Data sharing is not applicable to this study as no data sets were generated or analyzed during this study.

Authors’ Contributions
NY performed the data collection. SL and NY performed the prototype and design validation with participants. PJ, RN, and PDP supervised this project. NY drafted the manuscript. The paper was thoroughly reviewed by all authors, and they approved the final submitted research manuscript. They also assumed complete accountability for their contributions and ensured the academic integrity of their work.

Conflicts of Interest
PJ received salary from the ministry of Health of Quebec, Fonds de recherche en Santé du Québec and Ste-Justine Hospital to conduct the research and to finance the research.

References


Abbreviations

CDSS: clinical decision support system
EHR: electronic health record
PICU: pediatric intensive care unit
TVL: Tableau de visualisation de lits (beds visualization table)
Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Comparison of Blended Learning With Traditional Dermatology Learning for Medical Students: Prospective Evaluation Study

Cristiana Silveira Silva¹, MD, PhD; Cidia Vasconcellos², PhD; Murilo Barreto Souza³, PhD; Juliana Dumet Fernandes¹, PhD; Vitoria Regina Pedreira de Almeida Rego¹, MD

¹Department of Dermatology, Federal University of Bahia, Salvador, BA, Brazil
²Department of Dermatology, University of São Paulo, São Paulo, Brazil
³Department of Ophthalmology, Federal University of Bahia, Salvador, Brazil

Corresponding Author:
Cristiana Silveira Silva, MD, PhD
Department of Dermatology, Federal University of Bahia
Av. Milton Santos, s/nº - Ondina
Salvador, BA, 40170-110
Brazil
Phone: 55 713362850
Email: silveira.cristiana@gmail.com

Abstract

Background: Novel internet-based applications and associated technologies have influenced all aspects of society, ranging from commerce and business to entertainment and health care, and education is no exception. In this context, this study was designed to evaluate the impact of a dermatology e-learning program on the academic performance of medical students in dermatology.

Objective: The aim of this study is to develop a dermatology blended-learning course for undergraduate medical students, evaluate the knowledge gained by students exposed to this course, and compare the results to those of traditional teaching methods.

Methods: In this prospective study, we evaluated the performance of fourth-semester medical students at the Federal University of Bahia, Brazil. Students who had been in their second year of the medical course in 2019 were considered the control group, while students in their second year in 2020 were considered the blended or hybrid group. The first group attended traditional classes, using printed material (books and handouts), while the second group used our web-based course and e-book as a supplement in a hybrid web-plus-traditional fashion. Neither participants nor evaluators were blinded. The students in both groups were subjected to the same pre- and postcourse face-to-face, multiple-choice, paper-based evaluations, and we compared their performances. The content of the classes was the same for both groups. All didactic activities were developed by a team of certified dermatologists and professors from the university.

Results: A total of 129 students were selected and divided into 2 groups: the control group (n=57) and the hybrid group (n=72). The precourse tests did not indicate any difference between the control group (mean score 2.74, SD 1.25) and the hybrid group (mean score 3.2, SD 1.22 SD; \(P > 0.05\)). The hybrid group had better final-term grades (mean 8.18, SD 1.26) than the traditional group (mean 7.11, SD 1.04). This difference was statistically significant (\(P < 0.05\)).

Conclusions: This study explores pedagogical possibilities in the field of dermatology teaching for medical school students. The results suggest that the performance of undergraduate students who attended the course with additional e-learning material was superior when compared to the performance of those who participated in the traditional course alone.

(JMIR Form Res 2024;8:e49616) doi:10.2196/49616

KEYWORDS
dermatology; distance education; distance learning; e-learning; medical education; undergraduate medical education

Introduction

During medical school, dermatological teaching in various countries, including the United Kingdom, is usually restricted [1,2]. Students are exposed to the topic as part of short-term internships or as an optional discipline [1,2]. In some institutions, dermatology disciplines are not even offered [1,3]. Published surveys in different countries have demonstrated that
the amount of time devoted to dermatology in the medical student curriculum represents only 0.24%-0.3% of the 4 years of study [1]. In the last few years, even after curricular reformulation, the time devoted to teaching dermatology has decreased or remained the same [4-6]. McCleskey et al [4] found that only 10% of medical schools require a clinical dermatology rotation and that 93% of institutions offer dermatology as an elective rotation, usually a 4-week clerkship.

The current available time for dermatology training in medical schools worldwide is insufficient to learn about the various cutaneous diseases that students are likely to encounter in their future medical practice [5,7]. In view of this reality, the use of technologies that are able to optimize learning in dermatology has a great impact.

A meta-analysis evaluating the efficacy of educational interventions that improve diagnostic dermatological skills found that a blended curriculum that integrates multiple modalities of clinical dermatology teaching may be the most effective approach to meeting learning objectives [3]. The results observed by Lujan and DiCarlo [8] showed that first-year medical students learn through a variety of learning styles, with only 36.1% preferring a single way of acquiring new information. In recent years, we have noticed a growing interest among researchers in using new technologies to improve medical education [9-11]. The use of e-textbooks, podcasts, anatomical models, and virtual and interactive 3D computer models has positively impacted the educational experience of medical students [9]. Students exposed to interactive technology tools during their learning period demonstrate significant improvement on their performance tests [11].

Despite existing evidence that web-based teaching tools associated with interconnected content, when carefully selected, can assist the learning process, conventional teaching methods are still mainstream in medical teaching [12-14]. Teaching is mainly conducted in the form of hall lectures and laboratory sessions [1,12]. Despite large investments, there is a lack of sufficient evidence to support the effectiveness of digital interventions in the education of health professionals [15].

This study explores some pedagogical possibilities in the field of dermatology teaching for medical school students. It evaluates the use of web-based tools and an e-book developed specifically for this purpose, explores their impact on medical students’ learning, and compares this form of learning with traditional learning.

Methods

Overview

In this paper, we analyze the impact of web-based teaching tools on the performance of medical students at the Federal University of Bahia (UFBA), Brazil, and compare the results with those of traditional learning. Hence, we conducted a prospective study including medical students with computer literacy in the fourth semester at UFBA who were studying dermatology between June 2019 and June 2020. All the content was set in and developed in Brazil.

Students who had been in their second year of the medical course in 2019 were considered the control group, while students in their second year in 2020 were considered the blended group. The students were randomly allocated into the control or blended groups, and neither participants nor evaluators were blinded. All didactic activities were developed by a team of certified dermatologists and professors from UFBA.

All students participated in face-to-face activities. The classes included patient care in a general dermatology outpatient clinic. During the care, dermatological physical examination findings were emphasized, and the students were instructed to identify patients’ skin lesions and describe them according to the teaching material provided.

In the control group (traditional learning), after treating patients, students participated in an expository class structured into eight modules: (1) semiology, (2) leprosy, (3) syphilis, (4) atopic dermatitis, (5) skin viroisis, (6) pyodermitis, (7) superficial mycosis, and (8) skin cancer. Doubts about the modules were clarified on this occasion.

The hybrid activities were composed of 5 distinct stages. In the first stage, we made a photographic record of patients who had dermatological lesions during a medical consultation held at the dermatology outpatient clinic at UFBA. In the second stage, we wrote a book (Manual of Dermatology [16]) using the cases cataloged in the first stage. During the third stage, we planned and prepared the web-based course according to predetermined modules. For each module, a video lesson was made available, lasting an average of 30 minutes, and the Camtasia (TechSmith) program was used for this activity. The video lessons were formatted and published on the Moodle (Moodle HQ) platform. The fourth stage comprised an e-learning module that included an 8-week course administered simultaneously with face-to-face classes.

The students in both groups were subjected to the same pre- and postcourse face-to-face evaluations, and their performances were compared. A total of 40 multiple-choice questions were written in accordance with the recommendations of the National Council of Medical Examiners to compose the pre- and postcourse exams [17]. To evaluate the validity of the content, 2 independent dermatologists examined all questions. The subject of the tests was chosen in accordance with the British Association of Dermatologists’ Undergraduate Curriculum [18].

Students in the control and hybrid groups received identical evidence-based content, and the courses had the same 8-week duration. The e-learning course was developed using the open-source Moodle learning management system. Students logged in using individual usernames and passwords. A new text, video, and web-based discussion forum that addressed the same content as the face-to-face classes was available each week in an asynchronous mode. The students received weekly email notifications that a new class was available. In addition to face-to-face communication, students in the hybrid group could receive feedback on the discussion boards or by sending direct messages to the tutor. A 40-question multiple-choice test was given to all students in both groups before and after the courses, with scores ranging from 0 to 10.
In the fifth stage, we analyzed the results using Stata (version 13.1; StataCorp) and Microsoft Excel (version 2007; Microsoft Corporation). Initially, the studied variables were evaluated in a descriptive manner, with the data presented as mean (SD) or median (IQR). The Shapiro-Wilk test was used to test normality. Pre- and postcourse scores obtained for each group were compared (intragroup comparisons). The results obtained in the pre- and postcourse tests were also compared between the control and hybrid groups (intergroup comparisons). According to the normality test, a 2-tailed paired t test, or Wilcoxon signed rank test, was applied for intragroup comparisons and a 2-tailed t test, or Mann-Whitney U test, for intergroup comparisons. The internal consistency of the pre- and postcourse assessments was evaluated using Cronbach α coefficients.

Ethical Considerations

This study was approved by the Ethics Committee of the UFBA (1688.502) and conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all students. This work was not supported by any funding or external support, and no artificial intelligence resources were used.

Results

A total of 129 students were included in this study. The average age was 23 (SD 1.3) years, and 62 students were male. No significant differences were found between the 2 groups in relation to sex and age. The control group (n=57) used traditional classroom paper-based tool activities, while the hybrid group (n=72) used our e-learning course and e-book made specifically for this course in a hybrid web-plus-traditional fashion. Demographic data per group are presented in Table 1. All participants completed the study. The mean pretest score for the control group was 2.74 (SD 1.25) and for the hybrid group, the mean pretest score was 3.2 (SD 1.22; P>.05). The final posttest mean score was 7.11 (SD 1.04) for the control group and 8.18 (SD 1.26) for the hybrid group. The intragroup comparisons of pre- and postcourse scores obtained for each group were statistically significant (P<.05).

Table 1. Demographic data.

<table>
<thead>
<tr>
<th>Course</th>
<th>Male, n (%)</th>
<th>Female, n (%)</th>
<th>Age (years), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid group (n=57)</td>
<td>27 (47)</td>
<td>30 (53)</td>
<td>24 (1.2)</td>
</tr>
<tr>
<td>Control group (n=72)</td>
<td>35 (49)</td>
<td>37 (51)</td>
<td>22 (1.3)</td>
</tr>
</tbody>
</table>

Intergroup comparisons of the pretest scores demonstrated that there was no significant difference between the control and hybrid groups (P>.05), indicating that the baseline knowledge for each group was comparable (Figure 1). The results indicated increased scores in the hybrid group, implying the hybrid delivery method outperformed the traditional approach. A statistically significant difference in the postcourse scores between the 2 groups was achieved.

Figure 1. Boxplot with median (IQR) and range (minimum-maximum) illustrating the pre- and postcourse scores in both the conventional and hybrid groups.
**Discussion**

**Overview**

The results of this study expand earlier findings on how hybrid learning can enhance learning outcomes in medical education. Moreover, this model was found to increase the effectiveness of teaching and learning methods in improving knowledge acquisition, which is consistent with the results of several other studies.

Although dermatology is essentially a visual specialty with great potential to benefit from today’s digital technologies, the conventional way of teaching still prevails [13,19]. At our university, medical education follows a traditional lecture-based curriculum, and this was the first time that digital technology was used. Nearly all aspects of web-based education were new and had to be understood [20]. We know that e-learning offers medical schools powerful and flexible learning resources [21] and presents several advantages, including (1) increased monitoring of student progress in a simpler and more accurate manner [8]; (2) the possibility of watching classes several times at more convenient times and places [8]; and (3) allowance for more than one way of student-teacher communication by means of emails, chats, and online discussion forums [22,23]. This last point is an advantage from the students’ point of view—although it may come at the expense of teachers’ time, as it has the potential to consume more of their time when compared with classroom teaching alone (where teachers are only available during class time or office hours) [13,23]. Web-based teaching also allows medical training to continue even in difficult situations (eg, the COVID-19 pandemic), and the greatest benefit is the flexibility offered by teaching platforms [24].

While e-teaching has real advantages, as discussed above, it also comes with some drawbacks. For example, this method does not support direct contact with the teacher or the patient, which may limit the observation of certain diseases and their diagnosis. Moreover, the method is dependent on the availability of electronic devices with adequate internet access [9] and needs a highly educated, motivated, and expert core team of teachers [18]. Silva et al [13] found the same challenges in web-based courses and described a significant technical difficulty in producing educational material for distance learning. The authors also highlighted great difficulty in facilitating students’ engagement with each other and in assessing the acquisition of practical skills in dermatology.

The provision of high-quality e-learning is highly labor-intensive. Like Forde et al [25], we realized that the work spent on making web-based activities was more challenging than face-to-face teaching, especially when considering the design, organization, delivery, and engagement of participants in the discussion. A combination of both methods appears to be the best strategy [22,23,26]. In this study, these limitations were circumvented, as face-to-face activities were performed in both groups, and the students were given face-to-face contact time with both the teacher and patients seen at the clinic. Although some individuals report visual discomfort and others prefer reading a print book, both this study and the literature support the use of e-book technology in modern medical curriculum as an adjunct to traditional methods [9].

In this study, the full e-book content was available for download and could be accessed at any time, regardless of internet access. One of the main concerns about the switch to web-based lectures is the possible difficulty of lengthy readings on a screen and students’ ability to focus on reading. There is a great advantage to reading on the screen of an electronic device, as it allows for an increase in the font and size of the image, which facilitates assimilation of the content and helps individuals with reading difficulties [9]. Singer and Alexander’s [27] results indicated a clear preference of their participants for digital texts, as they generally achieved a better understanding when reading digitally [27]. However, the higher degree of satisfaction on the part of the student was not necessarily compatible with the results obtained in subsequent evaluations [24]. The vast majority approved the use of new technologies for dispensing the dermatological subject, and there were no complaints about this approach in this study.

It is currently believed that although reading on a computer screen may be more superficial and occasionally less accurate, it is the quality of the image presented to the reader that is crucial for the best use of the reading book [9]. Although there was a greater gain in knowledge in the group exposed to the distance e-learning associated with our e-book, some considerations must be made regarding the limitations and difficulties found in this study.

First, the evaluation was conducted in just one institution; ideally, more studies in multiple teaching centers with different realities from ours would be necessary for e-learning to consolidate itself as an effective form of education. Second, since the e-book was written especially for medical students, it is possible that its content made knowledge more accessible and didactic to the hybrid group, whereas the traditional group had to use renowned but conventional dermatology books. Third, we must mention the fact that the students in the traditional group spent 1 hour less per week on practical activities, totaling a reduction of 8 hours from their on-site internship due to the period spent in the in-person theoretical classes. Thus, students in the hybrid group received 8 hours more exposure to practical classes since the theoretical classes were attended at home. This difference may have favored the hybrid group in relation to obtaining better grades. In addition to better grades, increasing the time exposed to the discipline is one of the goals we strive to achieve in dermatological education.

The participants in the 2 groups had different admission years and were asked to maintain the contents and evaluations of the class confidential; however, we did not check for contamination between the 2 groups.

The field of education is destined to evolve. The professor is not the ultimate gatekeeper of definite knowledge; they also learn from students and need to incorporate feedback into the curriculum [28]. Despite this, the highest-quality clinical dermatology education will always require guided clinical exposure and feedback [18]. Innovative technologies cannot replace the need for enthusiastic and knowledgeable clinical teachers [1,28].

https://formative.jmir.org/2024/1/e49616

JMIR Form Res 2024 | vol. 8 | e49616 | p.670

(page number not for citation purposes)
Conclusion
This study explores pedagogical possibilities in the field of dermatology teaching for medical school students. The results suggest that the performance of undergraduate students who attended the course with additional e-learning material was superior to the performance of those who participated in the traditional course alone.

Data Availability
The data sets generated during and/or analyzed during this study are not publicly available due to privacy or ethical restrictions but are available from the corresponding author on reasonable request.

Authors' Contributions
CSS developed the web-based course, including an e-book and web-based classes; compiled the bibliography; and wrote the manuscript. CV developed the project and reviewed the paper. MBS participated in the creation of the web-based course and performed the statistical analysis of the data. JDF and VRPDAR helped write the book chapters, develop the web-based classes, and review the manuscript.

Conflicts of Interest
None declared.

References

Abbreviations

UFBA: Federal University of Bahia
Support for Chronic Pain Management for Breast Cancer Survivors Through Novel Digital Health Ecosystems: Pilot Usability Study of the PainRELife Mobile App

Marianna Masiero1,2, PhD; Chiara Filipponi1,2, MSc; Elisa Fragale2, MSc; Silvia Francesca Maria Pizzoli1,2, PhD; Elisabetta Munzone3, MD; Alessandra Milani4, MSc; Luca Guido5, MD, PhD; Vittorio Guardamagna5, MD; Sara Marceglia6, PhD; Roberto Prandin6, PhD; Marco Prenassi6, MSc; Annamaria Caruso6,7, MSc; Vania Manzelli7, MSc; Chiara Savino7, MSc; Costanza Conti8, PhD; Federica Rizzi8, PhD; Alice Casalino8, MSc; Giulia Candiani9, MSc; Francesca Memini9, MSc; Luca Chiveri10, MD; Andrea Luigi Vitali10, MSc; Massimo Corbo10, MD; Roberto Grasso1,2, MSc; Florence Didier2, MSc; Roberta Ferrucci1, PhD; Gabriella Pravettoni1,2, PhD

1Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy
2Applied Research Division for Cognitive and Psychological Science, European Institute of Oncology, Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy
3Division of Medical Senology, European Institute of Oncology, Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy
4Nursing School, European Institute of Oncology, Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy
5Division of Palliative Care and Pain Therapy, European Institute of Oncology, Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy
6Dipartimento di Ingegneria e Architettura, Università degli Studi di Trieste, Milan, Italy
7Nuvyta, Società a Responsabilità Limitata, Cologno Monzese, Italy
8Istituto di Management Sanitario, Milano, Italy
9Agenzia di comunicazione scientifica Zadig, Società a Responsabilità Limitata, Società benefit, Milan, Italy
10Dipartimento di Scienze Neuoriabilitative, Casa di Cura del Policlinico, Milan, Italy

Corresponding Author:
Marianna Masiero, PhD
Department of Oncology and Hemato-Oncology
University of Milan
Via Festa del Perdono 7
Milan, 20122
Italy
Phone: 39 0294372009
Email: marianna.masiero@unimi.it

Abstract

Background: Chronic pain is one of the most common and critical long-term effects of breast cancer. Digital health technologies enhance the management of chronic pain by monitoring physical and psychological health status and supporting pain self-management and patient treatment decisions throughout the clinical pathway.

Objective: This pilot study aims to evaluate patients’ experiences, including usability, with a novel digital integrated health ecosystem for chronic pain named PainRELife. The sample included patients with breast cancer during survivorship. The PainRELife ecosystem comprises a cloud technology platform interconnected with electronic health records and patients’ devices to gather integrated health care data.

Methods: We enrolled 25 patients with breast cancer (mean age 47.12 years) experiencing pain. They were instructed to use the PainRELife mobile app for 3 months consecutively. The Mobile Application Rating Scale (MARS) was used to evaluate usability. Furthermore, pain self-efficacy and participation in treatment decisions were evaluated. The study received ethical approval (R1597/21-IEO 1701) from the Ethical Committee of the European Institute of Oncology.

Results: The MARS subscale scores were medium to high (range: 3.31-4.18), and the total app quality score was 3.90. Patients with breast cancer reported reduced pain intensity at 3 months, from a mean of 5 at T0 to a mean of 3.72 at T2 ($P=.04$). The total number of times the app was accessed was positively correlated with pain intensity at 3 months ($P=.03$). The engagement ($P=.03$),
information \((P=0.04)\), and subjective quality \((P=0.007)\) subscales were positively correlated with shared decision-making. Furthermore, participants with a lower pain self-efficacy at T2 (mean 40.83) used the mobile app more than participants with a higher pain self-efficacy (mean 48.46; \(P=0.057\)).

**Conclusions:** The data collected in this study highlight that digital health technologies, when developed using a patient-driven approach, might be valuable tools for increasing participation in clinical care by patients with breast cancer, permitting them to achieve a series of key clinical outcomes and improving quality of life. Digital integrated health ecosystems might be important tools for improving ongoing monitoring of physical status, psychological burden, and socioeconomic issues during the cancer survivorship trajectory.

**International Registered Report Identifier (IRRID):** RR2-10.2196/51021

(JMIR Form Res 2024;8:e51021) doi:10.2196/51021

**KEYWORDS**
chronic pain; eHealth; cancer; decision-making; survivorship; self-efficacy; pain; oncology; health ecosystem; health ecosystems; breast; survivor; survivors; mHealth; mobile health; app; apps; applications; MARS

**Introduction**

In 2040, it is expected there will be approximately 26 million new cancer survivors in the United States, underscoring the growing importance of survivorship [1,2]. Mullan [3] defined survivorship as a process characterized by 3 different stages: acute survival, from diagnosis to active treatments; extended survival, from treatments to active-surveillance; and permanent survival, in which the probability of recurrence is low. As suggested by Mullan’s definition, cancer survivorship is a complex, multistep, dynamic process characterized by evolving needs and challenges. Pain, fatigue, and psychological distress (eg, anxiety, depression, worry, and fear of cancer recurrence) are typical long-term effects that deleteriously affect survivors’ engagement in work, personal, and social activities [4-6]. In particular, one of the most common and critical long-term effects of cancer in survivors is chronic pain. It has been linked with several physical, psychological, and socioeconomic sequelae. A study by Jiang et al [1] stated that, of 4526 cancer survivors, around 34.6% reported chronic pain, and 16.1% conveyed that it limited daily life and work activities. Notwithstanding, chronic pain throughout the survivorship trajectory is underinvestigated and undertreated [5,7,8]. Since chronic pain affects the quality of life (QoL) of patients with breast cancer [9], it is essential to design, test, and implement patient-driven interventions [10] that enable ongoing monitoring of physical and psychological health status, from the “hospital to the patient’s home,” and support pain self-management and patient treatment decisions throughout the entire clinical pathway. This might be particularly important in extended and permanent survivorship, reducing the risk of cancer survivors exiting the care system [5].

A growing body of evidence has shown that using digital health technologies integrated into dynamic ecosystems enhances the management of chronic pain by assessing patients’ physical health and psychological well-being and providing tailored interventions [11]. Digital health technologies aim to integrate various digital tools and technologies, including patient electronic health records, telemedicine, wearables, and mobile apps, into the health care system [12]. Overall, digital health technology encompasses both eHealth, which involves the use of the internet and related technologies to enhance health care systems through information and communication technology [13,14], and mobile health, which focuses on health practices supported by mobile devices [15]. The widespread use of digital health technologies has opened an innovative “window of opportunity” for managing chronic pain in a more personalized, accessible, and patient-centered way [8]. Evidence has highlighted that digital health technologies are valuable solutions for remotely collecting patient health data (eg, using patient-reported measures or wearable devices), improving symptom management, and decreasing patient appointments and hospitalizations [16]. Jongerius et al [17], in a systematic review, highlighted that digital health technologies are used in cancer clinical practice for the following different reasons: to stimulate the adoption of preventive behaviors, to increase early cancer identification, to manage cancer care, and to provide assistance to cancer survivors. Overall, digital health technology supports patients and the health care system to achieve several critical outcomes for better cancer clinical management [17-19]. For example, Zhu et al [19] reported that an app-based program named “Breast Cancer e-Support” was able to decrease symptomatology related to anticancer treatments and therefore improve self-efficacy and QoL. In addition, Maguire et al [20] designed and tested ASyMS (Advanced Symptom Management System) for the management of chemotherapy toxicity; it not only enables the evaluation and monitoring of patient symptomatology related to anticancer treatments but also provides tailored and evidence-based recommendations to manage symptoms [20].

Considering the specific case study of chronic pain in cancer survivorship, digital health technologies enhance access to nonpharmacological interventions; address pain-related mobility issues; improve patient networking and connections; foster self-management, self-efficacy, coping, and patient engagement; and facilitate communication among health care professionals of various specialties (eg, physicians, nurses, physiotherapists, psychologists) [18,21,22]. For example, Ranney et al [18] described that 89.5% of patients with chronic pain reported using digital health tools (eg, websites to search for health information, social media, and mobile apps), and such usage is associated with improved chronic pain coping mechanisms. Different digital health technologies have been designed to manage cancer and pain in the clinical pathway [23]. For example, an educational digital intervention called the “Pain
Education after Cancer Collaborative” (PECAN) project was developed for survivors of breast cancer; in the intervention, a decision tree system is used to provide tailored educational information to cancer survivors based on their answers to specific queries [24]. In addition, a Mobile Pain Coping Skills Training Protocol has been proposed to support patients’ understanding of the pain experience and strategies to cope with the pain [25]. More generally, Hauser-Ulrich et al. [26] recently developed a text-based chatbot named “SELMA” (PainSELFManagement) aimed to booster self-management of chronic pain in different types of diseases, supporting health care professionals in the delivery of evidence-based interventions. Moreover, digital health technologies could encourage patients to be more involved in their treatment decisions, through the implementation of specific decision aids [16,27-29]. Studies have stressed that the implementation of tailored decision aids in mobile apps increases patients’ awareness about treatment preferences (eg, pharmacological vs nonpharmacological treatments), reduces decisional conflict, and enhances adherence to treatments [30-32].

Even with the key role of digital health technologies in the cancer clinical pathway, few studies on digital and integrated health ecosystems are currently available [33]. Consequently, in this pilot study, our primary endpoint was the usability of the novel digital integrated health ecosystem, PainRELife, for managing and monitoring chronic pain in patients with breast cancer throughout the survivorship trajectory. Further, we aimed to evaluate the PainRELife ecosystem’s effectiveness at enhancing pain self-efficacy, improving shared decision-making, and reducing pain perception as secondary endpoints. The PainRELife ecosystem stands out as the first to seamlessly integrate all the essential components required for comprehensive pain management within a single platform. This includes features such as pain monitoring, physical and psychological assessment, e-diaries, exercises, educational resources, and decision aids. Furthermore, it incorporates a dedicated platform for health care providers and a robust big data cloud infrastructure. This holistic integration sets our ecosystem apart in the realm of pain management solutions.

**Methods**

**Study Design and Patient Recruitment**

**PainRELife Ecosystem**

This pilot study is nested in a national project titled “PainRELife, Sustainable and integrated big data ecosystem for continuity of care and decision support for patients with pain” (ID: 1173269). This national project guided the development and testing of an integrated health ecosystem for the management of chronic pain. Specifically, the PainRELife ecosystem consists of a cloud technology platform interconnected with electronic health records, which is named the Nu Platform, connected to the Fast Healthcare Interoperability Resources (FHIR) server for data analysis related to the patient care pathway. Health care providers use the Nu Platform to collect and store patient clinical data, and it enables the ongoing monitoring of patient health status (eg, pain, psychological well-being, and decision preferences about treatments; see Figures 1A and 1B), from diagnosis and active treatments to follow-ups (see Figures 1C and 1D). In addition, a big data infrastructure linked to the FHIR server enables a series of dynamic dashboards aimed at providing a systematic, intuitive outline of patient population features that might be used by researchers, clinicians, and health care stakeholders. The Nu Platform is associated with a mobile app for patients named PainRELife, which collects health care data. This technological solution permits dual communication between patients and health care professionals. Information collected by the mobile app is saved in the Nu Platform and overseen by health care professionals [7].

The PainRELife mobile app enables patient education and the collection of patient-reported outcomes. The mobile app is composed of an “educational section” that includes educational resources to improve learning about chronic pain throughout the cancer pathway (throughout the different phases of survivorship: acute, extended, and permanent; see Figure 2D) [3] and a “pain and psychological well-being assessment section” that contains a set of validated questionnaires (eg, pain intensity and interference, anxiety, and depression; see Figures 2A and 2B). Furthermore, the mobile app includes an e-diary (see Figure 2C) and exercises for pain and emotion-body mapping (see Figures 2E and 2F), enabling a holistic evaluation of psychological well-being and the pain experience. In addition, the mobile app incorporates a decision aid section, which is structured in 2 modules: profiling patients’ preferences and a decision tree (see Figure 2G). These modules are designed to empower patients with breast cancer by increasing their awareness of treatment preferences and facilitating shared decision-making regarding their care. The “profiling patients’ preferences” module aims to assist patients with evaluating and comprehending essential aspects of both pharmacological and nonpharmacological treatments. This includes understanding how interventions and treatments work to reduce pain or aid in recovery, identifying their advantages, and recognizing potential disadvantages. The decision tree module enables patients with breast cancer to tailor their health care preferences using the subjective expected utility approach [7] (see Figure 2G). This empowers patients with breast cancer to actively participate in the decision-making process, aligning their treatment choices with their unique needs and goals.
Figure 1. Health care professional interface on the Nu Platform: (A) home page displaying all available activities for health care professionals, (B) patient questionnaire list providing the measures used to assess psychological and physical status administered via the PainRELife mobile app, (C) patient list providing a directory of all patients registered on the Nu Platform, (D) clinical evaluation page offering access to detailed information on clinical events and therapeutic suggestions.

Figure 2. Patient interface on the PainRELife mobile app: (A) home page showing an overview of the mobile app sections, (B) pain and psychological well-being assessment section displaying the questionnaires that patients are required to complete, (C) patient’s e-diary, (D) educational section showing some of the educational content within the mobile app, (E) and (F) pain and emotion-body mapping exercises, (G) decision aid section showing an example for preferences for pharmacological and nonpharmacological treatments.

Participants
Participants of this pilot usability study included 25 patients with diagnosed breast cancer and pain (mean age 47.12, SD 8.41 years) admitted to the Division of Medical Senology and the Division of Pain Therapy and Palliative Care of the European Institute of Oncology (IEO). Participants were introduced to the mobile app after their clinic visit and instructed to use it for 3 months consecutively. Participants were recruited according to the following established set of inclusion criteria: >18 years
old, affected by breast cancer, has undergone surgical intervention for breast carcinoma, experiencing post-surgical pain (≥3 on a numeric rating scale [NRS]), and in possession of internet access and a personal smartphone. We excluded patients with breast cancer who had a previous or ongoing psychiatric or neurological disorder or other disease requiring active analgesic treatments. The inclusion and exclusion criteria were established considering that chronic pain is a common side effect (related to both the surgery and anticancer treatments) for patients with breast cancer (~60% experience it), and persistent acute pain after surgery is considered a risk factor for developing chronic pain during survivorship [34]. A full and detailed description of the research protocol of this pilot study was published previously [7].

**Instruments**

Patient sociodemographic and medical data were gathered through electronic medical records and a set of ad hoc items during the enrollment consultation. For the perceived pain assessment, the NRS was used to evaluate pain using a numerical range from 0 (no pain) to 10 (severe pain) [35]. Further, a set of validated self-measures was used to evaluate the primary and secondary endpoints. In detail, the Mobile Application Rating Scale (MARS) was used to evaluate the eHealth platform usability. MARS is a self-administered questionnaire with 29 items evaluating the following dimensions: engagement; functionality; aesthetics; quality of the information received; subjective perception of the app quality; impact of the mobile app on knowledge, attitudes, and probability to change user behaviors (in this specific case, it refers to behaviors related to pain management). The Cronbach $\alpha$ of the MARS is .90, indicating good internal consistency [36-38]. The Pain Self-Efficacy Questionnaire (PSEQ) is a self-administered questionnaire consisting of 10 items that evaluate self-efficacy in patients with pain. The Cronbach $\alpha$ is .94, indicating excellent internal consistency [39,40]. Finally, the 9-item Shared Decision-Making Questionnaire (SDM-Q-9) is a self-administered questionnaire comprising 9 items that evaluate a series of aspects related to the possibility of achieving a shared decision [41]. The Cronbach $\alpha$ is .94, indicating excellent internal consistency [42].

**Statistical Considerations**

A series of descriptive analyses were performed to depict the characteristics of the sample. In order to evaluate the primary endpoint, the mean score and its SD were calculated for each MARS subscale (engagement experienced while using the app; functionality; aesthetics; quality of the information received; subjective perception of the app quality; expected impact on knowledge, attitudes, and probability to change user behaviors) at 3 months; in addition, the total number of times the PainRELIFE mobile app was accessed by each participant was determined.

Further, a new variable named total app quality was created using the mean values of the engagement, functionality, aesthetics, and information quality MARS subscales. The final measurement of app quality was the average value of the 4 means [43]. Pearson correlational analysis was performed among all self-reports used (NRS, PSEQ, SDM-Q-9, MARS) and the total number of times the app was accessed during the 3 months of the study. A repeated measures ANOVA was performed to detect variation in pain intensity (NRS) from T0 (baseline) to T2 (3 months). Further, a new dichotomous variable named “frequency of use” was created considering the entire number of times the app was accessed (mean 22.92, SD 15.60; range 2-73) and the lowest number of times the PainRELIFE mobile app needed to be accessed (21 times) by participants to finalize the study’s tasks. The “frequency of use” variable permitted dividing the participants into groups based on higher and lower frequencies of access. Further, a Student t test was run to evaluate the difference between the “frequency of use” and PSEQ, SDM-Q-9, and MARS scores. Data were analyzed using SPSS version 26.0 (IBM Corp).

**Ethical Considerations**

This study received ethical approval in December 2021 (R1597/21-IEO 1701) from the Ethical Committee of the IEO and respects the Declaration of Helsinki and Good Clinical Practice Guidelines. All participants read and signed the informed consent form, which encompassed a comprehensive and exhaustive explanation of the primary and secondary endpoints, study procedures, and possible risks and benefits. Participants were not compensated and were able to withdraw their participation at any time during the study. Concerning privacy and confidentiality protection, all data collected were deidentified and anonymized, complied with national data protection legislation, and will be stored in the IEO databases for 10 years.

The authors affirm that human research participants provided informed consent for publication of their data.

**Results**

**Participant Characteristics**

The sociodemographic, cancer, and treatment characteristics are listed in Tables 1-3.
**Table 1.** Sociodemographic information of participating patients with breast cancer (n=25).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Results, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Cohabiting</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Widowed</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Single</td>
<td>5 (20)</td>
</tr>
<tr>
<td>Married</td>
<td>16 (64)</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
</tr>
<tr>
<td>PhD</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>8 (32)</td>
</tr>
<tr>
<td>High school</td>
<td>12 (48)</td>
</tr>
<tr>
<td>Primary school</td>
<td>3 (12)</td>
</tr>
</tbody>
</table>

**Table 2.** Diagnosis, cancer type, familiarity, and genetic mutation of participating patients with breast cancer (n=25).

<table>
<thead>
<tr>
<th>Cancer characteristics</th>
<th>Results, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Lobular carcinoma</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Ductal carcinoma</td>
<td>17 (68)</td>
</tr>
<tr>
<td>Ductal carcinoma in situ</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Mucinous carcinoma</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Occult carcinoma</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Cancer type</strong></td>
<td></td>
</tr>
<tr>
<td>Triple negative</td>
<td>2 (8)</td>
</tr>
<tr>
<td>HER2+&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5 (20)</td>
</tr>
<tr>
<td>Luminal</td>
<td>18 (72)</td>
</tr>
<tr>
<td><strong>Familiarity</strong></td>
<td></td>
</tr>
<tr>
<td>I&lt;sup&gt;+&lt;/sup&gt; breast</td>
<td>8 (32)</td>
</tr>
<tr>
<td>II&lt;sup&gt;+&lt;/sup&gt; breast</td>
<td>6 (24)</td>
</tr>
<tr>
<td>No familiarity</td>
<td>11 (44)</td>
</tr>
<tr>
<td><strong>Mutation</strong></td>
<td></td>
</tr>
<tr>
<td>BRCA1</td>
<td>2 (8)</td>
</tr>
<tr>
<td>BRCA1</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Negative</td>
<td>6 (24)</td>
</tr>
<tr>
<td>No testing</td>
<td>13 (52)</td>
</tr>
</tbody>
</table>

<sup>a</sup>HER2: human epidermal growth factor receptor 2.
Table 3. Surgery, medical treatments, and radiotherapy undergone by participating patients with breast cancer (n=25).

<table>
<thead>
<tr>
<th>Treatment characteristics</th>
<th>Results, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surgery</strong></td>
<td></td>
</tr>
<tr>
<td>Mastectomy</td>
<td>23 (92)</td>
</tr>
<tr>
<td>Axillary dissection</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Quadrantectomy</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Medical treatment</strong></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy + endocrine therapy</td>
<td>8 (32)</td>
</tr>
<tr>
<td>Chemotherapy + immune therapy</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Endocrine therapy</td>
<td>12 (48)</td>
</tr>
<tr>
<td>Immune + endocrine therapy</td>
<td>1 (4)</td>
</tr>
<tr>
<td>No treatment</td>
<td>2 (8)</td>
</tr>
<tr>
<td><strong>Radiotherapy</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8 (32)</td>
</tr>
<tr>
<td>No</td>
<td>17 (68)</td>
</tr>
</tbody>
</table>

Usability

The total MARS score (ranging from 1 to 5) provided overall medium-to-high mean values for each subscale (range 3.31-4.18; see Table 4) and a mean total app quality score of 3.90 (SD 0.506), suggesting generally good usability as evaluated by the participants. This was also confirmed by the total number of times the participants accessed the app during the entire study (mean 22.92, SD 15.60; range 2-73). In particular, 3 of 5 subscales had the highest scores: functionality (mean 4.14, SD 0.630), information (mean 4.18, SD 0.608), and behavioral change (mean 4.05, SD 0.666).

On the functionality subscale, 57% (15/23) of the participants judged that the mobile app is straightforward to use. Moreover, 91% (21/23) of the participants affirmed that the interactions are reliable and intuitive (ease of use: 8/23, 35% agree; 13/23, 57% strongly agree), positively evaluated the design (gestural design: 8/23, 35% agree; 12/23, 52% strongly agree), and evaluated the navigation properties as good (navigation: 12/23, 52% agree; 8/23, 35% strongly agree). However, some slight uncertainties were observed regarding the general performance of the mobile app, specifically moving between pages and sections (performance: 8/23, 35% undecided; see Table 5).

Concerning the distribution of responses in the information subscale, 78% (18/23) of the participants reported that the information in the mobile app is evidence-based (information: 9/23, 39% agree; 9/23, 39% strongly agree), relevant, focused on chronic pain in breast cancer and its management during the disease clinical pathway (quality of information: 9/23, 39% agree; 11/23, 48% strongly agree), and trustworthy (credibility: 22/23, 96% strongly agree). Further, the amount of information (quantity of information: 7/23, 30% agree; 9/23, 39% strongly agree) and how the information is reported using different setups (visual information: 11/23, 48% agree; 9/23, 39% strongly agree) were considered positive by the participants (see Table 6). Finally, most participants stated that the mobile app’s goals are achievable (goals: 11/23, 48% agree; 3/23, 13% strongly agree), even if 30% (7/23) reported some concerns.

Last, the distribution of responses in the behavioral change subscale revealed that 83% (19/23) of the participants strongly agreed that the mobile app had improved awareness about the issue of chronic pain in the cancer disease pathway, and 70% (16/23) agreed that the app increased chronic pain–related knowledge. Likewise, 70% (16/23) of the participants reported that the mobile app might support attitudes toward chronic pain (attitudes: 9/23, 39% agree; 7/23, 30% strongly agree; see Table 7).

In addition, most of the participants reported that the mobile app would potentially be helpful to bolster help-seeking behaviors (help seeking: 5/23, 22% agree; 9/23, 39% strongly agree) as well as support an intention to change (intention to change: 5/23, 22% agree; 9/23, 39% strongly agree). Still, 39% (9/23) showed concerns about the capacity to transform intention into a fundamental behavioral change (see Table 7). Overall, participants judged the app to be well-targeted (engagement subscale: mean 3.31, SD 0.617) and the app’s layout to be adequate (aesthetics subscale: mean 3.98, SD 0.849); likewise, the overall subjective quality was adequate (subjective quality subscale: mean 3.50, SD 0.494).
### Table 4. Mobile Application Rating Scale (MARS) subscale scores.

<table>
<thead>
<tr>
<th>MARS subscales</th>
<th>Results, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement</td>
<td>3.31 (0.617)</td>
</tr>
<tr>
<td>Functionality</td>
<td>4.14 (0.630)</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>3.98 (0.850)</td>
</tr>
<tr>
<td>Information</td>
<td>4.18 (0.608)</td>
</tr>
<tr>
<td>Subjective quality</td>
<td>3.50 (0.494)</td>
</tr>
<tr>
<td>Behavioral change</td>
<td>4.05 (0.666)</td>
</tr>
<tr>
<td>Total app quality</td>
<td>3.90 (0.506)</td>
</tr>
</tbody>
</table>

### Table 5. Functionality assessment of the PainRELife mobile app using the Mobile Application Rating Scale (MARS) [37,38] (n=23).

<table>
<thead>
<tr>
<th>Functionality assessment</th>
<th>Response, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Ease of use( ^a )</td>
<td>0</td>
</tr>
<tr>
<td>Gestural design( ^b )</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Navigation( ^c )</td>
<td>0</td>
</tr>
<tr>
<td>Performance( ^d )</td>
<td>0</td>
</tr>
</tbody>
</table>

\( ^a \)“How easy is it to learn how to use the app?”; “How clear are the menu labels/icons and instructions?”

\( ^b \)“Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?”

\( ^c \)“Is moving between screens logical/accurate/appropriate/untinterrupted; are all necessary screen links present?”

\( ^d \)“How accurately/fast do the app features (functions) and components (buttons/menus) work?”

### Table 6. Information assessment of the PainRELife mobile app using the Mobile Application Rating Scale (MARS) [37,38] (n=23).

<table>
<thead>
<tr>
<th>Information assessment</th>
<th>Response, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>Information( ^a )</td>
<td>0</td>
</tr>
<tr>
<td>Credibility( ^b )</td>
<td>0</td>
</tr>
<tr>
<td>Quality of information( ^c )</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Quantity of information( ^d )</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Visual information( ^e )</td>
<td>0</td>
</tr>
<tr>
<td>Goals( ^f )</td>
<td>2 (9)</td>
</tr>
</tbody>
</table>

\( ^a \)“Contains high-quality information (eg, text, feedback, measures, references) from a credible source.”

\( ^b \)“Does the app come from a legitimate source (specified in app store description or within the app itself)?”

\( ^c \)“Is app content correct, well written, and relevant to the goal/topic of the app?”

\( ^d \)“Is the extent of coverage within the scope of the app and comprehensive but concise?”

\( ^e \)“Is the visual explanation of concepts—through charts/graphs/images/videos, etc—clear, logical, correct?”

\( ^f \)“Does app have specific, measurable, and achievable goals (specified in the app store description or within the app itself)?”
Specifically, the information in the different sections and should be managed during the survivorship trajectory. Information provided on chronic pain and how chronic pain cancer appreciated both the quality and quantity of the health usability of the PainRELife mobile app with breast cancer provided a generally positive rating for the Considering the primary endpoint of this pilot study, patients with breast cancer reported that habitual use of the mobile app helped increase help-seeking behaviors for chronic pain (14/23, 61%), their general attitudes toward chronic pain, and their willingness to identify preeminent strategies for managing chronic pain. These results are particularly noteworthy considering that many studies have suggested that chronic pain syndrome in patients with breast cancer is commonly undiagnosed and not often considered by oncologists [8]. In addition, many cancer patients report poor knowledge about cancer-related chronic pain, available interventions, and possible health system resources [44]. Furthermore, the positive correlation between the total number of times the mobile app was accessed and pain intensity ($P = .03$) might indicate that patients with breast cancer who had a higher pain level might have utilized the mobile app to find evidence-based information and strategies to self-manage their pain. This possible interpretation might be linked to the difference in pain self-efficacy observed between the participants with higher versus lower frequency mobile app use. Specifically, participants with lower pain self-efficacy used the mobile app more than participants with a higher pain self-efficacy (mean 48.46, SD 7.90).

**Frequency of Use and Pain Self-Efficacy**

According to the Student $t$ test, younger participants used the mobile app less (mean 44.15, SD 7.11) than older participants (mean 50.33, SD 8.80; $t_{23}=-1.937$, $P = .03$; $d=0.77$). A difference in pain self-efficacy was observed between participants with higher versus lower frequency use ($t_{23}=1.644, P = .057$; $d=0.65$). The latter data indicate that, at T2, participants with a lower pain self-efficacy (mean 40.83, SD 14.58) used the mobile app more than participants with a higher pain self-efficacy (mean 48.46, SD 7.90).

**Pain Intensity and Shared Decision-Making**

The repeated measures ANOVA revealed that participants reported a reduction in pain intensity from T0 (mean 5, SD 1.68) to T2 (mean 3.72, SD 2.59; $F_{2,48}=3.407$, $P = .04$). A positive correlation was found between the total number of times the mobile app was accessed and pain intensity at T2 ($r=0.458$, $P = .03$).

No correlations were detected between the MARS subscales and PSEQ or NRS. A negative correlation was observed between the subjective quality subscale and the number of times the mobile app was accessed ($r=-0.498$, $P = .02$). Further, the engagement ($r=0.445$, $P = .03$), information ($r=0.427$, $P = .04$), and subjective quality ($r=0.548$, $P \leq .007$) subscales were positively correlated with shared decision-making.

**Discussion**

**Principal Findings**

Considering the primary endpoint of this pilot study, patients with breast cancer provided a generally positive rating for the usability of the PainRELife mobile app. Patients with breast cancer appreciated both the quality and quantity of the health information provided on chronic pain and how chronic pain should be managed during the survivorship trajectory. Specifically, the information in the different sections and modules were perceived as informative and comprehensible (20/23, 87%) and from credible sources of information (18/23, 78%). Most of the patients with breast cancer reported that habitual use of the mobile app helped increase help-seeking behaviors for chronic pain (14/23, 61%), their general attitudes toward chronic pain, and their willingness to identify preeminent strategies for managing chronic pain. These results are particularly noteworthy considering that many studies have suggested that chronic pain syndrome in patients with breast cancer is commonly undiagnosed and not often considered by oncologists [8]. In addition, many cancer patients report poor knowledge about cancer-related chronic pain, available interventions, and possible health system resources [44]. Furthermore, the positive correlation between the total number of times the mobile app was accessed and pain intensity ($P = .03$) might indicate that patients with breast cancer who had a higher pain level might have utilized the mobile app to find evidence-based information and strategies to self-manage their pain. This possible interpretation might be linked to the difference in pain self-efficacy observed between the participants with higher versus lower frequency mobile app use. Specifically, participants with lower pain self-efficacy used the mobile app more than participants with higher pain self-efficacy. Perhaps the participants with lower pain self-efficacy used the mobile app to find a strategy or way to manage their pain. Self-efficacy has a crucial role for patients with cancer, and studies have reported that it improves psychological well-being, reduces fear of cancer recurrence, enhances self-care, and improves management of symptoms such as pain [45,46]. Vinnikova et al [47] observed that individuals might use mobile apps to learn more about their health problems and monitor their physical and psychological status. Furthermore, considering that self-efficacy is an attribute of cancer pain self-management, the prevalent use of the mobile app by participants with a lower self-efficacy could explain the percentage of participants who reported concerns about intention to change and the capacity to transform intention into a fundamental behavioral change [48].
A second series of results are linked to the secondary endpoints. Participants reported a reduction in pain intensity at 3 months ($P=0.04$). We argue that the use of the mobile app might have relieved the pain experience, disease, and treatment-related symptomatology as observed in other previous studies [16,19].

One noteworthy finding is related to the association between specific features of the mobile app, evaluated with the MARS, and shared decision-making. Indeed, participants who provided higher positive evaluations about engagement ($P=0.03$), information ($P=0.04$), and subjective quality ($P=0.007$) also reported higher perceptions of having shared decisions along their care pathway. We argue that patients with breast cancer who feel involved in their treatment decisions are more engaged with the mobile app. For this type of patient, information retrieved in the mobile app might be used to reinforce and reiterate the ability to achieve a shared decision throughout their care pathway.

**Limitations**

Despite the interesting and challenging results, this pilot usability study had some limitations that must be considered. The primary limitation is the relatively small sample size of patients with breast cancer and the decision to use a single group to test usability. This decision might have caused a loss of pertinent information about the patients’ perceptions of the usefulness of this digital health technology. However, our sampling strategy is consistent with the pilot study design and methodological guidelines [49-51]. Related to this point, we must also mention that the statistical significance reported for pain self-efficacy might be considered borderline ($P=0.057$). However, the effect size is medium-to-large ($d=0.65$), which supports the statistical difference between the groups. We argue that the $P$ value might be due to the small sample size [52]. We also argue that this could be a significant result that has to be further investigated in successive studies, considering the positive effect of cancer pain self-management on QoL, when compared with pharmacological treatments such as opioid consumption [48].

In addition, the “frequency of use” variable had a moderately high SD (15.60). However, the distribution of the participants between the low frequency ($n=13$) and high frequency ($n=12$) groups was homogeneous and balanced. Furthermore, we hypothesized the presence of selection bias resulting from the inclusion criteria, which required participants to have internet access and a personal smartphone. This criterion may have limited the inclusion of certain vulnerable groups among patients with breast cancer, such as older adults or individuals with lower health literacy and socioeconomic challenges who could be at higher risk of undiagnosed chronic pain. Most of our patients with breast cancer also reported medium-to-low pain during the entire study and were in the acute and extended stages of the survivorship trajectory, which might have affected the frequency of mobile app use. Indeed, even if the total number of times the mobile app was accessed was relatively high and satisfactory, in the last month of the study, some participants decreased their total usage; 2 of 25 participants used the mobile app only at enrollment. The last limitation is related to the previous one and concerns the lack of assessment of the timing of mobile app use. Indeed, only the total number of times the mobile app was accessed was collected and evaluated. These limitations have been intensely discussed in the full protocol published elsewhere, and we plan to address them in future studies [7].

**Conclusions**

The data retrieved from this pilot study evaluating patients’ experiences using a novel and integrated health ecosystem for the management of chronic pain for breast cancer survivors are consistent with other studies highlighting that digital health technologies, when developed using a patient-driven approach, might be considered valuable tools for increasing the participation of patients with breast cancer in clinical care. In addition, these tools permit the achievement of critical clinical outcomes and improvement in QoL [4,8,22]. Moreover, health-integrated ecosystems permit secondary key outcomes such as reducing the burden on health care professionals and optimizing health system resources. Finally, we argue that digital integrated health ecosystems might be important devices for improving the ongoing monitoring of physical status, psychological burden, and socioeconomic issues during the cancer survivorship trajectory.

**Acknowledgments**

This work is supported by a grant from Regione Lombardia: “PainRELife, Sustainable and integrated big data ecosystem for continuity of care and decision support for patients with pain” (ID: 1173269). The European Institute of Oncology (Istituto Europeo di Oncologia [IEO], Italy) monitors the study’s scientific, legal, and ethical aspects. Participants are recruited at IEO.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

**Authors’ Contributions**

MM, SFMP, CF, and GP conceived and designed the study. GP and MM coordinated the study, GP and MM acquired legal authorizations, and CF and EF managed the participants. MM handled the drafting and writing of the manuscript. All authors read and approved the final manuscript.

**Conflicts of Interest**

VM is a partner and chief technology officer of Nuvyta, the software house that developed and sells Nuplatform.
References


Abbreviations

ASyMS: Advanced Symptom Management System
FHIR: Fast Healthcare Interoperability Resources
IEO: European Institute of Oncology
MARS: Mobile Application Rating Scale
NRS: numeric rating scale
PECAN: Pain Education after Cancer Collaborative
PSEQ: Pain Self-Efficacy Questionnaire
QoL: quality of life
SDM-Q-9: 9-item Shared Decision-Making Questionnaire

Please cite as:

Support for Chronic Pain Management for Breast Cancer Survivors Through Novel Digital Health Ecosystems: Pilot Usability Study of the PainRELife Mobile App
JMIR Form Res 2024;8:e51021
URL: https://formative.jmir.org/2024/1/e51021
doi:10.2196/51021
PMID:38306176

©Marianna Masiero, Chiara Filipponi, Elisa Fragale, Silvia Francesca Maria Pizzoli, Elisabetta Munzone, Alessandra Milani, Luca Guido, Vittorio Guardamagna, Sara Marceglia, Roberto Prandin, Marco Prenassi, Annamaria Caruso, Vania Manzelli, Chiara Savino, Costanza Conti, Federica Rizzi, Alice Casalino, Giulia Candiani, Francesca Memini, Luca Chiveri, Andrea Luigi Vitali, Massimo Corbo, Roberto Grasso, Florence Didier, Roberta Ferrucci, Gabriella Pravettoni. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Mental Health and Well-Being in Racial or Ethnic Minority Individuals After Using a Faith and Prayer Mobile App (Pray.com): Feasibility and Preliminary Efficacy Trial

Breanne Laird1, MS; Sabrina Zuniga2, MS; Joshua N Hook2, PhD; Daryl R Van Tongeren3, PhD; Lynda Joeman4, BS; Jennifer Huberty5, PhD

1Pray, Inc, Westlack Village, CA, United States
2Department of Psychology, University of North Texas, Denton, TX, United States
3Department of Psychology, Hope College, Holland, MI, United States
4Tombridge, United Kingdom
5Fit Minded, Inc, Phoenix, AZ, United States

Abstract

Background: Research is needed on how faith and prayer apps fit within the values of racial and ethnic minority (REM) groups, as well as whether such apps are effective in promoting mental health and well-being.

Objective: This study aims to determine the feasibility and preliminary effectiveness of using the mobile app Pray.com on mental health and well-being among REM participants.

Methods: This study was a single-group (N=77), 4-week feasibility trial in REM groups (65/77, 84% Black or African American). Participants were asked to use the Pray.com app at no cost for at least 5 times per week for 5 minutes per day. Participants completed questionnaires at the baseline and postintervention time points. Feasibility questionnaires were only completed at the postintervention time point, including qualitative interviews (n=15). The feasibility questions included acceptability (ie, satisfaction, intent to continue use, perceived appropriateness, and fit within culture), demand (ie, self-reported app use, expressed interest, and perceived demand), and practicality (ie, ease or difficulty of use, ability to use the app, and cost-effectiveness). Frequency and descriptive statistics were used to analyze feasibility outcomes. Changes in dependent variables were analyzed using paired-sample 2-tailed *t* tests. Partial correlations were conducted to explore the association between app use and outcomes, controlling for baseline scores.

Results: Participants reported (54/72, 75% responded with “very likely” or “likely” to the feasibility questions) that they perceived the Pray.com app as acceptable. These findings were supported by qualitative interviews (n=15). Most participants (62/72, 86%) did not meet the app use prescription but expressed interest in using the app in the future and perceived demand for it in their communities. In addition, participants reported that the app was easy to use and perceived it to be inexpensive (US $7.99). Participants reported improved mental health (ie, stress and depressive and anxiety symptoms) and well-being (ie, satisfaction with life, spiritual well-being, religious commitment, and racial or ethnic identity development) at postintervention despite relatively low average levels and high variability of app use (average total of 45.83, SD 111.90 min over the course of the study). Greater app use was significantly associated with improvements in mental health and spiritual well-being. However, app use and study methodology limitations suggest that the study results may not accurately capture the full impact of Pray.com use.

Conclusions: This is the first study to assess the feasibility of a faith and prayer app for mental health and well-being in a sample of REM individuals. Our findings suggest that the use of a faith and prayer app (ie, Pray.com) could be feasible and significantly
impactful for the improvement of mental health symptoms and well-being in REM individuals and their communities, especially Black and African American individuals with a Christian affiliation. Further research is warranted.

**KEYWORDS**

religion; spirituality; mobile apps; mental health; well-being; app; ethnic; technology; engagement; stress; depression; anxiety; quality of life; spiritual well-being; racial; spiritual practices; spiritual practice; mobile phone; mobile health; mHealth

**Introduction**

**Background**

Religious and spiritual identity can extend across and within various cultures, even worldwide, although religious experiences are largely influenced by and intersect with cultural context. Some have conceptualized religion as a form of culture [1], and previous research has examined how religious expressions differ by culture [2]. Religion can serve as a primary social identity [3] that intersects with other salient features of identity, such as one’s racial or ethnic identity [4]. Put simply, one’s religious experience and expression notably intersect with one’s cultural identity and context.

Despite the heterogeneity of religious expressions, researchers have identified certain core dimensions of religion, including cognitive, emotional, behavioral, and social features [5]. A substantial body of research suggests that religion enhances individual well-being [6-16]. However, less is known about the role of culture in the connection between religion and individual health and well-being.

For racial and ethnic minority (REM) populations, the religious or spiritual aspect of one’s identity might be particularly important. For example, Constantine and Sue [17] outlined a theoretical model describing optimal human functioning for people of color in the United States. One of the key premises of their argument was that one’s cultural context affects what is considered optimal human functioning for various groups and it is necessary to consider the impact of cultural values on the well-being of people of color. In their model, cultural values, beliefs, and practices such as collectivism, racial and ethnic pride, spirituality and religion, interconnectedness of mind, body, or spirit, and family and community are viewed as important predictors of optimal human functioning for people of color.

In this study, we focused on religion or spirituality as an important factor contributing to the mental health and well-being of REM groups. Indeed, research has supported the importance of religion in the lives of such groups (eg, African American and Latinx groups), and these individuals tend to engage in religious or spiritual practices at a higher rate than the US population overall [18,19]. For example, studies have found high levels of religious participation in African American individuals (ie, 89% are religious; 78% attend services regularly; and 90% pray, meditate, or use religious materials) [20]. Furthermore, previous research has supported a strong connection between religiousness and well-being in African American individuals [21,22]. For example, in a sample of African American individuals, racial or ethnic identity was positively related to both satisfaction with and meaning in life, and these relationships were partially mediated by religious commitment [23].

Similarly, engaging in religious practices has been linked to increased social support and psychological well-being and decreased perceived stress and physical pain in Latinx populations [24,25]. Furthermore, types of religious coping (eg, positive vs negative) affected symptoms of both depression and obsessive-compulsive disorder [24,26]. As religiosity tends to be more emphasized by Latinx immigrants in comparison with US-born Latinx populations [27-29], some scholars have hypothesized that a decline in religiosity may partially explain the Latino health paradox, in which more recently immigrated Latinx individuals tend to display better health outcomes than more assimilated and US-born Latinx individuals [30,31].

Regarding Asian and Asian American populations, religious identity was positively associated with self-esteem, positive affect, and meaning in life in a sample of Asian American adolescents [32]. In a sample of Chinese American and Korean American older adults, religious coping and support were associated with increased life satisfaction and decreased depressive symptoms [33]. In a sample of Korean adult immigrants, religious support was associated with higher well-being in the midst of experiencing financial hardship and difficulties with English [34]. Interviews with Filipino American individuals regarding their faith revealed that their experiences with religion and spirituality had both positive and negative effects on their well-being [35], although recent research has proposed that the negative effects of religion may be explained by scrupulosity in communities in which religious participation plays a large role in culture [26].

This high degree of religious participation by REM individuals is observed not only in the United States but also worldwide. For example, high rates of religious identification are observed in territories with high numbers of REM individuals, such as Latin America or the Caribbean, Middle East or North Africa, and sub-Saharan Africa [36]. Overall, it is clear that, for various REM populations, religious and spiritual identity are important factors that affect one’s quality of life. Thus, it is imperative to identify how religious identity, practices, and engagement empower REMs and assess how current practices can be implemented, changed, or improved to enhance religious and spiritual experiences and increase well-being in these populations.

**Different Pathways to Religious and Spiritual Well-Being**

Technological advances have permitted various ways for individuals to facilitate well-being through engagement in
religious and spiritual practice [37-39]. Although various media have previously encouraged several methods toward transcendent connection, the global health pandemic of 2020 accelerated the development of digital tools that allow individuals to enhance their well-being via religious and spiritual practices [37,38]. Many of these behavior changes have persisted, and individuals are engaging in technologically mediated religious and spiritual practices more regularly.

For example, individuals may interact with religious and spiritual content through mobile apps. One such app is Pray.com, which is an app that allows participants to engage in Christian-based faith and prayer content, such as reading the Bible, listening to sermons, or hearing stories or other creative content to engage with their spiritual faith. This app can be downloaded to any smartphone or internet-enabled (mobile) device, and it includes the opportunity to receive notifications to remain engaged in one’s religious practices. The Pray.com app was chosen for this study because of its popularity and potential reach but was not specifically designed for REM individuals.

Such research on the role of technology in religious and spiritual practices is nascent, and most research has been conducted on the Christian faith as it is a prominent part of public and political life in the United States [40]. Among the largest religious groups in the United States, Black Protestant denominations tend to use technology in their religious practices to a greater degree [41], yet digitally based interventions struggle to be representative of REM groups [42]. Thus, there is a dearth of research including REM diverse samples, and less research has focused on how such apps align with the cultural values of such participants. There is a need for research that explores how such apps fit (or do not fit) with the cultural values of REM participants as well as whether such apps are effective in promoting mental health and well-being. To our knowledge, this is the first study to assess the feasibility and preliminary efficacy of a faith and prayer app for mental health and well-being in REM individuals. Accordingly, the goal of this research was to examine how individuals self-identifying as belonging to REM groups engage with technologically mediated religious and spiritual practices to enhance their mental health and well-being.

Study Overview and Hypothesis

The purpose of this study was to determine the feasibility (ie, acceptability, demand, and practicality) and preliminary effectiveness of using the mobile app Pray.com on mental health (ie, stress and depressive and anxiety symptoms) and well-being (ie, satisfaction with life, spiritual well-being, religious commitment, and racial or ethnic identity development) among REM participants. We predicted that Pray.com would be feasible in REM individuals and be associated with improved mental health and well-being outcomes at 4 weeks.

Methods

Ethical Considerations

This study was approved by the institutional review board of Biola University (F22-013) and registered at ClinicalTrials.gov (NCT05626673). All participants provided electronic consent before taking part.

Study Design or Recruitment

This study was a single-group, 4-week feasibility trial. This method was chosen as this was a preliminary study of the feasibility and effectiveness of the app in a sample of REM individuals. Participants were recruited using convenience sampling through faith-based organizations (eg, churches and religious psychology organizations), personal contacts of the researchers, and social media between January 2023 and April 2023 and were directed to a link to complete the web-based eligibility survey.

Participants and Procedure

Potential participants completed a web-based eligibility questionnaire via Qualtrics (Qualtrics International, Inc). Potential participants were eligible if they (1) self-identified as belonging to an REM group, (2) were aged ≥18 years, (3) owned a smartphone and were willing to download a mobile app, and (4) were willing to engage in Christian-based religious practice on a mobile app. Eligible participants were directed to a web-based informed consent form and informed of the potential risks of taking part in the study. After signing the consent form, participants were asked to complete a series of web-based baseline questionnaires (time 1) related to mental health and well-being. Once the questionnaires were completed, participants were provided with instructions on how to download the Pray.com mobile app.

Intervention

Participants were asked to use the Pray.com app at no cost for at least 5 times per week for 5 minutes per day. The Pray.com app was chosen because of its popularity and potential reach. The app was not specifically designed for REM individuals. This dose was chosen as frequent private prayer has been associated with significant mental health benefits [43], and frequency may be more important than the duration of time spent in prayerlike practices. We chose a brief minimum period to increase adherence and consistency. Participants were contacted at time 2 (approximately 30 days after completing time 1) and invited to complete the web-based questionnaires again as well as a feasibility questionnaire about their experience using the app. After completing the time 2 questionnaires, participants were compensated with a digital gift card, debriefed about the purpose of the study, and invited to contact the researchers if they had any questions about their participation. A subset of participants (n=15) was invited (using random selection) to participate in a qualitative interview with a member of the research team (making this a mixed methods study). These qualitative interviews were conducted on the web via Zoom (Zoom Video Communications) and transcribed (and subsequently coded by the qualitative data analyst). The participants who completed the qualitative interviews were compensated with an additional digital gift card.
Measures

Feasibility (Time 2)

All feasibility measures were assessed at the postintervention time point (4 weeks). We measured feasibility using the guidelines by Bowen et al [44], including acceptability, demand, practicality, and preliminary efficacy (ie, trends in changes in stress, depressive and anxiety symptoms, satisfaction with life, spiritual well-being, religious commitment, and racial or ethnic identity development). Benchmarks for feasibility included (1) acceptability (ie, satisfaction, intent to continue use, perceived appropriateness, and fit within culture; ≥75% of participants reporting satisfaction with the app and that the app was appropriate and fit within their cultural identity or worldview), (2) demand (ie, self-reported app use, expressed interest, and perceived demand; ≥75% of participants adhering to the Pray.com prescription, expressing interest in future use, and perceiving the app as demanded within their community), and (3) practicality (ie, ease or difficulty of use, ability to use the app, and cost-effectiveness; ≥75% of participants reporting ease of use and that they could pay for the app).

The feasibility measures were developed by psychology and behavioral health researchers. Textbox 1 presents a list of the feasibility questions.

A subset of participants (n=15) completed a semistructured qualitative interview that focused on exploring their experiences using the app in a deeper way. The qualitative interview focused on topics similar to those in the feasibility questionnaire (eg, acceptability, demand, and practicality), as well as including an open-ended question asking for general feedback about the app. Textbox 2 presents a list of the interview questions.
Textbox 1. Feasibility questions.

Acceptability
- Satisfaction
  - How satisfied were you with the Pray.com app?
- Intent to continue use
  - How likely are you to continue to use Pray.com?
- Perceived appropriateness
  - How relevant was the Pray.com app to improving your mental health?
  - My racial/ethnic identity is well represented on the Pray.com app?
  - The material on the Pray.com app was a good fit with my racial/ethnic identity?
  - The material on the Pray.com app was a good fit with my cultural worldview?
- Fit within culture
  - How well did using the Pray.com app fit within your religious beliefs or worldview?
  - How well did using the Pray.com app fit within your cultural (ie, racial or ethnic) identity or worldview?

Demand
- Self-reported use
  - How often did you use the Pray.com app?
- Expressed interest
  - How interested are you in using the Pray.com app in the future?
- Perceived demand
  - How much demand do you think there would be for the Pray.com app within your religious community?
  - How much demand do you think there would be for the Pray.com app within your culture?

Practicality
- Ease or difficulty of use
  - How easy or difficult was it to use the Pray.com app?
- Ability to use the app
  - How would you rate your ability to use and navigate Pray.com?
- Cost
  - How cheap or expensive does this price seem to you?
  - How easy or difficult would it be to pay the monthly fee for the Pray.com app?
  - How likely would you be willing to pay the monthly fee for the Pray.com app?
Textbox 2. Interview questions.

Acceptability

- Satisfaction
  - What did you like the most about using the Pray.com app?
  - What did you like least about using the Pray.com app?

- Intent to continue use
  - What factors would influence your decision to continue using the Pray.com app in the future?
  - What factors would keep you from using the Pray.com app after the study is over?

- Perceived appropriateness
  - What about the Pray.com app was most relevant to helping you with your mental health?
  - What about the Pray.com app was most relevant to helping your spiritual life?
  - How did Pray.com contribute to how you feel about meditation/mindfulness?

- Fit within culture
  - What aspects of using the Pray.com app were a good fit within your religious beliefs or worldview?
  - What aspects of using the Pray.com app did not seem to fit within your religious beliefs or worldview?
  - What aspects of using the Pray.com app did not seem to fit within your cultural (ie, racial/ethnic) identity or worldview?
  - How well represented was your racial/ethnic group on the Pray.com app?

Demand

- Expressed interest
  - What aspects of the Pray.com app are you most interested in?
  - What aspects of the Pray.com app are you least interested in?

- Perceived demand
  - What could be improved about the Pray.com app to make it a better fit for you?
  - What could be improved about the Pray.com app to make it a better fit for your religious community?

Practicality

- Ease or difficulty of use
  - What aspects of the Pray.com app were difficult to use?
  - What aspects of the Pray.com app were easy to use?

- Ability to use the app
  - What aspects of the Pray.com app did you feel like you had the ability to use?
  - Was there anything about the Pray.com app that you felt like you did not have the ability to use?

- Other
  - Do you have any other feedback about your use of the Pray.com app?

Perceived Stress Scale (Time 1 and 2)

The Perceived Stress Scale-10 (PSS-10) [45] was used to measure participants’ subjective levels of stress. The PSS-10 is a 10-item measure in which participants rate items related to stress (eg, “how often have you been angered because of things that happened that were outside of your control?”) on a scale from 0=never to 4=very often within the previous month. A mean perceived stress score was calculated, with higher scores indicating higher levels of perceived stress. Previous research has demonstrated high reliability and construct validity of the PSS-10 [45,46], including in a sample of REM participants [47].
In this sample, the Cronbach α for the PSS-10 ranged from .81 to .84 across time points.

**Hospital Anxiety and Depression Scale (Time 1 and 2)**
The Hospital Anxiety and Depression Scale (HADS) [48] was used to measure participants’ symptoms of anxiety and depression. The HADS is a 14-item measure in which participants rate 7 items related to anxiety (eg, “I feel tense or ‘wound up.’”) and 7 items related to depression (eg, “I still enjoy the things I used to enjoy”) (reverse coded) on a scale from 0 to 3 (end points vary across items). Mean scores for anxiety and depression were calculated, with higher scores indicating more severe symptoms of anxiety and depression. Previous research has demonstrated evidence for the reliability and construct validity of the HADS [49], including cross-culturally [50]. In this study, the Cronbach α for the anxiety and depression subscales of the HADS ranged from .83 to .84 (anxiety) and from .79 to .84 (depression) across time points.

**Spiritual Well-Being Scale (Time 1 and 2)**
Participants’ spiritual well-being was assessed using the Spiritual Well-Being Scale (SWBS) [51,52]. The SWBS comprises 20 items divided into 2 subscales: religious well-being (10 items; eg, “I feel most fulfilled when I’m in close communion with God”) and existential well-being (10 items; eg, “I believe there is some real purpose for my life”). Participants rated each item on a scale from 1=strongly disagree to 6=strongly agree, with higher scores indicating higher levels of spiritual well-being. The SWBS has previously demonstrated high internal consistency and construct validity [51,52], including in samples with African American individuals [53]. In this study, the Cronbach α for the SWBS ranged from .91 to .92 across time points.

**Satisfaction With Life Scale (Time 1 and 2)**
Participants’ life satisfaction was measured using the Satisfaction With Life Scale (SWLS) [54]. Participants rated 5 items related to life satisfaction (eg, “In most ways my life is close to my ideal”) on a scale from 1=strongly disagree to 6=strongly agree, with higher scores suggesting higher life satisfaction. The SWLS has demonstrated high internal consistency and construct validity in previous research across various populations and contexts [54], including diverse populations [55]. In this study, the Cronbach α for the SWLS ranged from .73 to .83 across time points.

**Religious Commitment Inventory–10 (Time 1 and 2)**
Religious commitment was measured using the Religious Commitment Inventory–10 (RCI-10) [56]. Participants rated 10 items assessing religious commitment (eg, “I spend time trying to grow in understanding of my faith”) on a scale from 1=not true at all of me to 5=totaly true of me. Scores were measured using the mean of all 10 items, with higher scores suggesting higher levels of religious commitment. The RCI-10 has shown evidence of internal consistency and construct validity [56], including in diverse samples [56,57]. In this study, the Cronbach α for the RCI-10 ranged from .84 to .88 across time points.

**Multigroup Ethnic Identity Measure–Revised (Time 1 and 2)**
Racial or ethnic identity development was measured using the Multigroup Ethnic Identity Measure–Revised (MEIM-R) [58]. Participants rated 6 items assessing racial or ethnic identity development across 2 subscales—commitment (eg, “I have a strong sense of belonging to my own ethnic group”) and exploration (eg, “I have spent time trying to find out more about my ethnic group, such as its history, traditions, and customs”)—on a scale from 1=strongly disagree to 5=strongly agree, with higher scores indicating higher levels of commitment and exploration, respectively. The MEIM-R has demonstrated evidence of internal consistency and construct validity [58]. In this study, the Cronbach α for the MEIM-R ranged from .79 to .79 across time points.

**Statistical Analysis**
The data were analyzed using SPSS (version 28.0; IBM Corp). Frequency and descriptive statistics were used to analyze the feasibility outcomes and demographic characteristics. As not all participants answered every question (ie, it was a free-response survey), the sample sizes differed across analyses. We reported all available completed data. Self-reported use frequency was categorized as ordinal, reflecting the use of the app as (1) less than once per week, (2) up to 2 times per week, (3) 3 to 5 times per week, or (4) ≥5 times per week. All other feasibility questions were categorized as ordinal. Changes in dependent variables between time 1 and time 2 were analyzed using paired-sample 2-tailed tests. To assess the relationship between objective app use and changes in dependent variables, partial correlations were conducted to explore the association between app use and scores on time 2 variables controlling for scores on time 1 variables. P values of <.05 were considered significant.

With regard to the qualitative data, interview transcripts were imported into NVivo (QSR International) for coding and analysis. Thematic analysis methods were used based on the inductive or deductive approach proposed by Braun and Clarke [59]. In brief, this involved first identifying top-level themes based on the main issues and questions covered in the interviews (feasibility and mental health and well-being outcomes). Within these, emergent themes were identified inductively from the transcripts and labeled appropriately. An iterative process was used in which relevant extracts of interview data were allocated to these, with the coding being continually reviewed and revised until the analyst felt confident that it most accurately reflected the lived experiences of the research participants as reported in their interviews.

**Results**

Of the 77 final participants in our study, 72 (94%) completed time 2 measures. There were no significant differences in our main variables between completers and noncompleters.

**Participants**

We received 707 initial responses to the eligibility survey; however, the vast majority of these were found to be inauthentic responses (eg, bots and duplicate IP addresses). Accordingly,
we checked the data using elimination of duplicate IP addresses and consistent responses to demographic information. Then, we installed various data quality checks such as bot detection and reCAPTCHA (Google) and accuracy on 3 quality-check questions (eg. “Please select disagree”). This resulted in a final sample of 77 participants. Table 1 shows the sample demographic characteristics. Table 2 shows the religious or spiritual engagement at the beginning of the study.

Table 1. Sample demographics (N=77).

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>31 (5)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Cisgender women</td>
<td>23 (30)</td>
</tr>
<tr>
<td>Cisgender men</td>
<td>54 (70)</td>
</tr>
<tr>
<td><strong>Race or ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>6 (8)</td>
</tr>
<tr>
<td>Asian or Asian American</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>65 (84)</td>
</tr>
<tr>
<td>Latina, Latino, or Hispanic</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>2 (3)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>74 (96)</td>
</tr>
<tr>
<td>Gay</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Lesbian</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Religious affiliation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Christian (Catholic)</td>
<td>42 (55)</td>
</tr>
<tr>
<td>Christian (evangelical Protestant)</td>
<td>19 (25)</td>
</tr>
<tr>
<td>Christian (mainline Protestant)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Christian (Black Protestant)</td>
<td>5 (6)</td>
</tr>
<tr>
<td>Buddhist</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Hindu</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Jewish</td>
<td>1 (1)</td>
</tr>
<tr>
<td>None</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Religious identity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Stably religious</td>
<td>56 (73)</td>
</tr>
<tr>
<td>New identifier</td>
<td>8 (10)</td>
</tr>
<tr>
<td>Reidentifier</td>
<td>12 (16)</td>
</tr>
<tr>
<td>Never identified</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>
Table 2. Religious or spiritual engagement at the beginning of the study (N=77).

<table>
<thead>
<tr>
<th>Engagement characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Religious service attendance</strong></td>
<td></td>
</tr>
<tr>
<td>At least once a week</td>
<td>57 (74)</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>18 (23)</td>
</tr>
<tr>
<td>Seldom or never</td>
<td>2 (3)</td>
</tr>
<tr>
<td><strong>Prayer frequency</strong></td>
<td></td>
</tr>
<tr>
<td>At least once a week</td>
<td>54 (70)</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>19 (25)</td>
</tr>
<tr>
<td>Seldom or never</td>
<td>4 (5)</td>
</tr>
<tr>
<td><strong>Participation in prayer, scripture, or religious education groups</strong></td>
<td></td>
</tr>
<tr>
<td>At least once a week</td>
<td>43 (56)</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>29 (38)</td>
</tr>
<tr>
<td>Seldom or never</td>
<td>4 (5)</td>
</tr>
<tr>
<td><strong>Meditation frequency</strong></td>
<td></td>
</tr>
<tr>
<td>At least once a week</td>
<td>43 (56)</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>26 (34)</td>
</tr>
<tr>
<td>Seldom or never</td>
<td>8 (10)</td>
</tr>
<tr>
<td><strong>Scripture reading</strong></td>
<td></td>
</tr>
<tr>
<td>At least once a week</td>
<td>47 (61)</td>
</tr>
<tr>
<td>Once or twice a month</td>
<td>21 (27)</td>
</tr>
<tr>
<td>Seldom or never</td>
<td>8 (10)</td>
</tr>
</tbody>
</table>

Feasibility

Acceptability

All the acceptability benchmarks were met. Most participants (64/72, 89%) reported being very satisfied or satisfied with the app and that they were very likely or likely to continue to use the app in the future (63/72, 88%). No significant differences between men and women were observed regarding app acceptability.

The app was perceived as appropriate. Most participants reported that using the app was very relevant or relevant to improving their mental health (57/72, 79%) and spiritual lives (62/72, 86%). Most participants (60/72, 83%) reported that they strongly agreed or agreed that their racial or ethnic identity was well represented on the app and that the content was a good fit with their racial or ethnic identity (53/72, 74%) and cultural worldview (59/72, 82%). Most participants reported that they strongly agreed or agreed that Pray.com was a good fit with their religious beliefs or worldview (58/72, 81%) and cultural identity or worldview (55/72, 76%).

Participants reported that the app was more relevant to improving their spiritual life than their mental health ($P=.01$). There was no significant difference between fit with religious or cultural worldview ($P=.23$). Table 3 shows participant ratings of the acceptability of the app.
Table 3. Participant ratings of the acceptability of the app (N=72).

<table>
<thead>
<tr>
<th>Acceptability characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction, mean (SD)</strong></td>
<td>4.50 (0.69)</td>
</tr>
<tr>
<td>Very satisfied, n (%)</td>
<td>44 (61)</td>
</tr>
<tr>
<td>Satisfied, n (%)</td>
<td>20 (28)</td>
</tr>
<tr>
<td>Somewhat satisfied, n (%)</td>
<td>8 (11)</td>
</tr>
<tr>
<td>Dissatisfied, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Intent to continue use, mean (SD)</strong></td>
<td>4.47 (0.71)</td>
</tr>
<tr>
<td>Very likely, n (%)</td>
<td>43 (60)</td>
</tr>
<tr>
<td>Likely, n (%)</td>
<td>20 (28)</td>
</tr>
<tr>
<td>Somewhat likely, n (%)</td>
<td>9 (12)</td>
</tr>
<tr>
<td>Not likely, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Perceived appropriateness</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Relevance in improving mental health, mean (SD)</strong></td>
<td>4.14 (0.74)</td>
</tr>
<tr>
<td>Very relevant, n (%)</td>
<td>25 (35)</td>
</tr>
<tr>
<td>Relevant, n (%)</td>
<td>32 (44)</td>
</tr>
<tr>
<td>Somewhat relevant, n (%)</td>
<td>15 (21)</td>
</tr>
<tr>
<td>Irrelevant, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Very irrelevant, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Relevance in improving spiritual life, mean (SD)</strong></td>
<td>4.42 (0.88)</td>
</tr>
<tr>
<td>Very relevant, n (%)</td>
<td>44 (61)</td>
</tr>
<tr>
<td>Relevant, n (%)</td>
<td>18 (25)</td>
</tr>
<tr>
<td>Somewhat relevant, n (%)</td>
<td>7 (10)</td>
</tr>
<tr>
<td>Irrelevant, n (%)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Very irrelevant, n (%)</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Fit within culture</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Racial or ethnic identity well represented, mean (SD)</strong></td>
<td>4.29 (0.74)</td>
</tr>
<tr>
<td>Strongly agree, n (%)</td>
<td>33 (46)</td>
</tr>
<tr>
<td>Agree, n (%)</td>
<td>27 (38)</td>
</tr>
<tr>
<td>Somewhat agree, n (%)</td>
<td>12 (17)</td>
</tr>
<tr>
<td>Disagree, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Content good fit with racial or ethnic identity, mean (SD)</strong></td>
<td>4.15 (0.85)</td>
</tr>
<tr>
<td>Strongly agree, n (%)</td>
<td>31 (43)</td>
</tr>
<tr>
<td>Agree, n (%)</td>
<td>22 (31)</td>
</tr>
<tr>
<td>Somewhat agree, n (%)</td>
<td>18 (25)</td>
</tr>
<tr>
<td>Disagree, n (%)</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Content good fit with cultural worldview, mean (SD)</strong></td>
<td>4.29 (0.76)</td>
</tr>
<tr>
<td>Strongly agree, n (%)</td>
<td>34 (47)</td>
</tr>
<tr>
<td>Agree, n (%)</td>
<td>25 (35)</td>
</tr>
<tr>
<td>Somewhat agree, n (%)</td>
<td>13 (18)</td>
</tr>
<tr>
<td>Disagree, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Content good fit with religious beliefs or worldview, mean (SD)</strong></td>
<td>4.25 (0.87)</td>
</tr>
<tr>
<td>Very good fit, n (%)</td>
<td>35 (49)</td>
</tr>
<tr>
<td>Good fit, n (%)</td>
<td>23 (32)</td>
</tr>
</tbody>
</table>
### Acceptability characteristics

<table>
<thead>
<tr>
<th></th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somewhat good fit, n (%)</td>
<td>11 (15)</td>
</tr>
<tr>
<td>Poor fit, n (%)</td>
<td>3 (4)</td>
</tr>
</tbody>
</table>

**Fit with cultural identity or worldview, mean (SD)**

<table>
<thead>
<tr>
<th></th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good fit, n (%)</td>
<td>28 (39)</td>
</tr>
<tr>
<td>Good fit, n (%)</td>
<td>27 (38)</td>
</tr>
<tr>
<td>Somewhat good fit, n (%)</td>
<td>14 (19)</td>
</tr>
<tr>
<td>Poor fit, n (%)</td>
<td>3 (4)</td>
</tr>
</tbody>
</table>

### Demand

All demand benchmarks were met except for self-reported use (at least 75% of participants used the app at least 5 times per week for 5 min/d) as this could not be determined from the answer choices selected by the participants. Most participants (58/72, 81%) self-reported that they used the app between 3 and 5 times per week (35/72, 49%) or >5 times per week (23/72, 32%). However, objective use data showed that participants used the app for an average of 45.83 (SD 111.90) minutes and a total of 5.77 (SD 8.19) days throughout the entire study period, suggesting that app use may have been overreported. Regardless of app use, most participants (61/72, 85%) indicated that they were extremely interested or interested in continuing to use the app in the future and that they believed that there would be interest or a great deal of interest in the app in both their religious (59/72, 82%) and cultural communities (56/72, 78%). There was no significant difference between perceived demand in one’s religious community and in one’s cultural community ($P = .13$). No significant differences were observed between men and women regarding demand for the app. Table 4 shows participant ratings regarding demand for the app.

**Table 4.** Participant ratings regarding demand for the app (N=72).

<table>
<thead>
<tr>
<th>Demand characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frequency of use, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>&gt;5 times/wk, n (%)</td>
<td>3.11 (0.74)</td>
</tr>
<tr>
<td>3-5 times/wk, n (%)</td>
<td>23 (32)</td>
</tr>
<tr>
<td>1-2 times/wk, n (%)</td>
<td>35 (49)</td>
</tr>
<tr>
<td>Less than once/wk, n (%)</td>
<td>13 (18)</td>
</tr>
<tr>
<td><strong>Interest in future use, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Extremely interested, n (%)</td>
<td>33 (46)</td>
</tr>
<tr>
<td>Interested, n (%)</td>
<td>28 (39)</td>
</tr>
<tr>
<td>Somewhat interested, n (%)</td>
<td>11 (15)</td>
</tr>
<tr>
<td>Not interested, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Perceived demand</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Within participants’ religious community, mean (SD)</strong></td>
<td>4.29 (0.83)</td>
</tr>
<tr>
<td>A great deal, n (%)</td>
<td>36 (50)</td>
</tr>
<tr>
<td>Interested, n (%)</td>
<td>23 (32)</td>
</tr>
<tr>
<td>Somewhat interested, n (%)</td>
<td>11 (15)</td>
</tr>
<tr>
<td>Not really, n (%)</td>
<td>2 (3)</td>
</tr>
<tr>
<td><strong>Within participants’ cultural community, mean (SD)</strong></td>
<td>4.11 (0.82)</td>
</tr>
<tr>
<td>A great deal, n (%)</td>
<td>26 (36)</td>
</tr>
<tr>
<td>Interested, n (%)</td>
<td>30 (42)</td>
</tr>
<tr>
<td>Somewhat interested, n (%)</td>
<td>14 (19)</td>
</tr>
<tr>
<td>Not really, n (%)</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

### Practicality

Almost all participants (58/72, 81%) reported that it was very easy or easy to use the app and that they were very able or able to use and navigate the app (63/72, 88%). The monthly price of the app was perceived to be very cheap or cheap by most participants (55/72, 76%). Although only half (39/72, 54%) reported that paying the monthly fee of the app would be very
easy or easy, most participants (56/72, 78%) reported that they would be very likely or likely to pay the monthly fee. No significant differences were observed between men and women regarding the practicality of the app. Table 5 shows participant ratings regarding the practicality of the app.

Table 5. Participant ratings regarding the practicality of the app (N=72).

<table>
<thead>
<tr>
<th>Practicality characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Difficulty of use, mean (SD)</strong></td>
<td>1.76 (0.90)</td>
</tr>
<tr>
<td>Very easy, n (%)</td>
<td>35 (49)</td>
</tr>
<tr>
<td>Easy, n (%)</td>
<td>23 (32)</td>
</tr>
<tr>
<td>Somewhat easy, n (%)</td>
<td>10 (14)</td>
</tr>
<tr>
<td>Difficult, n (%)</td>
<td>4 (6)</td>
</tr>
<tr>
<td><strong>Ability to use and navigate the app, mean (SD)</strong></td>
<td>4.35 (0.70)</td>
</tr>
<tr>
<td>Very high ability, n (%)</td>
<td>34 (47)</td>
</tr>
<tr>
<td>Able, n (%)</td>
<td>29 (40)</td>
</tr>
<tr>
<td>Somewhat able, n (%)</td>
<td>9 (12)</td>
</tr>
<tr>
<td>Not able, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Perceived cost of the app, mean (SD)</strong></td>
<td>2.46 (0.86)</td>
</tr>
<tr>
<td>Very cheap, n (%)</td>
<td>29 (40)</td>
</tr>
<tr>
<td>Cheap, n (%)</td>
<td>26 (36)</td>
</tr>
<tr>
<td>Somewhat cheap, n (%)</td>
<td>8 (11)</td>
</tr>
<tr>
<td>Expensive, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Ease of paying monthly fee, mean (SD)</strong></td>
<td>2.35 (0.97)</td>
</tr>
<tr>
<td>Very easy, n (%)</td>
<td>16 (22)</td>
</tr>
<tr>
<td>Easy, n (%)</td>
<td>23 (32)</td>
</tr>
<tr>
<td>Somewhat easy, n (%)</td>
<td>26 (36)</td>
</tr>
<tr>
<td>Difficult, n (%)</td>
<td>6 (8)</td>
</tr>
<tr>
<td>Very difficult, n (%)</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Willingness to pay monthly fee, mean (SD)</strong></td>
<td>4.14 (0.83)</td>
</tr>
<tr>
<td>Very likely, n (%)</td>
<td>28 (39)</td>
</tr>
<tr>
<td>Likely, n (%)</td>
<td>28 (39)</td>
</tr>
<tr>
<td>Somewhat likely, n (%)</td>
<td>14 (19)</td>
</tr>
<tr>
<td>Unlikely, n (%)</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

**Mental Health and Well-Being Improvement Over Time**

Regarding mental health symptoms, participants reported significant improvement in perceived stress ($P=.004; d=0.35$), depression ($P=.02; d=0.29$), and anxiety ($P=.01; d=0.32$). Effect sizes were in the small to medium range. Regarding well-being, participants reported significant improvement in satisfaction with life ($P<.001; d=0.52$), spiritual well-being ($P<.001; d=0.48$), religious commitment ($P=.02; d=0.28$), and racial or ethnic identity development ($P=.048; d=0.24$). Effect sizes were in the small to medium range. Table 6 shows the $t$ test reporting of improvements.
Table 6. Paired-sample 1-tailed *t* test for mental health and well-being outcome improvement from time 1 and time 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time 1, mean (SD)</th>
<th>Time 2, mean (SD)</th>
<th><em>t</em> test (<em>df</em>)</th>
<th><em>P</em> value</th>
<th>Cohen <em>d</em> (effect size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>2.72 (0.66)</td>
<td>2.52 (0.71)</td>
<td>2.95 (76)</td>
<td>.004</td>
<td>0.35</td>
</tr>
<tr>
<td>Depression</td>
<td>0.84 (0.53)</td>
<td>0.70 (0.57)</td>
<td>2.46 (76)</td>
<td>.02</td>
<td>0.29</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.12 (0.59)</td>
<td>0.98 (0.58)</td>
<td>2.71 (76)</td>
<td>.01</td>
<td>0.32</td>
</tr>
<tr>
<td>Religious commitment</td>
<td>3.84 (0.71)</td>
<td>4.01 (0.61)</td>
<td>−2.35 (76)</td>
<td>.02</td>
<td>0.28</td>
</tr>
<tr>
<td>Spiritual well-being</td>
<td>4.47 (0.80)</td>
<td>4.89 (0.83)</td>
<td>−4.10 (76)</td>
<td>&lt;.001</td>
<td>0.48</td>
</tr>
<tr>
<td>Satisfaction with life</td>
<td>4.68 (1.26)</td>
<td>5.31 (0.97)</td>
<td>−4.43 (76)</td>
<td>&lt;.001</td>
<td>0.52</td>
</tr>
<tr>
<td>Racial or ethnic identity</td>
<td>3.96 (0.63)</td>
<td>4.09 (0.58)</td>
<td>−2.01 (76)</td>
<td>.048</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Mental Health and Well-Being Improvement With Objective App Use

Participants used the app during the study period (approximately 4 weeks) for an average of 45.83 (SD 111.90) minutes and a total of 5.77 (SD 8.19) days. Partial correlations were conducted, controlling for baseline variable scores, between total minutes and total days of app use and outcome variables. When controlling for the outcome variable at baseline, greater app use was associated with decreased stress and depressive and anxiety symptoms and increased spiritual well-being at the postintervention time point (Table 7). These results suggest that greater app use is associated with decreased mental health symptoms and increased spiritual well-being over time.

Table 7. Partial correlations between app use and dependent variables at the postintervention time point (N=77).

<table>
<thead>
<tr>
<th></th>
<th>Total minutes</th>
<th><em>P</em> value</th>
<th>Total days</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>−0.20</td>
<td>.10</td>
<td>−0.30</td>
<td>.01</td>
</tr>
<tr>
<td>Depression</td>
<td>−0.27</td>
<td>.02</td>
<td>−0.29</td>
<td>.01</td>
</tr>
<tr>
<td>Anxiety</td>
<td>−0.36</td>
<td>.002</td>
<td>−0.39</td>
<td>.001</td>
</tr>
<tr>
<td>Religious commitment</td>
<td>0.16</td>
<td>.20</td>
<td>0.22</td>
<td>.07</td>
</tr>
<tr>
<td>Spiritual well-being</td>
<td>0.26</td>
<td>.03</td>
<td>0.30</td>
<td>.01</td>
</tr>
<tr>
<td>Satisfaction with life</td>
<td>0.15</td>
<td>.22</td>
<td>0.20</td>
<td>.09</td>
</tr>
<tr>
<td>Racial or ethnic identity</td>
<td>0.12</td>
<td>.32</td>
<td>0.06</td>
<td>.67</td>
</tr>
</tbody>
</table>

Qualitative Analyses

Acceptability

When the interview participants (n=15) were asked what they liked most about the app, the types of responses fell into 2 main categories: general aspects and specific features.

General Aspects of the App

In total, 20% (3/15) of the participants reported that they mostly liked the wide range of topics and content available on the app, whereas 13% (2/15) referred to its relatability. Others (3/15, 20%) mentioned liking the most how convenient the app was to use via their phone, compared with hard-copy methods of accessing spiritual content, and how educational it was for them (2/15, 13%). Other participants mentioned liking the community aspect of the app (1/15, 7%) and the overall look of the app and its interface (1/15, 7%).

Specific Features of the App

A total of 27% (4/15) of the interviewees reported that they most liked the notifications that reminded them to pray regularly, and 13% (2/15) reported that they particularly liked the podcasts: “I’ve been getting addicted to the podcasts very early in the morning.”

In total, 27% (4/15) of the interview participants mentioned that they especially liked the daily scriptures or Bible passages. Other favorite features mentioned by individual participants included the daily motivational messages (1/15, 7%), daily prayer time (1/15, 7%), and streaks (1/15, 7%) enabling them to track their own progress. The participants were also asked which aspects of the app they were most interested in. In descending order of frequency in which they were mentioned, these were reported to be daily prayers (6/15, 40% of the participants), preaching and sermons (5/15, 33%), daily motivations (3/15, 20%), scriptures (3/15, 20%), podcasts (3/15, 20%), books (1/15, 7%), and storylines (1/15, 7%).

What Participants Liked the Least About the App

When asked what they liked the least about the Pray.com app, 60% (9/15) of participants reported that there was nothing they did not like about the app.

Among the 40% (6/15) who did report on features that they disliked or liked less, several of these were related to technical issues encountered when using the app (eg, delays in reminders, delays in content updates, and inability to save activity when exiting).

Individuals also commented on morning prayers being too long (1/6, 17%) and sermons being too short (1/6, 17%) and disliking the 21-Day Prayer Journeys (1/6, 17%).
aspects of Pray.com they were the least interested in, 40% (6/15) indicated that there were no aspects they were uninterested in. A total of 13% (2/15) said that they were the least interested in certain preachers or those they knew less about, and 7% (1/15) referred to content perceived as irrelevant to them, such as anxiety meditations. In addition, some participants said that they were least interested in movies (2/15, 13%), meditations (1/15, 7%), prayer journals (1/15, 7%), and prayers (1/15, 7%).

**Intent to Continue Use**

We asked the participants what factors would influence their decision to continue using the Pray.com app. In total, 60% (9/15) indicated that they would continue to use it because of the beneficial effect it had had on their spirituality, especially the notifications that reminded them to pray. Others stated that the convenience and ease of use of the Pray.com app were factors that influenced their intention to continue using it (7/15, 47%) and that range, quality of content, or continued access to content (3/15, 20%) would influence their continued use.

**Perceived Appropriateness**

Many participants mentioned that they had experienced positive impacts of the Pray.com app on their mental health. A total of 27% (4/15) reported that these benefits were due to a feeling of greater connectedness, either with God or with others who share their religious beliefs. In total, 20% (3/15) described how they had benefited from the ease of being able to locate mental health content on the app. A total of 20% (3/15) of the participants reported that their levels of stress or depression had decreased since they had started using the app. Some participants cited specific types of content that had helped their mental health: meditations (3/15, 20%), daily podcasts (2/15, 13%), scriptures (1/15, 7%), and songs (1/15, 7%).

The participants were also asked how the app had influenced their feelings about meditation and mindfulness. In response, 40% (6/15) indicated that they had an improved understanding or that using the app had changed their misconceptions about meditation and mindfulness. In total, 27% (4/15) of the participants described how the app had helped them enter a meditative state or how much they had enjoyed the meditations.

**Impacts on Spirituality**

When asked about the impact of the Pray.com app on their spirituality, 33% (5/15) of the participants indicated that it had helped them develop spiritual habits or increase their familiarity with the scriptures. A total of 27% (4/15) of the participants referred to having increased their frequency of praying since they had started using the app. Another 27% (4/15) of the participants explained how the app had given them a more spiritual focus or direction in their lives. In total, 13% (2/15) of the interviewees described ways in which the app had taught them to pray more effectively. A total of 13% (2/15) highlighted ways in which they had gained new spiritual insights or understanding through using the Pray.com app.

**Fit Within Culture**

When asked whether the Pray.com app was a good fit with their religious beliefs or worldview, 20% (3/15) of the participants commented that all aspects were a good fit. In total, 67% (10/15) of the participants specified ways in which particular content or preachers on the app were a good fit with their religious beliefs or worldview. When asked which aspects were not a good fit with their religious beliefs or worldview, 20% (3/15) cited specific content not relevant to them, such as parenting content, or that they or others in their community might not agree with, such as meditation or options for gift buying on the app.

**Fit Within Cultural and Ethnic Identity**

A total of 60% (9/15) of the interviewees specifically said that the Black community was well represented among the preachers. Although one participant expressed a perception that Black preachers only accounted for approximately 20% of all those on the app, he did not see this as a problem. A further 40% (6/15) of the interviewees similarly commented on how inclusive and nondiscriminatory Pray.com was. One participant mentioned that the app was a good fit with their cultural identity because of the type of language used, which he could relate to and understand.

**Demand**

In total, 80% (12/15) of the interviewees made suggestions for ways of improving the app to make it a better fit for them personally. These fell broadly into the following categories: features, technical improvements, and cost.

**Features**

Interviewees suggested more or different types of notifications, the ability to interact with others on the app (eg, in a forum or groups), or being able to post questions or comments to the preachers.

**Technical Improvements**

A total of 27% (4/15) of the participants said that technical improvements would make the app a better fit for them personally. These generally referred to specific technical problems they had encountered (eg, allowing the app to continue running while using the phone and inability to save content).

**Cost**

In total, 13% (2/15) of the participants suggested improvements related to the cost of subscriptions, including continued free access to the full app, a payment plan to make subscriptions more affordable, and a free trial option.

**Practicality (Ease of Use)**

Interviewees mentioned specific aspects or features of the app that were easy for them to use. A total of 40% (6/15) of the participants stressed that navigating the app was particularly easy for them. Other individuals commented that the podcasts or live prayers (3/15, 20%), meditations (1/15, 7%), website (1/15, 7%), and registration process (1/15, 7%) were easy to use. A small number of participants did describe some difficulties they had encountered when using the app. In total, 13% (2/15) of the participants explained that they had not realized that there was a notification option that they needed to activate. Others mentioned difficulty in locating the streak data (1/15, 7%) and using the display on the side of the podcasts (1/15, 7%).
Discussion

Principal Findings

Overview

The purpose of this study was to determine the feasibility (ie, acceptability, demand, and practicality) and preliminary effectiveness of using the mobile app Pray.com on mental health (ie, stress and depressive and anxiety symptoms) and well-being (ie, satisfaction with life, spiritual well-being, religious commitment, and racial or ethnic identity development) among REM participants. We also explored the associations between objective app use (total minutes and days) and the outcomes measured. In addition, follow-up interviews were conducted to provide further insight into feasibility outcomes. This is one of the first studies to assess the feasibility and preliminary effectiveness of a faith- and prayer-based app in REM individuals for the purpose of improving mental health and well-being.

Overall, participants reported high levels of satisfaction with the app, intent to continue use, appropriateness of the app, and fit within their culture. Interestingly, most participants (62/72, 86%) did not meet the app use prescription provided by the researchers (ie, at least 5 min/d, 5 d/wk) but expressed interest in using the app in the future and perceived demand for it in their communities. In addition, participants reported that the app was easy to use and perceived it to be inexpensive. Only approximately half (39/72, 54%) of the participants reported that paying the monthly fee would be easy for them, but most (56/72, 78%) said that they would be willing to pay. Even though most participants did not use the app as often as prescribed, the results indicated that app use was associated with significant mental health improvements over time. Critically, with regard to preliminary effectiveness, participants reported improved mental health (ie, stress and depressive and anxiety symptoms) and well-being (ie, satisfaction with life, spiritual well-being, religious commitment, and racial or ethnic identity development) over the course of the study despite relatively low average levels and high variability of app use. Greater app use was significantly associated with improvements in mental health and spiritual well-being. These findings suggest that Pray.com is feasible for REM adults. Future research is warranted to determine the efficacy of Pray.com on mental health and well-being outcomes in REM populations using randomized controlled trials.

Feasibility

Using a faith and prayer mobile app for mental health and well-being was found to be feasible in REM individuals, particularly for Black and African American adults (who constituted most of our sample). Most of the participants (64/72, 89%) expressed high satisfaction with the app and that the app was a good fit with their cultural and religious beliefs or worldview. A number of interviewees (6/15, 40%) also indicated that they perceived a fair representation of different cultural backgrounds in the app content. The findings related to cultural fit are promising as, historically, REM groups have not been well represented in apps, especially in apps that claim to have health benefits [60]. Ramos et al [61] have suggested that apps with a focus on improving mental health may have the ability to reduce unmet mental health needs if users believe that the content is appropriate for them and fits within their culture. Others have suggested that apps should be adapted to various cultures for increased benefits [42,62]. However, Black and African American populations are still underrepresented in mobile health research, limiting our ability to inform the development and design of apps for REM groups [63]. Our qualitative findings show that a small subset of participants reported that cultural fit could be increased if content such as parenting was improved or potential culturally controversial content such as meditation or options for gift buying on the app were addressed. This is one of the first studies to explore a faith-based app for mental health and well-being among REM groups. These initially promising findings suggest that more research in this area is warranted.

The findings on feasibility and cultural fit in our study are also exciting in light of the sometimes mixed reception of mindfulness and meditation in certain cultural and religious groups [64,65]. The positive reception and perceived cultural fit of the Pray.com app in a sample of religious REM individuals provides evidence that these apps may be an important avenue to help introduce the benefits of prayer, mindfulness, and meditation in a context that is experienced as a good fit with one’s cultural and religious worldview.

Although most participants (>54/72, 75%) reported high levels of feasibility across the board, one area in which responses were mixed was cost-effectiveness. Namely, although approximately three-quarters (56/72, 78%) of the participants expressed willingness to pay the monthly fee for the app, only approximately half (39/72, 54%) expressed that it would be easy for them to do so. This was supported by the qualitative data, in which some participants suggested improvements related to the cost of subscriptions (eg, a payment plan and free trial option). Individuals considering subscribing to an app for the purpose of improving or maintaining mental health and well-being should weigh the costs and potential benefits of subscribing to the app. Thus, when developing or implementing apps to provide access to faith and prayer content to various cultural groups, it is critical to keep cost-effectiveness and affordability in mind.

Previous research has demonstrated that overall product satisfaction and perceived quality are directly linked to increased purchase intentions, whereas increased product involvement may indirectly increase purchase intentions through overall satisfaction and perceived product quality [66,67]. Although only half of the sample expressed that it would be easy to pay the monthly fee, participants did say that paying the monthly fee was feasible and were interested in continuing use. Hence, despite some perceived expensiveness of the app, participants reported intent to continue use, likely based on high satisfaction with the app and its high relevance to improving mental and spiritual well-being.
Preliminary Effectiveness on Mental Health and Well-Being

Participants reported improved mental health as a result of using the Pray.com app. Specifically, participants reported significant improvements in mental health symptoms, including decreased stress, depression, and anxiety symptoms, over the course of the study. This was confirmed in the qualitative interviews; interviewees noted that their stress and depression levels decreased because of their increased hope and enjoyment of the present moment as encouraged by app content (eg, daily podcasts, scripture, and songs). In addition, the qualitative interviews suggested that it was easy to search for and access religious content related to stress, depression, and anxiety on the app. Prior work has demonstrated that spirituality can help operate as a buffer against stress and adversity [68]. To the degree that increased involvement with spiritual content can facilitate coping and help individuals in periods of stress, future research should explore potential mediators responsible for such benefits (eg, spiritual meaning [68]) and the contextual factors in which such benefits are maximized.

Interestingly, participants reported an initial hesitancy toward meditation, but interviewees reported that engagement with the meditation content on the app helped them enter a meditative state, which reportedly helped them face problems directly and better absorb information. This is an important finding given that some conservative Christians may hold negative views toward meditation because of its relationship with Buddhism [65]. For example, some more theologically conservative religious individuals may mistrust practices founded on traditions outside their own religion or view them as taboo. Accordingly, there might be hesitation to engage with content that does not squarely align with one’s religiously endorsed practices or that may not be prescribed by one’s religious teachings. Future work should explore the boundary conditions of these hesitations and what predicts such attitudes toward these practices.

Furthermore, the sample reported significantly greater well-being over time over the course of the study (ie, spiritual well-being, religious commitment, life satisfaction, and racial or ethnic identity development). Along these lines, interviewees expressed that engagement with the app (eg, podcasts, scripture reading, and meditations) positively affected their faith, spiritual well-being, and daily spiritual practices. Such findings are consistent with those of previous research suggesting that religious and spiritual identity and engagement are positively associated with well-being across REM groups [22,24,32]. Our findings are also consistent with those of other app-based mental health interventions that suggest that apps may serve as a self-management tool or adjunctive treatment [69,70]. However, app-based interventions have typically lacked diversity in their samples, and therefore, considering the preliminary nature of this study, outcome consistency cannot yet be determined [61].

The finding that participants reported higher levels of racial or ethnic identity over the course of the study was interesting and warrants further exploration. Religion and spirituality have been theorized to be important cultural factors in promoting positive psychological well-being in people of color [17]. For example, in a sample of African American individuals, religious commitment was related to higher levels of racial or ethnic identity and satisfaction with life [23]. It may be that the religious content on the app can provide a way to help encourage positive messages about one’s racial or ethnic identity and help people of color cope with and counter negative messages or narratives related to one’s racial/ethnic identity that are experienced in the broader culture (eg, racism and microaggressions). To be sure, future research sampling from a broader range of REM groups is important.

Given that the literature consistently demonstrates a significant association between religion or spirituality and well-being among REM populations, this app may provide a supplemental digital avenue to increase mental and spiritual well-being among REM groups, with the potential for greater improvement with increased use. This is especially important given the historical lack of access to mental health resources among African American and other REM individuals [71]. Future research should use rigorous methodological designs that rely on random assignment to experimental and control conditions, such as randomized clinical trials, to establish the causal effect of using a faith and prayer app for the purpose of improving mental health and well-being in REM groups.

Mental Health and Well-Being Improvement Association With Objective App Use

There was considerable variability in the degree to which participants used the app. Not all participants adhered to the instructions to use the app 5 minutes a day, 5 days per week. Greater objective app use was associated with decreased stress as well as depressive and anxiety symptoms. On average, participants used the app for 45.83 (SD 111.90) minutes and a total of 5.77 (SD 8.19) days throughout the course of the study. This is not consistent with what participants self-reported, and average app use across the study period was not consistent with what we asked participants to do (use the app for 5 min/d, 5 times/wk). Adherence to web-based and digital interventions is typically lower than that to in-person interventions [72], and future research should implement strategies to promote and ensure participation.

Examining well-being outcomes, objective app use was only significantly associated with an increase in spiritual well-being. Interestingly, improvements in satisfaction with life, religious commitment, and racial or ethnic identity development were not associated with app use. Religious commitment was also high in our sample at baseline as 73% (56/77) always identified with a religion and 26% (20/77) reported that they were new identifiers or reidentifiers, and it may take longer than 4 weeks for a religiously committed sample to observe significant changes in religious commitment. Most participants (57/77, 74%) attended in-person services at least once a week, and almost 25% (18/77, 23%) attended at least once or twice a month, whereas 56% (43/77) participated in a religious group at least once a week and 38% (29/77) participated in a religious group once or twice a month. Religious and spiritual practices outside of an app may be one reason why well-being outcomes significantly improved at the postintervention time point but only spiritual well-being was associated with app use. Future
research should consider testing the use of a technology-mediated religious or spiritual practice against samples who attend and do not attend in-person religious services or activities.

In addition, behavior change is complex and may affect individuals in a variety of ways [73]. It is possible that the amount of behavior requested (5 times a week for 4 weeks) was not appropriate for each participant. Although satisfaction with life, religious commitment, and racial or ethnic identity development outcomes were not significantly associated with app use, merely having access to technology-mediated religious or spiritual practices may have contributed to the significant improvements observed. Others have reported that perceived access may be no less valid than actual access and may be a stronger predictor of use than actual access [67,74]. Future research should assess the amount of time spent in digitally mediated religious and spiritual practices that are feasible and elicit changes in well-being outcomes such as satisfaction with life, religious commitment, and racial or ethnic identity development.

Limitations and Future Research
The results of this study should be interpreted within the context of its limitations. First, this was a feasibility study, and we found initial evidence to our hypothesis that our app was perceived as feasible by participants. Although small effect sizes were observed, clinical significance cannot be determined at this time. We encourage future research that uses a more rigorous methodology that relies on longer study periods and random assignment to experimental and control conditions, such as randomized clinical trials. A longer randomized controlled trial would be beneficial for establishing the causal effect of using a faith and prayer app for the purpose of improving mental health and well-being. Randomization and comparison with a control condition will also limit the potential of confounding factors such as possible COVID-19 impact as well as social and economic characteristics. In addition, all outcome measures were self-reported, and the findings should be interpreted based on the potential bias that self-reporting introduces, such as social desirability bias.

Second, a large percentage of participants who initially responded to our survey were determined to be inauthentic responses (eg, bots and duplicate IP addresses). This is a common problem when using the internet and social media to recruit participants. Once we realized the problem, we implemented additional strategies to ensure data quality, such as bot detection and reCAPTCHA and checking responses. Future research should implement these strategies from the beginning of data collection. In addition, it may be helpful to require a face-to-face meeting with potential participants before they enter the study.

Third, our sample predominantly comprised Black or African American individuals who were aged between 26 and 36 years and reported being affiliated with the Christian faith. Our findings may not be generalizable to other REM populations, age groups, or religious or spiritual faiths. For example, religion or spirituality may be experienced differently by older individuals [75], leading to different needs regarding interventions. In addition, as participants self-selected into the study, they may have had expectancy effects regarding the effectiveness of faith-based interventions and may have experienced confirmation bias regarding the effects of using the app. Future work including using purposeful sampling with greater racial or ethnic, denominational, and age range diversity would be desirable to test the generalizability of our findings.

In addition, information from these studies could be used to help increase the cultural and religious inclusivity of the app.

Fourth, because of the web-based and autonomous nature of the study, we were not able to control how many minutes and days the participants actually engaged with the app. At the onset of their participation, participants were instructed to “engage in the app at least 5x/week and at least 5 minutes/day.” Although our results suggest that greater app use was associated with improved mental health symptoms and increased spiritual well-being, the study results may not have captured the full impact of app use as accurately as a study design in a more controlled setting would have. Future research should strive to establish causal mechanisms of a faith-based app that may elicit improved outcomes and use engagement strategies to ensure adequate and appropriate app use to better capture the relationship between technology-mediated religious and spiritual practice and mental health symptoms and well-being. Future research should also collect subscription rates after the study to increase the validity of the feasibility findings.

Conclusions
In conclusion, this is the first study to assess the feasibility of a faith and prayer app for mental health (ie, stress, depression, and anxiety) and well-being (ie, religious commitment, spiritual well-being, satisfaction with life, and racial or ethnic identity) in a sample of REM individuals. Our findings suggest that the use of a faith and prayer app (ie, Pray.com) is feasible and may be significantly impactful for the improvement of mental health symptoms and spiritual well-being in REM individuals and their communities, especially Black and African American individuals. Participants also reported some critical feedback about some of the content as well as about the technological interface of the app. These findings should be interpreted based on the preliminary nature of the study and the context of its limitations. This study lays the foundation for future work to be conducted in REM groups to assess the impact of technology-mediated religious or spiritual practice on health and well-being.

Acknowledgments
The authors would like to thank Todd Hall of Biola University for his contribution in assisting with the institutional review board process.
Data Availability
The deidentified data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
Authors JNH and DRVT serve on a consulting basis for Pray.com research but receive no incentives from the growth of Pray.com. Author BL is a paid scientist at Pray.com but is not paid for the results of the research, only to conduct the research. Authors SZ and LJ declare they have no financial interests or incentives from the growth of Pray.com. Author JH discloses they have equity stake in Pray.com but equity is not dependent upon the results of the research.

References


40. Key findings from the global religions futures project. Pew Research Center. URL: https://www.pewresearch.org/religion/2022/12/21/key-findings-from-the-global-religious-futures-project/ [accessed 2024-01-10]


46. Key findings from the global religions futures project. Pew Research Center. URL: https://www.pewresearch.org/religion/2022/12/21/key-findings-from-the-global-religious-futures-project/ [accessed 2024-01-10]


Abbreviations

HADS: Hospital Anxiety and Depression Scale
MEIM: Multigroup Ethnic Identity Measure–Revised
PSS-10: Perceived Stress Scale–10
RCI-10: Religious Commitment Inventory–10
REM: racial and ethnic minority
SWBS: Spiritual Well-Being Scale
SWLS: Satisfaction With Life Scale

©Breanne Laird, Sabrina Zuniga, Joshua N Hook, Daryl R Van Tongeren, Lynda Joeman, Jennifer Huberty. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Attitudes Toward School-Based Surveillance of Adolescents’ Social Media Activity: Convergent Parallel Mixed Methods Survey

Colin Burke¹,²*, PhD; Cynthia Triplett¹,³*, MPH, MA; Caryn Kseniya Rubanovich³,⁴,⁵, MSc; Melissa M Karnaze³, PhD; Cinnamon S Bloss¹,³, PhD

1Herbert Wertheim School of Public Health and Longevity Science, University of California San Diego, La Jolla, CA, United States
2Department of Sociology, University of California San Diego, La Jolla, CA, United States
3Center for Empathy and Technology, Institute of Empathy and Compassion, University of California San Diego, La Jolla, CA, United States
4Joint Doctoral Program in Clinical Psychology, San Diego State University, San Diego, CA, United States
5Joint Doctoral Program in Clinical Psychology, University of California San Diego, La Jolla, CA, United States
*these authors contributed equally

Corresponding Author:
Cinnamon S Bloss, PhD
Herbert Wertheim School of Public Health and Longevity Science
University of California San Diego
9500 Gilman Drive MC0811
La Jolla, CA, 92093-0811
United States
Phone: 1 (858) 534 2230
Email: cbloss@eng.ucsd.edu

Abstract

Background: US schools increasingly implement commercially available technology for social media monitoring (SMM) of students, purportedly to address youth mental health and school safety. However, little is known about how SMM is perceived by stakeholders, including the students who are the focus of these efforts.

Objective: We aimed to assess attitudes toward SMM in schools among 4 stakeholder groups and examine reasons for holding supportive, neutral, or unsupportive views toward the technology. We also sought to explore whether any differences in attitudes were associated with binary sex, race, ethnicity, sexual orientation, or gender identity.

Methods: In October 2019, we conducted a convergent parallel mixed methods web-based survey of young adults (aged 18-22 y; n=206), parents (n=205), teachers (n=77), and school administrators (n=41) via Qualtrics web-based panels. We included Likert-type survey items to assess perceived benefits, risks, and overall support of SMM in schools and test for differences based on stakeholder group or demographic characteristics. We also included open-ended questions, and the responses to these items were analyzed using thematic content analysis of reasons given for holding supportive, neutral, or unsupportive views.

Results: The tests of group differences showed that young adults perceived lower benefit (P<.001) as well as higher risk (P<.001) and expressed lower overall support (P<.001) of the use of SMM in schools than all other stakeholder groups. Individuals identifying as nonheterosexual also perceived lower benefit (P=.002) and higher risk (P=.02) and expressed lower overall support (P=.02) than their heterosexual counterparts; respondents who identified as people of racial and ethnic minorities also perceived higher risk (P=.04) than their White counterparts. Qualitative thematic content analysis revealed greater nuance in concerns about SMM. Specifically, the primary reasons given for not supporting SMM across all stakeholder groups were (1) skepticism about its utility, (2) perceived privacy violations, and (3) fears of inappropriate or discriminatory use of the data. Within the young adult group in particular, concerns were also raised about (4) unintended and adverse consequences, including the erosion of trust between students and school institutions and administrators, and the chronic adverse effects of constant or prolonged surveillance. Thematic analysis also showed that individuals in every stakeholder group who indicated overall support of SMM were likely to cite the potential for enhanced school safety as the reason. Young adults’ overall stances toward SMM were the most polarized, either strongly for or strongly against SMM, and responses from teachers indicated similar polarization but more often favored support of SMM in schools.
Conclusions: This study found differing perspectives among stakeholder groups regarding SMM in schools. More work is needed to assess the ways in which this type of surveillance is being implemented and the range and complexity of possible effects, particularly on students.

KEYWORDS
social media; surveillance; privacy; public health; students; schools; social media monitoring; SMM; school safety; mental health; adolescents

Introduction

Background

In recent years, the United States has witnessed a troubling increase in both youth suicide [1] and incidents of school gun violence [2]. To address these concerning trends, public K-12 schools have implemented security measures, including the increased use of security cameras in schools as well as lockdown drills and protocols, based on data from the National Center for Education Statistics [1]. In addition, according to a recent survey, nearly 90% of school teachers reported that their school used technology during the 2021-2022 school year to track student activity on school-issued and personal devices, such as by accessing the content of students’ internet searches or remotely viewing students’ computer screens in real time [2].

As part of the increased surveillance of students, a growing number of schools have turned to commercially available social media monitoring (SMM) technology, which some companies claim will prevent harms on school campuses by monitoring students’ activity on the web, including, in some cases, activity that occurs outside of school hours [3-5].

SMM technology works by scanning public content posted by students on social media platforms such as Twitter, Facebook, and Instagram for certain words and phrases that might signal a threat of harm to oneself or others, in most cases without the explicit knowledge of students. The next step in the use of this SMM technology is the flagging of posts containing potentially problematic references to harmful behavior, such as suicide or self-harm, bullying, violence, or hate speech. When such posts are detected, the monitoring software alerts school officials, who can then notify teachers and parents or intervene by taking disciplinary measures or contacting school authorities or law enforcement. There is limited transparency regarding the specific workings of SMM (eg, algorithms, training data, and quality control) and limited evidence of its efficacy; for example, general overflagging of lesbian, gay, bisexual, transgender, queer, and similar minority (LGBTQ+)–related words has been noted across some SMM technologies, which raises issues of negative and disproportionate impact on certain groups of students and discrimination more broadly [3].

It has been difficult to determine exactly how many schools in the United States currently use SMM services. One of the first publicized uses of SMM technology was in September 2013, when the Glendale Unified School District in California hired the firm Geo Listening to monitor the social media content of 13,000 middle and high school students residing in their district [4,5]. Since 2013, the number of schools implementing SMM technologies has grown significantly. According to SmartProcure, a database for government purchase orders, in 2018, SMM services were purchased by schools to directly monitor the social media activity of >3 million students across 63 public school districts in the United States. This represented a 10-fold increase from 2013 when just 6 school districts were found by SmartProcure to have purchased such services [6].

More recently, Social Sentinel, a leading provider of SMM services, has claimed that it serves “thousands of schools in more than 35 states” [7]. The increasing frequency of anecdotal reports of SMM use in the media, usually in response to episodes of gun violence in schools, is consistent with this trend; for example, after the tragic shooting at Robb Elementary School in Uvalde, Texas, in May 2022, The Dallas Morning News reported that Uvalde was among at least 52 school districts in Texas alone that hired the firm Social Sentinel to monitor the social media content of its tens of thousands of students with the purported goal of preventing harm to students [7,8]. This statistic for just 1 SMM company may be a conservative estimate of school-based SMM across the United States.

Contrary to what might be expected, studies of general school surveillance practices suggest associations with decreased student perceptions of safety [9-12]; for instance, 1 study found that the use of security cameras outside of school was associated with higher perceived safety, but the use of cameras inside was associated with lower perceived safety, support, and equity [13]. Another study of American middle and high school students found that visible security measures (cameras, guards, and metal detectors) were associated with higher odds of students’ fear of exposures to violence, bullying and other harms at school [14]. Furthermore, a meta-analysis of qualitative studies found that students thought that the presence of closed-circuit television cameras often resulted in risky behaviors shifting from monitored areas to less-monitored areas (eg, hallways and restrooms) [15]. A 2021 survey conducted by the Center for Democracy and Technology also reported the “chilling” effects of web-based surveillance [16]. Specifically, 6 out of 10 students reported feeling uncomfortable expressing their true thoughts and feelings on the web if they knew they were being monitored. The report argued that chilling effects that curb exploration and self-expression could be especially problematic for minors and might also make students less likely to seek web-based resources for mental health, to their detriment [16].

Youth and minoritized communities are also likely to disproportionately experience unintended adverse effects of surveillance; for instance, the Center for Democracy and Technology report also speculated that web-based monitoring might pose a risk that LGBTQ+ students may be outed as a result of surveillance [16]. In addition, some minoritized youth
have more fraught relationships with institutions such as law enforcement or disciplinary frameworks; for example, students of color are known to face higher rates and severity of punishment than their White peers, and thus, if surveillance leads to punishment, the effects of surveillance are more likely to fall on them. In fact, there is anecdotal evidence that school-based SMM has led to false positives [17], particularly with students of color. In addition, SMM algorithms may not accurately process content written in nonstandard English or languages other than English, and therefore SMM can disproportionately single out and label as dangerous students who are more likely to use nonstandard English or slang [6].

Objectives

Although schools are increasingly deploying SMM technology, there has been no systematic assessment of how it is perceived by school stakeholders or how it might affect the students whom it purportedly aims to help. To address this gap, we conducted an exploratory survey assessing attitudes toward SMM in schools among 4 key stakeholder groups: school administrators, teachers, parents, and young adults. This survey included both closed-ended quantitative and open-ended qualitative questions to assess stakeholders’ attitudes. As a secondary aim, we also sought to statistically test for any differences in attitudes as a function of stakeholders’ self-reported gender, race, ethnicity, and sexual orientation.

Methods

Ethical Considerations

This study was reviewed and approved by the University of California, San Diego Office of Institutional Review Board Administration (191060) and received a waiver of signed consent. Each participant provided informed consent via radio button selection at the bottom of the web-based landing page that included written information about the study. Survey participants were compensated by Qualtrics.

Recruitment

From October 7 to 15, 2019, we conducted an 8-minute web-based survey of young adults and parents via Qualtrics web-based consumer panels, as well as of teachers and administrators via Qualtrics web-based business-to-business panels. Participants on these panels are recruited from various sources, including website recruitment, member referrals, targeted email lists, gaming sites, customer loyalty web portals, permission-based networks, and social media. Qualtrics validates consumer panel members’ names, addresses, and dates of birth via third-party measures, and panel members are subject to additional quality control measures such as LinkedIn matching, telephone calls to the participant’s place of business, and other third-party verification methods (provided by companies such as TrueSample, RelevantID, and Verity). Although we originally desired to have an equal representation of teachers and administrators, this was not feasible owing to cost for the recruitment service. We also oversampled parents and young adults with the reasoning that, to date, these groups have been largely absent from dialogue and decision-making pertaining to SMM.

Eligibility and Screening

Young adults were eligible if they were aged between 18 and 22 years and either a high school graduate or current high school student. Parents were eligible if they had children aged between 14 and 22 years. Teachers and administrators were eligible if they were employed in the education industry and were middle or high school teachers or administrators. For teachers and administrators, Qualtrics used a combination of the profiled information they had on file to target these professionals and screening questions at the beginning of the survey to confirm information for the specific survey respondents who qualified for the survey.

Survey Design

The survey measure included (1) screening questions (4-7 items, dependent on skip logic), (2) basic demographic questions (8 items), and (3) questions soliciting views about SMM (9 items) that were modeled after another survey on public views of genome editing [18]. In the last category (Multimedia Appendix 1), there were 7 items that were measured on a 7-point Likert-type scale ranging from 1=strongly disagree to 7=strongly agree: 3 items asked about the perceived efficacy of SMM for addressing school-related violence, bullying, and mental health issues (ie, potential benefits of SMM); 3 items asked about the level of concern for SMM as it relates to privacy, data misuse, and discrimination (ie, potential risks of SMM); and 1 item asked about the overall level of support of the use of SMM in schools. There were also 2 open-ended questions: “Please describe how you feel about middle schools and high schools monitoring students’ social media activity.” “Is there anything else you would like to share about this issue?” This combination of closed- and open-ended questions reflects our use of a convergent parallel mixed methods design in which the quantitative and qualitative data collection occurred concurrently.

Data Collection

We defined sample size quotas for the survey that aimed to collect responses from 200 young adults, 200 parents, 60 teachers, and 40 administrators. Qualtrics distributed the survey via a dashboard service whereby the survey would appear on a panel member’s dashboard if their profile indicated that they potentially met the inclusion criteria. We estimate that the survey was made available to between 14,400 to 16,000 individuals via this method, and 1600 individuals clicked a link to view the study information and consent page for the study. Of these 1600 individuals, 690 (43.13%) provided consent to participate in the study. Study data were collected and managed using the Qualtrics web-based survey platform. A soft launch to pilot-test the survey and check for any administration problems collected 30 responses that were used to establish quality benchmarks. Once data collection closed, responses were reviewed for completeness and quality in 2 phases. First, Qualtrics research panel staff filtered out all respondents who did not complete the survey or who completed the survey in less than half the median response time. Second, study team members filtered out any respondents who left gibberish in response to the open-ended questions.
Data Analysis

Overview

We generated descriptive statistics to summarize and compare the sociodemographic characteristics of study participants. Quantitative data analyses were conducted using SPSS software (version 28.0; IBM Corp), and significance was set at $P < .05$ for all analyses. To enhance interpretability, an SMM benefits score was created by summing responses on the first 3 survey items pertaining to potential benefits from SMM (ie, to help address mental health, bullying, and threats of harm or violence; range: 3-21, with higher scores suggesting greater perceived benefits). Across this set of survey items, the Cronbach $\alpha$ value was .853, indicating high internal consistency. An SMM risks score was created by summing responses on the next 3 survey items pertaining to the potential risks of SMM (ie, it potentially violates privacy, leads to abuse or misuse of information, and leads to potential discrimination; range: 3-21, with higher scores suggesting greater perceived risks), and the Cronbach $\alpha$ value was .828. The final quantitative survey item assessed overall support of SMM in schools.

Quantitative Analyses

We used 1-way analysis of covariance and Bonferroni-adjusted post hoc pairwise comparisons to test for statistically significant differences among the 4 stakeholder groups on (1) SMM benefits, (2) SMM risks, and (3) overall support, controlling for sex (female vs male), race and ethnicity (non-Hispanic White vs all other groups), and sexual orientation (heterosexual vs all other groups). Partial eta–squared ($\eta^2_p$) was used as a measure of effect size, and the values of 0.01, 0.06, and 0.14 represent small, medium, and large effects, respectively [19]. To examine our second question regarding differences in the perceptions of SMM as a function of demographic characteristics, a series of independent samples 2-tailed $t$ tests were conducted on the full sample to compare SMM benefits, SMM risks, and overall support of SMM by sex (female vs male), race and ethnicity (non-Hispanic White vs all other groups), and sexual orientation (heterosexual vs all other groups).

Qualitative Analyses

In analyzing the 2 open-ended questions, most participants answered such that their response to the second open-ended question was an extension of their answer to the first open-ended question. Thus, we considered responses to both items together and conducted thematic analysis with a contextualist lens [20]. Our inductive approach began by studying the short responses repeatedly to identify commonly discussed content, generating a list of 15 initial codes. A single coder then collated and refined the codes into a list of themes. As a final step, we integrated the data by merging the quantitative results with the qualitative results [21]. Specifically, the qualitative comments were sorted and read to identify similarities and differences within and among stakeholder groups, demographic groups, and levels of overall support of SMM in schools. In this way, we used triangulation to yield a more holistic understanding of the data and draw the conclusions set forth in the results presented.

Results

Sample Characteristics

A total of 690 individuals entered and consented to the survey. Of the 690 responses, 161 (23.3%) were identified as poor-quality completes and were removed from the data set. This yielded a final survey sample of 529 participants, which included young adults (n=206, 38.9%), parents (n=205, 38.8%), teachers (n=77, 14.6%), and school administrators (n=41, 7.8%). Table 1 provides descriptive statistics of the demographics of our sample.

Table 1. Descriptive statistics: Demographics of survey participants (n=529).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Administrators (n=41)</th>
<th>Teachers (n=77)</th>
<th>Parents (n=205)</th>
<th>Young adults (n=206)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>46.9 (12.9)</td>
<td>41.1 (10.6)</td>
<td>48.8 (10.9)</td>
<td>21.0 (1.3)</td>
</tr>
<tr>
<td>Sex: female, n (%)</td>
<td>32 (78)</td>
<td>48 (62.3)</td>
<td>119 (58)</td>
<td>117 (56.8)</td>
</tr>
<tr>
<td>Race and ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0.5)</td>
<td>8 (3.9)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (2.4)</td>
<td>3 (3.9)</td>
<td>10 (4.9)</td>
<td>12 (5.8)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>5 (12.2)</td>
<td>9 (11.7)</td>
<td>14 (6.8)</td>
<td>30 (14.6)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4 (9.8)</td>
<td>4 (5.2)</td>
<td>20 (9.8)</td>
<td>49 (23.8)</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>3 (1.5)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>30 (73.2)</td>
<td>62 (80.5)</td>
<td>170 (82.9)</td>
<td>132 (64.1)</td>
</tr>
<tr>
<td>&gt;1 race</td>
<td>3 (7.3)</td>
<td>2 (2.6)</td>
<td>6 (2.9)</td>
<td>11 (5.3)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (4.9)</td>
<td>1 (1.3)</td>
<td>4 (2)</td>
<td>10 (4.9)</td>
</tr>
<tr>
<td>Sexual orientation: heterosexual, n (%)</td>
<td>40 (97.6)</td>
<td>73 (94.8)</td>
<td>192 (93.7)</td>
<td>157 (76.2)</td>
</tr>
</tbody>
</table>
Quantitative Results

Stakeholder Group Comparisons

There were significant differences by stakeholder group in perceived SMM benefits, SMM risks, and overall support of SMM after controlling for sex, race, ethnicity, and sexual orientation. Table 2 provides the results of the analysis of covariance analyses, and Figure 1 shows the proportions within each group who were supportive, neutral, or unsupportive of SMM. Follow-up pairwise comparisons (data not shown) found that young adults were significantly different from all other groups and by comparison perceived lower benefit, higher risk, and less overall support of use of SMM in schools than parents, teachers, and administrators.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Administrators, mean (SD)</th>
<th>Teachers, mean (SD)</th>
<th>Parents, mean (SD)</th>
<th>Young adults, mean (SD)</th>
<th>F test (df)</th>
<th>P value</th>
<th>ηp²</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMM benefits (range: 3-21)</td>
<td>16.93 (3.12)</td>
<td>16.92 (3.37)</td>
<td>16.39 (3.74)</td>
<td>14.05 (4.48)</td>
<td>12.68 (3.519)</td>
<td>&lt;.001</td>
<td>0.068</td>
</tr>
<tr>
<td>SMM risks (range: 3-21)</td>
<td>12.20 (4.51)</td>
<td>12.83 (4.75)</td>
<td>13.60 (4.77)</td>
<td>15.21 (3.95)</td>
<td>7.11 (3.519)</td>
<td>&lt;.001</td>
<td>0.039</td>
</tr>
<tr>
<td>Overall support of SMM (range: 1-7)</td>
<td>4.76 (1.58)</td>
<td>4.92 (1.71)</td>
<td>4.91 (1.64)</td>
<td>3.84 (1.81)</td>
<td>14.02 (3.519)</td>
<td>&lt;.001</td>
<td>0.075</td>
</tr>
</tbody>
</table>

Figure 1. Overall opinion about the use of social media monitoring in schools by stakeholder group. Using data from question 7 of the survey, we collapsed all the agree categories (strongly agree, agree, and somewhat agree) into supportive and all the disagree categories (strongly disagree, disagree, and somewhat disagree) into not supportive; neutral refers to responses of neither agree nor disagree.

Demographic Group Comparisons

There were also significant differences by sexual orientation, race, and ethnicity. Specifically, nonheterosexual individuals perceived significantly lower benefit, higher risk, and less overall support of the use of SMM than their heterosexual counterparts. In addition, individuals identifying as people of racial and ethnic minorities perceived significantly lower benefit and higher risk than their non-Hispanic White counterparts. Tables 3, 4, and 5 provide the results of the 2-tailed t tests. There were no significant differences as a function of sex.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sex, mean (SD)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMM benefits (range: 3-21)</td>
<td>Male (n=210)</td>
<td>15.21 (4.17)</td>
<td>Female (n=316)</td>
</tr>
<tr>
<td>SMM risks (range: 3-21)</td>
<td>14.10 (4.30)</td>
<td>13.93 (4.72)</td>
<td>0.41 (524)</td>
</tr>
<tr>
<td>Overall support of SMM (range: 1-7)</td>
<td>4.38 (1.79)</td>
<td>4.56 (1.78)</td>
<td>−1.18 (524)</td>
</tr>
</tbody>
</table>
Table 4. Demographic group (sexual orientation) perceptions of social media monitoring (SMM) benefits, SMM risks, and overall support of SMM (n=529).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sexuality, mean (SD)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heterosexual (n=462)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMM benefits (range: 3-21)</td>
<td>15.86 (3.95)</td>
<td>3.19 (77.67)</td>
<td>.002</td>
</tr>
<tr>
<td>SMM risks (range: 3-21)</td>
<td>13.84 (4.55)</td>
<td>−2.4 (527)</td>
<td>.02</td>
</tr>
<tr>
<td>Overall support of SMM (range: 1-7)</td>
<td>4.55 (1.76)</td>
<td>2.3 (527)</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>All other groups (n=67)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Demographic group (race and ethnicity) perceptions of social media monitoring (SMM) benefits, SMM risks, and overall support of SMM (n=529).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Race and ethnicity, mean (SD)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Hispanic White (n=349)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMM benefits (range: 3-21)</td>
<td>15.85 (3.96)</td>
<td>1.94 (527)</td>
<td>.05</td>
</tr>
<tr>
<td>SMM risks (range: 3-21)</td>
<td>13.73 (4.68)</td>
<td>−2.08 (527)</td>
<td>.04</td>
</tr>
<tr>
<td>Overall support of SMM (range: 1-7)</td>
<td>4.53 (1.82)</td>
<td>0.78 (527)</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>All other groups (n=180)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Qualitative Results

Overall Stances Toward SMM and Stakeholder Group Differences

Generally, responses from the young adult cohort were the most polarized, with most respondents being either strongly in favor or strongly against the use of SMM in schools. Responses from teachers indicated similar polarization, but they were more commonly in favor of school-based SMM.

Reasons to Support SMM

Across all stakeholder groups, among those who indicated that they supported the use of SMM, the primary reason offered for this support was its potential utility to assist in identifying or preventing violence and bullying. One respondent noted as follows:

*Kids are bullied every single day and are taking their lives. They won’t hardly talk to anyone and it mostly happens on Facebook or Instagram and things of that sort. I believe it would be a good idea to slightly monitor social media.* [Young adult respondent 3MtgW]

Another respondent stated that SMM might be a valuable service “if watching children’s social post[s] could save a child from bullying, suicide or abuse” (Parent respondent 23gc8). Surveillance, according to respondents, could be 1 tool in a school or school district’s safety toolbox as “an extra layer of protection to make sure the school is safe” (Teacher respondent YQzyo) and “no different [than] installing metal detectors to screen for weapons” (Administrator respondent 3MGxE). Justification for respondents’ support ranged from feeling that SMM was “a necessary evil” (Parent respondent 27JO1) to feeling that SMM was “absolutely necessary” (Teacher respondent 3fjYb) and “a good and wise thing to do” (Young adult respondent BKwWJ).

Reasons for Concern About SMM

Overview

Across the stakeholder groups, individuals who were unsupportive of SMM cited similar reasons for their stance. Specifically, our qualitative analysis showed that the primary reasons given for not supporting SMM that were cited across all stakeholder groups were (1) skepticism about its utility, (2) perceived privacy violations, and (3) fears of inappropriate or discriminatory use of the data captured in SMM reports. Importantly, the critics of SMM also felt that SMM in schools could lead to (4) unintended and adverse consequences, such as the erosion of trusted relationships among students, parents, and schools or the chronic adverse effects of constant or prolonged surveillance. We expand on each of these areas in the following subsections, and additional relevant quotes are provided in Textbox 1.
Textbox 1. Examples of reasons provided for not supporting social media monitoring (SMM).

Skepticism About Utility of SMM

- Administrators: “Sometimes the content posted can also be figurative speech so you can’t really determine if they mean it or not. Example: ‘I’m so going to kill you.’ 99% of the time that term is used jokingly.” (Administrator respondent 6J00U)
- Teachers: “Well, I’m not sure how it would be done. It’s not easy to monitor...Hiring a social media monitoring staff seems just a bit too far for me. My school uses Twitter and Facebook to interact with students but there are so many app based forums which are far more popular with the youth. As is typical, us adults are ten years behind the kids in terms of tech.” (Teacher respondent 2YJac)
- Parents: “[I] post things on social media that I can understand but it may not exactly tell you who I am as a person. I’m not sure how true to light you can actually see a person as far as social media goes.” (Parent respondent 3GvVZ)
- Young adults: “[S]tudents would feel very violated and would try not to post revealing content or get around this by making another account with fake information.” (Young adult respondent 2rMEm)

Perceived Privacy Violations

- Administrators: “Invasion of privacy and over reach of responsibility.” (Administrator respondent 2CVC0)
- Teachers: “The harms and benefits of surveillance aren’t mutually exclusive. It’s a question of which are more important. In my opinion privacy and misuse are a bigger concern than mental health and bullying. There are other ways to deal with those problems. On the other hand, there are fewer ways to protect the students privacy and prevent schools from abusing their power. Who is watching the watcher?” (Teacher respondent 25G66)
- Parents: “What happens to the data and how are concerns then addressed? Schools often seem to do a poor job of intervening when kids are being bullied even without access to online social media accounts!” (Parent respondent 2vkC)
- Young adults: “I feel like it can help see bullying in schools and some threats to the school, but it does violate privacy. When a student steps into school their freedoms are revoked, but once they are out of school they have a right to privacy and monitoring would take away from that right.” (Young adult respondent 22Gtq)

Fears of Inappropriate or Discriminatory Use of Data

- Administrators: “I have concerns that middle and high school staffs are not fully equipped to deal with bullying and school violence. I fear that certain students from certain backgrounds will be discriminated against based on what they write/say on social media platforms.” (Administrator respondent 3241I)
- Teachers: “I am concerned how the information would be used—who would be ‘singled’ out for further actions?” (Teacher respondent 1GPQI)
- Parents: “I’m afraid schools will abuse it and students will get in trouble for every little thing school administration disapproves of.” (Parent respondent 3CUT0)
- Young adults: “I think it’d be a good idea but I also think that a lot of teachers would mistreat some students. I’ve had it happen to me before over something I didn’t even post and I can tell you it made me more depressed than before. The idea is good but you can’t trust teachers to be fair, they are only human and it could really hurt a student’s mental health when the teachers act like children themselves.” (Young adult respondent 2dM7k)

Skepticism About the Utility of SMM

The main question raised by those who did not support the use of SMM in schools was whether any SMM company, presumably operated by adults, could accurately and reliably catch troubling posts. Some felt that once students were aware of the monitoring, they would simply make their posts private, rendering the monitoring efforts ineffective. Stakeholders were also doubtful that teachers and administrators could correctly interpret students’ social media posts, given the loss of context and the challenge of deciphering irony in web-based messages. Finally, stakeholders questioned the overall feasibility of SMM because “adults are ten years behind the kids in terms of tech” (Teacher respondent 2YJac). On this point, another respondent wrote that “the threats are made through apps 9 times out of 10 where the message disappears” (Parent respondent yNNHHU), referring to apps that allow users to send text messages that are automatically deleted after a period of time or once read by the recipient.

Perceived Privacy Violations

A major concern expressed by those who were unsupportive of the use of SMM was that monitoring would violate student privacy because “kids have the right to a life outside of school” (Teacher respondent 2kRi) and such monitoring “removes a space where students can feel totally free to be themselves” (Young adult respondent 1mjXY). Other comments expressing concern ranged from those who simply stated that SMM was an invasion of privacy to those who felt discomfort with schools taking such actions:

"[I]t feels a bit weird. Like an over reaching of boundaries. It just feels not right for schools to be monitoring personal social media accounts usually meant for friends or family. [Young adult respondent 3QF2F]

The words “private” or “privacy” were explicitly mentioned in roughly a third of all young adult and administrator comments (72/206, 35% and 16/41, 39%, respectively), whereas parents and teachers mentioned these words less often (27/205, 13.2% and 15/77, 19%, respectively).
Fears of Inappropriate or Discriminatory Use of Data

The most frequently cited concern among those who were not supportive of SMM was the potential for SMM reports to be used in discriminatory ways (e.g., discriminatory punishment). However, these concerns were raised primarily by young adults and parents, with teachers and administrators citing this concern in only 2 instances. Respondents worried that SMM might “become too far reaching and subjective” (Parent respondent 3iqKX). Young adults, in particular, feared that they would get in trouble for simply “posting what they feel like or how they feel” (Young adult respondent 1dnr8) “because staff at the school may not agree with posts or get offended by them and cause them to feel negatively toward that student and treat them unfairly” (Young adult respondent WjMZk). More broadly, respondents across all stakeholder groups recognized the possibility for SMM to exacerbate unconscious or conscious biases.

Unintended and Adverse Consequences

Finally, respondents wondered how the use of SMM might have some unforeseen and adverse impact:

*It is a slippery slope...It seems like a good idea as far as safety, but I worry about what the information could be used for and if it will cause more trouble than good.* [Administrator respondent 2SIJs]

Young adult stakeholders, in particular, raised specific concerns about the potential for SMM to have the opposite of its intended effect (bold emphasis added by the authors); for instance, a young adult respondent noted the potential psychological and behavioral impact of SMM:

>*If I were to be monitored, I would simply not use social media at all. The idea of people overlooking my online presence is anxiety-inducing and should not be allowed.* [Young adult respondent 3lXio]

Similarly, another young adult respondent stated as follows:

*Monitoring students constantly can lead to a sense of paranoia as students are constantly being watched in real life, by their parents, teachers and if monitoring is enabled on social media.* [Young adult respondent 1lirF]

A different young adult respondent brought up the potential chilling effect of SMM on students:

*I feel like schools monitoring students’ social media activity is like being a helicopter parent which isn’t necessarily bad, but it may restrict the student’s freedom knowing they’re always watched, that if they say something someone doesn’t agree with, they may be punished.* [Young adult respondent 2uy40]

Another young adult respondent pointed out the potential strain on students’ relationship with educational institutions:

*That’s a big overstep, also considering developmental psychology of that age group that seems like it would not go over well at all with the students and would breed animosity towards the schools.* [Young adult respondent 32JeH]

All these examples indicate that young adults felt that SMM in schools, contrary to its stated purpose, might increase feelings of anxiety and paranoia, potentially leading to detrimental mental health outcomes and ultimately worsening student relationships with teachers, administrators, and the school system overall.

Discussion

Principal Findings

This study assessed attitudes toward SMM in schools and identified similarities and differences across groups of young adults, parents, teachers, and school administrators, as well as across demographic groups. We found that the young adults we surveyed perceived lower benefit as well as higher risk and expressed lower overall support of the use of SMM in schools than all other stakeholder groups. In addition, individuals identifying as nonheterosexual also perceived lower benefit as well as higher risk and expressed lower overall support than their heterosexual counterparts. Respondents identifying as people of racial and ethnic minorities also perceived higher risk than those identifying as White. Qualitative thematic analysis highlighted the nuances of stakeholder attitudes and found that individuals in every stakeholder group who indicated support of SMM were likely to cite enhanced school safety as the reason. Individuals who were unsupportive cited skepticism about the utility of SMM, perceived privacy violations, and fears of inappropriate or discriminatory use of data. Young adults, in particular, also raised concerns about unintended consequences, including the erosion of trust between students and school institutions and the chronic adverse effects of constant or prolonged surveillance. Taken together, this study provides some of the first empirical documentation of stakeholders’ attitudes toward the use of SMM technologies in schools and is a first step toward generating needed discourse around this emerging technology.

Although we anticipated a priori that young adults would express the most unfavorable views of SMM, the qualitative responses we received indicated that this group thoughtfully considered potential benefits as well as potential drawbacks of SMM in schools. Specifically, young adults across the board, including those who were neutral or generally supportive of SMM, raised concerns about privacy and discrimination. This suggests that even young adults who favor school-based SMM may be concerned about potential harms to students, including that it could lead to negative mental health outcomes, the opposite of its intent. These findings suggest that young people do see problems or concerning trends in their schools and see a need for intervention but are skeptical about whether SMM is an appropriate or effective solution to such problems. The young people in our survey also demonstrated a keen awareness of trends in social media and web-based communication that SMM service providers and clients need to be aware of when using SMM and interpreting social media content.

The diverging viewpoints between young adults and the other stakeholder groups is also an essential finding because students are the primary targets of SMM; yet, they have had very little decision-making power in the implementation of these
surveillance systems. The skepticism among young people toward these technologies further problematizes their general absence from the decision-making process, especially when schools that use SMM services often do so without students’ consent and, in some cases, without their knowledge [22]. Our results highlight how schools that consider or implement SMM should at a minimum engage in dialogue with students and recent graduates and consider how to make surveillance practices and policies more transparent. More research is also needed to better understand young adults’ concerns about currently evolving technologies and surveillance methods to minimize potential harm to students. This is particularly relevant for those who are not yet adults and may have fewer legal protections should school-based actions be taken against them based on social media data received from SMM companies. Similarly, constructing policies and ethical standards for SMM in schools would require bridging any gaps between the perceptions and knowledge of young adults and those of other stakeholder groups, perhaps by developing shared conceptual frameworks.

Expressed skepticism about SMM efficacy is also particularly salient, given that SMM continues to operate as a quickly moving target. The tragic school shooting in Uvalde, Texas, provides an unfortunate example of a case when SMM in schools did not function as claimed. The Dallas Morning News reported that according to records from GovSpend, an organization that tracks state and local government spending, the Uvalde Consolidated Independent School District was among at least 52 Texas school districts that hired Social Sentinel to monitor the social media activity of its tens of thousands of students [7,8]. However, like other SMM service providers, Social Sentinel only monitors public social media activity and, consequently, was unable to detect the shooter’s private communications related to the shooting [23]. The failure of SMM to prevent this tragedy has raised questions about the efficacy of such technologies and whether the potential harms of SMM might outweigh the potential good [24]. Our findings suggest that any cost-benefit analysis of SMM in schools must directly probe perceived costs and benefits from the members of all stakeholder groups and seek to recruit individuals across the spectrum of attitudes because, although there were commonalities in attitudes expressed across participants, the groups did diverge on issues.

In particular, discrimination was more important to parents and young adults than to administrators, and the group approximating the population considered vulnerable of school-attending youth—young adults—provided richer descriptions of unintended consequences, of which other stakeholder groups and SMM companies need to be aware. It will also be important to gather insights from the public and individuals situated within the technology industry and predictive sciences who can provide expert opinion on what constitutes efficacy regarding purported SMM benefits.

We found that SMM perceptions also significantly differed by respondent sexual orientation such that nonheterosexual respondents saw fewer benefits (P=.002) and greater risks of SMM (P=.02), leading to less overall support of SMM (P=.02). LGBTQ+ individuals, as a group, have been reported to be frequent social media users [25], more so than heterosexual individuals [26] or the general public [27]. Previous literature has underscored the importance of social media for LGBTQ+ individuals. Specifically, social media is used as a space for identity exploration, social support, making platonic or romantic connections, and finding resources [28,29]. Moreover, social media has been described as a “safe space” [30] for LGBTQ+ youth. Given the levels of anonymity [31] or privacy settings [32] that social media can afford, LGBTQ+ individuals can manage how or whether to disclose their identities as well as express themselves more fully with less fear of stigma or marginalization than with in-person interactions. Through this lens, SMM might continue to disproportionately affect LGBTQ+ individuals and jeopardize the safety and anonymity they feel in using social media. We also found that the perceptions of SMM risks differed by race and ethnicity such that people racial and ethnic minorities respondents perceived greater risks. This finding may be explained by the “racial discipline gap” [33] or the disproportionate rate of school disciplinary sanctions against students of color. Given the long history of differential disciplinary treatment (eg, suspensions and expulsions) of students belonging to racial and ethnic minority groups compared with their non-Hispanic White counterparts, respondents may have concerns about the potential inequitable disciplinary actions taken as a result of SMM surveillance.

More broadly, some of these concerns have also been underscored by a recent US Congressional investigation [34] of 4 educational technology companies, which found that their surveillance platforms may be misused for disciplinary purposes, that surveillance often occurs around the clock (with alerts sometimes bypassing school personnel and going straight to law enforcement), and that parents are not adequately informed. This investigation concluded that “these surveillance products may continue to put students’ civil rights, safety, and privacy at risk” [34] and called on “the federal government...to track the potential impacts of student surveillance technology on students in protected classes...and work to ensure that products used by schools maintain student safety and privacy.” Our study seeks to answer this call to action by generating new insights about stakeholder perceptions of SMM. Moreover, although this study takes 1 step in this direction, more discussion among key stakeholder groups is essential to enhancing awareness and understanding of these technologies and their potential consequences. Failure to do so could have significant consequences, including the erosion of trust among stakeholders, such as students, parents, and educational institutions. Youth mental health and school safety are both urgent and increasingly complex societal challenges, and SMM represents an effort to look to science and technology for a solution. However, per the “technologies of humility” espoused by Jasanoff [35], we must reflect carefully on the ethical dimensions of this landscape and seek to understand and alleviate vulnerability to harm and be mindful of the distribution of risks and benefits.

The need for greater discourse in this area has also been amplificed by the COVID-19 pandemic because the blurring of educational and digital spaces has led, and likely will continue to lead, to greater adoption of digital monitoring technologies. The expansion of school administrators’ guardianship and jurisdiction over students beyond school grounds and into digital
spaces is likely to continue in this environment, meaning that critical scholarship and further investigation into these technologies are urgently needed. This urgency is further underscored by the lack of tangible and effective solutions to ongoing issues around youth violence, bullying, and suicidality. The immense pressure on schools to address such issues will likely lead to the further adoption of SMM without fully considering the potential consequences and harm that may result from such decisions.

This work has some limitations. First, we designed this as an exploratory survey, and thus the items were not validated for specific populations. Future surveys that use validated measures could more meaningfully probe associations between attitudes toward SMM and the characteristics of students. One future direction could be to use the insights from the qualitative data to inform the creation of more specific survey items assessing the perceived benefits and risks of SMM. Second, our findings reflect a sample of convenience, and future studies should seek to obtain nationally representative samples and samples with higher response rates. Third, although the data collected by our open-ended questions were valuable for our analysis, future studies might use in-depth interviews and focus groups to gain a deeper understanding of stakeholder attitudes and beliefs. Finally, this work did not survey current middle or high school students, which would be a fruitful approach in future work to gain more direct insight into young people’s perspectives and attitudes toward SMM.

Conclusions

The results of this study reveal commonalities as well as divergences across stakeholder groups, both in surveyed attitudes regarding SMM in schools and open-ended responses provided directly by participants. The results also highlight the need for greater inclusion of individuals identifying as members of marginalized groups. Future research could examine what steps can be taken to foster greater inclusion of these groups in dialogue and decisions regarding the use of SMM in schools and investigate the real and potential harms and consequences of the use of SMM technologies for those being surveilled.

Acknowledgments

This work was supported by the Center for Empathy and Technology and the National Human Genome Research Institute (R01 HG008753; principal investigator: C Bloss). The authors thank the participants of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Social media monitoring exploratory survey. [DOCX File, 30 KB - formative_v8i1e46746_app1.docx]

References


https://formative.jmir.org/2024/1/e46746

JMIR Form Res 2024 | vol. 8 | e46746 | p.716

(page number not for citation purposes)


18. Join the debate: genome editing. What is Biotechnology. URL: https://www.whatisbiotechnology.org/index.php/survey/index/6707a773b/nature [accessed 2022-12-12]


34. Warren, Marky investigation finds that EdTech student surveillance platforms need urgent federal action to protect students. Office of U.S. Senator. URL: https://www.warren.senate.gov/oversight/letters/warren-markey-raise-concerns-that-student-
Original Paper

Internet Use and Effects on Mental Well-being During the Lockdown Phase of the COVID-19 Pandemic in Younger Versus Older Adults: Observational Cross-Sectional Study

Chou Chuen Yu1*, BSc, MSc, PhD; Nien Xiang Tou1*, BSc, PhD; James Alvin Low1,2*, MBBS, MRCP (UK)

1Geriatric Education and Research Institute, Singapore, Singapore
2Khoo Teck Puat Hospital, Singapore, Singapore
*all authors contributed equally

Corresponding Author:
Chou Chuen Yu, BSc, MSc, PhD
Geriatric Education and Research Institute
2 Yishun Central 2
Singapore, 768024
Singapore
Phone: 65 68078024
Email: yu.chou.chuen@geri.com.sg

Abstract

Background: Majority of individuals, including both younger and older adults, had to adapt to digital means to cope with lockdown measures and pandemic-induced lifestyle changes during the COVID-19 pandemic. While internet accessibility was beneficial during the pandemic, existing literature suggests that excessive use could lead to the rise of problematic internet use in adolescents and younger adults. However, the effects on older adults remain unclear.

Objective: This study aimed to examine differences in internet use during the lockdown phase of the COVID-19 pandemic and explore how age differences in mental health could be explained by time spent on the internet.

Methods: A door-to-door survey of a nationally representative sample of 602 adults in Singapore was carried out using computer-assisted personal interviewing during the early phase of the COVID-19 pandemic (October to November 2020). Participants were categorized into younger (21-59 years old) and older (60 years or above) age groups. We assessed self-reported measures of depression, anxiety, and stress; psychosocial adaptability; ability to perform essential activities; social support; health status; digital media use patterns, and time spent on the internet. Procedures complied with existing safe distancing measures.

Results: Older adults reported being less able to use digital platforms to meet needs and acquire information updates compared with younger adults during the lockdown period of the pandemic. Older adults spent significantly less time on the internet for both work and personal uses per day (mean 146.00 min, SD 9.18 min) compared with younger adults (mean 433.27 min, SD 14.32 min). Significant age differences in depression, anxiety, and stress were found, with younger adults showing poorer mental health. Mediation analysis showed that age differences in depression, anxiety, and stress were partially explained by time spent on the internet. These variables together explained 43%, 40%, and 40% of the variances in depression, anxiety, and stress scores, respectively.

Conclusions: The findings showed that younger adults spent significantly more time on the internet compared with older adults during the lockdown phase of the pandemic. They were also ahead in their ability to use digital resources to meet needs and engage socially compared with older adults. Despite this, the mental health of younger adults was poor, and this was partially accounted for by the amount of time spent on the internet. Since past research suggests that excessive time spent on the internet could lead to disordered use, the benefits brought by digital technologies could have been attenuated during the lockdown phase of the pandemic. Considering this potential negative effect, it is imperative to educate both young and old adults in the appropriate use of information and communication technology.

(JMIR Form Res 2024;8:e46824) doi:10.2196/46824
KEYWORDS
COVID-19; digital divide; well-being; older adults; information and communication technology; internet of things; online; mental health; lockdown; depression; stress; anxiety; digital technology; pandemic

Introduction

COVID-19 has brought about significant changes to the lives of people, with negative impacts on well-being. During the initial stages of the COVID-19 pandemic, “spatial distancing” was a promoted practice that entailed keeping a safe distance between individuals and reducing the number of close interactions individuals have with one another [1]. In a bid to protect lives from this pandemic, many governments around the world subsequently proceeded to initiate lockdowns that mandated the restriction of people’s movements and confined citizens to their homes, limiting, if not halting, unnecessary interactions [2]. Even after lockdowns, various forms of regulations and recommendations on spatial distancing were maintained. As a consequence of such measures, stress and anxiety greatly affected individuals, families, and the society as a whole [3-5]. Even in late 2022, lockdown measures had not been fully discarded and continued to be practiced in some countries [6]. A common trend across the world was that time spent isolated at home increased significantly for most individuals with a resultant loss of daily routines [2]. Even when mandatory movement restrictions were lifted, different waves of new COVID-19 infections meant that people often found themselves being confined to their homes owing to stay-at-home orders or self-sequestration [7]. With most physical activities being curtailed, running activities online became the new normal.

The use of information and communication technology (ICT) was critical during the COVID-19 crisis. ICT not only allowed for the dissemination of timely COVID-19–related information to the public to act upon, but also made it possible to work and study remotely. Arguably, the psychological impact of isolation was mitigated by ICT as friends and families were kept connected despite the physical restrictions imposed by lockdowns [8]. ICT provided access to various forms of entertainment and even materials guiding physical exercises [9]. Access to entertainment through ICT is important as it can help to alleviate the stress of daily living [8,10]. Although a growing number of older adults have been adopting ICT [11,12], the digital exclusion of older adults, otherwise known as the “grey digital divide,” has been an ongoing global issue during the pandemic [13,14]. Older adults face problems in a variety of basic tasks (eg, booking of tickets, claiming benefits online, and gaining access to health care services through appointments) and face exclusion because they cannot connect with their peers through online platforms owing to limited digital skills [14]. Because of the effects of restrictions in social gatherings and spatial distancing during the pandemic, this grey digital divide puts older adults at a disadvantage and may lead to feelings of social isolation and possibly exacerbate health disparity among older adults [15].

While the benefits of ICT during the pandemic seem apparent, especially if used for the purpose of communication, information, and task performance [7,16], excessive use of ICT can be problematic. For instance, the use of ICT without moderation for online gambling, viewing pornography, playing video games, viewing social media, and shopping may lead to higher risks of disordered use [2,17,18]. Research has shown that the disordered use of the internet, also known as problematic internet use, can cause emotional distress and significantly affect different domains in one’s life, including personal, family, and social relations. It can also lead to adverse effects on work or education and other areas of functioning [2,18]. In the current COVID-19 pandemic, an increase in problematic internet use has been reported and excessive internet use has been suggested as a means to cope with the enforced sedentary norms [19,20] in part due to lockdown measures and possibly pandemic-induced life changes [21,22]. The negative consequences of excessive internet use on mental health have been reported in studies during the current pandemic [23,24]. Research conducted around the same time as the inception of this study has shown that there was an increase in problematic internet use during the COVID-19 pandemic [25-32]. However, other than the studies conducted in Taiwan [28] and Japan [29], most of these studies adopted a convenience sampling approach in data collection, and the inferences made from these studies could be limited owing to the sampling bias and poor generalizability [33] associated with this sampling approach. Moreover, the populations of interest in most of these studies were younger adults and adolescents. A recent review on problematic internet use during the COVID-19 pandemic affirmed our aforementioned observation [34]. The authors of the review called for future studies comparing age influences. Only the study carried out in Japan by Oka et al [29] examined age differences. The study found that internet gaming–related problems in younger adults (<30 years) increased during the pandemic, and the numbers were much higher than those for more mature adults. Properly sampled studies comparing the effects of internet use on well-being between younger and older adults were largely missing during the lockdown period of the pandemic.

In view of the identified gaps, this study aimed to examine (1) differences in internet use between younger and older adults during the lockdown phase of the COVID-19 pandemic, and (2) the influence of internet use on the relationship between age and mental health. Given what we know about the internet use of younger adults during the pandemic, we hypothesized that younger adults would have greater use of the internet as compared with older adults, and such a difference in digital use would affect the relationship between age and mental health. This study was conducted in Singapore, an island country and city-state in maritime Southeast Asia. According to official statistics [35], internet use in Singapore has increased from 58% in 2019 to 81% in 2021 among residents aged 60 years or older. The findings from this study may contribute to our understanding of the role of internet use for mental health issues among older adults in countries with increasing internet adoption rates.
Methods

Study Design, Setting, and Participants
This study employed an observational cross-sectional study design. Residents aged 21 years or above were recruited using stratified random sampling with stratification based on housing type, geographical region, gender, and age group. A door-to-door survey was conducted by experienced interviewers between October 17, 2020, and November 27, 2020, not long after partial lockdown restrictions were gradually lifted [20]. The in-person surveying approach ensured inclusion of study participants who might not have access to the internet to mitigate selection bias. The questionnaire administered by the interviewers was worded in the English language, which is the main language in Singapore. To be included in the study, participants were required to have resided in Singapore during the lockdown phase of the pandemic (known locally as the circuit breaker (CB) period between April 7, 2020, and June 1, 2020), in which spatial distancing measures were imposed (known locally as safe distancing). They were also required to be able to speak English and be aged 21 years or older. Interviewers were trained by a geriatrician to exclude older adults who exhibited possible signs of cognitive impairment (eg, drowsiness, agitation, and incongruent language) during the process of taking informed consent. Additionally, among those aged above 70 years, the Abbreviated Mental Test (AMT) [36] was used to screen for possible signs of dementia and other cognitive impairments, and participants were excluded if they failed 1 of the 3 items in the AMT. Safe distancing rules were adhered to during the data collection period.

Ethical Considerations
Ethics approval was obtained from the National Healthcare Group Domain Specific Review Board (2020/00973), and written informed consent was obtained from all participants. Participants were reimbursed with grocery vouchers worth SGD 10 (USD 7.50) for participation. All data collected were deidentified prior to analysis.

Survey Measures

Digital Platform Use
Digital platform use was a single item measure. Participants responded to the question “What did you use digital media for during the CB [circuit breaker] period?” Participants selected from a list of 7 common uses: (1) food delivery, (2) online banking, (3) grocery shopping, (4) online shopping (excluding groceries), (5) online entertainment, (6) social media, and (7) online telecommunication. Participants could also specify, in free-text format, other uses beyond the 7 listed if required. The option “did not use” was available if participants did not make use of any of the platforms.

Perform Essential Activities
The ability to perform essential activities was measured using a 6-item measure. The individual items have been reported in the Results section given the focus on the activities related to internet use (eg, “I was able to use online services to settle what I needed to do [eg, online banking and filling application forms],” “I was able to use telecommunication platforms for work or education purposes,” and “I was able to use online platforms to obtain my supplies [eg, groceries and buy take outs] whenever there was a need to”). Participants responded on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The scale has good internal consistency (α=.81).

Use of the Internet Before and After the Pandemic
Use of the Internet before and after the pandemic was determined by whether participants reported the use of any of the 9 online services commonly used in the local setting (eg, Redmart, Shopee, Taobao, FairPrice Online, Foodpanda, Grabfood, Cold Storage online, Deliveroo, and Lazada) before as well as after or during the pandemic. Participants could also report any other platforms that they used, which were not listed as options in the survey, in free-text format.

Time Spent on the Internet
Time spent on the internet was measured using the following question that was asked in the context of the pandemic: “On average, how much time do you spend on the internet per day (for both work and personal uses)?” Participants reported the number of hours and minutes.

Mental Health
Mental health status was measured using the shortened version of the Depression, Anxiety, and Stress Scale (DASS-21) [37]. The DASS-21 consists of three 7-item subscales designed to measure levels of depression (α=.87), anxiety (α=.76), and stress (α=.87). A sample item for depression is as follows: “I felt that I had nothing to look forward to.” A sample item for anxiety is as follows: “I felt scared without any good reason.” A sample item for stress is as follows: “I found myself getting agitated.” Items were rated on a 4-point Likert scale ranging from 0 (did not apply to me at all) to 4 (applied to me very much or most of the time). Details of interpreting the scores and the use in this study have been described previously [38].

Psychosocial Adaptability
Psychosocial adaptability was an 8-item composite measure (eg, “I was able to adjust my regular social activities to my satisfaction,” “I was able to adjust the way I interact with those I lived with to my satisfaction,” and “I was able to adjust to how I spend my free time [eg, hobbies and entertainment] to my satisfaction”). Participants responded on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The internal consistency of the scale was good (α=.82).

Social Support
Social support was measured using the Resilience Scale for Adults subscale [39]. Sample items in this 3-item measure included “I have some close friends/family members who really care about me” and “I always have someone who can help me when needed.” Participants responded on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The scale has good internal consistency (α=.86).
Health Status
Health status was a single-item measure from the 36-Item Short Form Survey (SF-36) [40]. Participants responded to the question “In general, would you say your health is” using the following options: 1 (poor), 2 (fair), 3 (good), 4 (very good), and 5 (excellent).

Other Measures
Data on the background characteristics of the participants were collected, including (1) age, (2) gender, (3) marital status, (4) nationality, (5) ethnicity, (6) religion, (7) education level, and (8) occupation.

Power Analysis
Based on an a priori power analysis (G*Power 3.1.9.7) using a power of 0.80 and an error probability of 0.05, a sample size of 300 participants was required for each group to detect a between-group difference of a small to moderate effect size.

Results
Participant Characteristics
A total of 602 participants (Table 1) were recruited for the study (mean age 53.30 years, SD 16.26 years). Of these 602 participants, 302 were categorized as younger (21-59 years old; mean age 39.87 years, SD 11.46 years) and the other 300 were categorized as older (60 years or above; mean age 66.82 years, SD 5.84 years). The majority of the younger adults completed tertiary education (213/302, 70.5%) and were employed (234/302, 77.5%), while the majority of the older adults completed secondary education (145/300, 48.3%) and were largely retired (145/300, 48.3%) or were no longer in employment (44/300, 14.7%).
Table 1. Descriptive characteristics of the study sample stratified by age group (younger vs older).

<table>
<thead>
<tr>
<th>Variable</th>
<th>All adults (N=602)</th>
<th>Younger adults&lt;sup&gt;a&lt;/sup&gt; (n=302)</th>
<th>Older adults&lt;sup&gt;b&lt;/sup&gt; (n=300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>53.30 (16.26)</td>
<td>39.87 (11.46)</td>
<td>66.82 (5.84)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>302 (50.2)</td>
<td>133 (44.0)</td>
<td>169 (56.3)</td>
</tr>
<tr>
<td>Female</td>
<td>300 (49.8)</td>
<td>169 (56.0)</td>
<td>131 (43.7)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>431 (71.6)</td>
<td>188 (62.3)</td>
<td>243 (81.0)</td>
</tr>
<tr>
<td>Single</td>
<td>124 (20.6)</td>
<td>103 (34.0)</td>
<td>21 (7.0)</td>
</tr>
<tr>
<td>Divorced</td>
<td>22 (3.7)</td>
<td>11 (3.6)</td>
<td>11 (3.7)</td>
</tr>
<tr>
<td>Widowed</td>
<td>25 (4.2)</td>
<td>0 (0.0)</td>
<td>25 (8.3)</td>
</tr>
<tr>
<td>Nationality, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singaporean</td>
<td>565 (93.9)</td>
<td>274 (90.7)</td>
<td>291 (97.0)</td>
</tr>
<tr>
<td>Permanent resident</td>
<td>37 (6.2)</td>
<td>28 (9.3)</td>
<td>9 (3.0)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>404 (67.1)</td>
<td>198 (65.6)</td>
<td>206 (68.7)</td>
</tr>
<tr>
<td>Malay</td>
<td>91 (15.1)</td>
<td>49 (16.2)</td>
<td>42 (14.0)</td>
</tr>
<tr>
<td>Indian</td>
<td>82 (13.6)</td>
<td>43 (14.2)</td>
<td>39 (13.0)</td>
</tr>
<tr>
<td>Other</td>
<td>25 (4.2)</td>
<td>12 (4.0)</td>
<td>13 (4.3)</td>
</tr>
<tr>
<td>Religion, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buddhism</td>
<td>181 (30.1)</td>
<td>95 (31.5)</td>
<td>86 (28.7)</td>
</tr>
<tr>
<td>Islam</td>
<td>106 (17.6)</td>
<td>58 (19.2)</td>
<td>48 (16.0)</td>
</tr>
<tr>
<td>Christianity (non-Roman Catholic)</td>
<td>81 (13.5)</td>
<td>34 (11.3)</td>
<td>47 (15.7)</td>
</tr>
<tr>
<td>Hinduism</td>
<td>60 (10.0)</td>
<td>28 (9.3)</td>
<td>32 (10.7)</td>
</tr>
<tr>
<td>Roman Catholic</td>
<td>48 (8.0)</td>
<td>15 (5.0)</td>
<td>33 (11.0)</td>
</tr>
<tr>
<td>Taoism</td>
<td>15 (2.5)</td>
<td>7 (2.3)</td>
<td>8 (2.7)</td>
</tr>
<tr>
<td>Sikhism</td>
<td>3 (0.5)</td>
<td>2 (0.7)</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>No religion</td>
<td>108 (17.9)</td>
<td>63 (20.9)</td>
<td>45 (15.0)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary or below</td>
<td>56 (9.3)</td>
<td>6 (2.0)</td>
<td>50 (16.7)</td>
</tr>
<tr>
<td>Secondary</td>
<td>200 (33.2)</td>
<td>55 (18.2)</td>
<td>145 (48.3)</td>
</tr>
<tr>
<td>Postsecondary (nontertiary)</td>
<td>54 (9.0)</td>
<td>28 (9.3)</td>
<td>26 (8.7)</td>
</tr>
<tr>
<td>Tertiary or above</td>
<td>292 (48.5)</td>
<td>213 (70.5)</td>
<td>79 (26.3)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>345 (57.3)</td>
<td>234 (77.5)</td>
<td>111 (37.0)</td>
</tr>
<tr>
<td>Not in employment</td>
<td>109 (18.1)</td>
<td>65 (21.5)</td>
<td>44 (14.7)</td>
</tr>
<tr>
<td>Retired</td>
<td>148 (24.6)</td>
<td>3 (1.0)</td>
<td>145 (48.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Younger adults refer to participants aged 21-59 years.

<sup>b</sup>Older adults refer to participants aged 60 years or above.

**Internet Use Patterns**

Up to a third of older adults (102/300, 34.0%) in the study sample reported not using digital media to communicate or run errands during the CB measurement period compared with only 2.7% (8/302) of younger adults who did not do so. Among older adults who did not use the internet (102/300, 34.0%), an overwhelming majority did not have tertiary education (94/102, 92.2%). For older adults, the top 3 reasons for using the internet (Table 2) were social media (154/300, 51.3%), telecommunication (120/300, 40.0%), and online banking (111/300, 37.0%), whereas for younger adults, the top reasons...
were social media (253/302, 83.8%), online banking (234/302, 77.5%), food delivery (211/302, 69.9%), and telecommunication (210/302, 69.5%).

Table 2. Comparison of the frequency of internet use patterns between younger and older adults in the study sample during the COVID-19 pandemic.

<table>
<thead>
<tr>
<th>Internet use</th>
<th>All adults (N=602), n (%)</th>
<th>Younger adults(^a) (n=302), n (%)</th>
<th>Older adults(^b) (n=300), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social media</td>
<td>407 (67.6)</td>
<td>253 (83.8)</td>
<td>154 (51.3)</td>
</tr>
<tr>
<td>Online banking</td>
<td>345 (57.3)</td>
<td>234 (77.5)</td>
<td>111 (37.0)</td>
</tr>
<tr>
<td>Online telecommunication</td>
<td>330 (54.8)</td>
<td>210 (69.5)</td>
<td>120 (40.0)</td>
</tr>
<tr>
<td>Online entertainment</td>
<td>288 (47.8)</td>
<td>189 (62.6)</td>
<td>99 (33.0)</td>
</tr>
<tr>
<td>Online food delivery</td>
<td>288 (47.8)</td>
<td>211 (69.9)</td>
<td>77 (25.7)</td>
</tr>
<tr>
<td>Online shopping</td>
<td>254 (42.2)</td>
<td>195 (64.6)</td>
<td>59 (19.7)</td>
</tr>
<tr>
<td>Online grocery</td>
<td>203 (33.7)</td>
<td>149 (49.3)</td>
<td>54 (18.0)</td>
</tr>
<tr>
<td>Did not use</td>
<td>110 (18.3)</td>
<td>8 (2.7)</td>
<td>102 (34.0)</td>
</tr>
</tbody>
</table>

\(^a\)Younger adults refer to participants aged 21-59 years.  
\(^b\)Older adults refer to participants aged 60 years or above.

Use of Resources to Meet Needs

Overall, agreement scores on the ability to use digital platforms to meet needs were lower for older adults (mean 3.37, SD 1.14) than for younger adults (mean 4.17, SD 0.78) \((t_{600}=10.04; P<.001; d=0.82)\), and the difference was large in magnitude. The ability to use digital resources for information updates related to the pandemic was lower for older adults (mean 3.28, SD 1.20) than for younger adults (mean 4.22, SD 0.75) \((t_{600}=10.50; P<.001; d=0.86)\), and the difference was large in effect size.

More specifically, for older adults who were able to use the internet (Table 3), the agreement scores in their ability to do so were lower compared with the scores for younger adults in the areas of using telecommunication \((t_{427}=7.55; P<.001; d=0.76)\), obtaining supplies \((t_{477}=8.97; P<.001; d=0.82)\), using online services \((t_{503}=8.17; P<.001; d=0.74)\), and changing appointments \((t_{498}=3.53; P<.001; d=0.31)\). The effect sizes of the differences were large on the whole, except for the last variable.

Table 3. Differences in study sample mean scores between younger and older adults for various essential activities that were conducted over the internet during the COVID-19 pandemic.

<table>
<thead>
<tr>
<th>Variable</th>
<th>All adults (N=602)</th>
<th>Younger adults(^a) (n=302)</th>
<th>Older adults(^b) (n=300)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observations, n</td>
<td>Score, mean (SD)</td>
<td>Observations, n</td>
</tr>
<tr>
<td>Buy takeaway</td>
<td>587</td>
<td>4.16 (0.58)</td>
<td>296</td>
</tr>
<tr>
<td>Run errands</td>
<td>590</td>
<td>4.07 (0.69)</td>
<td>298</td>
</tr>
<tr>
<td>Use telecommunication</td>
<td>429</td>
<td>3.98 (1.01)</td>
<td>272</td>
</tr>
<tr>
<td>Change appointments</td>
<td>500</td>
<td>3.86 (0.78)</td>
<td>250</td>
</tr>
<tr>
<td>Use online services</td>
<td>505</td>
<td>3.84 (1.05)</td>
<td>287</td>
</tr>
<tr>
<td>Obtain supplies</td>
<td>479</td>
<td>3.72 (1.09)</td>
<td>272</td>
</tr>
</tbody>
</table>

\(^a\)Younger adults refer to participants aged 21-59 years.  
\(^b\)Older adults refer to participants aged 60 years or above.

Use of the Internet Before and After the Pandemic

Among older adults, there were minimal changes in the nonuse of the internet for online shopping for essential items (Table 4) before and during or after the pandemic (186/300, 62.0% and 183/300, 61.0%, respectively). This relationship was similar among younger adults, although a much smaller proportion of younger adults did not use the internet for online shopping for essential items before and during or after the pandemic (45/302, 14.9% and 39/302, 12.9%, respectively).
Table 4. Differences in the frequency of the use of the internet before and after the COVID-19 pandemic for obtaining essential items as self-reported by younger and older adults in the study sample.

<table>
<thead>
<tr>
<th>Internet use pattern</th>
<th>All adults (N=602), n (%)</th>
<th>Younger adults(^a) (n=302), n (%)</th>
<th>Older adults(^b) (n=300), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of the internet before the pandemic</td>
<td>371 (61.6)</td>
<td>257 (85.1)</td>
<td>114 (38.0)</td>
</tr>
<tr>
<td>Use of the internet after the pandemic</td>
<td>380 (63.1)</td>
<td>263 (87.1)</td>
<td>117 (39.0)</td>
</tr>
</tbody>
</table>

\(^a\)Younger adults refer to participants aged 21-59 years.
\(^b\)Older adults refer to participants aged 60 years or above.

**Time Spent on the Internet**

There was a significant negative relationship between time spent on the internet and age \((r=−0.63; P<.001)\), indicating that increasing age is associated with decreasing time spent on the internet. On average, older adults indeed spent significantly less time on the internet for both work and personal uses per day (mean 146.00 min, SD 9.18 min) compared with younger adults (mean 433.27 min, SD 14.32 min) \((t_{500}=16.84; P<.001; d=1.38)\).

**Age, Time Spent on the Internet, and Distress**

Zero-ordered bivariate correlations revealed that there was a negative relationship between age and the well-being indicators of depression scores \((r=−0.36; P<.001)\), anxiety scores \((r=−0.22; P<.001)\), and stress scores \((r=−0.19; P<.001)\).

There was also a significant positive association between time spent on the internet and depression scores \((r=0.31; P<.001)\), anxiety scores \((r=0.23; P<.001)\), and stress scores \((r=0.28; P<.001)\). Time spent on the internet was also positively associated with stress concerns \((r=0.08; P=.04)\).

On interpreting the findings, depression scores suggested that those who spent more time on the internet tended to experience less positive feelings, feeling nothing to look forward to, difficulty in working up initiative, etc. For anxiety scores, those who spent more time on the internet tended to feel worried and scared. Likewise, for stress scores, those who spent more time on the internet tended to experience difficulty in winding down, overreaction to situations, and difficulty in relaxing. The effect sizes for these indicators ranged from low to medium.

**Use of the Internet and Mental Health**

Links between internet use patterns in this sample and mental health were explored. The findings showed that mental health was significantly better among those who did not use the internet for social media compared with those who did (Table 5). The effect sizes of the differences however were small \((d_{\text{depression}}=−0.27; d_{\text{anxiety}}=−0.17; d_{\text{stress}}=−0.31)\). This was similarly the case among those who did not use the internet for online shopping compared with those who did (Table 6), with the effect sizes being small \((d_{\text{depression}}=−0.44; d_{\text{anxiety}}=−0.17; d_{\text{stress}}=−0.31)\).

**Main Effects of Time Spent on the Internet and Age on Distress**

The main effects of time spent on the internet and age on distress were examined using regression modeling. The regression models were statistically significant for depression \((F_{5,501}=39.64; P<.001)\), anxiety \((F_{5,501}=19.84; P<.001)\), and stress \((F_{5,501}=31.40; P<.001)\). The models explained 28%, 17%, and 24% of the variances in depression, anxiety, and stress, respectively (see Table 7 for the adjusted and unadjusted models).

To facilitate comparison of the effect sizes, all variables were standardized. Examining the variables in the models, time spent on the internet and age were significant predictors of depression, anxiety, and stress. This was the case even after controlling for the effects of individual adaptability, social support, and health status. Standardized regression coefficients showed that the...
magnitude of the effect of time spent on the internet was higher than that of age for anxiety (β_age = -0.16; β_time spent on the internet = 0.18) and stress (β_age = -0.17; β_time spent on the internet = 0.21), but this effect was lower for depression (β_age = -0.24; β_time spent on the internet = 0.20). Overall, the findings suggested that distress was associated with decreasing age, and likewise, this was the case for those who spent more time on the internet.

To examine if the effect of age on mental health was influenced by time spent on the internet, mediation analysis was conducted using the approach advocated by Baron and Kenny [41]. The bootstrapping method with 1000 resamples to estimate the 95% CI was additionally conducted to investigate the significance of indirect effects [42]. A significance level of \( P < .05 \) was used for all analyses.

To facilitate interpretation of the effect sizes, all variables were standardized to reflect the strength of correlations. Results on the partial mediating role of time spent on the internet for the effect of age on mental health are shown in Figure 1.

In all models, mediation analyses showed that there was a significant indirect effect of age on mental health through time spent on the internet. For depression, the indirect effect of time spent on the internet was as follows: \( ab = -0.08; \ z = -2.66; P = .008 \). The mediation effect accounted for 23% of the total effect (β = -0.36). For anxiety, the indirect effect of time spent on the internet was as follows: \( ab = -0.09; \ z = -2.15; P = .03 \). The mediation effect was moderate and accounted for 39% of the total effect (β = -0.24). For stress, the indirect effect of time spent on the internet was as follows: \( ab = -0.11; \ z = -3.20; P = .001 \). The mediation effect was moderate and accounted for 42% of the total effect (β = -0.22). Examining the \( r^2 \) of the models, all variables together explained 43% of the variance in depression, 40% of the variance in anxiety, and 40% of the variance in stress.

Overall, the effect of age on mental health was partially explained by time spent on the internet. On interpreting the \( r^2 \) of the models and effect sizes, a conclusion could be drawn that the mediating role of time spent on the internet for the effect of age on mental health was relatively substantial, especially for anxiety and stress. However, as full mediation did not occur, for time spent on the internet, which played a substantial role, the effect through age was not a major causal pathway.

### Table 7. Main effects of time spent on the internet and age on distress using multivariate regression analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Depression scores</th>
<th>Anxiety scores</th>
<th>Stress scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted Model 1</td>
<td>Adjusted Model 1</td>
<td>Unadjusted Model 1</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>t (df)</td>
<td>P value</td>
</tr>
<tr>
<td>Age</td>
<td>-0.28</td>
<td>-5.68</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>(602)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time spent on the internet</td>
<td>0.13</td>
<td>2.76</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>(602)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^a \)Adjusted for adaptability, social support, and health status.

### Figure 1. Mediation models illustrating the partial mediating role of time spent on the internet for the effect of age on mental health (mental health was indexed by depression, anxiety, and stress individually). All presented effects are standardized. The a-path is the direct effect, b-path is the direct effect, c1-path is the direct effect, and c-path is the total effect.
Discussion

Principal Findings

This study examined the relationship among age, internet use, and mental health during the lockdown phase of the COVID-19 pandemic. Specifically, we compared internet use between younger and older adults and examined the mediating role of internet use in the relationship between age and mental health. In support of our hypothesis, older adults were reported to spend less time on the internet as compared with their younger counterparts. In addition, internet use was found to partially mediate the effects of age on depression, anxiety, and stress.

Given that ICT is ubiquitous in today’s world, digital exclusion of older adults is emerging as an imperative concern amid rapidly aging populations. Previous studies have reported poor adoption of ICT among older adults [43-46], but findings specific to older adults during the early phase of the COVID-19 pandemic were sparse. The findings of this study extend this empirical evidence by demonstrating that such an age-related digital divide was pervasive during the COVID-19 pandemic. In support of our hypothesis, internet use among older adults during the lockdown period was found to be less as compared with that among their younger counterparts. Specifically, the use of digital platforms for essential services and entertainment purposes was less prevalent among older adults than among younger adults. While age has been known to be a predictor of internet use [47,48], the large effect size ($d=1.38$) found in this study suggests that the magnitude of such a grey digital divide is substantial and warrants more attention. Given that higher education is also associated with greater internet use [47], our results may be confounded by lower education levels among older adults in the study sample. Such a difference may also be plausibly attributed to the negative relationship between age and technology acceptance [49] or fear of stereotype threats regarding technology use in older adults [50].

Imposed spatial distancing measures during the COVID-19 pandemic have proliferated the adoption of technology to help individuals meet essential needs and stay connected with one another. This has been claimed to amplify the digital divide between younger and older adults [15,51,52]. Indeed, our findings showed that older adults not only used less internet than their younger peers, but also reported to be less able to use online services to run essential activities during the lockdown phase of the pandemic. This suggests that older adults were less capable of adapting to digital means to meet their basic needs. Our findings corroborate the results of another study reporting that the inability of older adults to use digital devices limited their access to transportation, medical care, and food supplies during the pandemic [52]. Consequently, this has raised concerns that the negative impacts of the pandemic on mental health may be disproportionate in certain groups, such as older adults, who may risk being excluded from the society due to such a digital divide [53].

While lockdown measures were well intended to mitigate infections during the COVID-19 pandemic, they had detrimental effects on the psychological well-being of individuals [5]. The results of this study suggest that both age and internet use are significant predictors of mental health, and importantly, such associations persisted even after accounting for adaptability, social support, and health status. As purported by the socioemotional selectivity theory [54], we found that older age was associated with better mental health. Specifically, we previously reported that older adults had lower depression, anxiety, and stress levels as compared with younger adults [38]. Our results are consistent with those of other studies that reported better mental health in older adults during the pandemic as compared with younger adults [54-57]. In contrast, higher internet use during the lockdown period was found to be associated with poorer mental health. While digital technologies certainly could aid individuals in the continuation of their work and daily essential activities during the lockdown, they could also be used in a disordered manner [17,18]. Our results suggest that it could be the case of the latter since mental distress has been observed to be higher in those with higher internet use (ie., spent more time on the internet and used more social media and online shopping). Indeed, the risk of problematic internet use was reported to have increased during the pandemic [2].

Many studies conducted during the earlier phase of the COVID-19 pandemic focused on the effects of internet use on younger adults and adolescents who were often conveniently sampled [25,28,30-32]. In this study, a door-to-door survey using a stratified random sampling method ensured that potential sampling bias issues were circumvented. This study ensured that there were adequate responses sampled from participants in the older age group (ie., over 60 years of age) and therefore contributed to the literature by showing the effects of internet use in not only younger adults but also older adults. This study was therefore able to address calls [34] for research comparing the linkages among age, internet use, and mental well-being during the COVID-19 pandemic. Importantly, our statistical modeling suggests that internet use acts as a partial mediator for the effects of age on depression, anxiety, and stress levels. This suggests that the inverse relationship between age and mental distress could be partially explained by the amount of time spent on the internet by younger adults. As mentioned earlier, younger adults were found to spend more time on the internet as compared with older adults. Such increased use may be problematic and consequently have negative effects on mental health. Indeed, problematic internet use during the COVID-19 pandemic has been reported to be associated with poorer mental well-being [58,59]. Studies have reported that pandemic-related stress is associated with tendencies toward problematic digital use [59-61]. Given that younger adults were found to have greater stress concerns during the pandemic [38], it may be the case that younger adults experienced greater stress and thus engaged in greater problematic internet use, which resulted in poorer mental health. More research is recommended to investigate this. Our findings support the recommendations to mitigate the risk of problematic internet use during the pandemic [2].

Despite the negative relationship between internet use and mental health, it is important to note that not all types of digital use had detrimental consequences on mental health, and ICT use for the right purposes could be potentially beneficial. For example, a recent scoping review revealed that internet use for
communication purposes was associated with better mental health for older adults during the COVID-19 pandemic [62]. Indeed, older adults have reported the importance of ICT to help them maintain their social connections during the lockdown period [63-65]. This study showed that social media use was the highest among the various uses of the internet even for older adults, and a future study can consider investigating how older adults use social media and assessing the benefits they derive from it during the pandemic. Given that some older adults may experience loneliness due to lack of physical contact, especially under the exceptional circumstances imposed by the pandemic, the use of ICT can help to allay the negative effects of social isolation. Thus, in view of the large digital divide between younger and older adults found in this study, greater efforts are necessary to close such a divide, including advocating for age-sensitive design of technologies [49,52] and deterring stereotype threats associated with technology use [50]. Considering the potential negative relationship between internet use and mental health, it is also imperative to educate older adults in using ICT appropriately.

Limitations

Notwithstanding the contributions of this study to existing literature, it is important to acknowledge some limitations. First, the study sample consisted of community-dwelling adults, and thus, the findings may not be generalizable to specific subgroup populations who may be more vulnerable. The relationship between age and mental distress observed in this study may differ if other groups of older adults are included, such as those residing in nursing homes or experiencing diminished mental capacity.

Second, the interpretation of this study’s results is limited by its outcome measures. Although this study established the partial mediating role of internet use for the effects of age on depression, anxiety, and stress, the study did not collect measures that would be important indicators of disordered use of the internet in younger adults, including online gambling, viewing pornography, and playing video games. Nonetheless, there is evidence from various studies indicating that increased time spent on the internet during the earlier phase of the COVID-19 pandemic was the main contributory factor for a number of mental health problems, such as depression, anxiety, and stress [66-68]. The findings from this study are therefore aligned with the existing research and add to this field by presenting data from a developed country with high digital adoption rates. However, a future study can consider examining how the pandemic could have possibly exacerbated problematic internet use and whether time spent on the internet is indeed an adequate proxy. The digital divide and its relationship with mental health can be influenced by a myriad of factors going beyond problematic internet use. There is a possibility that increased internet use in younger adults during the lockdown period was attributed to having to work remotely from office, and the distress experienced could therefore be attributed to this new form of work arrangement rather than problematic internet use per se. Given that the study was making reference to the first lockdown period, younger adults may not have adapted to this form of working arrangement despite the known benefits of ICT, including continuation of their work and running of other essential activities.

Finally, all measures of this study were collected from participants’ retrospective recollection of their experiences during the lockdown period. Since time spent on the internet was not based on any objective indicator, we cannot rule out the possibility that this or other attitudinal responses are subject to recall bias.

Conclusion

This study showed that older adults lagged behind younger adults in the use of digital resources during the pandemic, which could have helped them in communication and socialization, and the findings support existing literature on the poor adoption of ICT among older adults. This study further contributes to the literature by showing how through mediation modeling, the negative relationship between increasing age and mental distress appears to be partially explained by the amount of time spent on the internet by younger adults. Without moderate use, the benefits brought by digital technologies could have been attenuated during the lockdown phase of the pandemic. It is imperative to educate both young and old adults in the appropriate use of ICT.

Acknowledgments

The authors gratefully acknowledge the support of Dr Siew Pei Ong, Vivienne Le, Emily Koh, Robin Choo, and Bernard Tang in this study. This work was supported by Geriatric Education and Research Institute Limited under its COVID-19 research grant (CG002).

Data Availability

The data sets used or analyzed during this study are available from the corresponding author on reasonable request. The data sets used cannot be shared openly as the participants have not consented to share their data openly, and this is therefore not allowed under the approved study protocol.

Authors’ Contributions

CCY and JAL contributed to the study concept and design. CCY and NXT contributed to the data analysis and interpretation of the results. All authors were involved in the writing of the manuscript, and have critically read, reviewed, and approved the final manuscript.
Conflicts of Interest

None declared.

References


Abbreviations

- **AMT**: Abbreviated Mental Test
- **CB**: circuit breaker
- **DASS-21**: Depression, Anxiety, and Stress Scale
- **ICT**: information and communication technology

©Chou Chuen Yu, Nien Xiang Tou, James Alvin Low. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e46824 JMIR Form Res 2024 | vol. 8 | e46824 | p.732
Explainable Machine Learning Classification to Identify Vulnerable Groups Among Parenting Mothers: Web-Based Cross-Sectional Questionnaire Study

Akiko Hanai\(^1,2,3\), OTR, PhD; Tetsuo Ishikawa\(^1,2,3,4\), PhD; Shoko Sugao\(^5\), PsyD, PhD; Makoto Fujii\(^6\), RN, PHN, PhD; Kei Hirai\(^5\), PhD; Hiroko Watanabe\(^6\), RM, RN, PhD; Masayo Matsuzaki\(^7\), RM, RN, PHN, PhD; Goji Nakamoto\(^6\), PhD; Toshihiro Takeda\(^8\), MD, PhD; Yasuji Kitabatake\(^9\), MD, PhD; Yuichi Itoh\(^10\), PhD; Masayuki Endo\(^11\)*, MD, PhD; Tadashi Kimura\(^11\), MD, PhD; Eiryo Kawakami\(^1,2\)*, MD, PhD

\(^1\)Medical Data Mathematical Reasoning Team, Advanced Data Science Project, RIKEN Information R&D and Strategy Headquarters, RIKEN, Yokohama, Japan
\(^2\)Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan
\(^3\)Institute for Datability Science, Osaka University, Suita, Japan
\(^4\)Department of Extended Intelligence for Medicine, The Ishii-Ishibashi Laboratory, Keio University School of Medicine, Tokyo, Japan
\(^5\)Graduate School of Human Sciences, Osaka University, Suita, Japan
\(^6\)Division of Health Sciences, Graduate School of Medicine, Osaka University, Suita, Japan
\(^7\)Department of Reproductive Health Nursing, Graduate School of Health Care Sciences, Tokyo Medical and Dental University, Tokyo, Japan
\(^8\)Department of Medical Informatics, Graduate School of Medicine, Osaka University, Suita, Japan
\(^9\)Department of Pediatrics, Graduate School of Medicine, Osaka University, Suita, Japan
\(^10\)Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University, Sagamihara, Japan
\(^11\)Department of Obstetrics and Gynecology, Osaka University Graduate School of Medicine, Suita, Japan

* these authors contributed equally

Corresponding Author:
Tetsuo Ishikawa, PhD
Medical Data Mathematical Reasoning Team
Advanced Data Science Project
RIKEN Information R&D and Strategy Headquarters, RIKEN
W507/509 West Research Building,
1-7-22 Suehiro-cho, Tsurumi-ku
Yokohama, 230-0045
Japan
Phone: 81 455039455
Email: tetsuo.ishikawa@riken.jp

Abstract

Background: One life event that requires extensive resilience and adaptation is parenting. However, resilience and perceived support in child-rearing vary, making the real-world situation unclear, even with postpartum checkups.

Objective: This study aimed to explore the psychosocial status of mothers during the child-rearing period from newborn to toddler, with a classifier based on data on the resilience and adaptation characteristics of mothers with newborns.

Methods: A web-based cross-sectional survey was conducted. Mothers with newborns aged approximately 1 month (newborn cohort) were analyzed to construct an explainable machine learning classifier to stratify parenting-related resilience and adaptation characteristics and identify vulnerable populations. Explainable k-means clustering was used because of its high explanatory power and applicability. The classifier was applied to mothers with infants aged 2 months to 1 year (infant cohort) and mothers with toddlers aged >1 year to 2 years (toddler cohort). Psychosocial status, including depressed mood assessed by the Edinburgh Postnatal Depression Scale (EPDS), bonding assessed by the Postpartum Bonding Questionnaire (PBQ), and sleep quality assessed by the Pittsburgh Sleep Quality Index (PSQI) between the classified groups, was compared.

Results: A total of 1559 participants completed the survey. They were split into 3 cohorts, comprising populations of various characteristics, including parenting difficulties and psychosocial measures. The classifier, which stratified participants into 5
groups, was generated from the self-reported scores of resilience and adaptation in the newborn cohort (n=310). The classifier identified that the group with the greatest difficulties in resilience and adaptation to a child’s temperament and perceived support had higher incidences of problems with depressed mood (relative prevalence [RP] 5.87, 95% CI 2.77-12.45), bonding (RP 5.38, 95% CI 2.53-11.45), and sleep quality (RP 1.70, 95% CI 1.20-2.40) compared to the group with no difficulties in perceived support. In the infant cohort (n=619) and toddler cohort (n=461), the stratified group with the greatest difficulties had higher incidences of problems with depressed mood (RP 9.05, 95% CI 4.36-18.80 and RP 4.63, 95% CI 2.38-9.02, respectively), bonding (RP 1.63, 95% CI 1.29-2.06 and RP 3.19, 95% CI 2.03-5.01, respectively), and sleep quality (RP 8.09, 95% CI 4.62-16.37 and RP 1.72, 95% CI 1.23-2.42, respectively) compared to the group with no difficulties.

**Conclusions:** The classifier, based on a combination of resilience and adaptation to the child’s temperament and perceived support, was able identify psychosocial vulnerable groups in the newborn cohort, the start-up stage of childcare. Psychosocially vulnerable groups were also identified in qualitatively different infant and toddler cohorts, depending on their classifier. The vulnerable group identified in the infant cohort showed particularly high RP for depressed mood and poor sleep quality.

**(JMIR Form Res 2024;8:e47372) doi:10.2196/47372**

**KEYWORDS**

explainable machine-learning; unsupervised clustering; perceived support; resilience; adaptation; mother’s health; mobile phone; machine learning; web-based; parenting; postpartum; antenatal; survey; mother; women; newborn; psychosocial; infant; parents; children; depression; digital health; maternal

**Introduction**

One life event that requires extensive resilience and adaptation is parenting. However, parenting difficulties have become highly varied and complex, especially for mothers, who are increasingly expected to balance their family and social roles [1,2]. The dynamics of parenting are further complicated by the rise of working parents, leading to evolving parenting environments and an amplified demand for social support [3]. The challenges of parenting have been compounded by the SARS-CoV-2 pandemic, which brought about blockades, physical distancing, and social isolation, adversely impacting mental health. These challenges are exacerbated by factors such as economic insecurity and the increased burden of childcare and housework [4].

A meta-analysis shows that 19.2% of women experience a major depressive episode during the first 3 months post partum, with most of these episodes occurring after delivery [5]. For this reason, most medical institutions in Japan evaluate the mother’s mental status during the 1-month health examination. Common risk factors of postpartum depression were high life stress, the lack of social support, current or past abuse, prenatal depression, and marital or partner dissatisfaction, which are not only limited to the early postpartum child-rearing periods [6]. Psychosocial stressors during long child-rearing periods are even more varied and complex, and individual mother’s personalities and cognitive characteristics must be comprehensively considered to support safe child-rearing practices [7,8]. Hence, we developed the Comprehensive Scale for Parenting Resilience and Adaptation (CPRA) to assess parenting difficulties based on environmental factors, maternal personality traits, and a mother’s perception of her child [2]. Thus, the CPRA could be used to evaluate the complexities of parenting difficulties.

Timely access to face-to-face social support has proven to be psychologically and socially challenging for mothers who struggle with parenting [9]. Studies have indicated that prior parenting knowledge and experience lead to resilience, as primiparous women have significantly lower resilience to perinatal depressive mood than multiparous women [10,11]. A web-based platform could be easily accessible for busy and vulnerable mothers. Thus, a web-based assessment and timely intervention could provide parenting resources to such inexperienced mothers. Web-based screening and communication are expected to play a significant role not only for parenting under the “new normal” lifestyle due to the SARS-CoV-2 pandemic but also for parents who do not prefer face-to-face communication.

The Japanese government supports the Society 5.0 project, which aims to integrate cyber and physical spaces (also known as the cyber-physical space) where data science can contribute to task executions and coordination [12]. In this project, we created a web-based screening system linked to an app that calculates gestational age and provides parenting information services. This system enabled researchers to collect data from mothers in various parenting environments across Japan without requiring hospital visits. However, the real-world data obtained by this system were heterogeneous. Machine learning is considered more suitable than model-driven conventional statistical methods for classifying characteristics and developing interventions based on complex data, even with a small sample size [13]. Especially in mental health research, machine learning is used for diagnosis, treatment support, research and development, and clinical work management. The increase in the number of studies suggests that machine learning may be useful for the detection and diagnosis of mental health conditions such as depression, schizophrenia, and Alzheimer disease [14]. A systematic review screened 482 papers and evaluated 11 papers that used machine learning to predict postpartum depression; it found that although a solid conclusion was not achieved since the algorithms and data sets used were heterogeneous, all studies reached an area under the curve greater than 0.7, indicating that predicting postpartum depression by machine learning is feasible [15]. Machine learning methods also allow the use of blood test data from before the onset of perinatal depression to identify high-risk populations who are the most in need of preventive interventions [16].
The disadvantages of machine learning include its complex algorithms and reproducibility. For example, k-means clustering is a machine learning–based algorithm that groups similar data by analyzing its structures and patterns; each data point is assigned to a cluster to minimize variance within the cluster while maximizing it between clusters. However, reproducing these clusters in a different data set can be challenging. Therefore, we used explainable k-means clustering to overcome these limitations [17].

The purpose of this study is to create a vulnerability classifier based on cross-sectional survey data on maternal resilience and adaptation and to examine the identified psychosocially vulnerable groups. This study also attempts to determine the psychosocial status of mothers during other child-rearing periods, by confirming whether the classifier can be applied to infant and toddler populations to identify similarly vulnerable groups.

Methods

Study Design and Setting

This study included a cross-sectional survey and machine learning stratification. Recruitment, opt-in informed consent, and data collection were conducted through a web application that rewarded participants with the equivalent cost of the internet resources used to complete the survey. We used a Japanese web service company called Milcare to conduct the survey. For those who downloaded Milcare’s app, which is used to record the number of weeks of pregnancy, an advertisement for the survey was shown in the app for the duration of our survey setting. The system was designed to provide survey participants incentives such as gift cards and childcare goods through the app. The app was available to Japanese smartphones and had been downloaded approximately 5000 times at the time of the survey. All data reported by the participants in the web-based survey were included as outcomes. The response period was from January to June 2020. A maternal personality traits classifier was created from the newborn cohort using explainable k-means clustering. The classifier was then applied to the infant cohort and the toddler cohort, and the features of each group were analyzed.

Participants

The target population of this study was parenting mothers with newborns to 2-year-old toddlers. The eligibility criteria for the participants were (1) mothers with children aged <2 years; (2) web-based consent to participate in the survey through the Milcare web service; (3) ability to read, write, and understand Japanese while using the internet; and (4) ability to complete a questionnaire using a smartphone. The exclusion criteria were those who (1) completed the survey in less than 8 minutes (the minimum time required to read and answer all questions) and (2) abandoned the survey during the process.

Ethical Considerations

Ethical review approval was obtained because this study involves human subjects: Osaka University Hospital Observational Research Ethics Review Committee (19Z290-2) and Research Ethics Review Committee of RIKEN (2020-23(2)). Data for the study were analyzed anonymously and no individuals were identified. In addition, consent for primary data acquisition and secondary use was obtained.

Patient and Public Involvement

Although some of the researchers have parenting experience, neither patients with depressive moods nor the public were involved in the design, planning, conduct, or reporting of this study.

Assessment

The mothers’ resilience and adaptation were assessed using the CPRA. The CPRA consisted of 5 domains (child’s temperament and health, environmental resources, perceived support, mother’s cognitive and behavioral characteristics, and psychological adaptation to parenting) and 21 subscales (Table 1). Responses were collected on a 5-point Likert scale with higher scores indicating increased parenting difficulty [8].

Depressive mood was assessed using the Edinburgh Postnatal Depression Scale (EPDS). In Japan, an EPDS score of 9 or above is used as the cutoff for clinical screening, which has a sensitivity of 75% to 82% and a specificity of 93% to 95% for high-risk depression [18-20]. Therefore, we defined an EPDS score of 9 or above to indicate depressive mood in this study. The Sense of Coherence (SOC) scale was used to assess how one understands, manages, and feels emotional meaning when experiencing stress [21]. The Postpartum Bonding Questionnaire (PBQ) was used to assess mother-infant bonding, with a high score (≥13) indicating impaired bonding [22]. The Pittsburgh Sleep Quality Index (PSQI) was used to assess sleep quality, where a high score (≥5.5) indicates poor sleep quality [23,24].

https://formative.jmir.org/2024/1/e47372

JMIR Form Res 2024 | vol. 8 | e47372 | p.735

(page number not for citation purposes)
Table 1. Comprehensive Scale for Parenting Resilience and Adaptation domains and subscales.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Abbreviation</th>
<th>Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child’s temperament and health</td>
<td>Child</td>
<td>• Child’s temperament and health</td>
</tr>
<tr>
<td>Environmental resources</td>
<td>Environment</td>
<td>• Child-rearing or long-term care burden</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Parental autonomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Partner autonomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Partner temperament</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Relationship with the medical staff</td>
</tr>
<tr>
<td>Perceived support</td>
<td>Support</td>
<td>• Husband’s or partner’s support</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lack of psychological support from husband or partner</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Parental support</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sufficient social support</td>
</tr>
<tr>
<td>Mother’s cognitive and behavioral characteristics</td>
<td>Cognitive</td>
<td>• Attachment problems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Emotional control</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Inattentiveness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Simultaneous or overall processing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Social intolerance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Systemization urge</td>
</tr>
<tr>
<td>Psychological adaptation to parenting</td>
<td>Psychological</td>
<td>• Lack of self-confidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Possibility of coping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Love for the child</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Self-esteem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Self-responsibility</td>
</tr>
</tbody>
</table>

Explainable Clustering With Decision Tree

To classify the characteristics of resilience and adaptation difficulties, a clustering algorithm, explainable k-means clustering, was used. Explainable k-means clustering is an unsupervised clustering algorithm, characterized by its use of decision trees with minimum leaf size for data set partitioning. These trees are meticulously designed to contain a constrained number of nodes and leaves, adhering to the principle of parsimony. Specifically, the number of leaves is deliberately set to correspond with the desired number of clusters (k) in the k-means algorithm. This strategic design enhances both the interpretability and comprehensibility of the clustering process. Consequently, this framework facilitates the application of classifiers to external populations, provided that these populations are quantified on analogous scales. Such an approach underscores the algorithm’s use in ensuring coherent and interpretable clustering outcomes, which are pivotal for the analysis of heterogeneous data sets [17]. For estimating the number of groups, gap statistics analysis was conducted using the CPRA data set from the newborn cohort. Gap statistics analysis compares the change in cluster dispersion with that expected under an acceptable, reference null distribution using the output of any clustering algorithm [25]. Using these results, participants in the newborn cohort were stratified into groups using an unsupervised explainable k-means algorithm with a decision tree, depending on the 5 CPRA domains. Last, the clustering-based anomaly detection algorithm, the so-called decision tree, was applied to the infant cohort and the toddler cohort.

Comparison of Mothers’ States in the Identified Groups

We described the characteristics of the CPRA domain scores and the psychological assessment scores (EPDS, PBQ, and PSQI) for each group. The stratification of groups in each cohort was plotted and the mean score of the 21 CPRA subscales for each of the 5 stratified groups was calculated and plotted as a radar chart.

Statistical Analysis

We analyzed the characteristics of the stratified groups in each cohort, using the group with no difficulties as a reference. The prevalence of depressive moods (EPDS score ≥ 9), bonding problems (PBQ score ≥ 13), and sleep problems (PSQI score ≥ 5.5) was used for logistic regression analysis with relative prevalence (RP), 95% CI, and P value. The distributions of the CPRA and psychosocial scores of the stratified groups in the cohorts were compared using the Kruskal-Wallis test. The results of the Kolmogorov-Smirnov test showed deviations from the normal distribution for many variables, and the Levene test may have violated the assumption of equal variability in the EPDS and PBQ (Multimedia Appendix 1). Therefore, the Kruskal-Wallis test was performed as a multiple-group comparison.

All data from participants who completed the survey within the specified time frame were included in the analysis. Participants’ characteristics were expressed using descriptive statistics, and P < .05 was considered statistically significant.

All statistical analyses were performed using R (version 3.6.2; R Foundation for Statistical Computing). We used the ExKMC package to produce explainable k-means clustering and the ggstatsplot and ggplot packages for data visualization, all of
which are available on GitHub (GitHub Inc) [17,26]. The ExKMC package automatically produces the shallowest tree depth.

**Results**

**Participant Characteristics**

From the web-based recruitment, 1559 participants completed the web survey. Data from those who took at least the minimum required time to read and answer the questions were analyzed. The participants were divided into the newborn cohort (n=310; mean age 31, SD 4.60 years), infant cohort (n=619; mean age 31.5, SD 4.15 years), and toddler cohort (n=461; mean age 32.1, SD 4.16 years). The newborn cohort was analyzed to generate a classifier, and the classifier was applied to the infant cohort and toddler cohort (Figure 1). The CPRA domain scores and the psychosocial assessments (EPDS, PBQ, and PSQI) in each cohort are represented in Table 2.

Overall, parenting difficulties as assessed by the CPRA tended to be more severe during the newborn period, which is up to 1 month post partum (Table 2). Statistically significant differences were observed in the environmental and psychological domains of the CPRA, as well as for the EPDS, PBQ, and PSQI (all \( P < .001 \)).

**Table 2.** Comprehensive Scale for Parenting Resilience and Adaptation and psychosocial scores by cohort.

<table>
<thead>
<tr>
<th>Scores</th>
<th>Newborn (n=310), median (IQR)</th>
<th>Infant (n=619), median (IQR)</th>
<th>Toddler (n=461), median (IQR)</th>
<th>( P ) value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>2.20 (1.80-2.80)</td>
<td>2.20 (1.80-2.60)</td>
<td>2.20 (1.80-2.60)</td>
<td>.005</td>
</tr>
<tr>
<td>Cognitive</td>
<td>2.36 (2.00-2.71)</td>
<td>2.29 (1.93-2.64)</td>
<td>2.36 (2.00-2.71)</td>
<td>.01</td>
</tr>
<tr>
<td>Environment</td>
<td>2.35 (2.05-2.75)</td>
<td>2.25 (1.95-2.55)</td>
<td>2.25 (2.00-2.60)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Psychological</td>
<td>2.47 (2.05-2.95)</td>
<td>2.26 (1.95-2.74)</td>
<td>2.42 (2.05-2.79)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Support</td>
<td>2.47 (1.88-2.93)</td>
<td>2.27 (1.80-2.73)</td>
<td>2.40 (2.00-2.80)</td>
<td>.003</td>
</tr>
<tr>
<td>EPDS(^b)</td>
<td>9 (5-13)</td>
<td>5 (2-9)</td>
<td>6 (3-10)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>PBQ(^c)</td>
<td>10 (6-18)</td>
<td>9 (6-14)</td>
<td>12 (7-17)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>PSQI(^d)</td>
<td>7 (5-9)</td>
<td>7 (5-9)</td>
<td>5 (4-8)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\) Kruskal-Wallis rank sum test.
\(^b\) EPDS: Edinburgh Postnatal Depression Scale.
\(^c\) PBQ: Postpartum Bonding Questionnaire.
\(^d\) PSQI: Pittsburgh Sleep Quality Index.
Explainable Clustering

Using gap analysis to decide the optimal number of the clustering revealed that 5 groups were suitable for k-means clustering (grouping) of the participants based on the similarities in CPRA scores. Using the explainable k-means method, the participants in the newborn cohort were stratified into 5 groups based on multivariate factors, and a decision tree explaining how to stratify these groups was generated using 2 evaluation axes: perceived support–related and child’s temperament–related difficulties (Figure 2). Using the classifier as a decision tree algorithm, 5 groups were stratified in each of the 3 cohorts. According to the classifier rules, Group 0 represented those with no perceived support–related difficulties, Group 1 represented those with moderate perceived support–related difficulties (Support+), Group 2 represented those with moderate perceived support–related difficulties and child’s temperament–related difficulties (Support+Child+), Group 3 represented those with moderate perceived support–related difficulties and severe child’s temperament–related difficulties (Support+Child++), and Group 4 represented those with severe perceived support–related and child’s temperament–related difficulties (Support++Child++; Figure 2). A split of similar proportions was achieved, with the maximum difference of 5% occurring in Group 4. To illustrate the distribution of the stratified groups in each cohort, we focused on the child and support scores, which were used as features for classifier generation; plotted them on the vertical and horizontal axes, respectively; and colored each group to visualize them (Figure 3). The CPRA scores for each cohort were concentrated at 2-3 points, but the distribution of outliers was different for each cohort (Figure 3). The 21 subscales of the CPRA for each stratified group in each cohort were plotted in a radar chart (Figure 4).

Figure 2. Decision tree explanation for k-means clustering. CPRA: Comprehensive Scale for Parenting Resilience and Adaptation.

Figure 3. Comprehensive Scale for Parenting Resilience and Adaptation domains scores (child’s temperament [C] and perceived support [S]) in the 3 cohorts (newborn, infant, and toddler), colored by the stratified groups.
Regression Analysis and Characteristics of the Detected Groups

Psychosocial conditions in each cohort were plotted as density plots for the EPDS, PBQ, SOC, and PQSI scores, colored according to the stratified groups (Figure 5). In the newborn cohort (n=310), Group 4 (Support++Child++) was the group with the greatest difficulties, having higher incidences of problems with depressed mood (RP 5.87, 95% CI 2.77-12.45), bonding (RP 5.38, 95% CI 2.53-11.45), and sleep quality (RP 1.70, 95% CI 1.20-2.40) compared to the group with no difficulties in perceived support (Group 1; Table 3). In the infant cohort (n=619), the stratified group with the greatest difficulties had higher incidences of problems compared to the group with no difficulties, with depressed mood (RP 9.05, 95% CI 4.36-18.80) and sleep quality (RP 8.69, 95% CI 4.62-16.37) having a greater RP than the newborn cohort, but a smaller RP for bonding (RP 1.63, 95% CI 1.29-2.06; Table 3). In the toddler cohort (n=461), the stratified group with the greatest difficulties had higher incidences of problems with depressed mood (RP 4.63, 95% CI 2.38-9.02), bonding (RP 3.19, 95% CI 2.03-5.01), and sleep quality (RP 1.72, 95% CI 1.23-2.42) compared to the group with no difficulties.

Figure 5. Mothers’ condition in the stratified groups. Postnatal depressive mood was assessed using the EPDS, postpartum bonding was assessed using the PBQ, sense of coherence was assessed using the SOC, and sleep quality was assessed using the PSQI. 1N: newborn cohort; 2I: infant cohort; 3T: toddler cohort; C: child’s temperament; EPDS: Edinburgh Postnatal Depression Scale; PBQ: Postpartum Bonding Questionnaire; PSQI: Pittsburgh Sleep Quality Index; S: perceived support; SOC: Sense of Coherence.
Table 3. Comparison of the frequency of problems with depressive mood, postpartum bonding, and sleep quality.

<table>
<thead>
<tr>
<th>Cohort and group</th>
<th>Depressive mood (EPDS≥9)</th>
<th>Postpartum bonding (PBQ≥9)</th>
<th>Sleep quality (PSQI≥5.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count, n (%)</td>
<td>RP (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newborn cohort (n=310)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 (no difficulties; n=44)</td>
<td>6 (14)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 (S++; n=47)</td>
<td>15 (32)</td>
<td>2.34 (1.00-5.49)</td>
<td>.05</td>
</tr>
<tr>
<td>2 (S+C++; n=94)</td>
<td>43 (46)</td>
<td>3.35 (1.54-7.29)</td>
<td>.002</td>
</tr>
<tr>
<td>3 (S+++; n=50)</td>
<td>37 (74)</td>
<td>5.43 (2.53-1.62)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4 (S+++C++; n=75)</td>
<td>60 (80)</td>
<td>5.87 (2.77-12.45)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Infant cohort (n=619)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 (no difficulties; n=102)</td>
<td>7 (6.9)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 (S++; n=101)</td>
<td>10 (9.9)</td>
<td>1.44 (0.57-3.64)</td>
<td>.44</td>
</tr>
<tr>
<td>2 (S+C++; n=202)</td>
<td>50 (24.8)</td>
<td>3.61 (1.70-7.67)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3 (S+++; n=111)</td>
<td>50 (45.0)</td>
<td>6.56 (3.12-13.81)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4 (S+++C++; n=103)</td>
<td>64 (62.1)</td>
<td>9.05 (4.36-18.80)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Toddler cohort (n=461)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 (no difficulties; n=60)</td>
<td>8 (13)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1 (S++; n=76)</td>
<td>8 (11)</td>
<td>0.79 (0.31-1.98)</td>
<td>.61</td>
</tr>
<tr>
<td>2 (S+C++; n=146)</td>
<td>50 (34.2)</td>
<td>2.57 (1.30-5.09)</td>
<td>.007</td>
</tr>
<tr>
<td>3 (S+++; n=90)</td>
<td>37 (41)</td>
<td>3.08 (1.55-6.15)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4 (S+++C++; n=89)</td>
<td>55 (62)</td>
<td>4.63 (2.38-9.02)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

a EPDS: Edinburgh Postnatal Depression Scale.
b PBQ: Postpartum Bonding Questionnaire.
c PSQI: Pittsburgh Sleep Quality Index.
d RP: relative prevalence (compared with Group 0).
e N/A: not applicable.
f S: perceived support.
g C: child’s temperament.

Discussion

Principal Findings

The classifier based on the resilience and adaptation assessment (CPRA) stratified the mothers in the newborn cohort, the start-up stage of childcare, into 5 groups and identified vulnerable psychosocial groups. The score of resilience to the child’s temperament and perceived support was selected as an important feature for building an explainable classifier, instead of environmental resources, the mother’s cognitive-behavioral characteristics, and psychological adaptation to parenting. Depending on the decision tree–based explainable classifier, psychosocially vulnerable groups were also identified in the qualitatively different infant and toddler cohorts. The identified group with high sensitivity to the temperament of their children and difficulty in perceiving support showed high RP for depressed mood and poor sleep quality, especially in the infant...
cohort. This group was particularly vulnerable to severe psychosocial problems in the infant cohort.

Limitations
The study data were collected using a web-based self-reported survey to acquire large-scale nationwide data. Therefore, the degree of objectivity, the participants’ backgrounds, and family characteristics (i.e., the partner’s housework load and the child’s developmental disabilities) were not guaranteed. The possibility that false responses were treated as true values cannot be excluded. Nevertheless, a web-based survey method with limited psychological barriers was chosen because it increased the ease of participation, so that even mothers who are too busy to make time for a formal interview or have difficulties with interpersonal conversations could be represented. This study saw a greater prevalence of perinatal depressive mood (EPDS>9; 480/1390, 34.5%) than that reported by a previous face-to-face Japanese study (15,506/108,431, 14.3%). This disparity may be due to the anonymity of the survey environment, which allowed mothers to share grievances beyond supporting the image of a good, tolerant mother. We are conducting a face-to-face observational study and analyzing the target population gap caused by the data collection environment, as the question of whether web-based or in-person data accurately reflect the true state of a participant must be considered in the context of increasing web-based communication to minimize error.

The lack of background information also limits the analysis of individual factors in this study. Environmental variables could have a significant impact on the mother’s family environment, but the CPRA could not assess “home-town delivery,” a common practice in Japan where mothers return to their parent’s homes before and after childbirth. Future research using the CPRA with more detailed data on individual characteristics, such as the home environment or family structure, will be needed. In addition, those who had difficulty perceiving support and were sensitive to their child’s temperament are particularly prone to serious psychosocial problems in the infant cohort, but it is difficult to determine whether this is a bias in the cross-sectional survey population or whether those with resilience difficulties are more specific to the infant period. Therefore, we are conducting a longitudinal survey with the environmental information to examine the effects of the environmental setting or the time period.

The accuracy of the clustering algorithm will be limited because of the nature of this algorithm, with a trade-off between explainability and accuracy [17]. In addition, no general method of predetermining k is established [27]. Although the EPDS, PBQ, SOC, and PSQI scores were significantly different across the 5 groups, especially the groups with no versus greatest difficulties in resilience and adaptation, we believe that further investigation is needed to use the partition score as a clear cut-off point for detecting persons with psychosocial disabilities.

Comparison With Prior Work
Previous studies have reported that personality traits can influence perinatal depressive mood. For example, a study involving 15,012 mothers, including 13.1% with depressive moods (an EPDS score≥9 in 1 month was defined as “postpartum depressive symptoms” in that study), indicated that increased neuroticism and reduced extraversion were associated with postpartum depressive symptoms [28]. Regarding interventions for perceived support or parenting resources, a review showed that online peer support groups offered informational and emotional support and positively impacted maternal mental health [27]. A path analysis of web-based survey data collected from Japanese parents of children aged 3-5 years revealed that childcare support had no direct positive effect on children’s social development; however, the benefits of childcare support were mediated by its impact on parents’ psychological state and parenting style, which improved child social development [29]. This indicates that supporting mothers, according to their adaptive tendencies, may positively affect their children’s development. A study of 398 Australian women from the prenatal period to 1 year post partum suggested that the dual intervention of social support and the recognition of prenatal depressive symptoms is a promising strategy to prevent persistent depressive symptoms [30]. This is because digital technology supports the core values associated with psychosocial intervention and fulfills the “ancillary values” that constrain how coproduction operates [31]. To assess the effects of such web-based psychosocial intervention for mothers, we are conducting a randomized controlled trial of a web-based screening and feedback program using the same assessment instruments.

Our study was conducted before and after April 2020, and although the pandemic may have had an impact, we did not detect any significant differences when comparing the different survey periods (Multimedia Appendix 2). This may be because the study was completed in June 2020; however, the data set collected from 5 countries during the COVID-19 pandemic from July to December 2020 also showed that the symptoms of maternal depression and anxiety can be predicted using machine learning algorithms and that efficient tools can be used to predict maternal depression and anxiety [32].

Conclusions
The classifier, generated from the data from the most stressful and confusing period (newborn cohort) using an unsupervised clustering algorithm with enhanced explanatory and applicability, could be adapted to another parenting cohort (infant and toddler cohort), and psychosocially vulnerable groups were detected. Mothers with high sensitivity to the temperament of their children and difficulties in perceiving support would be prone to depressed mood and poor sleep quality, especially in the infant cohort. To overcome the study’s limitations, further research in other study designs is needed. Considering the additional research, the classifier will help support child-rearing in the Society 5.0 era according to the resilience and adaptation characteristics of individual mothers and is likely to contribute to the implementation of web-based child-rearing support.

https://formative.jmir.org/2024/1/e47372 JMIR Form Res 2024 | vol. 8 | e47372 | p.741
(page number not for citation purposes)
Acknowledgments
This work was supported by MEXT “Innovation Platform for Society 5.0” Program grant (JPMXP0518071489). TI received grants from the Japan Society for the Promotion of Science KAKENHI grants (JP20K21837 and JP21K02356). We thank all the participants who provided research data and all those who were involved in managing the project. The project was conducted as part of the Ministry of Education, Culture, Sports, Science, and Technology, Society 5.0 Project PJ1-1 Watching Over the First 1000th Day of Life consortium.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
EK, SS, and ME were responsible for the survey protocol. AH, EK, and TI were responsible for all the analyses. All authors were responsible for primary data collection. AH, EK, and TI were accountable for the initial draft of the study, and they revised the initial draft. All authors contributed to the critical revision of the study, approved the final version, and agreed to be accountable for the content. ME and EK are the guarantors of this study.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Results of Kolmogorov-Smirnov test and Levene test.
[DOCX File, 13 KB - formative_v8i1e47372_app1.docx ]

Multimedia Appendix 2
Comparison of psychological measures before and during pandemic.
[DOCX File, 14 KB - formative_v8i1e47372_app2.docx ]

References

Abbreviations

**CPRA:** Comprehensive Scale for Parenting Resilience and Adaptation

**EPDS:** Edinburgh Postnatal Depression Scale

**PBQ:** Postpartum Bonding Questionnaire

**PSQI:** Pittsburgh Sleep Quality Index
RP: relative prevalence
SOC: Sense of Coherence
Tumor Immunotherapy–Related Information on Internet-Based Videos Commonly Used by the Chinese Population: Content Quality Analysis

Chen-xu Ni¹, MM, MS; Yi-bo Fei¹, MD, PhD; Ran Wu¹, MM, MS; Wen-xiang Cao¹, MM, MS; Wenhao Liu¹, MM, MS; Fang Huang¹, MM, MS; Fu-ming Shen¹, MD, PhD; Dong-jie Li¹, MD, PhD
Shanghai Tenth People’s Hospital, Shanghai, China

Corresponding Author:
Dong-jie Li, MD, PhD
Shanghai Tenth People’s Hospital
301 Middle Yanchang Road
Shanghai, 200072
China
Phone: 86 021 66302570
Email: djili@tongji.edu.cn

Abstract

Background: Tumor immunotherapy is an innovative treatment today, but there are limited data on the quality of immunotherapy information on social networks. Dissemination of misinformation through the internet is a major social issue.

Objective: Our objective was to characterize the quality of information and presence of misinformation about tumor immunotherapy on internet-based videos commonly used by the Chinese population.

Methods: Using the keyword “tumor immunotherapy” in Chinese, we searched TikTok, Tencent, iQIYI, and BiliBili on March 5, 2022. We reviewed the 118 screened videos using the Patient Education Materials Assessment Tool—a validated instrument to collect consumer health information. DISCERN quality criteria and the JAMA (Journal of the American Medical Association) Benchmark Criteria were used for assessing the quality and reliability of the health information. The videos’ content was also evaluated.

Results: The 118 videos about tumor immunotherapy were mostly uploaded by channels dedicated to lectures, health-related animations, and interviews; their median length was 5 minutes, and 79% of them were published in and after 2018. The median understandability and actionability of the videos were 71% and 71%, respectively. However, the quality of information was moderate to poor on the validated DISCERN and JAMA assessments. Only 12 videos contained misinformation (score of >1 out of 5). Videos with a doctor (lectures and interviews) not only were significantly less likely to contain misinformation but also had better quality and a greater forwarding number. Moreover, the results showed that more than half of the videos contain little or no content on the risk factors and management of tumor immunotherapy. Overall, over half of the videos had some or more information on the definition, symptoms, evaluation, and outcomes of tumor immunotherapy.

Conclusions: Although the quality of immunotherapy information on internet-based videos commonly used by Chinese people is moderate, these videos have less misinformation and better content. Caution must be exercised when using these videos as a source of tumor immunotherapy–related information.

(JMIR Form Res 2024;8:e50561) doi:10.2196/50561

KEYWORDS
immunotherapy; internet videos; quality; misinformation; health informatics; Chinese
Introduction

Background
Tumor immunotherapy is an innovative treatment today. After the implementation of China’s new medical insurance rates in 2022, the monthly treatment cost of immunotherapy has entered the “thousand era,” which greatly improves the accessibility of drugs. However, tumor immunotherapy has obvious uncertainty and complexity [1]. Accurate transmission of immunotherapy information to the population is important to the survival and quality of life of patients with cancer [2]. The study found that patients were open to video education and found it helpful and worth watching [3].

The world’s population is increasingly referring to health-related internet-based information as it represents an easily accessible educational tool [4,5]. The Chinese population, overseas Chinese individuals, and people who master Chinese worldwide prefer web-based video applets or websites, such as videos on TikTok, Tencent, iQIYI, or BiliBili [6]. These sites, similar to YouTube, are popular for their rich content, convenient log-in methods, quick sharing, and 24-hour multiplatform seamless application experiences. Recently, the originality, interactivity, and sociable nature of TikTok and BiliBili have provided the younger generation a better user experience and sense of engagement while seeking health information [7]. The penetration and usage of TikTok and BiliBili are also on the rise among some older age groups [8]. However, the medical content available on the internet is controversial and has not been properly examined. Di Bello et al [9] reported that YouTube videos have contributed to the spread of misinformation by underestimating the role of information on immunotherapy for urological tumors in a multimodality approach and missing the findings of published clinical trials. Not only were audiences not availing of accurate therapy, but also they were opting for therapies that may be harmful, which could lead to other complications [10,11].

Objectives
This study aims to report an evaluation of the quality, reliability, and content of videos related to tumor immunotherapy on the internet among the Chinese population. Our findings could serve as a guide for health care providers and awareness campaigns.

Methods

Ethical Considerations
Ethics approval was not required as this descriptive study was conducted by examining publicly accessible videos on the internet. Also, no human participants or animals were included in this study. The study data are anonymous. This study was registered in the Chinese Clinical Trial Registry (ChiCTR2400081071).

Search Strategy and Data Collection
Using the keyword “肿瘤免疫治疗” (“tumor immunotherapy” in Chinese), we searched TikTok, Tencent, iQIYI, and BiliBili on March 5, 2022, which yielded 1820, 395, 400, and 1000 results for each search, respectively. The videos were sorted in accordance with the video’s default “the most viewed” sorting parameter, and the first 50 videos per website were evaluated.

Inclusion and Exclusion Criteria
A total of 200 videos were considered from all the searches. Duplicate videos, paid videos, and videos not related to tumor immunotherapy were excluded. After the screening, we obtained 118 videos for further data extraction and analysis (Figure 1).

Figure 1. Search strategy and video screening procedure.

Variables Extracted
Basic information obtained included the URL, video duration, likes, forwarding number, subscription, comments, and upload date. Profiles of the uploaders were recorded and classified under 5 categories: lectures, interviews, health-related animations, academic institutions or universities, and news agencies. The extracted data were recorded in Excel (Microsoft Corp).

Scoring System
The videos were evaluated independently by 2 authors (CN and Y-BF). The raters were blinded to each other's ratings (they could not consult each other). We reviewed the screened 118 of 3615 videos on TikTok, Tencent, iQIYI, and BiliBili on “tumor immunotherapy,” using the Patient Education Materials Assessment Tool (PEMAT)—a validated instrument for obtaining consumer health information (Multimedia Appendix 1) [12]. Moreover, we adopted 6 questions from Goobie et al [13] to evaluate the videos' content. These 6 questions ask to what degree a video addresses the definition of a disease, its signs and symptoms, risk factors, evaluation, management, and outcomes. Each aspect was scored on a 3-item scale: 0=not addressed, 1=partially addressed, and 2=sufficiently addressed.

The DISCERN quality criteria [14] and the JAMA (Journal of the American Medical Association) Benchmark Criteria [15] were used for assessing the quality and reliability of the health information. The modified version of the original DISCERN questionnaire was used to assess the reliability and quality of
the health information. It consists of five questions, each with a “yes/no” answer (yes=1 point; no=0 points; maximum score=5): (1) Is the video clear and complete? (2) Are reliable sources of information used? (3) Is the information presented balanced and unbiased? (4) Are additional sources of information listed for reference? (5) Are uncertain areas mentioned? The JAMA assessment is used to evaluate web-based videos and resources on the basis of 4 criteria: authorship, attribution, disclosure, and currency (1 point each).

- Authorship (1 point): the video should include authors, contributors, and contact information.
- Attribution (1 point): the references and sources should be listed properly.
- Disclosure (1 point): conflicts of interest, financing, sponsorship, advertising, support, and video ownership should be disclosed.
- Currency (1 point): the dates on which the videos were published and updated should be indicated.

After the scores are calculated, a score of 4 indicates that the source is of high quality.

We assessed the presence of misinformation using an analogous 5-point Likert scale [16,17]. Videos were independently coded by 2 authors with random coding checks to verify intercoder reliability. Each video was rated separately, and its mean score was calculated.

**Statistical Analysis**

The mean, median, IQR, and SD were used as descriptive statistics for continuous variables. To identify differences among the variables extracted, the Mann-Whitney U test was performed. The intraclass correlation coefficient was determined to ensure interrater reliability. A P value of less than .05 was considered significant. Statistical analysis was performed by using the GraphPad Prism 8 (GraphPad Software, Inc).

**Results**

The 118 videos about tumor immunotherapy mostly uploaded by channels dedicated to lectures, health-related animations, and interviews (Table 1; median length 5 minutes; 93, 79% uploaded in and after 2018). The median forwarding number and number of likes was 12 and 15, respectively. However, the median understandability and actionability of the videos were 71% and 71%, respectively. Overall, the quality of information was moderate to poor in 54% of videos (overall DISCERN scores of 1-3 out of 5) and 64% of videos (overall JAMA scores of 1-2 out of 4).

Only 12 videos contained misinformation (score >1 out of 5). Videos with a doctor (published by channels dedicated to lectures and interviews) not only were significantly less likely to contain misinformation but also had better quality and a greater forwarding number. Videos on Tencent and BiliBili had lesser misinformation than TikTok and iQIYI. Regarding DISCERN criteria and JAMA Benchmark Criteria, the quality of information on TikTok and iQIYI was higher than that on BiliBili and Tencent.

Moreover, our results show that more than half of the videos contain little or no content on the risk factors and management of tumor immunotherapy. Overall, over half of the videos had some or more information on the definition, symptoms, evaluation, and outcomes of tumor immunotherapy (Table 2). The overall scores for all internet videos are presented in Figure 2.
**Table 1.** Characteristics of internet-based videos about immunotherapy.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of the video (minutes), median (IQR)</td>
<td>5.0 (1.0-118.2)</td>
</tr>
<tr>
<td><strong>Year of publication of the video, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Before 2018</td>
<td>25 (21)</td>
</tr>
<tr>
<td>2018 and after</td>
<td>93 (79)</td>
</tr>
<tr>
<td>Forwarding number, median (range)</td>
<td>12 (0-364)</td>
</tr>
<tr>
<td>Likes, median (range)</td>
<td>15 (0-1613)</td>
</tr>
<tr>
<td>Comments, median (range)</td>
<td>0 (0-215)</td>
</tr>
<tr>
<td>Subscription, median (range)</td>
<td>0 (0-1473)</td>
</tr>
<tr>
<td><strong>Publisher type, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Lecture</td>
<td>51 (43)</td>
</tr>
<tr>
<td>Interview</td>
<td>22 (19)</td>
</tr>
<tr>
<td>News agency</td>
<td>9 (8)</td>
</tr>
<tr>
<td>Health-related animation</td>
<td>33 (27)</td>
</tr>
<tr>
<td>Academic institution or university</td>
<td>3 (3)</td>
</tr>
<tr>
<td><strong>Overall DISCERN scores, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 (2)</td>
</tr>
<tr>
<td>2</td>
<td>8 (7)</td>
</tr>
<tr>
<td>3</td>
<td>44 (37)</td>
</tr>
<tr>
<td>4</td>
<td>33 (28)</td>
</tr>
<tr>
<td>5</td>
<td>6 (5)</td>
</tr>
<tr>
<td><strong>DISCERN scores, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>TikTok</td>
<td>3.0 (1.0)</td>
</tr>
<tr>
<td>Tencent</td>
<td>2.5 (1.8)</td>
</tr>
<tr>
<td>iQIYI</td>
<td>3.3 (0.6)</td>
</tr>
<tr>
<td>BiliBili</td>
<td>2.7 (1.6)</td>
</tr>
<tr>
<td><strong>PEMAT^a scores (%), median (IQR)</strong></td>
<td></td>
</tr>
<tr>
<td>Understandability</td>
<td>75 (22-100)</td>
</tr>
<tr>
<td>Actionability</td>
<td>71 (0-100)</td>
</tr>
<tr>
<td><strong>Misinformation score, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6 (5)</td>
</tr>
<tr>
<td>2</td>
<td>5 (4)</td>
</tr>
<tr>
<td>3</td>
<td>0 (0)</td>
</tr>
<tr>
<td>4</td>
<td>1 (1)</td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Misinformation score, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Lecture</td>
<td>0.1 (0.6)</td>
</tr>
<tr>
<td>Interview</td>
<td>0.2 (0.5)</td>
</tr>
<tr>
<td>Health-related animation</td>
<td>0.3 (0.7)</td>
</tr>
<tr>
<td><strong>Misinformation score, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>TikTok</td>
<td>0.4 (0.8)</td>
</tr>
<tr>
<td>Tencent</td>
<td>0.09 (0.3)</td>
</tr>
<tr>
<td>iQIYI</td>
<td>0.2 (0.8)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
</tr>
<tr>
<td>BiliBili</td>
<td>0.1 (0.4)</td>
</tr>
<tr>
<td>JAMA&lt;sup&gt;b&lt;/sup&gt; overall score, n (%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17 (14)</td>
</tr>
<tr>
<td>2</td>
<td>47 (40)</td>
</tr>
<tr>
<td>3</td>
<td>16 (13)</td>
</tr>
<tr>
<td>4</td>
<td>4 (3)</td>
</tr>
<tr>
<td>JAMA score, mean (SD)</td>
<td></td>
</tr>
<tr>
<td>TikTok</td>
<td>1.8 (0.8)</td>
</tr>
<tr>
<td>Tencent</td>
<td>1.3 (1.4)</td>
</tr>
<tr>
<td>iQIYI</td>
<td>1.9 (0.7)</td>
</tr>
<tr>
<td>BiliBili</td>
<td>1.3 (1.1)</td>
</tr>
</tbody>
</table>

<sup>a</sup>PEMAT: Patient Education Materials Assessment Tool.

<sup>b</sup>JAMA: Journal of the American Medical Association.

**Table 2.** Completeness of the content of videos on the internet.

<table>
<thead>
<tr>
<th>Content</th>
<th>Definition, n (%)</th>
<th>Symptoms, n (%)</th>
<th>Risk factors, n (%)</th>
<th>Evaluation, n (%)</th>
<th>Management, n (%)</th>
<th>Outcomes, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No content (0 points)</td>
<td>17 (14)</td>
<td>15 (13)</td>
<td>53 (45)</td>
<td>7 (6)</td>
<td>46 (39)</td>
<td>30 (25)</td>
</tr>
<tr>
<td>Little content (0.5 points)</td>
<td>8 (7)</td>
<td>15 (13)</td>
<td>14 (12)</td>
<td>7 (6)</td>
<td>12 (10)</td>
<td>8 (7)</td>
</tr>
<tr>
<td>Some content (1 point)</td>
<td>12 (10)</td>
<td>19 (16)</td>
<td>20 (17)</td>
<td>40 (34)</td>
<td>35 (30)</td>
<td>50 (43)</td>
</tr>
<tr>
<td>Most content (1.5 points)</td>
<td>25 (21)</td>
<td>25 (21)</td>
<td>17 (14)</td>
<td>26 (22)</td>
<td>11 (9)</td>
<td>17 (14)</td>
</tr>
<tr>
<td>Extensive content (2 points)</td>
<td>60 (48)</td>
<td>44 (37)</td>
<td>14 (12)</td>
<td>38 (32)</td>
<td>14 (12)</td>
<td>13 (11)</td>
</tr>
</tbody>
</table>

**Figure 2.** Completeness of content in internet-based videos.

**Discussion**

We screened 118 videos on “tumor immunotherapy” from TikTok, Tencent, iQIYI, and BiliBili commonly used by the Chinese population. Chinese websites or applets uploaded videos related to tumor immunotherapy for the first time in 2011, and the number of videos has significantly increased since 2018. The median duration of the videos was 5 minutes, which is acceptable to the public.

Numerous studies have evaluated videos on YouTube only and not on other networks [18-20]. Our study evaluated information about tumor immunotherapy on the most popular Chinese
websites or applets, using validated instruments to evaluate the quality of information. Videos on BiliBili and TikTok had a significantly greater forwarding number and likes than those on iQIYI and Tencent; a possible reason is that there is no advertisement played before videos on BiliBili and TikTok.

Health care providers should recommend trustworthy sources of information to patients and should actively participate in social media for dissemination of evidence-based medicine. There is a great need for accurate tumor immunotherapy–related content that is also understandable and actionable. Suggestions for content creators include discussing both the benefits and risks of management alternatives, refraining from the use of medical terminology, and presenting the viewer with clear action items. Meanwhile, patients should be wary of internet-based videos. Misinformation, albeit well-intentioned, may be disseminated when a poorly informed patient advises others. Patients should talk to their physicians not only about immunotherapy but also their need for more information.

In conclusion, although the quality of tumor immunotherapy–related information on internet-based videos commonly used by Chinese people is moderate, it has less misinformation and better content. Caution must be exercised when using these videos as a source of tumor immunotherapy–related information.

Acknowledgments
This work was supported by a research project designed by the Chinese Pharmaceutical Association Hospital Pharmacy department (CPA-Z05-ZC-2023002), program for research-oriented physicians of Shanghai Tenth People’s Hospital (grant 2023LCYJFZRC002), and Chongming 2022 “Science and Technology Innovation Action Plan” (CKY2022-24).

Data Availability
The data sets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Full scores for PEMAT measures of understandability and actionability. PEMAT: Patient Education Materials Assessment Tool. [DOCX File, 17 KB - formative_v81e50561_app1.docx]

References


Remote Self-Administration of Cognitive Screeners for Older Adults Prior to a Primary Care Visit: Pilot Cross-Sectional Study of the Reliability and Usability of the MyCog Mobile Screening App

Stephanie Ruth Young1, PhD; Elizabeth McManus Dworak1, PhD; Greg Joseph Byrne1, MA; Callie Madison Jones1, BA; Lihua Yao1, PhD; Julia Noe1ani Yoshino Benavente2, MPH; Maria Varela Diaz1, MS; Laura Curtis2, MS; Richard Gershon1, PhD; Michael Wolf2, MPH, PhD; Cindy J Nowinski1, MD, PhD

1Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
2Center for Applied Health Research on Aging, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Stephanie Ruth Young, PhD
Department of Medical Social Sciences
Feinberg School of Medicine
Northwestern University
625 North Michigan Avenue 21st Floor
Chicago, IL, 60611-3008
United States
Phone: 1 3125033303
Email: stephanieruth.young@northwestern.edu

Abstract

Background: Routine cognitive screening is essential in the early detection of dementia, but time constraints in primary care settings often limit clinicians’ ability to conduct screenings. MyCog Mobile is a newly developed cognitive screening system that patients can self-administer on their smartphones before a primary care visit, which can help save clinics’ time, encourage broader screening practices, and increase early detection of cognitive decline.

Objective: The goal of this pilot study was to examine the feasibility, acceptability, and initial psychometric properties of MyCog Mobile. Research questions included (1) Can older adults complete MyCog Mobile remotely without staff support? (2) Are the internal consistency and test-retest reliability of the measures acceptable? and (3) How do participants rate the user experience of MyCog Mobile?

Methods: A sample of adults aged 65 years and older (N=51) self-administered the MyCog Mobile measures remotely on their smartphones twice within a 2- to 3-week interval. The pilot version of MyCog Mobile includes 4 activities: MyFaces measures facial memory, MySorting measures executive functioning, MySequences measures working memory, and MyPictures measures episodic memory. After their first administration, participants also completed a modified version of the Simplified System Usability Scale (S-SUS) and 2 custom survey items.

Results: All participants in the sample passed the practice items and completed each measure. Findings indicate that the Mobile Toolbox assessments measure the constructs well (internal consistency 0.73 to 0.91) and are stable over an approximately 2-week delay (test-retest reliability 0.61 to 0.71). Participants’ rating of the user experience (mean S-SUS score 73.17, SD 19.27) indicated that older adults found the usability of MyCog Mobile to be above average. On free-response feedback items, most participants provided positive feedback or no feedback at all, but some indicated a need for clarity in certain task instructions, concerns about participants’ abilities, desire to be able to contact a support person or use in-app technical support, and desire for additional practice items.

Conclusions: Pilot evidence suggests that the MyCog Mobile cognitive screener can be reliably self-administered by older adults on their smartphones. Participants in our study generally provided positive feedback about the MyCog Mobile experience and rated the usability of the app highly. Based on participant feedback, we will conduct further usability research to improve support functionality, optimize task instructions and practice opportunities, and ensure that patients feel comfortable using MyCog
Mobile. The next steps include a clinical validation study that compares MyCog Mobile to gold-standard assessments and tests the sensitivity and specificity of the measures for identifying dementia.

*(JMIR Form Res 2024;8:e54299) doi:10.2196/54299*

**KEYWORDS**
cognitive screening; cognitive; cognition; psychometric; usability; feasibility; early detection; dementia; Alzheimer’s disease, Alzheimer’s; Alzheimer’s disease and age-related dementia; mHealth, mobile health apps; detection; screening; mobile health; mobile phone; app; apps; applications; applications; user experience; smartphone; smartphones; gerontology; geriatric; geriatrics; older adult; older adults; elder; elderly; older person; older people; ageing; aging; aged

**Introduction**

Primary care visits provide an important opportunity to detect pathological cognitive decline in the early stages [1,2], yet less than half of all cases are detected in primary care [3]. Medicare covers cognitive screening as part of the Annual Wellness Visit for adults aged 65 years or older, however, primary care clinics face several barriers to conducting regular cognitive screenings with their patients, including constraints on time and clinic staff [4]. Completing a screening remotely before a primary care visit offers several benefits to both patients and clinicians [5]. Patients can complete the screening at their leisure in the privacy of their own homes, and providers can review the results before a visit. Critically, a complete screener before a visit benefits all stakeholders (eg, clinicians, patients, and support staff) by saving time to address other important issues in person [6].

Mobile apps offer an ideal mechanism for many older adults to complete at-home cognitive screeners. More than 60% of older adults in the United States own a smartphone [7], and over 30% regularly use mobile health apps [8]. Moreover, low-income and minority groups are more likely to access their personal health information on smartphones compared to other electronic devices [9]. A small body of emerging research supports the feasibility of self-administered cognitive screeners on personal smartphones in research contexts [10-12]. The cognitive assessments in these studies vary in administration frequency, length, and structure but tend to find high levels of adherence (70% or higher), receive positive feedback in exit surveys, and show convergent validity with established cognitive screening measures [11,13,14]. However, no screeners to date have been validated for clinical use before a primary care visit [15]. Further research is needed to determine if older adults will be able to access the app and complete cognitive screeners independently, if the data collected from these screeners are reliable, and how older adults will perceive the app from a usability and acceptability perspective.

To encourage broader cognitive screening practices within primary care, the National Institute on Aging funded MyCog Mobile (1R01AG074245-01), a cognitive screening app that participants can self-administer remotely on personal smartphones and sends results directly to their primary care provider’s electronic health record. MyCog Mobile is the smartphone-based counterpart to MyCog, a tablet-based app that was developed for in-person self-administration in clinical settings [16]. MyCog Mobile uses 2 measures from MyCog adapted for remote assessment on a smartphone: Picture Sequence Memory (called MyPictures in the mobile app), which measures episodic memory, and Dimensional Change Card Sorting (called MySorting in the mobile app), which measures executive functioning. When combined with self-report, these 2 measures have demonstrated good sensitivity and specificity to detect cognitive impairment [16]. To expand the breadth of cognitive domains assessed, the pilot version of MyCog Mobile also includes 2 additional measures that are not in the original MyCog tablet app: a measure of working memory (MySequences) and a measure of memory for faces (MyFaces). We modeled each of the MyCog Mobile measures on existing mobile measures in the Mobile Toolbox [17], a comprehensive research platform and assessment library that allows for remote cognitive measurement on a personal smartphone (see Measures section).

MyCog Mobile is unique for 2 important reasons. First, it is a cognitive screener meant to be used to help primary care providers make appropriate referrals and care recommendations, as opposed to a pure research measure such as the Mobile Toolbox. Second, MyCog Mobile is intended to be self-administered in a completely unsupervised remote setting, as opposed to the MyCog tablet app which is used in clinics under staff supervision. As such, MyCog Mobile underwent an extensive human-centered design process in which the platform and measures were optimized to be used by older adults in this context [5]. To ensure that the MyCog Mobile measures can be reliably self-administered by older adults in a remote setting, we piloted the screener in a sample of 51 adults aged 65 years or older who completed the measures on their personal iOS (Apple) smartphones. This pilot study will inform a subsequent construct and clinical validations, in which the sensitivity and specificity of the screener will be tested against clinical gold standards in a sample of healthy adults and used to differentiate healthy adults from those with cognitive impairment. Primary research questions for this pilot study include (1) Can older adults complete MyCog Mobile remotely on their smartphones? (2) Are the internal consistency and test-retest reliability of the measures acceptable? and (3) How do participants rate the user experience of MyCog Mobile?

**Methods**

**Ethical Considerations**

The research procedures were reviewed and approved by Northwestern University’s institutional review board (STU00214921). All participants provided informed consent and were compensated with a US $50 Visa gift card for their participation in this study. The data presented in this paper are
anonymous and free of identifiers that could be linked to specific participants.

Sample

We collaborated with a third-party market research agency to recruit older adults (N=51; Table 1) to take the measures on their smartphones twice, about 2 to 3 weeks apart. The agency contacted potential participants in their large database of thousands of older adults who had previously indicated interest in participation in research studies. Sample recruitment was broadly stratified by age, gender, racial and ethnic identity, and highest level of education. Inclusion criteria included (1) aged 65 years or older; (2) ownership of an iOS smartphone version 14 or higher; (3) being English-speaking; and (4) willing to complete the measures twice within approximately 2 to 3 weeks.

Table 1. Descriptive samples and sample demographics of pilot study participants.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Total sample (N=51), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>74.20 (6.25; 65-90)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>29 (57)</td>
</tr>
<tr>
<td>Men</td>
<td>22 (43)</td>
</tr>
<tr>
<td>Racial identity</td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>9 (18)</td>
</tr>
<tr>
<td>White</td>
<td>42 (82)</td>
</tr>
<tr>
<td>Ethnic identity</td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino (any race)</td>
<td>5 (10)</td>
</tr>
<tr>
<td>Not Hispanic or Latino (any race)</td>
<td>46 (90)</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
</tr>
<tr>
<td>HS\textsuperscript{a} diploma or GED\textsuperscript{b}</td>
<td>17 (33)</td>
</tr>
<tr>
<td>Some college</td>
<td>10 (20)</td>
</tr>
<tr>
<td>4-year college degree</td>
<td>12 (23)</td>
</tr>
<tr>
<td>Graduate or professional degree</td>
<td>12 (23)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}HS: high school.
\textsuperscript{b}GED: General Educational Development.

Procedure

Participants were asked to download the MyCog Mobile app onto their devices and complete the 4 activities in the battery and answer 2 demographic questions (age and education level). They received an email from this study’s staff with instructions to download the app and information on how to contact staff for support if needed. The app shows 2 brief intro screens (Figure 1) and then the cognitive screening begins with the learning trial of MyFaces. Participants then complete, in order, MySorting, MySequences, 2 demographics questions (age and education level), the recall subtests of MyFaces (see below), and, finally, MyPictures. After finishing their baseline MyCog Mobile assessment, participants completed a usability survey to provide feedback on their experiences. Participants were asked to self-administer MyCog Mobile a second time within 2 to 3 weeks of their baseline administration.
Figure 1. MyCog Mobile introduction screens.

Measures

MyFaces

MyFaces is an associative memory test originally developed by Rentz and colleagues [18] to predict cerebral amyloid beta burden. The MyCog Mobile version of this task was adapted from the Mobile Toolbox Faces and Names test, which was also based on the original test [17]. Participants are first shown 12 pictures of people paired with their names. After an approximately 5- to 10-minute delay, participants’ memories are tested in 3 subtests: the first subtest (recognition) asks the participant to select the person they saw in the learning trial from 3 options. The second subtest (first letter) asks participants to indicate the first letter of the name of the person presented on the screen (Figure 2). The third subtest (name matching) asks participants to select the name of the person presented from among 3 possible response options. A raw accuracy score is given for each of the 3 subtests.
**MySorting**

MySorting is a measure of executive function and cognitive flexibility adapted from the MyCog Dimensional Change Card Sorting [16] and the Mobile Toolbox Shape-Color Sorting test [17]. Respondents are asked to sort images across 2 dimensions—shape and color—as quickly as they can. The relevant dimension for sorting is indicated by a cue word ("shape" or "color") that appears on the screen (Figure 3). Scores are given for accuracy and response speed.

**Figure 2.** MyFaces first letter subtest example screen (face censored for publication).

**Figure 3.** MySorting example screen.
**MySequences**

MySequences is a measure of working memory adapted from the Mobile Toolbox sequences test [17]. MySequences requires participants to remember strings of letters and numbers and arrange them in order, with the letters in alphabetical order first and then the numbers in ascending numerical order (Figure 4). Trials begin with strings of 3 alphanumeric characters and increase in length, reaching a maximum difficulty of 10 characters. Scores reflect the number of correct trials.

**Figure 4.** MySequences response entry example screen.

**MyPictures**

MyPictures is a measure of episodic memory adapted from the MyCog Picture Sequence Memory [16] and the arranging pictures task in the Mobile Toolbox [17]. A series of images depicting independent, nonsequential activities is presented in a specific order and placed in specific, sequential locations on the screen. Following this presentation, the images are scrambled, and the participant is asked to recall the original position of the images accordingly (Figure 5). There are 2 trials. Scores are given for exact match (the number of pictures in the correct positions) as well as adjacent pairs (the number of correctly ordered pairs of pictures next to each other) on each trial.
**Simplified System Usability Scale**

The Simplified System Usability Scale (S-SUS) is a modified version of the original System Usability Scale designed for adults aged 65 years and older with or without cognitive impairments [19]. Participants rate their level of agreement with statements about their experience using MyCog Mobile on a 5-point Likert scale. The original System Usability Scale has demonstrated evidence of its internal consistency, sensitivity to changes, and concurrent validity with other usability measures.

**Custom Usability Items**

We also asked participants to respond to 2 additional 5-point Likert-scale items regarding their experience using MyCog Mobile: “the time to complete the MyCog Mobile Cognitive Screening was” (1=shorter than I expected, 3=about as much time as I expected, and 5=longer than I expected); and “how would you rate the experience of completing MyCog Mobile overall?” (1=very bad, 3=neutral, and 5=very good). Participants also provided feedback on the experience in 3 free-response items: (“what would you do if the app wasn’t working or you weren’t sure what to do next?”; “is there anything you would change about using the MyCog Mobile App to improve the experience?”; and “is there anything else you would like us to know about your experience using the MyCog Mobile app?”).

**Analysis**

All analyses were conducted in R (R Core Team, 2023), and packages and codes are available on the Open Science Framework [20]. With 51 participants, we had 80% power to detect effect sizes of 0.38 or greater, which was adequate to evaluate our primary outcome of reliability metrics. Internal consistency was assessed using various methods that aligned with each task’s paradigm. For MySorting and MySequences, we calculated median Spearman-Brown correlations between bootstrapped random split-half coefficients for the accuracy scores. For MyPictures, we used the Pearson correlation between trial 1 and trial 2 adjacent pairs’ scores to calculate the Spearman-Brown split-half reliability ($\frac{2r}{1+r}$). For MyFaces, we used a look-up table to find expected a posteriori scores and SDs based on the sum of the accuracy scores across the 3 subtests [21,22] and then calculated the empirical and mean marginal reliabilities [23]. We considered internal consistency coefficients of 0.70 or greater to be acceptable [24]. We used intraclass correlations (ICCs) to evaluate test-retest reliability for each of the measures. ICCs and practice effects are reported for the MySorting total score, MySequences total score, MyPictures sum of adjacent pairs’ scores across trials 1 and 2, and the total score across all 3 subtests for MyFaces. We considered ICCs less than 0.50 to be poor, 0.50 to 0.75 acceptable, 0.75 to 0.90 good, and above 0.90 excellent [25]. Practice effects were evaluated through paired 2-tailed $t$ tests of baseline and retest scores. CIs (95%) that contained 0 were considered to indicate nonsignificant practice effects.

We also conducted exploratory analyses of the relations between test performance, usability, and education, respectively, using Spearman $\rho$ correlations. Spearman $\rho$ correlations were used...
over Pearson $r$ correlations because we were interested in monotonic relationships between variables rather than strictly linear ones. Correlations with age were not conducted due to the restriction of age range by study design. To assess the usability of the screener, we examined the score distributions on the S-SUS and custom Likert-scale items and qualitatively evaluated the results from custom usability survey items. A total score greater than 70 out of 100 possible points is considered above average and an acceptable level of usability [26,27]. Further, 2 authors independently reviewed and coded the free-response items. Codes were then reconciled, grouped, and categorized by representative themes. Although we counted each code’s frequency, the survey free-response items were an informal method of gathering feedback rather than a formal quantitative or qualitative study, and our analysis is exclusively descriptive.

**Results**

**Overview**

Most participants completed both administration time points within 15 days (mean $\text{days}_{\text{between}}$ 15.09 days, SD 2.08; range 13.12-22.38). Further, 2 participants did not complete the second MyCog Mobile assessment, leaving a sample of 49 participants for test-retest reliability analyses.

**Psychometric Properties**

Internal consistency and test-retest reliability statistics were acceptable or better for each measure based on a priori cutoff criteria (Table 2). Test-retest reliability was moderate for each measure. Mean scores were not significantly different between baseline and retest except for MyFaces, which demonstrated a mean improvement of 4.70 (SD 1.06) in the total score across all 3 subtests at the second administration. The performance demonstrated moderate correlations with education level on each of the measures except MyPictures, which did not demonstrate significant correlations with education.

**Table 2.** MyCog Mobile measures reliability, practice effects, and correlation with education.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Internal consistency (95% CI)</th>
<th>Test-retest reliability (ICCb) (95% CI)</th>
<th>Practice effects (ΔM) (95% CI)</th>
<th>Education (p) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MyFaces</td>
<td>0.73 (0.63 to 0.82)</td>
<td>0.61 (0.40 to 0.76)</td>
<td>4.70 (2.62 to 6.79)</td>
<td>0.33 (0.05 to 0.56)</td>
</tr>
<tr>
<td>MySorting</td>
<td>0.90 (0.83 to 0.94)</td>
<td>0.71 (0.54 to 0.82)</td>
<td>1.75 (0.97 to 4.49)</td>
<td>0.45 (0.19 to 0.66)</td>
</tr>
<tr>
<td>MySequences</td>
<td>0.91 (0.85 to 0.95)</td>
<td>0.65 (0.46 to 0.78)</td>
<td>1.77 (0.43 to 3.96)</td>
<td>0.36 (0.08 to 0.58)</td>
</tr>
<tr>
<td>MyPictures</td>
<td>0.81 (0.73 to 0.94)</td>
<td>0.70 (0.53 to 0.82)</td>
<td>0.44 (0.94 to 1.82)</td>
<td>0.06 (0.22 to 0.33)</td>
</tr>
</tbody>
</table>

aSpearman-Brown corrected split-half correlations are reported for MySorting, MySequences, and MyPictures while empirical reliability is reported for MyFaces. Test-retest analyses are based on a sample of n=49.
bICC: intraclass correlation.

**Usability**

The mean overall usability rating on the S-SUS was acceptable (mean 73.17, SD 19.27). Ratings were not significantly correlated with education or performance on any of the measures. Analysis of the S-SUS items demonstrated Likert scale ratings in generally favorable directions (ie, positively worded items were greater than the neutral rating of 3, and negatively worded items were less than 3; Table 3). On additional custom Likert-scale items, participants indicated the time it took to complete the S-SUS was slightly less than expected on average, and the overall experience was positive.
Table 3. Usability ratings of MyCog Mobile.

<table>
<thead>
<tr>
<th>Measure and item</th>
<th>Descriptive range</th>
<th>Rating, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simplified System Usability Scale</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would use the MyCog Mobile app before an appointment with my doctor</td>
<td>Neutral to agree</td>
<td>3.53 (1.25)</td>
</tr>
<tr>
<td>The MyCog mobile app is too complex for me</td>
<td>Disagree to strongly disagree</td>
<td>1.98 (1.16)</td>
</tr>
<tr>
<td>The MyCog mobile app was easy to use</td>
<td>Agree to strongly agree</td>
<td>4.05 (0.97)</td>
</tr>
<tr>
<td>I really need help from someone to use the MyCog mobile app</td>
<td>Disagree to strongly disagree</td>
<td>1.67 (0.105)</td>
</tr>
<tr>
<td>The various parts of the MyCog mobile app were well integrated</td>
<td>Neutral to agree</td>
<td>3.78 (0.97)</td>
</tr>
<tr>
<td>The MyCog Mobile app was confusing for me</td>
<td>Disagree to strongly disagree</td>
<td>1.98 (1.01)</td>
</tr>
<tr>
<td>Learning to use the MyCog Mobile app was quick for me</td>
<td>Neutral to agree</td>
<td>3.75 (1.15)</td>
</tr>
<tr>
<td>The MyCog mobile app was hard to use</td>
<td>Disagree to strongly disagree</td>
<td>1.90 (1.04)</td>
</tr>
<tr>
<td>I felt confident using the MyCog mobile app</td>
<td>Neutral to agree</td>
<td>3.98 (1.12)</td>
</tr>
<tr>
<td>I will need to learn a lot before using the MyCog mobile app</td>
<td>Neutral to disagree</td>
<td>2.04 (1.11)</td>
</tr>
</tbody>
</table>

**Additional Likert-scale questions**

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The time to complete MyCog Mobile was...</td>
<td>As much time as expected or less</td>
</tr>
<tr>
<td>How would you rate the experience of completing MyCog Mobile overall?</td>
<td>Good to very good</td>
</tr>
</tbody>
</table>

**Free-Response Feedback**

On the first free-response item, “what would you do if the app wasn’t working, or you weren’t sure what to do next?” participants’ responses indicated several strategies they would use for help (Table 4). Most expected to be able to directly contact someone for support through an email or phone call. Several participants indicated they would use an in-app help resource or search for help resources online. Although most participants indicated they would try to solve the problem, 3 participants stated they would not finish MyCog Mobile if they encountered a difficulty (eg, “[I would] disregard it and continue on as I had before the app”).

On the second free-response item, “is there anything you would change about using the MyCog Mobile App to improve the experience?” most participants did not offer any suggestions or gave positive feedback (Table 4). Several participants suggested the instructions for the cognitive tests needed clarification (eg, “some of the exercises were not well explained or confusing. Particularly the ones where random letters and numbers were given, and they had to be reorganized. Simplifying the instructions would be helpful”). Some participants expressed concerns about their memories in response to this question or commented on the difficulties of the test items (eg, “I just wish I was smarter [and] had a better memory”). Regarding visual accessibility of the app, 2 participants indicated difficulty with the print size, and 1 indicated difficulty with the visual contrast of the tasks. Of note, 1 participant remarked they would prefer to complete MyCog Mobile in the clinic (eg, “while the app itself was easy & straightforward to download and access the survey material, I would most likely defer its home use and prefer an ‘in doctor’s office’ cognitive testing experience”). However, another participant remarked on how easy it would be to complete MyCog Mobile before the appointment (eg, “I really don’t see that anything was difficult. The app would work very well prior to a doctor visit.”). Further, 3 participants offered feedback on the process of participating in this pilot study, which will be considered for future study administration but is not relevant to the MyCog Mobile user experience specifically.

On the final free-response item, “is there anything else you would like us to know about your experience using the MyCog Mobile app?” most participants did not provide any feedback. As with the previous free-response item, several participants expressed concerns about their abilities (eg, “I found it to be quite challenging, especially since my memory isn’t what it once was.”). Some commented the instructions were confusing (eg, “no at 1st it was sort of confusing once I got into it, it was easier”), and 2 wanted more opportunities to practice before starting the live items. (eg, “I would like to see more practice questions to help the user feel more relaxed and confident”). Further, 2 participants wanted more explanation around the purpose of the test (eg, “perhaps a brief description of what the test is designed for, for example: to test mental recall, to test cognitive ability, to test onset of dementia or Alzheimer’s”). Only 1 participant reported difficulties loading the app for this study. Conversely, many participants provided positive feedback (eg, “it was simple and easy to use”).

https://formative.jmir.org/2024/1/e54299
Table 4. Free-response feedback on MyCog Mobilea.

<table>
<thead>
<tr>
<th>Item and type of response</th>
<th>Frequency, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>“What would you do if the app wasn’t working, or you weren’t sure what to do next?”</td>
<td></td>
</tr>
<tr>
<td>Contact a support administrator</td>
<td>27</td>
</tr>
<tr>
<td>Self-troubleshoot or restart the app</td>
<td>13</td>
</tr>
<tr>
<td>Use in-app help or search online</td>
<td>7</td>
</tr>
<tr>
<td>Not finish or give up</td>
<td>3</td>
</tr>
<tr>
<td>“Is there anything you would change about using the MyCog mobile app to improve the experience?”</td>
<td></td>
</tr>
<tr>
<td>No changes suggested or positive feedback</td>
<td>32</td>
</tr>
<tr>
<td>Clarify task instructions</td>
<td>7</td>
</tr>
<tr>
<td>Concerns about own abilities or test difficulty</td>
<td>5</td>
</tr>
<tr>
<td>Visibility issues</td>
<td>3</td>
</tr>
<tr>
<td>Concerns related to study administration</td>
<td>3</td>
</tr>
<tr>
<td>Preference for in-person experience</td>
<td>1</td>
</tr>
<tr>
<td>“Is there anything else you would like us to know about your experience using the MyCog mobile app?”</td>
<td></td>
</tr>
<tr>
<td>No feedback</td>
<td>24</td>
</tr>
<tr>
<td>Positive feedback</td>
<td>14</td>
</tr>
<tr>
<td>Concerns about own abilities or test difficulty</td>
<td>5</td>
</tr>
<tr>
<td>Additional practice items</td>
<td>2</td>
</tr>
<tr>
<td>Purpose of test unclear</td>
<td>2</td>
</tr>
<tr>
<td>Difficulty loading app</td>
<td>1</td>
</tr>
</tbody>
</table>

aParticipant responses could be coded for multiple themes; therefore, the frequency should not sum to the total sample size.

Discussion

Principal Findings

The findings suggest most healthy older adults can reliably complete the MyCog Mobile screener remotely on their smartphones. The 4 performance measures that comprise the MyCog Mobile screener demonstrated acceptable internal consistency and test-retest reliability. The performance demonstrated positive correlations with education as expected, except for MyPictures, which did not correlate with education. Participants in our sample rated the usability of MyCog Mobile as above average and rated the experience “Good” to “Very Good” overall. They indicated the time to complete MyCog Mobile was about as long as they expected. These results provide evidence of the feasibility and acceptability of remote self-administration of the MyCog Mobile cognitive screener and support its further evaluation in larger clinical samples to understand its diagnostic accuracy and construct validity.

The feedback on free-response items indicated that most participants had a positive user experience and revealed several actionable insights for the next iteration of the MyCog Mobile app. First, participants expect a dedicated support representative to be available if they have difficulty using the app. Clinics that implement MyCog Mobile into their workflows will have to consider how to best respond to the needs of patients. For example, clinics may choose to dedicate resources for support or inform patients that the screener is optional, and they may defer the use of the app until their clinic visit if they encounter problems. Participants also indicated that they expect a help resource within the app. Currently, participants can use the “Pause” icon to stop the activities and review the instructions. However, an additional button labeled “Help” may be easier to navigate for older adults. Participants indicated that they would use several strategies to troubleshoot on their own if they encountered difficulties (eg, restarting the app); however, clinics should expect some participants not to finish MyCog Mobile if problems arise. For these patients, clinics will have to default to their previous screening workflows (eg, using in-person screeners like the MyCog tablet app or a traditional paper-and-pencil screener).

Concerning what could be improved with the app, some participants offered feedback on the instructions for the activities and asked for more opportunities to practice before completing live items. Although this feedback came from a minority of patients (7/51, 14%), we will conduct further cognitive interviewing to ensure instructions are optimized for all users. Currently, participants are only allowed to try practice items again if they respond incorrectly. However, adding the opportunity to try the practice again even if the item is correct may be helpful for participants. Further, 2 participants also reported trouble reading the print on a smartphone screen. To address this, we increased the font size of the print to maximize the readability of the text which will be implemented in the subsequent MyCog Mobile validation studies.

https://formative.jmir.org/2024/1/e54299 JMIR Form Res 2024 | vol. 8 | e54299 | p.761 (page number not for citation purposes)
Several participants provided feedback that reflected insecurities about their abilities or performance on the test. Some reported the items were too difficult, however, the item difficulties cannot be changed to preserve the validity of the test. Instead of changing the items, steps could be taken to assure patients that it is normal for the items to be challenging. Based on feedback in a previous study, we designed the introduction screen (Figure 1) to alleviate potential concerns about the purpose of the test. For the next iteration of MyCog Mobile, we will collect participant feedback on how to optimize the introduction screen to make patients feel comfortable and assured when completing the screener at home.

Limitations

The generalizability of our findings is limited by the relative homogeneity of our small sample about racial and ethnic identities. Representation of racial identities other than White was low or nonexistent, and representation of Hispanic or Latino populations was relatively small. Findings will need to be replicated in these populations in future studies to ensure MyCog Mobile has equal validity evidence for these groups. Moreover, due to the constraints of the grant, we developed the first version of MyCog Mobile for iOS devices (iPhones) only. iPhones are among the most expensive smartphones, which may have biased our sample toward higher-income participants (though we did not collect income information). Future work will focus on developing and validating MyCog Mobile for Android.

In this small pilot study, we were not able to conduct qualitative interviews with patients but rather gave them opportunities to provide feedback via free-response survey items. While free-response items can capture a breadth of spontaneous viewpoints, they may not achieve the depth or nuanced understanding of participants’ experiences and perspectives that can be gleaned from qualitative interviews. Consequently, our findings might not encompass the subtleties or the full range of participant experiences with MyCog Mobile.

The recruitment of older adults with cognitive impairments was outside of the scope of our pilot study; however, it is important to note that MyCog Mobile has yet to be researched in these populations. We expect older adults who are currently struggling with cognitive decline will likely have difficulty using MyCog Mobile, and the app may be more appropriate for participants who are cognitively intact or in the early stages of cognitive decline. The forthcoming clinical validation of MyCog Mobile will provide valuable information about the sensitivity and specificity of the measures for detecting cognitive decline as well as the feasibility of using the app with cognitively impaired populations.

Further, 1 potential limitation of the MyCog Mobile app is older adults’ familiarity with mobile health apps in general. Although smartphone ownership is increasingly more common across age groups, some older adults still do not own smartphones or feel confident using them. The participants in our sample rated the user experience highly; however, the acceptability of a remote cognitive screening app is likely to be lower in a general population sample that has not chosen to participate in a highly controlled research study. Based on the results of the clinical validation, we will conduct a field test of MyCog Mobile, in which we will collect feedback on the acceptance of the app in real-world contexts.

Comparisons With Prior Work

Our findings are consistent with the small body of research on the feasibility and acceptability of smartphone apps for cognitive screening but also offer some novel contributions. Several studies have examined repeated cognitive assessments on smartphones in research contexts for older adults, although these have primarily examined adherence [11,13,28]. Further, 1 ecological momentary assessment study found that both cognitively normal and older adults with mild cognitive impairment were able to complete cognitive assessments on their smartphones with an adherence rate of 85% in the context of a research study [29]. We observed a 96% adherence rate in our pilot study (albeit with only 2 administration time points), but it is unclear if patients will respond the same way when MyCog Mobile is used in the context of a real-world primary care visit, even if they are asked to complete the activities once annually.

Research on attitudes toward cognitive screening in primary care suggests that most older adults are open to cognitive screening in primary care if they perceive there is a benefit [30]. In our sample, the average response to “I would use the MyCog Mobile app before an appointment with my doctor” skewed positive, but 10 (20%) out of 51 participants responded with “Disagree” or “Strongly Disagree.” Future iterations of the app will focus on communicating the benefits of cognitive screenings to older adults, especially in the absence of clinical staff to explain the assessments when they are taken at home. Moreover, clinics should expect there to be a portion of patients who do not complete MyCog Mobile before the visit and will need to complete usual-care cognitive screenings in the clinic. MyCog Mobile is not intended to replace all in-person screening practices, but rather supplement such practices. Likewise, MyCog Mobile is not intended to provide a clinical diagnosis, but rather to identify potential cognitive impairment, and lead to appropriate referrals for more comprehensive evaluation. For the portion of patients who are willing and able to complete MyCog Mobile on their smartphones before their appointment, clinics can use in-person time to focus on other important aspects of the visit.

Conclusions

Pilot evidence suggests the MyCog Mobile cognitive screener can be reliably self-administered by older adults on their smartphones. Participants in our study generally provided positive feedback about the MyCog Mobile experience and rated the usability of the app highly. Based on participant feedback, we will conduct further usability research to improve support functionality, optimize task instructions and practice opportunities, and make patients feel comfortable using MyCog Mobile. Additional next steps include a clinical validation study that compares MyCog Mobile to gold-standard assessments and tests the sensitivity and specificity of the measures for identifying cognitive impairment.
References


Abbreviations

ICC: intraclass correlation
S-SUS: Simplified System Usability Scale

©Stephanie Ruth Young, Elizabeth McManus Dworak, Greg Joseph Byrne, Callie Madison Jones, Lihua Yao, Julia Noelani Yoshino Benavente, Maria Varela Diaz, Laura Curtis, Richard Gershon, Michael Wolf, Cindy J Nowinski. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Vision-Language Model for Generating Textual Descriptions From Clinical Images: Model Development and Validation Study

Jia Ji¹, MSc; Yongshuai Hou², PhD; Xinyu Chen³, BSc; Youcheng Pan², PhD; Yang Xiang², PhD

¹Shenzhen Institute of Information Technology, Shenzhen, China
²Peng Cheng Laboratory, Shenzhen, China
³Harbin Institute of Technology, Shenzhen, China

Corresponding Author:
Youcheng Pan, PhD
Peng Cheng Laboratory
No. 2 Xingke 1st Street
Shenzhen, 518000
China
Phone: 86 18566668732
Email: pan.youcheng4@gmail.com

Abstract

Background: The automatic generation of radiology reports, which seeks to create a free-text description from a clinical radiograph, is emerging as a pivotal intersection between clinical medicine and artificial intelligence. Leveraging natural language processing technologies can accelerate report creation, enhancing health care quality and standardization. However, most existing studies have not yet fully tapped into the combined potential of advanced language and vision models.

Objective: The purpose of this study was to explore the integration of pretrained vision-language models into radiology report generation. This would enable the vision-language model to automatically convert clinical images into high-quality textual reports.

Methods: In our research, we introduced a radiology report generation model named ClinicalBLIP, building upon the foundational InstructBLIP model and refining it using clinical image-to-text data sets. A multistage fine-tuning approach via low-rank adaptation was proposed to deepen the semantic comprehension of the visual encoder and the large language model for clinical imagery. Furthermore, prior knowledge was integrated through prompt learning to enhance the precision of the reports generated. Experiments were conducted on both the IU X-RAY and MIMIC-CXR data sets, with ClinicalBLIP compared to several leading methods.

Results: Experimental results revealed that ClinicalBLIP obtained superior scores of 0.570/0.365 and 0.534/0.313 on the IU X-RAY/MIMIC-CXR test sets for the Metric for Evaluation of Translation with Explicit Ordering (METEOR) and the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) evaluations, respectively. This performance notably surpasses that of existing state-of-the-art methods. Further evaluations confirmed the effectiveness of the multistage fine-tuning and the integration of prior information, leading to substantial improvements.

Conclusions: The proposed ClinicalBLIP model demonstrated robustness and effectiveness in enhancing clinical radiology report generation, suggesting significant promise for real-world clinical applications.

(JMIR Form Res 2024;8:e32690) doi:10.2196/32690

KEYWORDS
clinical image; radiology report generation; vision-language model; multistage fine-tuning; prior knowledge

Introduction

Radiology reports offer essential textual descriptions of radiographs and play a pivotal role in the medical diagnosis and treatment process. Their precise interpretation can directly influence patient outcomes, underscoring the gravity of each assessment. However, even for seasoned radiologists, interpreting these images can be a meticulous task, often taking several minutes per image. In an era where timely medical intervention can be lifesaving, streamlining this process becomes imperative. Recognizing the immense potential to ease the workload of the health care sector and improve patient care, there has been a growing interest in the research for automatic radiology report generation.
As shown in Figure 1, several attempts have been made in the medical field to create medical reports from images. In the early stage, most researchers used traditional deep learning methods, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to produce radiology reports. IU X-RAY proposed by Demner-Fushman et al [1] was a significant step in this direction. In addition, Shin et al [2] innovatively applied a CNN-RNN model for structured report creation. Wang et al [3] used a nonhierarchical CNN-long-short term memory approach, emphasizing both semantic and visual cues. Vinyals et al [4] introduced visual attention mechanisms in the realm of image captioning with CNN-RNN structures. Subsequently, radiology report creation has evolved to adopt transformer-based models [5,6]. The Knowledge-Driven Encode, Retrieve, and Generate (KERV) model [7,8] and its extended Knowledge-Driven Encode, Retrieve, and Generate, then Retrieve and Generate (KERV2) model [9] have been proposed to improve model performance. Particularly, the current best model, ClinicalBLIP [10], is a transformer-based model with a large knowledge base that can generate accurate medical reports.

Recently, vision-language models (VLMs) [12-15] have become leading approaches, which use pretrained transformer models to handle both visual and textual data at the same time. These models are very good at understanding and creating content based on images and texts. One key feature is cross-modal learning [16,17], where VLMs learn to match specific image patterns with their related descriptions or findings. This understanding helps in making reports that are more relevant and accurate. VLMs have the potential to greatly improve radiology report generation by increasing accuracy, making processes faster, and ensuring consistency. However, it is important to address challenges related to data quality, integration, and rules when using VLMs in clinical settings. Thus, designing an effective fine-tuning method to boost VLM’s knowledge and understanding of medical images and texts is a very interesting research direction.

In this study, we fine-tune a medical VLM named ClinicalBLIP through a multistage fine-tuning strategy for the radiology report generation task. First, a joint optimization method that combines self-attention fine-tuning via low-rank adaptation (LoRA) [18] with layer normalization [19] is proposed to enhance the understanding of clinical images by a general visual encoder. The training target is the text generation loss of the large language model (LLM) without introducing extra clinical image-text pairs for further pretraining. Second, the joint fine-tuning process for both the image-text transformation layer and the multilayer perceptron (MLP) layer of the language model is designed to allow the LLM to draw upon its internal capability to generate the final report. In addition, we further incorporate the prior information to light the specialized clinical knowledge inherent in the LLM. Also, the clinical tag and brief description of the image as a text prompt are fed into the model for training and prediction. Experiments were conducted on the IU X-RAY [1] and MIMIC-CXR [20] data sets. We compared the proposed model with 11 competitor methods and analyzed the performance in several aspects. It is demonstrated that the proposed ClinicalBLIP achieved state-of-the-art performance for training and prediction. Experiments were conducted on the IU X-RAY and MIMIC-CXR data sets. We compared the proposed model with 11 competitor methods and analyzed the performance in several aspects. It is demonstrated that the proposed ClinicalBLIP achieved state-of-the-art performance for training and prediction. Experiments were conducted on the IU X-RAY and MIMIC-CXR data sets. We compared the proposed model with 11 competitor methods and analyzed the performance in several aspects. It is demonstrated that the proposed ClinicalBLIP achieved state-of-the-art performance for training and prediction. Experiments were conducted on the IU X-RAY and MIMIC-CXR data sets. We compared the proposed model with 11 competitor methods and analyzed the performance in several aspects. It is demonstrated that the proposed ClinicalBLIP achieved state-of-the-art performance for training and prediction. Experiments were conducted on the IU X-RAY and MIMIC-CXR data sets. We compared the proposed model with 11 competitor methods and analyzed the performance in several aspects. It is demonstrated that the proposed ClinicalBLIP achieved state-of-the-art performance for training and prediction. Experiments were conducted on the IU X-RAY and MIMIC-CXR data sets. We compared the proposed model with 11 competitor methods and analyzed the performance in several aspects. It is demonstrated that the proposed ClinicalBLIP achieved state-of-the-art performance for training and prediction. Experiments were conducted on the IU X-RAY and MIMIC-CXR data sets. We compared the proposed model with 11 competitor methods and analyzed the performance in several aspects. It is demonstrated that the proposed ClinicalBLIP achieved state-of-the-art performance for training and prediction.

Methods

Data Set

We evaluated our proposed method on the IU X-RAY [1] and MIMIC-CXR [20] data sets. Both data sets have been automatically deidentified.

The IU X-ray data set comprises 7470 images and 3955 reports. The images consist of chest x-rays originating from Indiana University. Each report in the data set primarily encompasses multiple attributes such as comparison, indications, findings, and impressions. Reports with empty findings were excluded, resulting in 3337 remaining reports. Subsequently, we divided the remaining reports into training and testing sets in a 4:1 ratio, yielding 2668 reports for training and 669 reports for testing.

The MIMIC-CXR data set was created by the Massachusetts Institute of Technology. The images are sourced from 65,379 patients who presented to the Beth Israel Deaconess Medical Center Emergency Department between 2011 and 2016. We used 152,173 medical reports for training and 1196 reports for testing. In this data set, each data entry comprises a specific report and 1 to 3 corresponding images.
Overview of the Proposed Method

Our work aims to transform clinical radiographs, accompanied by additional information, into textual descriptions that convey the same semantic meaning as the images. To achieve this, we introduce the ClinicalBLIP model, as depicted in Figure 2. This model comprises three core modules: (1) a visual encoder for converting clinical images into semantic representations; (2) query transformer (Q-Former), a crucial component for bridging the image-text gap; and (3) a LLM for generating textual reports based on queries learned from Q-Former and textual prompts.

Figure 2. Overview of the proposed ClinicalBLIP model. LLM: large language model; Q-Former: query transformer.

Initially, we briefly introduce the structure and pretraining of the ClinicalBLIP, which draws inspiration from Li et al [21], especially how Q-Former as an intermediate module effectively connects visual and textual data. Subsequently, we delve into the details of how to effectively fine-tune the task of radiology report generation.

Q-Former to Bridge the Modality Gap

Q-Former is designed to link a fixed image encoder with a standard LLM. Notably, it can extract a consistent set of features from the visual encoder, regardless of the input image resolution. As shown in Figure 3, the model is composed of two primary transformer submodules: (1) an image transformer for direct interaction with the visual encoder and (2) a text transformer that serves as both encoder and decoder. The efficacy of the Q-Former is greatly influenced by learnable query embeddings, which facilitate self-attention and cross-attention layer interactions. These embeddings also enable communication with text through similar attention mechanisms. During its 2 pretraining phases, that is, vision-language representation learning and vision-to-language generative learning, Q-Former uses distinct attention masks for specific tasks, controlling the interaction between queries and text.

Figure 3. Model architecture of query transformer (Q-Former).

Vision-Language Representation Learning From Visual Encoder

In the representation learning phase, Q-Former, connected to a frozen visual encoder, undergoes pretraining with image-text pairs. The objective here is to train the model to enable queries to extract visual representations corresponding to the text. Inspired by Li et al [22], 3 pretraining tasks are jointly optimized, using the same input format and model parameters. As illustrated in Figure 3, these tasks include image-text contrastive learning, image-grounded text generation, and image-text matching. Image-text contrastive learning aligns
image and text representations by contrasting the similarity of a positive image-text pair against that of negative pairs. Image-grounded text generation encourages the Q-Former to compel the queries to extract visual features that contain the whole information of the text. Image-text matching seeks to capture fine-grained alignment between image and text representations through a binary classification task. Each task uses a specific attention-masking strategy to control the interaction between queries and text.

**Vision-to-Language Generative Learning From LLM**

During the generative pretraining phase, Q-Former, connected to a frozen LLM, leverages its language generation capabilities. A fully connected layer is used to linearly project the output query embeddings to match the dimension of the LLM’s text embedding. These embeddings then act as visual prompts, guiding the LLM based on the visual representation captured by Q-Former. Since Q-Former has been trained to extract visual representations that are informative for language, it effectively serves as an information filter, providing only the most relevant information to the LLM and excluding unnecessary visual details. This setup reduces the load on the LLM to learn vision-language alignment, mitigating the risk of the catastrophic forgetting problem.

**General Vision-Language Instruction Tuning**

Following the pretraining phases, as in Dai et al [23], Q-Former undergoes a vision-language instruction tuning process. Here, the LLM integrates visual encodings from Q-Former with additional instruction text tokens. The instruction interacts with the query embeddings through the Q-Former’s self-attention layers. This interaction aids in extracting relevant image features, which are further provided to optimize the LLM for following instructions. Both quantitative and qualitative analyses confirm the effectiveness of the instruction tuning process in achieving vision-language zero-shot generalization.

**Effective Fine-Tuning of Radiology Report Generation**

To enhance the performance of a general visual encoder and an LLM for medical image understanding and report generation, various aspects need careful consideration. As shown in Figure 4, a multistage parameter fine-tuning approach is used to improve model performance, namely visual encoder enhancement and vision-language joint training.

In the first stage, the model’s weights are adjusted to focus more on relevant features within medical images. This refinement aids in understanding critical elements such as lesions, organs, and more. Concurrently, layer normalization is applied to maintain a consistent response across varying image scales and brightness levels. The primary objective here is the generation loss of the LLM, aiming to improve the quality of final reports by enhancing the visual encoder’s ability to use visual information more effectively during report generation, without the need for additional medical text data for further pretraining.

In the second stage, the joint training process encompasses the fusion of visual and textual inputs, and crucially, the incorporation of the attention layer and MLP layer of the LLM. The model simultaneously processes information from the visual encoder and textual sources. The attention layer enables dynamic focus on specific regions of medical images, aligning with features crucial for report generation. Meanwhile, the MLP layer transforms the combined visual-textual data, boosting the model’s ability to generate contextually accurate and coherent medical reports. The whole approach ensures full use of the model’s attention and transformation capabilities, yielding medically precise and linguistically sound reports, thus effectively bridging the gap between visual and textual data.

Moreover, general LLMs often struggle with the absence of specialized medical domain knowledge in the medical report generation task. To mitigate the issue, we incorporate the prior information into the model during the second stage. Specifically, medical tags related to the medical image and a brief image description are embedded as text prompts. In training, these prompts are linked to corresponding medical images, facilitating the model’s comprehension of the image content. This association enables the model to better learn medical
domain-specific terms and concepts. This embedding of text prompts guides the model with domain knowledge, addressing its limitations in the medical field. During prediction, these prompts provide additional contextual information, enabling the model to better understand medical images, identify features within them, and express the medical reports in a more professional manner.

In our research, we analyze 2 relevant data sets: IU X-RAY [1] and the MIMIC-CXR [20]. Each of them has unique prior information as input. The IU X-RAY data set enriches the model with essential prior information, including “problem,” “image,” and “indication.” The input template for the IU X-RAY data set is formatted as follows:

“Problems: {problem} \n Image: {image} \n Indication: {indication} \n”

exemplified by “Problems: normal \n Image: Chest, 2 views, frontal and lateral \n Indication: Pruritic \n”.

In contrast, the MIMIC-CXR data set lacks direct access to similar prior information. To maintain consistency, we use the CheXBERT [24] model to extract medical observations from the reports within the MIMIC-CXR data set. The input template for this data set is formatted as follows:

“Symptoms of existence: {} \n Symptoms of non-existence: {} \n”

illustrated by “Symptoms of existence: Cardiomegaly, Atelectasis \n Symptoms of non-existence: Edema, Consolidation \n”.

**Experimental Settings**

We adopt the InstructBLIP [23] as the base model, in which contrastive language-image pretraining [13] and Flan-T5-XL [25] are used as visual encoders and LLM structures, respectively. In the training phase, we integrated LoRA [18] into both the visual encoder and the language model. This integration of LoRA was strategically implemented within the query projection and value projection stages during self-attention operations, enhancing the model’s ability to capture and leverage relational information. For the training process, we configured our settings as follows: a batch size of 3 was used, and gradient accumulation was carried out over 4 steps to facilitate stable and efficient training. The initial learning rate for the Q-Former parameters was set to $1 \times 10^{-4}$, while the initial learning rate for the LoRA-related parameters was established at $5 \times 10^{-4}$. To dynamically adapt the learning rate during training, we used a cosine decay learning rate scheduler, optimizing the convergence and fine-tuning process. Furthermore, to enhance the training efficiency and minimize memory consumption, we utilized float16 precision, a half-precision training technique, which effectively balances training speed and model performance. This comprehensive approach allowed us to train our model effectively, incorporating LoRA’s enhancements for improved performance and robustness. All the experiments are conducted on a graphics processing unit (NVIDIA V100).

To evaluate the performance of the ClinicalBLIP model, we compared our method with the following 11 state-of-the-art methods. R2GEN [9] is a memory-driven radiology report generation model with a relational memory to record the information from the previous generation processes and a layer normalization mechanism to incorporate the memory. CA [26] is a contrastive attention model to capture and depict abnormalities by comparing the input image with known normal images. CMCL [27] is a novel competence-based multimodal curriculum learning framework to alleviate data bias by efficiently using limited medical data for medical report generation. Posterior-and-Prior Knowledge Exploring-and-distilling [28] is an effective approach to exploring and distilling posterior and prior knowledge for radiology report generation. R2GEN enhanced with cross-modal memory networks [29] is a radiology report generation model with cross-modal memory networks in which a memory matrix is used to record the alignment and interaction between images and texts, and another memory is used to perform querying and responding to obtain the shared information across modalities. ALIGNTRANSFORMER [30] is a radiology report generation model to alleviate the data bias problem and model the very long sequence. Knowledge Matters [31] is a novel radiology generation framework assisted by general and specific knowledge. Meshed-Memory Transformer [32] is a simple but effective progressive text generation model to produce the radiology report by incorporating high-level concepts into the generation process. Reinforcement Learning Over a Cross-Modal Memory (CMM-RL) [33] is an enhanced radiology report generation model with reinforced cross-modal alignment to alleviate the requirement of annotated supervision while facilitating interactions across modalities. Cross-Modal Contrastive Attention (CMCA) [34] is a novel model to capture both visual and semantic information from similar cases. Observation-Guided Radiology Report Generation (ORGAN) [35] is a generation framework that can maintain the clinical consistency between radiographs and generated free-text reports.

We adopted natural language generation metrics to evaluate the methods. Specifically, we selected Bilingual Evaluation Understudy (BLEU) [36], Metric For Evaluation of Translation with Explicit Ordering (METEOR) [37], and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [38]. BLEU-1, BLEU-2, BLEU-3, BLEU-4, METEOR, and ROUGE-L are reported.

BLEU is primarily used to evaluate the quality of machine-generated translations by comparing them to 1 or more reference translations. It computes a precision-based metric by counting the number of n-grams (contiguous sequences of n items, usually words) in the generated translation that matches any reference translation. In this work, BLEU is used to evaluate the generated text report.

METEOR is based on the harmonic mean of unigram precision and recall, with recall weighted higher than precision. It incorporates features not found in other metrics, such as stemming and synonymy matching, along with standard exact word matching. The metric was designed to address some of the issues found in the more popular BLEU metric and to produce a good correlation with human judgment at the sentence or segment level.
**Results**

**Quantitative Evaluation**

Tables 1 and 2 provide the quantitative results of the IU X-RAY and MIMIC-CXR test sets, respectively. The detailed results show that the ClinicalBLIP model exhibited robust performance when compared with other methods across the IU X-RAY and MIMIC-CXR data sets. For the IU X-RAY data set, as shown in Table 1, although ClinicalBLIP was slightly inferior to the competitor methods on some individual metrics, it significantly surpassed the competitor methods on most metrics. With a BLEU-A score of 0.296, it boasted an improvement of roughly 6.9% over its nearest competitor, CMCA, which had a BLEU-A score of 0.277. This showcases ClinicalBLIP’s enhanced capability in producing reports that are more aligned with the reference. Moreover, when assessing the METEOR metric, which provides insights into the robustness of generation, ClinicalBLIP achieved a score of 0.570. This was approximately 1.7 times higher than CMCA’s 0.209, reflecting ClinicalBLIP’s superior relevance to the generated report. The ROUGE-L metric further solidified this observation; ClinicalBLIP’s score of 0.534 was about 33.8% higher than ORGAN’s score of 0.399, suggesting that ClinicalBLIP consistently maintained a high level of linguistic quality and relevance in its results.

For the MIMIC-CXR data set, as shown in Table 2, there were areas where ClinicalBLIP did not have the highest score, but its comprehensive performance remains commendable. The BLEU-A score for ClinicalBLIP stood at 0.162, which, while marginally behind ORGAN’s score of 0.184, indicates a competitive translation quality. However, ClinicalBLIP made a strong comeback in the METEOR metric, recording a score of 0.365, which is approximately 1.25 times higher than ORGAN’s score of 0.162. This underlines ClinicalBLIP’s proficiency in generating semantically relevant reports. Furthermore, with a ROUGE-L score of 0.313, ClinicalBLIP managed to surpass ORGAN by roughly 6.8%, emphasizing its consistent linguistic excellence.

In summary, while individual metrics might have seen close competition, the overall trend clearly indicates the comprehensive strength of the ClinicalBLIP model. Its consistently high scores across various data sets and metrics demonstrate its versatility and reliability in the realm of clinical report generation.

---

**Ethical Considerations**

This study complied with all relevant ethical regulations. All the publicly available data sets have been deidentified and anonymized. With institutional review board approval (OHSRP#5357) by the National Institutes of Health Office of Human Research Protection Programs, the IU X-RAY data set was made publicly available by Indiana University, and no informed consent was necessary [1]. The MIMIC-CXR data set was originally approved by the institutional review board of the Beth Israel Deaconess Medical Center and the requirement for individual patient consent was waived [20].
Table 1. The BLEU\textsuperscript{a}, METEOR\textsuperscript{b}, and ROUGE-L\textsuperscript{c} scores of the generated reports by various methods on the IU X-RAY data set.

<table>
<thead>
<tr>
<th>Methods</th>
<th>IU X-RAY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLEU-1</td>
</tr>
<tr>
<td>R2GEN</td>
<td>0.470</td>
</tr>
<tr>
<td>CA\textsuperscript{f}</td>
<td>0.492</td>
</tr>
<tr>
<td>CMCL\textsuperscript{g}</td>
<td>0.473</td>
</tr>
<tr>
<td>PPKED\textsuperscript{h}</td>
<td>0.483</td>
</tr>
<tr>
<td>M2TR\textsuperscript{i}</td>
<td>0.475</td>
</tr>
<tr>
<td>R2GENCMN\textsuperscript{j}</td>
<td>0.486</td>
</tr>
<tr>
<td>ALIGNTRANSFORMER</td>
<td>0.484</td>
</tr>
<tr>
<td>KNOWMAT\textsuperscript{k}</td>
<td>0.496</td>
</tr>
<tr>
<td>CMM-RL</td>
<td>0.494</td>
</tr>
<tr>
<td>CMCA\textsuperscript{l}</td>
<td>0.496</td>
</tr>
<tr>
<td>ORGAN\textsuperscript{m}</td>
<td>0.510</td>
</tr>
<tr>
<td>ClinicalBLIP</td>
<td>0.433</td>
</tr>
</tbody>
</table>

\textsuperscript{a}BLEU: Bilingual Evaluation Understudy.
\textsuperscript{b}METEOR: Metric for Evaluation of Translation With Explicit Ordering.
\textsuperscript{c}ROUGE-L: Recall-Oriented Understudy for Gisting Evaluation-L.
\textsuperscript{d}BLEU-A: average of the BLEU-2/3/4 scores.
\textsuperscript{e}N/A: not available.
\textsuperscript{f}CA: contrastive attention.
\textsuperscript{g}CMCL: competence-based multimodal curriculum learning.
\textsuperscript{h}PPKED: Posterior-and-Prior Knowledge Exploring-and-distilling.
\textsuperscript{i}M2TR: Meshed-Memory Transformer.
\textsuperscript{j}R2GENCMN: R2GEN enhanced with cross-modal memory networks.
\textsuperscript{k}KNOWMAT: Knowledge Matters.
\textsuperscript{l}CMCA: Cross-Modal Contrastive Attention Model.
\textsuperscript{m}ORGAN: Observation-Guided Radiology Report Generation Framework.
### Ablation Study

We also conducted an ablation study to analyze the impact of fine-tuning on different modules, such as the original InstructBLIP (without any fine-tuning on this task), LLM, visual encoder, and prior information, and show the results in Table 2. Based on the ablation study results presented in Table 3, several observations can be made regarding the performance of different methods on the IU X-RAY data set. The ClinicalBLIP method achieved a BLEU score of 0.296, a METEOR score of 0.570, and a ROUGE-L score of 0.534, indicating its robust performance across the metrics. When the effective tuning was removed, namely InstructBLIP, there was a significant drop in all metrics, especially in the BLEU score, which dropped to a mere 0.011. This highlights the importance of effective tuning for the model’s performance. Similarly, removing prior information also led to a decline in performance, with the METEOR metric showing a noticeable drop, to 0.339. The removal of LLM tuning and visual encoder tuning resulted in reduced scores, but this was not as drastic as in the former cases. The BLEU score dropped to 0.149 and 0.245, respectively, while the METEOR score was 0.458 and 0.513 for the same conditions.

In summary, effective fine-tuning and prior information played a vital role in achieving optimal performance, and LLM tuning and visual encoder tuning were also important components for enhancing the model’s results. All the components together contributed to the best results.

### Table 2. The BLEU\(^a\), METEOR\(^b\), and ROUGE-L\(^c\) scores of the generated reports by various methods on the MIMIC-CXR data set.

<table>
<thead>
<tr>
<th>Methods</th>
<th>MIMIC-CXR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLEU-1</td>
</tr>
<tr>
<td>R2GEN</td>
<td>0.353</td>
</tr>
<tr>
<td>CA(^e)</td>
<td>0.350</td>
</tr>
<tr>
<td>CMCL(^f)</td>
<td>0.344</td>
</tr>
<tr>
<td>PPKED(^g)</td>
<td>0.360</td>
</tr>
<tr>
<td>M2TR(^h)</td>
<td>0.353</td>
</tr>
<tr>
<td>R2GENCMN(^i)</td>
<td>0.378</td>
</tr>
<tr>
<td>ALIGNTRANSFORMER</td>
<td>0.378</td>
</tr>
<tr>
<td>KNOWMAT(^k)</td>
<td>0.363</td>
</tr>
<tr>
<td>CMM-RL</td>
<td>0.381</td>
</tr>
<tr>
<td>CMCA(^l)</td>
<td>0.360</td>
</tr>
<tr>
<td>ORGAN(^m)</td>
<td>0.407</td>
</tr>
<tr>
<td>ClinicalBLIP</td>
<td>0.332</td>
</tr>
</tbody>
</table>

---

\(^a\)BLEU: Bilingual Evaluation Understudy.

\(^b\)METEOR: Metric for Evaluation of Translation With Explicit Ordering.

\(^c\)ROUGE-L: Recall-Oriented Understudy for Gisting Evaluation-L.

\(^d\)BLEU-A: average of the BLEU-2/3/4 scores.

\(^e\)CA: contrastive attention.

\(^f\)CMCL: Competence-Based Multimodal Curriculum Learning.

\(^g\)PPKED: Posterior-and-Prior Knowledge Exploring-and-distilling.

\(^h\)M2TR: Meshed-Memory Transformer.

\(^i\)R2GENCMN: R2GEN enhanced with cross-modal memory networks.

\(^j\)N/A: not available.

\(^k\)KNOWMAT: Knowledge Matters.

\(^l\)CMCA: Cross-Modal Contrastive Attention Model.

\(^m\)ORGAN: Observation-Guided Radiology Report Generation Framework.
Table 3. Experimental results of ablation study on the IU X-RAY test set.

<table>
<thead>
<tr>
<th>Methods</th>
<th>BLEU-A&lt;sup&gt;a&lt;/sup&gt; score</th>
<th>METEOR&lt;sup&gt;b&lt;/sup&gt; score</th>
<th>ROUGE-L&lt;sup&gt;c&lt;/sup&gt; score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClinicalBLIP with all fine-tuning</td>
<td>0.296</td>
<td>0.570</td>
<td>0.534</td>
</tr>
<tr>
<td>ClinicalBLIP without effective tuning</td>
<td>0.011</td>
<td>0.096</td>
<td>0.057</td>
</tr>
<tr>
<td>ClinicalBLIP without prior information</td>
<td>0.091</td>
<td>0.339</td>
<td>0.283</td>
</tr>
<tr>
<td>ClinicalBLIP without LLM&lt;sup&gt;d&lt;/sup&gt; tuning</td>
<td>0.149</td>
<td>0.458</td>
<td>0.412</td>
</tr>
<tr>
<td>ClinicalBLIP without visual encoder tuning</td>
<td>0.245</td>
<td>0.513</td>
<td>0.474</td>
</tr>
</tbody>
</table>

<sup>a</sup>BLEU-A: the average of the BLEU-2/3/4 scores.
<sup>b</sup>METEOR: Metric for Evaluation of Translation With Explicit Ordering.
<sup>c</sup>ROUGE-L: Recall-Oriented Understudy for Gisting Evaluation–L.
<sup>d</sup>LLM: large language model.

Discussion

Principal Results

Our proposed model, ClinicalBLIP, achieved the best METEOR and ROUGE-L scores and competitive BLEU scores on the test sets of both IU X-RAY and MIMIC-CXR. The primary outcomes of this study are to (1) propose a multistage fine-tuning strategy that separately enhances the visual encoder and the LLM’s understanding of medical image and text, allowing the LLM to harness the knowledge acquired during the pretraining process and (2) incorporate the medical tags of medical images and brief introductions of these images in the form of prompts into the model’s training and prediction processes, the large model can effectively combine the introduced text-based prior knowledge with medical images to generate a more accurate report. Experimental results demonstrate that ClinicalBLIP has great potential to help medical experts facilitate radiology report generation and improve the efficiency of decision-making for clinical diagnosis and treatment.

Case Study

In addition to quantitative evaluations, we conducted an extensive set of qualitative case studies to analyze the generated report. Figure 5 shows 4 cases selected from the generated reports on the MIMIC-CXR test set.

By comparing the prediction and the gold standard, it can be found that case 1 and case 2 are good cases. For case 1, although the prediction and the gold standard are not exactly the same, there are differences in the order of symptom descriptions and word choices; the deep semantic meanings expressed by the two are basically consistent. However, the gold standard provides more details than the prediction, which also explains why the BLEU score is not ideal in certain situations. For case 2, both the prediction and the gold standard reports are closely aligned and convey the same overall findings. The patient’s chest x-ray does not reveal any significant abnormalities. This is a good case as it highlights the consistency and accuracy of radiological interpretation.

Besides the first 2 good cases, there are also areas that need improvement and enhancement. Cases 3 and 4 in Figure 5 show 2 bad cases. For case 3, both the prediction and the gold standard state that the heart is within the normal size, and the lungs appear clear with no signs of pleural effusion or pneumothorax. However, the prediction mentions mild anterior wedging of a midthoracic vertebral body with slight degenerative changes along the midthorax. In contrast, the the gold standard report mentions degenerative changes in the thoracic spine but does not specify the location or type of degeneration. The discrepancies in the description of the bony structures between the prediction and the gold standard report could also be of concern. Different types and locations of degenerative changes can have different clinical implications. For case 4, while the prediction and the gold standard largely align on most observations, there are subtle differences in phrasing. For instance, the prediction mentions the cardiomesothelial silhouette is normal in size, whereas the the gold standard emphasizes the normal contours of the heart and mediastinum. Such subtle linguistic variations can potentially lead to misunderstandings in diagnosis or interpretation, especially in critical medical decisions. Therefore, even though the general assessments align, precision in wording remains essential.

In summary, it is crucial to ensure that automatic or artificial intelligence–based predictions in radiology are meticulously validated and cross-referenced with expert opinions to ensure patient safety and accurate diagnosis.
Comparison With Prior Work

In the medical or clinical field, there has been a surging interest in developing artificial intelligence applications for image captioning, that is, radiology report generation. Most studies have focused on improving the quality of the generated report by using cross-modal memory to facilitate the generation process [28], reinforcing learning to align the cross-modal information [32], and planning and iterative refinement for long text generation [25]. However, these methods have not explored the capabilities of large VLMs for this task. In this study, we successfully applied large VLMs to the radiology report generation task by designing effective multistage fine-tuning strategies and incorporating prior knowledge mechanisms. We validated our approach on multiple task data sets and achieved state-of-the-art performance.

Limitations and Future Work

Although ClinicalBLIP has made significant strides and shown promising outcomes, there are still some unresolved issues. As mentioned above, ClinicalBLIP has discrepancies in terminological expressions in some cases compared to the the gold standard and sometimes lacks or misinterprets comprehensive details in certain descriptions. Therefore, in future work, we will continue to optimize ClinicalBLIP, considering the integration of reasoning techniques like chain of thoughts into the fine-tuning process. This aims to enhance the model’s semantic consistency in professional expressions and provide more detailed descriptions while also verifying the model’s generalization capabilities on more data sets. Moreover, we will seek collaboration from professional practitioners, including both directions for model improvement and methods for model evaluation.

Conclusions

In this study, the ClinicalBLIP model was introduced, leveraging large VLMs for radiology report generation. Tested on the IU X-RAY/MIMIC-CXR data sets, ClinicalBLIP significantly outperformed several competitor methods in METEOR and ROUGE scores, showcasing its potential to enhance automatic report generation in clinical radiology and streamline patient care processes.
Acknowledgments

This work was jointly supported by grants from the Research Project of Shenzhen Institute of Information Technology (grant SZHT2022KJ050), and the project was funded by the China Postdoctoral Science Foundation (grant 2023M741843), and the Natural Science Foundation of China (grant 62106115).

Authors’ Contributions

JJ and XC proposed the methods, designed and carried out the experiments, and drafted the manuscript. YH supervised the research and participated in the study design. YP critically revised the manuscript and made substantial contributions to interpreting the results. YX provided guidance and reviewed the manuscript. All authors provided feedback and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

BERT: bidirectional encoding representation of transformer
BLEU: Bilingual Evaluation Understudy
CMCA: Cross-Modal Contrastive Attention
CMM-RL: Reinforcement Learning Over a Cross-Modal Memory
CNN: convolutional neural network
LLM: large language model
LoRA: low-rank adaptation
METEOR: Metric for Evaluation of Translation With Explicit Ordering
MLP: multilayer perceptron
ORGAN: Observation-Guided Radiology Report Generation
Q-Former: query transformer
RNN: recurrent neural network
ROUGE: Recall-Oriented Understudy for Gisting Evaluation
VLM: vision-language model

©Jia Ji, Yongshuai Hou, Xinyu Chen, Youcheng Pan, Yang Xiang. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
A Digitally Enabled Combined Lifestyle Intervention for Weight Loss: Pilot Study in a Dutch General Population Cohort

Rahul Gannamani1,2*, MD; José Castela Forte1,3*, MD; Pytrik Folkertsma1,4, MSc; Sven Hermans1, MSBA; Sridhar Kumaraswamy1, BTech, MSc, MBA; Sipko van Dam1,4, PhD; Niels Chavannes5,6, Prof Dr; Hendrikus van Os5,6, MD, PhD; Hanno Pijl7, Prof Dr; Bruce H R Wolffenbuttel4, Prof Dr

1Ancora Health BV, Groningen, Netherlands
2Department of Neurology, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands
3Department of Clinical Pharmacy and Pharmacology, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands
4Department of Endocrinology, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands
5Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden University, Leiden, Netherlands
6National eHealth Living Lab, Leiden, Netherlands
7Department of Endocrinology, Leiden University Medical Center, Leiden University, Leiden, Netherlands
*these authors contributed equally

Corresponding Author:
José Castela Forte, MD
Ancora Health BV
Hereplein 34
Groningen, 9711LM
Netherlands
Phone: 31 628218360
Email: jose@ancora.health

Abstract

Background: Overweight and obesity rates among the general population of the Netherlands keep increasing. Combined lifestyle interventions (CLIs) focused on physical activity, nutrition, sleep, and stress management can be effective in reducing weight and improving health behaviors. Currently available CLIs for weight loss (CLI-WLs) in the Netherlands consist of face-to-face and community-based sessions, which face scalability challenges. A digitally enabled CLI-WL with digital and human components may provide a solution for this challenge; however, the feasibility of such an intervention has not yet been assessed in the Netherlands.

Objective: The aim of this study was two-fold: (1) to determine how weight and other secondary cardiometabolic outcomes (lipids and blood pressure) change over time in a Dutch population with overweight or obesity participating in a pilot digitally enabled CLI-WL and (2) to collect feedback from participants to guide the further development of future iterations of the intervention.

Methods: Participants followed a 16-week digitally enabled lifestyle coaching program rooted in the Fogg Behavior Model, focused on nutrition, physical activity, and other health behaviors, from January 2020 to December 2021. Participants could access the digital app to register and track health behaviors, weight, and anthropometrics data at any time. We retrospectively analyzed changes in weight, blood pressure, and lipids for remeasured users. Surveys and semistructured interviews were conducted to assess critical positive and improvement points reported by participants and health care professionals.

Results: Of the 420 participants evaluated at baseline, 53 participated in the pilot. Of these, 37 (70%) were classified as overweight and 16 (30%) had obesity. Mean weight loss of 4.2% occurred at a median of 10 months postintervention. The subpopulation with obesity (n=16) showed a 5.6% weight loss on average. Total cholesterol decreased by 10.2% and low-density lipoprotein cholesterol decreased by 12.9% on average. Systolic and diastolic blood pressure decreased by 3.5% and 7.5%, respectively. Participants identified the possibility of setting clear action plans to work toward and the multiple weekly touch points with coaches as two of the most positive and distinctive components of the digitally enabled intervention. Surveys and interviews demonstrated that the digital implementation of a CLI-WL is feasible and well-received by both participants and health care professionals.
Conclusions: Albeit preliminary, these findings suggest that a behavioral lifestyle program with a digital component can achieve greater weight loss than reported for currently available offline CLI-WLs. Thus, a digitally enabled CLI-WL is feasible and may be a scalable alternative to offline CLI-WL programs. Evidence from future studies in a Dutch population may help elucidate the mechanisms behind the effectiveness of a digitally enabled CLI-WL.

(JMIR Form Res 2024;8:e38891) doi: 10.2196/38891

KEYWORDS
lifestyle intervention; prevention; obesity; overweight; weight loss; digital health; intervention; weight; pilot; digital; data; Fogg Behavior Model

Introduction

Background

The morbidity and mortality burden associated with obesity, diabetes, and cardiovascular disease (CVD) continues to increase globally [1]. The prevalence of diabetes and CVD has nearly doubled since 1990 to over 536 and 520 million cases worldwide, respectively [1,2]. Obesity, as a major risk factor for CVD and diabetes, is also associated with a decrease in life expectancy between 5 and 20 years, depending on the severity of the condition and comorbid disorders [3]. In addition, overweight-related diseases are expected to give rise to treatment costs equivalent to 8.4% of health spending in Organization for Economic Cooperation and Development countries [4].

A substantial portion of the risk of diabetes, CVD, and obesity is attributable to modifiable lifestyle factors such as an unhealthy diet, lack of physical activity, and smoking, which subsequently lead to metabolic imbalances and overweight or obesity [1-3]. Overweight and obesity are often the direct result of a disturbed energy balance, generated through a combination of the above-mentioned factors as well as excessive dietary energy consumption [5,6]. Despite being common knowledge, the principles of living a healthy lifestyle are insufficiently followed by the general population. Indeed, recent data show that over 50% of the Dutch population do not meet daily activity guidelines and that many elements of an individual's environment are often obesogenic [7,8]. In addition, other health behavioral aspects such as deficient sleep and poor stress management also increase the risk of developing excess weight [9-11]. It is therefore unsurprising that overweight and obesity rates among the general population of the Netherlands keep increasing: in 2021, almost 50% of Dutch adults were living with overweight and 13% were living with obesity [12].

Therefore, successful approaches to preventing and treating overweight/obesity, CVD, and associated risk factors ought to tackle at least the four essential lifestyle components of nutrition, activity, sleep, and stress management, along with smoking and alcohol consumption when possible [13-15]. Preferably, this is done in a personalized way, since different individuals will be exposed to different triggers and face different challenges that lead to excess weight or failure to lose weight [16,17].

Combined lifestyle interventions (CLIs), especially those targeting weight loss (CLI-WLs), provide a potential solution for people who are overweight or obese to initiate and maintain healthier lifestyle behaviors [18]. Since 2019, some CLI-WL programs have been covered by basic insurance in the Netherlands for individuals meeting prespecified criteria such as obesity in case of a BMI≥30 or overweight and increased risk of comorbidities, including diabetes and CVD [19]. Dutch CLI-WL programs are 2-year programs delivered at varying intensities throughout this period, which consist of guidance mostly on physical activity and nutrition [18]. These programs consist of both individual and group sessions to educate participants, allowing them to share experiences and provide support [18]. The program is typically carried out completely offline by lifestyle coaches, physiotherapists, and dietitians accredited to deliver CLI-WL to patients referred by general practitioners [18]. Internationally, interventions similar to the Dutch CLI-WL have been shown to be effective in terms of weight reduction and health improvements, even when compared with standard care or pharmacological treatment [20]. However, reports on the effectiveness of Dutch CLI-WL programs since their inclusion in the basic insurance package have been inconsistent, with the effects of the interventions either falling short of expectations or not structurally translating to sustained weight loss in the medium to long term [21-23].

A growing number of digital programs that can support individuals and providers in addressing health risks and conditions are being developed and made available to the public [25]. These so-called digital therapeutic solutions stem from an emerging field at the cross-section between health care and technology. Similar to pharmacological therapies or hardware medical devices, these therapeutics are evidence-based software products for the prevention, management, and treatment of health conditions. Several studies suggest that interventions including digital elements such as remote data monitoring and the possibility to digitally communicate with providers are associated with higher engagement from participants and subsequently with better health outcomes [26,27].

The incorporation of these digital elements—which are considered critical components of successful behavior change programs—can therefore help overcome the cited barriers to engagement in face-to-face programs such as accessibility, transportation, and scheduling [24,25,28]. In fact, digital therapeutic solutions, ranging from digital weight loss programs based on intensive dietary coaching to tracking and gamification apps, have been shown to induce weight loss to varying degrees, surpassing what offline interventions achieve [29,30]. Therefore, deploying a digitally enabled CLI-WL could broaden access to prevention and care, while delivering superior or at least comparable outcomes to those reported for currently available, completely offline CLI-WL programs in the Netherlands.
to the novelty of digitally enabled CLI-WL programs in the Netherlands, there is a research gap that we seek to bridge with this research.

**Objectives**

The objective of this pilot study was to provide a real-world contribution to the discussion surrounding the feasibility of digitally enabled CLI-WL programs in the Netherlands. In this study, we assessed the changes in weight and concomitant cardiometabolic risk factors of a cohort of Dutch adults with overweight or obesity and cardiometabolic risk who participated in the pilot of a digitally enabled CLI-WL. In addition, we also conducted surveys and semistructured interviews to assess the critical positive and improvement points reported by participants and health coaches to further guide the development of future iterations of the intervention.

**Methods**

**Study Sample**

As of December 2021, 420 users were enrolled in the Ancora Health Lifestyle program through employer health programs or as a direct-to-consumer option. A study on a subset of this cohort was published previously in *JMIR Cardio* [31]. Body weight, blood pressure, and lipids (total and low-density lipoprotein [LDL] cholesterol) were measured at baseline. Participants who were classified as obese (BMI ≥ 30) or overweight (BMI ≥ 25) with one or more cardiometabolic risk factors were asked to participate in a 16-week digitally enabled CLI consisting of digital and human components. The cardiometabolic risk factors considered were a diastolic blood pressure ≥ 80 mmHg or systolic blood pressure ≥ 130 mmHg, dyslipidemia (LDL cholesterol ≥ 3.0 mmol/L, total cholesterol ≥ 5.1 mmol/L, high-density lipoprotein cholesterol < 1 mmol/L, or triglycerides ≥ 1.8 mmol/L), or prediabetes (hemoglobin A1c ≥ 5.7% or fasting glucose ≥ 5.6 mmol/L). An overview of the study flow is given in Figure 1.

**Combined Lifestyle Intervention**

After baseline measurements, users were provided access to a web-based digital app where they could register and track health behaviors, weight, and anthropometrics data at any time. This app is a certified Class I medical device that generates evidence-based lifestyle recommendations spanning nutrition, physical activity, sleep improvement, stress reduction, and tracking behaviors (such as logging weight) that are tailored to the individual’s prevailing risk profile. Other preliminary evidence of the health benefits of lifestyle interventions delivered to a subset of this cohort with the Ancora Health platform for health goals other than weight loss (eg, lipids, blood pressure, and nutrient imbalances) has been published previously [32-34].

The intervention was initiated with a 30-minute intake consultation conducted by video call. The intake consultation provided counseling on health risks, recommendation of targeted lifestyle medicine actions, and a “handshake” to undertake these actions for the following period. During the intervention, coaching was delivered digitally through one-on-one chat-based contact with optional audio/video calls alongside this format. Coaching was delivered by a health care professional with a background in either lifestyle coaching, nutrition, physiotherapy, or psychology, depending on the prevailing behavioral coaching required. This coaching was complemented by weekly progress reports. This approach builds on the Fogg Behavioral Model.
FMATF (FBM) [35]. The FBM posits that behavior change occurs when users are prompted to perform target behaviors that they are sufficiently motivated and sufficiently able to perform, with a trade-off between the level of motivation and level of ability. In this intervention, the FBM was implemented proactively through digital coaching: coaches used motivational interviewing techniques to manage/positively influence participants’ motivation levels and adjusted the difficulty of target behaviors on an ongoing basis in line with the participants’ motivation and/or skill level. Moreover, coaches helped users with tips/tricks and strategies to overcome any barriers encountered.

Measurements at Intake, During the Program, and After the Program

Upon enrollment to the program, participants underwent a baseline assessment where a comprehensive lifestyle questionnaire, a blood biochemistry panel, and physical measurements were collected using the InBody model 570 for body composition and the InBody BIOBP750 cuff for blood pressure. After the baseline assessment, users could access the digital web app to register and track their health behaviors and modify weight data at any time during the intervention. At follow-up after the intervention, which participants could voluntarily enroll for, the subset of blood biochemistry parameters found to be abnormal at baseline and the lifestyle questionnaire and physical measurements were reassessed. Weight and other cardiometabolic risk factors were remeasured at the health center for participants who were able to return for a remeasurement. Participants not able to return were asked to self-report their weight after receiving instructions from their lifestyle coaches (namely to measure it in the morning, before eating) (Figure 1). For self-reporting participants, other markers were not remeasured. According to the definition used in the criteria for participation in a CLI-WL in the Netherlands, we classified BMI values between 25 and 30 as “overweight” and values greater than or equal to 30 as “obese.” Changes from baseline in weight, blood pressure, and lipid markers were calculated by subtracting the end values from the first reported values, and the percent change was calculated by dividing the observed change by the baseline value.

After the entire cohort completed their intervention, the first author conducted semistructured interviews with 17 participants and 6 of the coaches delivering the intervention, either physically or via audio/video in Dutch or English. Semistructured interviews allow for greater flexibility for both the interviewer and participant than traditional, structured interviewing, while simultaneously providing greater direction in the interview process than completely unstructured interviews. The user research framework reported in this study was based on guidelines published by the Medical Research Council for process evaluations [36] and the Conceptual Framework for Implementation Research [37,38]. The focus of the interviews was acceptability and accessibility, two critical process indicators in the adoption of digital interventions. While there is no consensus definition of acceptability, it can be broadly defined as “people’s affective attitudes toward a new digital health intervention,” “usage intentions or actual usage,” and “satisfaction after having engaged with the intervention” [39]. Since most available acceptability measures in pilot or feasibility studies of digital health interventions lack a theoretically or empirically established cutoff, it has been suggested that 1 to 5 ratings and accompanying free-text responses may provide a sufficiently precise measure of acceptability [40].

The interview guide had questions for the coaches and participants focused on their experience coaching or partaking in the intervention, and on what they thought were the most important positive and improvement points for the quality of the intervention.

Statistical Analysis

Descriptive statistics were calculated to characterize the population at baseline in terms of demographics and clinical parameters. Paired remeasurement versus baseline changes in weight, blood pressure, and cholesterol in participants who were remeasured after a median of 10 months were assessed with the Student t test or Wilcoxon signed-rank test depending whether or not the data were normally distributed. All categorical variables are reported as percentages and continuous variables are reported as mean and SD. The $\chi^2$ test and analysis of variance were used to evaluate differences in categorical and continuous variables, respectively, at the cohort level. We considered $P<.05$ to indicate a statistically significant difference between cohorts and in pre- and postintervention measurements. All data analyses were performed using R software v4.0.3.

Themes and coach/participant opinions were registered following an inductive process after the main findings of the interviews conducted by the first author were discussed in a multidisciplinary setting with the coaches, with no preexisting framework or theoretical constructs used to classify data [41,42].

Ethics Approval

The study was declared exempt from institutional review board approval through a waiver issued by the Medical Ethical Committee of the University Medical Centre Groningen (waiver number METC#2021/488).

Results

Baseline Characteristics

Baseline characteristics of the total study sample are shown in Table 1. We found that 208 of the 420 participants (49.5%) were classified as either obese or overweight with one or more cardiometabolic risk factors. Of these, 53 participants enrolled in the pilot study. These individuals were older, had higher lipid levels and blood pressure, as well as higher weight and BMI compared to the rest of the cohort (Table 1).
Baseline Values and Changes in Weight and Cardiometabolic Risk Factors

We analyzed weight data for the entire group of 53 participants in the pilot and for the two subsets of participants with overweight and obesity separately (Table 2). In the entire group, the average weight loss achieved was 3.7 kilograms, which equates to an average of 4.2% body weight loss (P<.001). In total, 25 individuals (47%) achieved a reduction of at least 5% body weight. In the 37 participants with overweight, the mean weight loss was 2.9 kilograms (3.5% change). For the 16 participants with obesity, weight loss was higher at –5.4 kilograms and –5.6% body weight with a mean baseline weight of 97.3 kilograms (P<.001; Table 2).

We also analyzed the changes in cardiometabolic risk factors in participants whose weight was remeasured on location and had abnormal lipid or blood pressure levels at baseline (Table 3). In these participants, total cholesterol, LDL cholesterol, systolic blood pressure, and diastolic blood pressure significantly decreased. In participants with obesity, changes in total and LDL cholesterol as well as in systolic and diastolic blood pressure were not significant, which was attributed to the very small sample size. In participants with overweight, total cholesterol, LDL cholesterol, and diastolic blood pressure were significantly decreased, whereas there was no significant decrease in systolic blood pressure.

None of the participants who were remeasured on location, based on information provided in the medical and lifestyle questionnaire at baseline and follow-up, had initiated blood pressure– or cholesterol-lowering medication during the lifestyle intervention.

Table 2. Changes in weight after the combined lifestyle intervention for weight loss.

<table>
<thead>
<tr>
<th>Group</th>
<th>Preintervention weight (kg), mean (SD)</th>
<th>Postintervention weight (kg), mean (SD)</th>
<th>Absolute (relative) weight change, kg (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire cohort (N=53)</td>
<td>87.2 (11.2)</td>
<td>83.5 (12.0)</td>
<td>–3.7 (–4.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Participants classified as overweight</td>
<td>82.8 (8.3)</td>
<td>79.9 (10.4)</td>
<td>–2.9 (–3.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>n=37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants classified as obese</td>
<td>97.3 (10.5)</td>
<td>91.9 (11.3)</td>
<td>–5.4 (–5.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>n=16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Changes in lipid profile and blood pressure after the combined lifestyle intervention in individuals with abnormal baseline values.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Preintervention, mean (SD)</th>
<th>Postintervention, mean (SD)</th>
<th>Absolute (relative) change, kg (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total cholesterol (mmol/L)</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire remeasured group (N=20)</td>
<td>6.38 (1.00)</td>
<td>5.73 (0.92)</td>
<td>–0.65 (–10.2)</td>
<td>.02a</td>
</tr>
<tr>
<td>Participants with overweight (n=18)</td>
<td>6.46 (1.01)</td>
<td>5.71 (0.97)</td>
<td>–0.75 (–11.6)</td>
<td>.01a</td>
</tr>
<tr>
<td>Participants with obesity (n=2)</td>
<td>5.75 (0.74)</td>
<td>5.88 (0)</td>
<td>0.13 (2.3)</td>
<td>&gt; .99a</td>
</tr>
<tr>
<td><strong>LDL b cholesterol (mmol/L)</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire remeasured group (N=24)</td>
<td>4.34 (0.86)</td>
<td>3.78 (0.88)</td>
<td>–0.56 (–12.9)</td>
<td>.007</td>
</tr>
<tr>
<td>Participants with overweight (n=20)</td>
<td>4.46 (0.86)</td>
<td>3.89 (0.85)</td>
<td>–0.57 (–12.8)</td>
<td>.02</td>
</tr>
<tr>
<td>Participants with obesity (n=4)</td>
<td>3.71 (0.56)</td>
<td>3.18 (0.88)</td>
<td>–0.53 (–14.3)</td>
<td>.17</td>
</tr>
<tr>
<td><strong>Systolic blood pressure (mmHg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire remeasured group (N=22)</td>
<td>146 (15)</td>
<td>141 (16)</td>
<td>–5 (–3.5)</td>
<td>.04a</td>
</tr>
<tr>
<td>Participants with overweight (n=18)</td>
<td>148 (16)</td>
<td>143 (16)</td>
<td>–5 (–3.4)</td>
<td>.12a</td>
</tr>
<tr>
<td>Participants with obesity (n=4)</td>
<td>139 (6)</td>
<td>131 (14)</td>
<td>–9 (–6.5)</td>
<td>.14</td>
</tr>
<tr>
<td><strong>Diastolic blood pressure (mmHg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire remeasured group (N=26)</td>
<td>93 (11)</td>
<td>86 (12)</td>
<td>–7 (–7.5)</td>
<td>.003a</td>
</tr>
<tr>
<td>Participants with overweight (n=19)</td>
<td>94 (11)</td>
<td>86 (12)</td>
<td>–8 (–8.5%)</td>
<td>.003</td>
</tr>
<tr>
<td>Participants with obesity (n=7)</td>
<td>90 (10)</td>
<td>85 (12)</td>
<td>–5 (–5.6%)</td>
<td>.18a</td>
</tr>
</tbody>
</table>

*aBased on the Wilcoxon signed-rank test owing to baseline data not following the normal distribution.
bLDL: low-density lipoprotein.

Quantitative and Qualitative Participant Feedback

The subjective, qualitative feedback collected through surveys (n=37) and semistructured interviews (n=17) from pilot participants is presented in Textbox 1.

The quantitative results of the participant survey that was filled out at the end of the intervention period are presented in Table 4. The highest scoring items were linked to the feeling of involvement in the program (item 1, score 4.5/5), the interaction with the coaches (item 9, 4.6/5), and the knowledge displayed by the coaches (item 10, 4.4/5). Importantly, participants also overwhelmingly expressed the wish to continue working toward their health goals after the intervention (item 4, 4.7/5). The items that scored the lowest were linked to the need for one-on-one coaching and the importance of coaching in achieving the proposed health goals (items 7 and 8, 3.7/5 and 3.8/5, respectively).

The subjective feedback provided during semistructured interviews by the coaches (N=6) is presented in Textbox 2. In general, interviewees valued several aspects as the most positive and differentiating factors of the digitally enabled CLI-WL. These included the possibility of setting clear action plans (goals) for participants to work toward; the promotion of healthy eating, exercise, and other lifestyle habits as opposed to enforcing strict diets; and the speedy, problem solving–oriented interactions across the multiple weekly touch points. Both improvement points were related to the intervention curriculum, with health care professionals suggesting improvements in the food-tracking capabilities of the app as well as an expansion of the coaching resources.
Textbox 1. Summarized feedback collected from pilot participants through surveys and semistructured interviews regarding the acceptability and accessibility of the program.

Positive points
- Coaches were friendly, engaged, and approachable for guidance
- Proactive check-ins
- Good, personal advice; coaches helped find solutions to overcome barriers
- Program was (positively) challenging
- Effectively motivated for behavioral change
- Provided support in making effective lifestyle changes

Improvement points
- Expand the features in the app (tracking, connectivity, reminders, personalized content)
- More coaching conversations instead of chat messages

Table 4. Quantitative results of the participant survey (N=37).

<table>
<thead>
<tr>
<th>Item</th>
<th>Scorea</th>
</tr>
</thead>
<tbody>
<tr>
<td>I felt involved in the program right from the start</td>
<td>4.5</td>
</tr>
<tr>
<td>I found the stepwise approach to behavioral change in the program</td>
<td>4.1</td>
</tr>
<tr>
<td>easy to follow</td>
<td></td>
</tr>
<tr>
<td>I was motivated to work toward my health goals</td>
<td>4.4</td>
</tr>
<tr>
<td>I would like to continue working toward my health goals</td>
<td>4.7</td>
</tr>
<tr>
<td>I feel my health and well-being have improved since I participated</td>
<td>3.9</td>
</tr>
<tr>
<td>in the Ancora program</td>
<td></td>
</tr>
<tr>
<td>I found the content of the program relevant and engaging</td>
<td>4.0</td>
</tr>
<tr>
<td>Human coaching was important for me to overcome my health challenges</td>
<td>3.7</td>
</tr>
<tr>
<td>Human coaching was important for me to achieve my health missions</td>
<td>3.8</td>
</tr>
<tr>
<td>I found my coach(es) friendly and empathic</td>
<td>4.6</td>
</tr>
<tr>
<td>I found my coach(es) knowledgeable</td>
<td>4.4</td>
</tr>
</tbody>
</table>

a Scoring runs from 1 ("disagree completely") to 5 ("agree completely"), with 3 being “neutral.”

Textbox 2. Summary of the feedback regarding subjective intervention parameters “acceptability and accessibility” provided by coaches (N=6).

Positive points
- Possibility to set clear goals for the intervention period based on a holistic assessment of the participant’s health
- Integrative approach that promotes healthy eating and exercise as opposed to strict diets, including specific elements such as strength-training advice, stress management, and good sleep habits
- Multiple touch points weekly between the Health Engagement team to enable rapid problem resolution and positive experience sharing
- Multidisciplinary expertise for knowledge transfer
- Protocolized digital coaching
- One-on-one coaching with real-life examples

Improvement points
- More insight into the participants’ daily lifestyle patterns during the intervention period
- Expand the database of coaching resources (ie, materials coaches have available to support participants)
Discussion

Principal Results

In this study of 53 participants using the pilot version of a digitally enabled CLI-WL, we observed a mean weight loss of 3.7 kilograms (or 4.2% body weight reduction), with weight loss of more than 5 kilograms (or 5.6% body weight reduction) in individuals with obesity compared to baseline values at a median of 10 months after 16 weeks of online coaching. Both total cholesterol as well as LDL cholesterol decreased by over 10%, and systolic and diastolic blood pressure decreased by 5 and 7 mmHg, respectively. Our process evaluation analysis through surveys and interviews showed that digital implementation of a CLI-WL in the Netherlands is feasible and well-received by both participants and coaches.

Comparison With Prior Work

Evidence for the efficacy and safety of digital (or digitally enabled) and blended CLI-WL and the reduction of cardiometabolic and cardiovascular risk has been building up over the last 5 to 10 years [43,44]. As stated previously, some of the world’s most widely adopted commercial digital therapeutic programs for weight loss have reported achieving reductions in weight varying from 2.0% to 6.8% [30,31,45]. In addition, recent reviews showed that primarily mobile digital interventions targeting overweight and obese populations with high cardiovascular risk can be at least as effective as offline programs in terms of meaningful change of lifestyle and weight loss [22]. Several offline CLI-WL programs were shown to be effective to varying degrees in the Netherlands. Across three of the four interventions available in Dutch health care, weight loss achieved at 1 year varied between 2.9 kilograms and 2.2 kilograms [22,23,46]. Only one of these interventions reported follow-up data at 2 years, where participants lost an average of only 1.5 kilograms [47]. Another study reported an average weight loss of 2.5 kilograms at 18 months [23]. While we currently do not have such long-term data, a digitally enabled intervention that was similar to this CLI-WL (with a total duration of 12 months, including a maintenance phase), delivered in a real-world context, achieved long-term weight reductions of 7.2% and 7.6% in participants with overweight and obesity, respectively [48].

While none of the currently available CLI-WLs provided data on improvement in concomitant cardiometabolic risk factors, it is worthwhile to contextualize the changes achieved in blood pressure and lipids in this pilot study. Evidence for beneficial effects of healthful lifestyle modifications on blood pressure is solid, with several studies suggesting that lifestyle adaptation is preferable to pharmacological treatment in early stages of the disease [49]. Keeping in mind the observational nature of the results reported in this study, these do surpass the results of other recently evaluated digital therapeutic tools, which yielded reductions between 2.4 and 4.3 mmHg in mean blood pressure in randomized trials [50,51]. Similarly, the effect of lifestyle programs on cholesterol levels is well-established, with reductions varying from 7% to 9% to 20% for interventions of different intensity and complexity [52]. Interestingly, for web-based interventions, meta-analyses have shown total and LDL cholesterol improvements of approximately 0.15 mmol/L [53]. When compared to the results achieved previously in another cohort of individuals who participated in an Ancora Health digital lifestyle intervention, this group of individuals with overweight and obesity showed higher baseline lipid and blood pressure values and equal or greater reductions after the program [32,33].

In light of this evidence, it is worth discussing potential reasons why this blended CLI-WL achieved better results compared to those of previous studies. In a study evaluating the features of digitally enabled weight management programs, success was linked to the ability to promote behavioral change [54]. This finding is unsurprising, but what is interesting is that the study further broke down this ability to drive behavioral change to 20 features that were essential to the program. This set of features included specific goal-setting for weight, diet, caloric balance, and physical activity, as well as educational focus on various skills such as regulating eating patterns, time management, and nutritional label reading. In addition, this program also included the development of more general skills such as learning to exercise at certain target heart rates, problem solving, stress reduction, and psychological advice on how to cope with negative thinking and social cues [54]. Interestingly, these features include some of the most positive feedback points gathered during our participant survey, such as the appreciation for the stepwise approach to behavioral change in the program, the program content being relevant and engaging, and the digital support provided by the human coaching for overcoming challenges. In the qualitative assessment, participants further mentioned highly valuing the possibility to set clear goals for the intervention and that the coaching included elements that go beyond those of regular weight loss programs, such as advice on strength training, stress management, and sleep habits.

Other studies have focused on the potential advantages of digitally enabled interventions, such as the ability for participants to access educational information more easily and at their own discretion [44,55]. Unlike offline interventions driven primarily by one health care provider with a specific focus, digitally enabled programs can easily provide information covering a range of topics required to lead a balanced lifestyle across all relevant lifestyle domains. This seems to be especially powerful when apps also provide users with tools to help them track changes in weight and BMI. For instance, in-app actions such as self-monitoring of weight and the consistency of such health behavior tracking, as well as app engagement as measured by log-ins, predicted weight loss even in older populations [44,55]. These elements again come back in the qualitative feedback from participants highlighted in Textbox 2, who highly valued the multidisciplinary nature of the content and of the coaching, and suggested that even more attention should be paid to data collected about participants’ daily lifestyle patterns. In another study comparing a mobile app–based weight loss program with a traditional offline weight loss program, success in the weight loss intervention was linked to the digitalization of components common to the traditional program [56], including providing online food diaries rather than paper diaries or when the same curriculum delivered by the dietitian in person was delivered digitally. One such app-based program featured

https://formative.jmir.org/2024/1/e38891

Gannamani et al

JMIR Form Res 2024 | vol. 8 | e38891 | p.785

(page number not for citation purposes)
regular contact moments with a dietitian via either text messages or video calls, while allowing them to record their food, activity, and weight, as well as providing access to educational materials and a group chat [57].

**Future Perspectives for Digital CLIs in The Netherlands**

In the face of increasing rates of obesity, CVD, and (pre)diabetes, it is clear that there is a need for CLI-WLs to be deployed at scale in the Netherlands. However, current market penetration of these insurance-reimbursed programs is extremely low [58]. As the pathogenesis of obesity and (pre)diabetes is multifactorial, the etiology and environmental triggers vary among individuals; moreover, socioeconomic circumstances and barriers to lifestyle change also vary at an individual level. Yet, insurance-reimbursed CLI-WL programs currently offered in the Dutch market lack a data-driven approach to translating user profiles to personalized care pathways, with coaches bearing the burden of translating one-size-fits-all guidelines to a user’s context through their consultations. As suggested by the health care professionals interviewed in this study, supporting the delivery of CLI-WL programs with eHealth solutions such as those described in this research could help tailor the intervention to individual characteristics and needs, as well as tackle the practical issues of lack of time and capacity faced by care providers.

In fact, the Dutch National Institute for Public Health and the Environment (Rijksinstituut voor Volksgezondheid en Milieu [RIVM]) defines several important, evidence-based factors for the successful deployment of a (combined) lifestyle intervention program [59]. Interestingly, although these recommendations have been designed for offline programs, some of them are much more easily achievable with digitally enabled interventions. For instance, the RIVM recommends that participants of a CLI-WL who fail to achieve significant weight loss in the first month should receive increased support since early weight loss is an important predictor of long-term success. Through its digital capabilities, the CLI-WL described herein allows for weekly monitoring of progress and enables coaches to review data and progress remotely. In addition, coaches are also able to contact the participant proactively to understand and address barriers in the absence of progress. Importantly, one full-time coach can effectively coach more than 100 participants across multiple regions while upholding the recommendation of 15 to 20 minutes of coaching time weekly. This allows for greater reach than possible with offline lifestyle interventions, which are often episodic, delivered at low frequency, and limited to participants within close proximity to the health coach. Importantly, a recent Dutch study showed that the Dutch population is generally sufficiently technology-literate, even across socioeconomic strata, and welcoming of digital technologies, which encourages the deployment of these interventions [60]. Lastly, the RIVM underscores the need for participants to develop self-efficacy skills to sustain behavioral change over the long term. This is in line with previous studies demonstrating that interventions applying theoretical frameworks or models for behavioral change—some of which were technology-based—were more effective at increasing adherence to healthy lifestyle habits than standard advice [61,62]. This program represents the first instance in which the FBM was used as part of a CLI-WL. The positive health outcomes and subjective feedback from users and health care professionals suggest that this approach holds promise and ought to be further explored for future iterations.

**Strengths and Limitations**

This study has several limitations. The first is the observational nature of the study. The second limitation is the small sample size of the remeasured population. Third, participants were not followed up for the duration of a standard CLI-WL (2 years); rather, these preliminary outcomes were reported after a median of 10 months after the 16-week intervention. As such, the results of the 16-week program should be compared with caution to those of other CLI-WLs that maintained participants engaged for the entirety of the 2-year period. Fourth, participants who could not return to the health center for a remeasurement were asked to self-report their weight after 10 months. As for all self-reported outcome measures, and despite careful instructions provided by the coach, measurement errors cannot be excluded. We are currently evaluating several possibilities to overcome this, such as the delivery of a personalized program box with a connected scale and wearable device. Lastly, the goal of CLI-WLs is to provide sustained positive health outcomes after 2 years (which is also the duration of the insured intervention in public health programs). Therefore, our 10-month results show that the weight loss can be sustained for a medium-long period of time, but we do not yet have data of 2-year follow-up for these participants. More research will also be conducted to evaluate the impact of interventional elements on engagement and health outcomes, such as to identify which elements are most effective for a digitally enabled CLI-WL.

Conversely, one strength of the study is that medication information was gathered at baseline and follow-up. This allows us to verify that weight loss and the improvements reported in lipids and blood pressure did not come from cholesterol- or blood pressure–lowering medications. In addition, by gathering detailed feedback from a majority of participants, we were able to identify critical points for improvement, which contributed to the further maturing of the intervention and informed the development of the mobile app designed to support it. Lastly, this is the first study to report on a real-world application of a digitally enabled intervention targeting health behavior change to promote weight loss in a cohort of Dutch adults with overweight and cardiometabolic risk factors.

**Conclusions**

In conclusion, a digitally enabled CLI showed sustained weight loss in individuals with overweight and obesity at the 10-month follow-up. Albeit preliminary, these findings suggest that a behavioral lifestyle program with a digital component deployed for this intervention can achieve greater weight loss than previously reported for currently available offline CLI-WLs in the Netherlands. Whereas larger-scale observational studies and randomized trials of other digital interventions have been conducted and shown evidence of effectiveness for digitally enabled CLI-WL in other countries, this study is a demonstration that such an approach can also be effective in the Netherlands. This is an important first step, as the feasibility of these
programs remains a hot topic in public health discussions. Based on the qualitative feedback collected from pilot participants and health care professionals, this first version of the program is being further developed and evaluated. Evidence from future studies in a Dutch population may help elucidate the mechanisms behind the effectiveness of a digitally enabled CLI-WL.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
RG had the original ideation of the manuscript, and made significant contributions to data collection and the drafting of the manuscript. JCF was the main contributor to the first draft of the manuscript. PF was the main contributor to the data analysis. SH contributed to ideation and discussion. SK was involved in the original study ideation. SvD contributed to the Methods, Results, and Discussion sections of the manuscript. NC, HvO, and HP contributed to the Introduction and Discussion sections of the manuscript. BHRW contributed to the Methods and Discussion sections of the manuscript. All authors gave input toward and approved the final manuscript.

Conflicts of Interest
RG, JCF, PF, SH, SK, and SvD are employed by Ancora Health BV. Additionally, JCF, PF, RG, SK, SH, and SvD own shares of Ancora Health BV. As the funder, Ancora Health BV provided support in the form of salaries for all employees. The other authors have no conflicts of interest to declare.

References


50. van Rijnzoo_vd_Goor_eindrapport_coaching_op_leefstijlcool_2018.pdf [accessed 2023-01-05]


60. Agachi E, Bijmolt THA, Mierau JO, van Ittersum K. Adoption of the website and mobile app of a preventive health program across neighborhoods with different socioeconomic conditions in the Netherlands: longitudinal study. JMIR Hum Factors 2022 Feb 02;9(1):e32112 [FREE Full text] [doi: 10.2196/32112] [Medline: 35107433]


Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLI</td>
<td>combined lifestyle intervention</td>
</tr>
<tr>
<td>CLI-WL</td>
<td>combined lifestyle intervention for weight loss</td>
</tr>
<tr>
<td>CVD</td>
<td>cardiovascular disease</td>
</tr>
<tr>
<td>FBM</td>
<td>Fogg Behavioral Model</td>
</tr>
<tr>
<td>LDL</td>
<td>low-density lipoprotein</td>
</tr>
<tr>
<td>RIVM</td>
<td>Rijksinstituut voor Volksgezondheid en Milieu (Dutch National Institute for Public Health and the Environment)</td>
</tr>
</tbody>
</table>

Edited by A Mavragani; submitted 26.09.22; peer-reviewed by A Bucher, S Zheng; comments to author 27.12.22; revised version received 04.05.23; accepted 25.09.23; published 08.02.24.

Please cite as:

A Digitally Enabled Combined Lifestyle Intervention for Weight Loss: Pilot Study in a Dutch General Population Cohort
JMIR Form Res 2024;8:e38891
doi: 10.2196/38891
PMID:38329792

©Rahul Gannamani, José Castela Forte, Pytrik Folkertsma, Sven Hermans, Sridhar Kumaraswamy, Sipko van Dam, Niels Chavannes, Hendrikus van Os, Hanno Pijl, Bruce H R Wolffenbuttel. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Predictors of Mental Health Literacy in a Sample of Health Care Major Students: Pilot Evaluation Study

Pia Tohme¹, PhD; Nour Abi Fadel¹, MSc; Nour Yaktine²,³, MA; Rudy Abi-Habib¹, PhD

¹Department of Social and Education Sciences, Lebanese American University, Beirut, Lebanon
²Department of Psychology, Saint Joseph University of Beirut, Beirut, Lebanon
³American University of Beirut, Beirut, Lebanon

Corresponding Author:
Rudy Abi-Habib, PhD
Department of Social and Education Sciences
Lebanese American University
Beirut Campus
Chouran
Beirut, 5056
Lebanon
Phone: 961 01786456 ext 2733
Email: rudy.abihabib@lau.edu.lb

Abstract

Background: The numerous mental health awareness campaigns during the COVID-19 pandemic have shifted our understanding and perception of mental health.

Objective: The purpose of this study is to evaluate predictors of mental health literacy (MHL), that is, one’s knowledge and beliefs about mental disorders. We evaluate whether digital health literacy, empathy, and mentalizing contribute to MHL.

Methods: Our sample consisted of 89 health care major students, aged between 17 and 32 years, studying at a university in Lebanon. The Mental Health Literacy Scale for Healthcare Students (MHLS-HS), the eHealth Literacy Questionnaire (eHLQ), the Basic Empathy Scale (BES), and the Reflective Functioning Questionnaire-8 (RFQ-8) were used.

Results: Multiple regression analyses revealed that the Engagement in Own Health subscale of digital health literacy constituted a predictor of MHL. While empathy and mentalizing did not directly predict MHL, they were found to predict components of MHL.

Conclusions: This is the first study to evaluate digital health literacy, empathy, and mentalizing as predictors of MHL in Lebanon, a country where mental health is still considered taboo. Moreover, this pilot study is the first to provide some support for the predictive role of some digital health literacy subscales on MHL in light of the rise of the digital era following the COVID-19 pandemic.

(JMIR Form Res 2024;8:e43770) doi:10.2196/43770

KEYWORDS
awareness; COVID-19; digital health literacy; digital health; disorder; empathy; health literacy; literacy; mental health literacy; mental health; mentalizing; questionnaire; students

Introduction

Overview

Despite the negative impact of the COVID-19 pandemic on both physical and mental health, the global health crisis has reshaped our perception and understanding of mental health, reducing mental health stigma and normalizing mental health-related discussions [1]. In a way, the COVID-19 era has promoted mental health literacy (MHL), a concept defined as
Addressing this gap, this study aims to explore predictors of MHL in a sample of health care major students. More specifically, we evaluate whether the availability of practical tools such as digital health technologies or psychological characteristics such as empathy or mentalizing can contribute to MHL in health care major students.

The Role of Digital Health in Promoting MHL

In today’s digital age, health care providers have been increasingly making use of technology in their professional practice [4]. The use of digital health has increased at an even higher rate during the COVID-19 pandemic, as it has become a necessity at the individual, institutional, and social levels [5,6]. Beyond being a highly effective solution to working through health problems during the pandemic, digital health—in other words, accessing health-related information on the web—has the potential to shift the health care paradigm from mere treatment to prevention [7,8]. In Lebanon, Tohme et al [9] found that mental health professionals already had experience delivering web-based consultations before the COVID-19 pandemic, and, despite preferring face-to-face sessions, mental health practitioners reported numerous benefits of using digital health tools.

Initially adopted as a short-term solution to overcome obstacles imposed during the COVID-19 pandemic (eg, lockdown), the use of digital technologies for health has enhanced digital health literacy [10], meaning the ability to seek, find, understand, and assess health-related information from electronic sources in order to address and solve a health problem [11]. Digital health literacy allows individuals to communicate health information and make informed decisions that promote well-being. In a systematic review evaluating the association between health literacy, digital health literacy, and physical health outcomes [12], digital health literacy was found to be associated with perceived and reported better communication with health care providers, health-promoting behaviors, and self-management of health needs [13,14]. As for mental health, Lincoln et al [15] found that low health literacy was associated with negative mental health outcomes such as depressive symptoms. Whether these results apply to digital health literacy is still unknown. Research has yet to establish an association between digital health literacy and mental health outcomes, an association that could be mediated by MHL.

As an extension of health literacy, MHL is a relatively new construct, focusing specifically on accessing web-based information about mental health, namely emotional and psychological difficulties [2]. Hence, most studies on digital health literacy focus on its association with physical health outcomes only. To our knowledge, no study has yet evaluated whether this association applies to mental health. It can be argued, however, that digital health literacy can be a crucial component in promoting MHL. Understanding the potential benefits of digital health literacy in promoting MHL would be useful for policymakers, as it would allow them to devise policies and interventions focusing on promoting the use of digital health in the health care sector. This would equip health care providers with the tools needed to work in a digitized health sector [16,17], contributing to MHL in the general public and potentially leading to better psychological outcomes.

Empathy and Mentalizing as Predictors of MHL

Another line of research focused on pinpointing the psychological factors that could predict MHL, such as empathy. Empathy, or putting oneself in others’ shoes, is understood as one’s ability to understand what the other is feeling and to match that emotion [18]. As a concept, empathy entails both cognitive and affective elements that lead to a sense of emotional understanding [19]. The emotional component of empathy relates to one’s emotional response to the other’s experience, while cognitive empathy involves imagining the other person’s mental state, feelings, and perspective. Both components have been shown to generate a more positive attitude toward mental disorders [20] and are important components of MHL [21]. Indeed, evaluating the association between empathy and MHL, Furnham and Sjokvist [22] established a positive correlation between the 2 constructs, showing that individuals who experience empathy are more likely to be knowledgeable about mental disorders, as they tend to be more interested in reading and learning about mental disorders. Mendenhall and Frauenholtz [23] have also shown that having children diagnosed with mood disorders such as bipolar disorder could promote MHL. More recently, Piper et al [24] confirmed this association by showing that in older people, being in close proximity to someone with a mental disorder predicted MHL. These findings show that familiarity with mental disorders and the ability to recognize mental disorder symptoms could be influenced by one’s personal experience with one’s children and loved ones, their curiosity to know more about mental disorders, and possibly other factors, such as being emotionally present for others. It can be argued that, in turn, this familiarity could decrease stigma toward mental disorders and increase the level of knowledge of symptoms and their appropriate treatments. Hence, exposure to mental disorders (eg, knowing someone with a mental disorder) could relate to greater knowledge and appreciation of mental illness [21], hence being significantly correlated to MHL.

Although empathy and mentalizing are similar constructs, mentalizing refers to one’s ability to envision one’s own mental states as well as others’. These include feelings, thoughts, intentions, and beliefs focusing on a more interpersonal level, thinking about how others’ feelings are affecting us, and how, in turn, this can modify our response to them [25]. Mentalizing does not necessarily entail empathy, as research shows that people diagnosed with psychosis are often unable to empathize with others but are capable of mentalizing [26]. Moreover, although empathy contains a cognitive component, cognitive empathy relates to one’s ability to attribute emotions rather than cognitions, such as in the case of mentalizing. Although the 2 constructs are dissociable in nature, both have common underlying features. As such, it could be argued that mentalizing, just like empathy, could also predict MHL. Indeed, the ability to understand others’ minds and mental states can be central to promoting MHL, since understanding others’ mental states could also mean understanding their mental health status and experiences. Given the fact that displaying empathy has been shown to be associated with increased MHL [21] and that
empathy and mentalization are similar constructs in nature, an association between mentalization and MHL has thus been hypothesized. However, this association has yet to be established in the literature.

MHL is a relatively new concept, hence the gap in the literature as to what contributes to its development. Indeed, the few studies evaluating the predictors of MHL have mainly focused on demographic factors such as age, gender, and level of education [23]. For instance, in Lebanon, the only study evaluating predictors of MHL in a sample of university students revealed that education in psychology was a strong predictor of MHL [27]. However, given the importance of digital health literacy in promoting well-being, this pilot study aims to understand whether digital health literacy contributes to MHL in health care major students. Moreover, we aim to explore psychological factors such as empathy and mentalizing as predictors of MHL. Evaluating these predictors in a population of health care major students would help set the stepping stone for future research, aiming to understand whether interventions and trainings are needed to promote the use of digital health in the health care system. Finally, our results could provide initial support in identifying psychological features that promote MHL in order for health care professionals to not only self-cultivate these features but also target these elements when working with patients in order to increase MHL and promote psychological well-being.

**Methods**

**Participants**

In this study, purposive sampling was used, and the study sample consisted of 89 university students. Participants consisted of both undergraduate and graduate health care major students (ie, psychology, premed tracks, nutrition, pharmacy, and medicine), recruited from the Lebanese American University, a private university in Beirut, Lebanon. Students were aged between 17 and 32 (mean 19.64, SD 2.01) years, and most of the sample (66/89, 75%) identified as women. Inclusion criteria included being a Lebanese university student, being aged 18 years or older, and being fluent in English.

**Measures**

In order to evaluate whether digital health literacy, empathy, and mentalizing predicted our outcome variable (ie, MHL), 4 questionnaires were used. Demographic variables that were collected included age, gender, and academic major.

The Mental Health Literacy Scale for Healthcare Students (MHLS-HS) [3] is a self-report questionnaire used to measure MHL in health care major students. Rated on a 5-point Likert scale (1=strongly disagree and 5=strongly agree), the MHLS-HS comprises 26 items and the following 5 subscales: maintenance of positive mental health (10 items), recognition of mental illness (4 items), attitude to mental illness stigma (6 items), help-seeking efficacy (3 items), and help-seeking attitude (3 items). The MHLS-HS is scored by summing the item scores, with higher scores indicating a better MHL. The MHLS-HS has shown good internal consistency, with α values ranging between .70 and .90 across subscales [3].

The eHealth Literacy Questionnaire (eHLQ) [27] is a self-report questionnaire that measures digital health literacy. The 35-item scale is rated on a 4-point Likert scale (1=strongly disagree and 4=strongly agree) and consists of seven subscales measuring the following components of digital health literacy: (1) using technology to process health information, (2) engagement in own health, (3) ability to actively engage with digital services, (4) the ability to feel safe and in control, (5) motivation to engage with digital services, (6) access to digital services that work, and (7) digital services that suit individual needs. Each subscale contains 5 items, except for subscale 6 containing 6 items and subscale 7 containing 4 items. The average score is calculated for each subscale, with higher scores indicating better digital health literacy. The eHLQ has been found to have good psychometric properties with composite reliability above 0.7 for all 7 subscales [27].

The Basic Empathy Scale (BES) [28] is a 20-item self-report scale that is used to measure empathy in its cognitive and affective elements. The first factor relates to cognitive empathy and is comprised of 9 items, while the second factor relates to affective empathy and is comprised of 11 items. Items are rated on a 5-point Likert scale (1=strongly disagree and 5=strongly agree). Scores are calculated by computing the average for each subscale, with higher scores indicating higher self-reported empathy. Overall empathy is calculated by summing the averages of the 2 subscales. The BES has demonstrated good internal consistency, with α=.79 for the cognitive empathy subscale and α=.85 for the affective empathy subscale [28].

The Reflective Functioning Questionnaire-8 (RFQ-8) [29] originally consisted of a 54-item self-report scale measuring mentalizing capacities. It is rated on a 7-point Likert scale (1=strongly disagree and 7=strongly agree) and is comprised of 2 subscales: uncertainty about mental states (RFQu) and certainty about mental states (RFQc). High scores on RFQc and low scores on RFQu reflect genuine mentalizing, while low scores on RFQc reflect hypermentalizing and high scores on RFQu reflect hypomentalizing, both indicating failure to mentalize. The RFQ-54 has demonstrated good internal consistency, with α=.67 for RFQc and α=.63 for RFQu. A shorter version of the RFQ-54, consisting of 8 items, was created by Fonagy et al [29] for research purposes and was used in this study.

**Procedure**

Data collection took place between October 2020 and December 2020, at a time when Lebanon was under lockdown due to the COVID-19 pandemic. Hence, data collection took place on the web, using Google Forms. Participation took between 15 and 20 minutes to complete.

**Data Analysis**

The aim of this exploratory pilot study was to investigate predictors of MHL. For this purpose, we ran a hierarchical multiple regression with MHL as the dependent variable and digital health literacy (eHLQ, Model 1), empathy (eHLQ and BES, Model 2), and mentalizing (eHLQ, BES, and RFQ, Model 3) as the independent variables. SPSS (SPSS Inc) was used for all analyses.
Moreover, since the literature shows that MHL entails recognition of mental illness, leading to lower stigma toward mental health, as well as help-seeking behaviors [2,3], we ran 3 additional hierarchical multiple regressions with the “Recognition of Mental Illness,” “Attitude Toward Mental Illness Stigma,” and “Help-Seeking Attitude” subscales as the dependent variables, and digital health literacy (eHLQ, Model 1), empathy (eHLQ and BES, Model 2), and mentalizing (eHLQ, BES, and RFQ, Model 3) as the independent variables.

**Ethical Considerations**

This study received ethical approval from the university institutional review board (LAU.SAS.PT5.27/Oct/2020). The survey was circulated on social media, including an information sheet. Participants interested in taking part e-signed the consent form before accessing the questionnaires. All data were anonymous, with no identifiers linking responses to the participant’s identity. Participants were informed that participation is voluntary and that they could drop out at any time. There was no compensation for participation.

**Results**

The hierarchical regression analysis predicting MHL revealed that in the first model, “Engagement in Own Health” was a significant predictor of MHL, with $F_{7,81}=2.19; P=.04$ and accounted for 16% of the variation in MHL. Introducing empathy (Model 2) explained an additional 4% of variation, though the model was not statistically significant ($F_{9,79}=1.75; P=.18$). Finally, introducing mentalizing (Model 3) did not explain any additional variation in the model, $F_{2,77}=0.23; P=.80$ (Table 1).

The hierarchical regression looking for predictors of the “Recognition of Mental Illness” subscale of MHL revealed that the first model, digital health literacy, was not significant ($F_{7,81}=9.60; P=.47$). Introducing the 2 empathy subscales explained an additional 8% of variation, and this change in $R^2$ was significant ($F_{2,79}=3.48; P=.04$), with affective empathy found to be a significant predictor. Adding the 2 mentalizing subscales to the regression model explained an additional 7% of variation, and this change in $R^2$ was significant ($F_{2,77}=3.58; P=.03$), with affective empathy and RFQc found to be significant predictors. Together, all factors explained 22% of the variation in the “Recognition of Mental Illness” subscale (Table 2).

The hierarchical regression looking for predictors of the “Attitude to Mental Illness Stigma” subscale revealed that in the first model, including the digital health literacy subscales, was not significant ($F_{7,81}=1.83; P=.09$). Introducing the 2 empathy subscales also led to a nonsignificant model (Model 2) with $F_{9,79}=2.09; P=.05$. When all factors were included (Model 3), they were found to significantly predict 27% of the variation in the “Attitude to Mental Illness Stigma” subscale ($F_{11,77}=2.54; P=.009$), with the “Being Motivated to Engage with Digital Services” subscale of digital health literacy and RFQu found to be significant predictors (Table 3).

The hierarchical regression looking for predictors of the “Help-Seeking Attitude” subscale of MHL revealed that in the first model, the “Ability to Process Information,” “Engagement in Own Health,” “Feeling Safe and in Control,” and “Being Motivated to Engage with Digital Services” subscales of digital health literacy were significant predictors in the regression model ($F_{7,81}=4.28; P<.001$) and accounted for 27% of the variation in the model. Introducing the 2 empathy subscales (Model 2) and the 2 mentalizing subscales (Model 3) did not lead to significant changes in $R^2$ with $F_{2,79}=0.1; P=.89$ and $F_{2,77}=1.12; P=.31$, respectively (Table 4).
Table 1. Summary of hierarchical regression analysis for variables predicting mental health literacy (MHL) total score.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>R</th>
<th>R^2</th>
<th>ΔR^2</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.39</td>
<td>0.16</td>
<td>0.16</td>
<td>.04</td>
</tr>
<tr>
<td>Ability to process information</td>
<td>−3.72</td>
<td>3.31</td>
<td>−.20</td>
<td>−1.12</td>
<td>.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>7.56</td>
<td>2.82</td>
<td>.36</td>
<td>2.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>.92</td>
<td>2.79</td>
<td>.05</td>
<td>0.33</td>
<td>.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>2.67</td>
<td>2.09</td>
<td>.15</td>
<td>1.28</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>1.39</td>
<td>3.06</td>
<td>.08</td>
<td>0.46</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>−2.99</td>
<td>2.69</td>
<td>−.16</td>
<td>−1.11</td>
<td>.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>2.08</td>
<td>3.09</td>
<td>.12</td>
<td>0.68</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.44</td>
<td>0.19</td>
<td>0.04</td>
<td>.18</td>
</tr>
<tr>
<td>Ability to process information</td>
<td>−3.01</td>
<td>3.31</td>
<td>−.16</td>
<td>−0.91</td>
<td>.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>6.64</td>
<td>2.84</td>
<td>.31</td>
<td>2.33</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>−0.03</td>
<td>2.82</td>
<td>−.00</td>
<td>−0.01</td>
<td>.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>2.47</td>
<td>2.08</td>
<td>.14</td>
<td>1.19</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>1.31</td>
<td>3.04</td>
<td>.08</td>
<td>0.43</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>−1.26</td>
<td>2.83</td>
<td>−.07</td>
<td>−0.45</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>0.31</td>
<td>3.23</td>
<td>.02</td>
<td>0.09</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>−3.23</td>
<td>1.73</td>
<td>−.24</td>
<td>−1.87</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>1.58</td>
<td>1.53</td>
<td>.13</td>
<td>1.03</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.20</td>
<td>0.00</td>
<td>.80</td>
</tr>
<tr>
<td>Ability to process information</td>
<td>−2.55</td>
<td>3.42</td>
<td>−.14</td>
<td>−0.75</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>6.86</td>
<td>2.89</td>
<td>.32</td>
<td>2.37</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>0.01</td>
<td>2.91</td>
<td>.00</td>
<td>0.00</td>
<td>.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>2.46</td>
<td>2.15</td>
<td>.14</td>
<td>1.14</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>1.37</td>
<td>3.07</td>
<td>.08</td>
<td>0.45</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>−1.19</td>
<td>2.94</td>
<td>−.06</td>
<td>−0.41</td>
<td>.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>−0.21</td>
<td>3.36</td>
<td>−.01</td>
<td>−0.06</td>
<td>.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>−2.73</td>
<td>1.90</td>
<td>−.20</td>
<td>−1.43</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>1.05</td>
<td>1.78</td>
<td>.08</td>
<td>0.59</td>
<td>.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: certainty subscale</td>
<td>0.35</td>
<td>2.82</td>
<td>.02</td>
<td>0.12</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: uncertainty subscale</td>
<td>−1.31</td>
<td>2.20</td>
<td>−.08</td>
<td>−0.59</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Table 2. Summary of hierarchical regression analysis for variables predicting the Recognition of Mental Illness Subscale Score.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>R²</th>
<th>ΔR²</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to process information</td>
<td>−0.11</td>
<td>0.21</td>
<td>−.09</td>
<td>−0.53</td>
<td>0.28</td>
<td>0.08</td>
<td>.47</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>−0.18</td>
<td>0.18</td>
<td>−.14</td>
<td>−1.02</td>
<td>.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>0.18</td>
<td>0.18</td>
<td>.16</td>
<td>1.03</td>
<td>.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>0.04</td>
<td>0.13</td>
<td>.03</td>
<td>0.27</td>
<td>.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>0.06</td>
<td>0.19</td>
<td>.06</td>
<td>0.31</td>
<td>.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>−0.32</td>
<td>0.17</td>
<td>−.29</td>
<td>−1.89</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>0.13</td>
<td>0.19</td>
<td>.12</td>
<td>0.65</td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Ability to process information</td>
<td>−0.05</td>
<td>0.21</td>
<td>−.05</td>
<td>−0.26</td>
<td>0.39</td>
<td>0.15</td>
<td>.08</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>−0.27</td>
<td>0.18</td>
<td>−.21</td>
<td>−1.50</td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>0.10</td>
<td>0.18</td>
<td>.09</td>
<td>0.59</td>
<td>.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>0.02</td>
<td>0.13</td>
<td>.01</td>
<td>0.12</td>
<td></td>
<td></td>
<td>.91</td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>0.06</td>
<td>0.19</td>
<td>.06</td>
<td>0.29</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>−0.18</td>
<td>0.18</td>
<td>−.16</td>
<td>−1.00</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>−0.02</td>
<td>0.20</td>
<td>−.02</td>
<td>−0.09</td>
<td>.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>−0.28</td>
<td>0.11</td>
<td>−.35</td>
<td>−2.63</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>0.11</td>
<td>0.09</td>
<td>.15</td>
<td>1.19</td>
<td>.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>Ability to process information</td>
<td>−0.04</td>
<td>0.20</td>
<td>−.04</td>
<td>−0.22</td>
<td>0.47</td>
<td>0.22</td>
<td>.07</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>−0.31</td>
<td>0.17</td>
<td>−.24</td>
<td>−1.78</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>0.01</td>
<td>0.17</td>
<td>.01</td>
<td>0.04</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>−0.05</td>
<td>0.13</td>
<td>−.05</td>
<td>−0.40</td>
<td>.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>0.05</td>
<td>0.18</td>
<td>.05</td>
<td>0.27</td>
<td>.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>−0.08</td>
<td>0.18</td>
<td>−.07</td>
<td>−0.44</td>
<td>.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>−0.02</td>
<td>0.20</td>
<td>−.02</td>
<td>−0.11</td>
<td>.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>−0.32</td>
<td>0.11</td>
<td>−.39</td>
<td>−2.79</td>
<td>.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>0.09</td>
<td>0.11</td>
<td>.12</td>
<td>0.86</td>
<td>.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: certainty subscale</td>
<td>0.39</td>
<td>0.17</td>
<td>.31</td>
<td>2.38</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: uncertainty subscale</td>
<td>0.24</td>
<td>0.13</td>
<td>.24</td>
<td>1.85</td>
<td>.07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Summary of hierarchical regression analysis for variables predicting the Attitude to Mental Illness Stigma Subscale Score.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>R</th>
<th>R^2</th>
<th>ΔR^2</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to process information</td>
<td>-0.00</td>
<td>0.27</td>
<td>-0.00</td>
<td>-0.00</td>
<td></td>
<td>0.37</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>0.12</td>
<td>0.23</td>
<td>0.07</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>0.19</td>
<td>0.23</td>
<td>0.13</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>-0.18</td>
<td>0.17</td>
<td>-0.12</td>
<td>-1.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>-0.65</td>
<td>0.25</td>
<td>-0.48</td>
<td>-2.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>-0.15</td>
<td>0.22</td>
<td>-0.09</td>
<td>-0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>0.35</td>
<td>0.26</td>
<td>0.24</td>
<td>1.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to process information</td>
<td>-0.01</td>
<td>0.27</td>
<td>-0.01</td>
<td>-0.05</td>
<td></td>
<td>0.44</td>
<td>0.19</td>
<td>0.06</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>0.19</td>
<td>0.23</td>
<td>0.15</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>0.23</td>
<td>0.23</td>
<td>0.15</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>-0.13</td>
<td>0.17</td>
<td>-0.09</td>
<td>-0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>-0.68</td>
<td>0.25</td>
<td>-0.49</td>
<td>-2.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>-0.24</td>
<td>0.23</td>
<td>-0.16</td>
<td>-1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>0.39</td>
<td>0.27</td>
<td>0.27</td>
<td>1.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>0.20</td>
<td>0.14</td>
<td>0.18</td>
<td>1.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>0.11</td>
<td>0.13</td>
<td>0.10</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to process information</td>
<td>0.12</td>
<td>0.27</td>
<td>0.08</td>
<td>0.45</td>
<td></td>
<td>0.52</td>
<td>0.27</td>
<td>0.07</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>0.28</td>
<td>0.23</td>
<td>0.16</td>
<td>1.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>0.27</td>
<td>0.23</td>
<td>0.18</td>
<td>1.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>-0.11</td>
<td>0.17</td>
<td>-0.08</td>
<td>-0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>-0.66</td>
<td>0.24</td>
<td>-0.48</td>
<td>-2.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>-0.25</td>
<td>0.23</td>
<td>-0.16</td>
<td>-1.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>0.24</td>
<td>0.26</td>
<td>0.17</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>0.36</td>
<td>0.15</td>
<td>0.33</td>
<td>2.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>-0.04</td>
<td>0.14</td>
<td>-0.04</td>
<td>-0.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: certainty subscale</td>
<td>-0.03</td>
<td>0.22</td>
<td>-0.02</td>
<td>-0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: uncertainty subscale</td>
<td>-0.47</td>
<td>0.17</td>
<td>-0.34</td>
<td>-2.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4. Summary of hierarchical regression analysis for variables predicting the Help-Seeking Attitude Subscale Score.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>R</th>
<th>R²</th>
<th>ΔR²</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to process information</td>
<td>-0.63</td>
<td>0.29</td>
<td>-0.35</td>
<td>-2.13</td>
<td>0.52</td>
<td>0.27</td>
<td>0.27</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>0.53</td>
<td>0.25</td>
<td>0.26</td>
<td>2.09</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>-0.26</td>
<td>0.25</td>
<td>-0.14</td>
<td>-1.05</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>0.59</td>
<td>0.19</td>
<td>0.36</td>
<td>3.21</td>
<td>0.32</td>
<td></td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>0.66</td>
<td>0.27</td>
<td>0.41</td>
<td>2.41</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>0.39</td>
<td>0.24</td>
<td>0.22</td>
<td>1.61</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>-0.30</td>
<td>0.27</td>
<td>-0.18</td>
<td>-1.09</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.89</td>
</tr>
<tr>
<td>Ability to process information</td>
<td>-0.63</td>
<td>0.29</td>
<td>-0.36</td>
<td>-2.10</td>
<td>0.52</td>
<td>0.27</td>
<td>0.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>0.54</td>
<td>0.26</td>
<td>0.27</td>
<td>2.11</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>-0.25</td>
<td>0.26</td>
<td>-0.14</td>
<td>-0.99</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>0.61</td>
<td>0.19</td>
<td>0.36</td>
<td>3.21</td>
<td>0.02</td>
<td></td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>0.65</td>
<td>0.28</td>
<td>0.41</td>
<td>2.36</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>0.37</td>
<td>0.26</td>
<td>0.21</td>
<td>1.43</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>-0.29</td>
<td>0.29</td>
<td>-0.18</td>
<td>-0.99</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>0.05</td>
<td>0.16</td>
<td>0.04</td>
<td>0.29</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>0.02</td>
<td>0.14</td>
<td>0.02</td>
<td>0.17</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.31</td>
</tr>
<tr>
<td>Ability to process information</td>
<td>-0.72</td>
<td>0.31</td>
<td>-0.41</td>
<td>-2.35</td>
<td>0.54</td>
<td>0.29</td>
<td>0.02</td>
<td>.02</td>
</tr>
<tr>
<td>Engagement in own Health</td>
<td>0.53</td>
<td>0.26</td>
<td>0.26</td>
<td>2.04</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to actively engage with digital services</td>
<td>-0.19</td>
<td>0.26</td>
<td>-0.11</td>
<td>-0.76</td>
<td>0.45</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Feel safe and in control</td>
<td>0.65</td>
<td>0.19</td>
<td>0.39</td>
<td>3.38</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivated to engage with digital services</td>
<td>0.64</td>
<td>0.28</td>
<td>0.40</td>
<td>2.34</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to digital services that work</td>
<td>0.29</td>
<td>0.26</td>
<td>0.16</td>
<td>1.10</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital services that suit individual needs</td>
<td>-0.19</td>
<td>0.30</td>
<td>-0.12</td>
<td>-0.64</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective empathy</td>
<td>-0.03</td>
<td>0.17</td>
<td>-0.02</td>
<td>-0.15</td>
<td>0.88</td>
<td></td>
<td></td>
<td>.39</td>
</tr>
<tr>
<td>Cognitive empathy</td>
<td>0.14</td>
<td>0.16</td>
<td>0.12</td>
<td>0.87</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: certainty subscale</td>
<td>-0.33</td>
<td>0.25</td>
<td>-0.16</td>
<td>-1.29</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective functioning: uncertainty subscale</td>
<td>0.08</td>
<td>0.19</td>
<td>0.05</td>
<td>0.44</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Discussion

With the global health crisis of the COVID-19 pandemic and the numerous lockdowns across the globe, health care providers have recently switched to web-based health care. Moreover, with the increase in mental health awareness campaigns, the COVID-19 era has drastically shifted our perception of mental health, reducing the stigma surrounding mental disorders. In this digital age of mental health awareness, this pilot study aimed to explore predictors of MHL in health care major students, specifically evaluating whether digital health literacy contributes to MHL. Moreover, we explored psychological factors such as empathy and mentalizing in an attempt to pinpoint predictors of MHL.

Our results suggest that “Engagement in Own Health,” a digital health literacy component, is a predictor of overall MHL. This confirms previous findings showing that health engagement improves patient activation, meaning patients’ ability to manage their health [30]. This ability goes hand in hand with MHL, as it entails recognizing, managing, and seeking treatment for mental disorders [2]. Moreover, the “Ability to Process Information,” “Engagement in Own Health,” “Feeling Safe and in Control,” and “Being Motivated to Engage with Digital Services” subscales of digital health literacy were found to predict MHL components. This partially confirms our hypothesis on the role of digital health literacy in promoting MHL, indicating that some constructs of digital health play a vital role in promoting MHL. In that sense, digital health could be hypothesized to play a role in decreasing stigma surrounding
mental disorders, allowing individuals to become more open to seeking help. This is especially important in a country such as Lebanon, where mental health remains taboo, thus preventing individuals from seeking help in an attempt to avoid criticism [31]. Our results hint at the importance of digital health practices and services in the health care system and call for future research to further explore whether specialized digital health trainings for health care major students as well as the general public could increase MHL. This is of special importance given the cost-effectiveness of digital services, making them more easily implemented, particularly given the current socioeconomic crisis in Lebanon.

While empathy has been shown to contribute to MHL [21-22], our results did not fully support this hypothesis. It is worth noting, however, that both empathy and mentalizing predicted components of MHL, namely “Recognition of Mental Illness” and “Attitude to Mental Illness Stigma.” This partially confirms findings on the role of empathy in promoting MHL [22-24]. Moreover, these findings support our hypothesis on the role of mentalizing in promoting MHL, a relationship that, to our knowledge, was never explored in the literature. Indeed, mentalizing refers to one’s capacity to think in terms of mental states underlying behaviors [25]; it can therefore be argued that this capacity facilitates people’s awareness of difficulties with emotion regulation, thus pinpointing signs of distress or mental health problems. Given that mentalizing was found to play a protective factor against stigma as it promotes thinking about the potential negative effects of being subjected to prejudice and stigma in general [32], and more specifically in relation to mental illness, we argue for the need for further research exploring this correlation in a larger, more representative sample.

While a recent study has evaluated demographic factors as predictors of MHL in Lebanon [33], to our knowledge, this study is the first to evaluate digital health literacy, empathy, and mentalizing as predictors of MHL in Lebanon. It is also globally the first to hint that mentalizing could constitute a predictor of MHL, making mentalizing a new variable of interest. Moreover, it is the first to examine the positive impact of digital health literacy during the COVID-19 pandemic, an era that has considerably shifted health care practices from face-to-face to web-based. However, it is important to interpret the results in light of some limitations. First of all, since data collection took place in the midst of the COVID-19 pandemic, the sample size was small. Indeed, not many students participated in the study, possibly due to a lack of motivation and a multitude of web-based demands. Second, since the survey was disseminated on the web, only students comfortable with technology took part in the study. This may have biased our results, especially since we evaluate digital health literacy as a predictor of MHL. Indeed, Tohme et al [9] have shown that those who are familiar with social media platforms are more likely to seek help on the web through digital health channels. Finally, the MHLS-HS [3] is new, and its psychometric properties are not fully evaluated. For that, the results of this study would need to be replicated on a larger sample after having validated the scale.

In summary, our findings provide initial support for arguing the role of digital health literacy in fostering and promoting MHL. These findings should be replicated using different measures of digital literacy and MHL in a larger, more representative sample. If findings were to be supported, they could impact recommendations at the institutional, social, and personal levels. At the institutional level, it could hint that digital health literacy practices will become part of university curricula. As for the social level, it could give policymakers support to raise awareness about the importance of digital health and to offer digital health trainings to the general population, including people from different backgrounds, age groups, and socioeconomic status. Finally, at the personal level, health care providers could make a case for the use of digital health, such as telepsychotherapy, mobile health, and telehealth. Since this was a pilot study, in order to gain further insight and be able to generalize our results, it is recommended to replicate this study while collecting data from a larger sample, a sample that is not limited to students or people who have access to and are familiar with technology. Finally, given the significant correlations between health literacy and mental health outcomes [15], as well as our results highlighting the predictive role of some digital health literacy subscales on MHL, future research should evaluate whether digital health literacy can contribute to better mental health outcomes, an association that could be mediated by MHL.

Acknowledgments
All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article. All authors have participated in the research design and manuscript preparation. All authors have approved the final version of the manuscript.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.
References


Abbreviations

BES: Basic Empathy Scale

eHLQ: eHealth Literacy Questionnaire

MHL: mental health literacy

MHLS-HS: Mental Health Literacy Scale for Healthcare Students

RFQ-8: Reflective Functioning Questionnaire-8

RFQu: Reflective Functioning Questionnaire-8—uncertainty about mental states

Please cite as:

Tohme P, Abi Fadel N, Yaktine N, Abi-Habib R

Predictors of Mental Health Literacy in a Sample of Health Care Major Students: Pilot Evaluation Study

JMIR Form Res 2024;8:e43770

URL: https://formative.jmir.org/2024/1/e43770
doi:10.2196/43770
PMID:38329801

©Pia Tohme, Nour Abi Fadel, Nour Yaktine, Rudy Abi-Habib. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete
bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Exploring User Perspectives on Brief Reflective Questioning Activities for Stress Management: Mixed Methods Study

Ananya Bhattacharjee1, BSc; Pan Chen1, BSc; Abhijoy Mandal1; Anne Hsu2, PhD; Katie O'Leary3, PhD; Alex Mariakakis1, PhD; Joseph Jay Williams1, PhD

1Department of Computer Science, University of Toronto, Toronto, ON, Canada
2School of Electronic Engineering and Computer Science, Queen Mary University, London, United Kingdom
3Google, Seattle, WA, United States

Corresponding Author:
Ananya Bhattacharjee, BSc
Department of Computer Science
University of Toronto
40 St George St
Toronto, ON, M5S 2E4
Canada
Phone: 1 647 619 6982
Email: ananya@cs.toronto.edu

Abstract

Background: Current online interventions dedicated to assisting individuals in managing stress and negative emotions often necessitate substantial time commitments. This can be burdensome for users, leading to high dropout rates and reducing the effectiveness of these interventions. This highlights an urgent need for concise digital activities that individuals can swiftly access during instances of negative emotions or stress in their daily lives.

Objective: The primary aim of this study was to investigate the viability of using a brief digital exercise, specifically a reflective questioning activity (RQA), to help people reflect on their thoughts and emotions about a troubling situation. The RQA is designed to be quick, applicable to the general public, and scalable without requiring a significant support structure.

Methods: We conducted 3 simultaneous studies. In the first study, we recruited 48 participants who completed the RQA and provided qualitative feedback on its design through surveys and semistructured interviews. In the second study, which involved 215 participants from Amazon Mechanical Turk, we used a between-participants design to compare the RQA with a single-question activity. Our hypotheses posited that the RQA would yield greater immediate stress relief and higher perceived utility, while not significantly altering the perception of time commitment. To assess these, we measured survey completion times and gathered multiple self-reported scores. In the third study, we assessed the RQA’s real-world impact as a periodic intervention, exploring engagement via platforms such as email and SMS text messaging, complemented by follow-up interviews with participants.

Results: In our first study, participants appreciated the RQA for facilitating structured reflection, enabling expression through writing, and promoting problem-solving. However, some of the participants experienced confusion and frustration, particularly when they were unable to find solutions or alternative perspectives on their thoughts. In the second study, the RQA condition resulted in significantly higher ratings ($P=.003$) for the utility of the activity and a statistically significant decrease ($P<.001$) in perceived stress rating compared with the single-question activity. Although the RQA required significantly more time to be completed ($P<.001$), there was no statistically significant difference in participants’ subjective perceived time commitment ($P=.37$). Deploying the RQA over 2 weeks in the third study identified some potential challenges to consider for such activities, such as the monotony of doing the same activity several times, the limited affordances of mobile phones, and the importance of having the prompts align with the occurrence of new troubling situations.

Conclusions: This paper describes the design and evaluation of a brief online self-reflection activity based on cognitive behavioral therapy principles. Our findings can inform practitioners and researchers in the design and exploration of formats for brief interventions to help people with everyday struggles.

(JMIR Form Res 2024;8:e47360) doi: 10.2196/47360
KEYWORDS
reflection; mental health; stress; reflective questioning activity; RQA; brief intervention; computer-mediated communication; email; SMS text messaging; mobile phone

Introduction

Background

Computer-mediated communication (CMC) platforms offer accessible resources to assist people in managing their stress and negative emotions [1,2]. Nevertheless, current online interventions can be time-consuming and inconvenient [3], necessitating users to commit to a series of hour-long sessions to achieve optimal results. Social media groups and SMS text messaging programs also require a substantial time commitment from users to deliver maximum benefit [3,4]. Although research demonstrates that these programs can be as effective as in-person therapy [5,6], the considerable time investment required may lead to high dropout rates. Consequently, the convenience of online resources is paramount in enhancing their efficacy and user engagement [7].

Therefore, we investigate whether it might be possible to construct a brief digital activity (as simple as answering questions in a web form) that people can easily reference or practice when they experience negative thoughts and emotions in their daily lives. Through a simple interface with a series of questions, we explore whether a brief reflective questioning activity (RQA) could prompt people to reflect on a stressful situation. This process of articulating thoughts and emotions has the potential to enhance an individual’s understanding of their personal challenges and foster a sense of self-agency [8,9], eventually strengthening their belief in their own ability to manage stress and negative emotions [10]. Brief activities such as RQAs, which require minimal effort and may provide tangible benefits, can also serve as a stepping stone to more extensive treatments [11,12]. This approach has tremendous potential in terms of convenience as well because such RQAs can be delivered to anyone anytime via email, app, and SMS text message. We posit that activities such as this can be made generalizable enough so that they can be adapted to fit the unique needs and preferences of individuals from diverse backgrounds and situations; for example, an individual experiencing stress at work may use reflective questioning to reflect on their thoughts and emotions related to a difficult conversation with a coworker, and another individual may adapt the same activity to reflect on their feelings after a breakup or a family conflict.

In our work, we draw on insights from clinical psychology and human-computer interaction literature on how to design brief RQAs that are helpful for people to manage psychological well-being and adopt healthy behaviors [13-15]. Murgraff et al [15] demonstrated that a persuasive 2-page pamphlet distributed at the beginning of an 8-week study period and informing female university students about recommended drinking limits could effectively reduce unhealthy drinking behaviors. Carney et al [16] used a similar intervention to support adolescent users of substances and their caregivers. These studies suggest that extensive interventions are not always necessary to foster healthy behavior; providing a brief guideline with crucial information and actionable practices for self-directed application can be beneficial too.

Our work is focused on the goal of promoting self-reflection, a crucial component of cognitive behavioral therapy (CBT) [17] and psychology in general. One can understand self-reflection as a person’s conscious effort to understand and reevaluate their own thoughts regarding any situations, thoughts, or feelings [18,19]. Self-reflection is often the driving force that converts one’s intentions into action [20]. Furthermore, it allows an individual to view situations from a different perspective, enabling them to understand the opinions of others [20,21]. In recent years, researchers have incorporated many reflective activities into mental health and behavior change interventions, particularly through mobile phone apps that show users summaries of their mood or physical activity [22-24]. Other digital tools have attempted to promote self-reflection through conversational agents [25,26]. As evident with the recent emergence of chatbots such as ChatGPT [27,28], conversational agents continue to become more sophisticated in parallel with advances in natural language processing, but they are still limited in their ability to have nuanced and empathetic conversations [27]. Furthermore, the literature suggests that back-and-forth conversations are not always necessary to elicit self-reflection because asking probing questions with the words why or how can be enough to increase one’s own understanding of a problem [29,30].

However, there are several reasons to speculate that brief RQAs may not effectively help individuals manage their stress. First, responses for self-reflection may not provide people with something concrete or tangible (eg, new information or social validation) and might require repeated exposure to yield benefits that people can see [31,32]. Moreover, it is unclear whether people would see value in answering reflective questions and whether an extended series of questions would add much value. Answering a static set of questions could not only be perceived as a waste of people’s time but also surface more negative emotions without a conversational partner to give input. Furthermore, people might prefer knowing that their thoughts and emotions are being shared with another person rather than relying on themselves to gain benefits.

Drawing upon these potential opportunities and challenges, we set the following guiding principles for our exploration:

- **Minimal time commitment:** the activity should be simple enough so that people can complete it in 15 minutes—the equivalent of a midday coffee break at work or a fraction of a person’s morning routine.
- **Applicability to the general public:** the activity should not be targeted toward a particular domain, culture, or population. In other words, the activity should be generalizable to the point where people can adapt it to their own context and situation.
• Scalability: the activity should be implemented and deployed in a way that does not need a significant support structure. This means that the activity should not require a live conversational partner or intensive scaffolding (eg, tutorial videos).

To investigate the feasibility, challenges, and opportunities in the design of digital RQAs, we created a design probe that asks people to answer a series of 9 questions to reflect on a troubling situation. The questions in our RQA are intended to help people articulate their thoughts and emotions about the situation using principles from CBT [33]. We leveraged thought records [34] and behavioral chaining analysis [10], which are techniques that encourage people to connect their thoughts, experiences, and emotions to identify triggers that generate negative patterns and come up with alternative ways of thinking.

We provide insights into the design of our RQA and how it was experienced by users, which we hope will inform the design of future interventions with similar goals. We gathered these observations through 3 studies. For our first study, we used a convenience sample of crowdworkers and university students to administer the RQA and obtain qualitative feedback on the design of the activity. In our second study, we investigated whether the perceived benefits of going through an RQA outweighed the additional time commitment required to answer a series of probing questions. In our third and final study, we investigated the potential impact of the RQA when delivered repeatedly over a 2-week period in a real-world context; we also explored the implications of distributing the RQA over email versus SMS text message. The design of our RQA was kept constant across all 3 studies so that we could maintain consistency across evaluations and determine which observations held true across the different scenarios.

We found that the structured analysis supported by our RQA helped people reduce their stress and identify solutions for improvement. Although our RQA consisted of 9 questions, people did not complain about the time commitment required to complete it and generally wrote thoughtful responses to the prompts. However, deploying the RQA over the course of 2 weeks raised some potential challenges, including the monotony of doing the same activity several times, the limited affordances of mobile phones, and the importance of having the prompts align with the occurrence of new troubling situations. These highlight design considerations and opportunities for researchers and practitioners to consider as they develop their own digital RQAs, such as giving users control over the frequency of prompts and automated question personalization.

Main Contribution

In summary, our main contribution is an investigation into whether people see value in a brief digital RQA without a conversational partner for interaction or advice. We deliver this contribution in four parts: (1) the creation of an RQA probe that people can complete on their computer or mobile phone to reflect upon a stressful situation; (2) insights into the value and pitfalls of RQAs gathered via surveys completed by, and interviews with, 42 Amazon Mechanical Turk (AMT) participants and 6 university students; (3) evidence that people see value in an RQA compared with a baseline activity via a comparison study run on AMT with 215 participants; and (4) observations and design considerations from a 2-week deployment of our RQA using different CMC platforms.

Methods

Overview

In this section, we first discuss the design of our RQA and then describe the logistics of the 3 studies we conducted. The studies were conducted simultaneously with the same RQA design to explore different aspects of the intervention. Study 1 involved gathering feedback on the qualities of the RQA from a broad demographic using surveys and semistructured interviews. In study 2, the perceived benefits of the RQA were compared with those of a shorter baseline activity with the goal of determining whether the additional time commitment required to complete the RQA was justified by the benefits of the intervention. Study 3 aimed to explore how people would perceive the RQA during their everyday lives and how best to prompt engagement using email and SMS text messaging.

The Design of Our RQA

Our research team, which consists of graduate students and faculty members with experience in psychology and human-computer interaction, was guided by existing CBT resources to create an RQA that helps people reflect on a troubling situation in their lives. We first reviewed popular CBT apps and websites intended for personally guided use (eg, Youper [35], Depression CBT Self-Help Guide [36], KokoBot [37], and Woebot [38]) to identify the techniques they used to provide benefits to users. In particular, we found that these resources leverage several components of a CBT exercise called a thought record [10]. A thought record is a worksheet with a grid that includes 5 columns: situation, thoughts, emotions, behaviors, and alternative thoughts. The exercise aims to encourage behavioral chaining—a process through which people draw connections between their thoughts and emotions to identify triggers and irrational thoughts—revealing potential opportunities to reframe their way of thinking [10,39].

Researchers have identified several benefits to thought records and behavioral chaining. Thought records can help people recall memories of prior events that were initially assumed to be unimportant [40]. Identifying the full timeline of an event can help people recognize their own faulty behavior patterns, thus preparing them for similar events in the future [41]. Moreover, informal exposure to negative experiences can increase one’s ability to tolerate troubling situations [42] or recover from problematic behaviors (eg, binge drinking and self-harming) [10]. Thought records are typically introduced as CBT homework assignments that patients can complete between visits with a trained professional, providing them with the scaffolding to complete the activity on their own.

Our RQA attempts to distill this exercise into a brief guided activity that can be completed on a person’s computer or mobile phone without the need for external support. After writing a collection of brief questions to encapsulate these concepts, we iteratively added, removed, revised, and reordered the questions until we reached the RQA structure shown in Table 1. Our
primary design goal was to give people a structured activity they could use independently to organize their thoughts. Inspired by thought records and behavioral chaining, our activity guides users through the following line of thinking: trigger → thought → feeling → behavior [9,10].

We first start by asking the user to think about a stressful situation and write about it in as much detail as they like. Prior work suggests that this sort of open-ended question allows users to open up about their problems and be comfortable with the activity [9]. The next 5 questions (Q2-Q6) become more specific, asking users to identify the most important stressor, the most troubling thoughts and feelings, and the behaviors that come from these thoughts and feelings. The seventh question then asks users to retype the details of the situation in a structured format. Beyond leading people through the process of behavioral chaining, these questions allow users to iterate upon their initial thoughts regarding their stressful situation. The structured format in the seventh question is also designed to help users draw connections between several components of their situation. This leads to the eighth question, which challenges the user’s mental process by asking them whether they believe that the trigger justifies their thoughts. Doing so can help people identify flaws in their logic or possible cognitive distortions [10]. The final question asks the user to explore alternative ways of thinking that would enable them to see the problem from a different perspective and induce a different emotion [43].

We presented our RQA to 4 clinical psychologists with expertise in CBT to validate its construction and help us consider the best ways of evaluating it. The psychologists verified that our RQA is aligned with activities that would be used in psychotherapy, but they also remarked that the questions focused on advanced techniques that were usually introduced only after several sessions of evaluation and psychoeducation. They suspected too that people might find the activity too lengthy or that people might not know how to respond to some of the questions; 1 psychologist even posited that >2 questions might be excessive for an online format without a conversational partner. The study that follows in this paper demonstrates that although these concerns were warranted, participants found value in the additional line of questioning.

Table 1. The questions that compose our reflective question activity. The design of these questions is influenced by thought records and behavioral chaining. Before seeing these questions, participants were provided with the following prompt: “Think of a particular situation where you felt stressed or had a negative emotion, which you can try to reflect on as you go through this activity. It could be a current situation, one in the past, or one you anticipate in the future.”

<table>
<thead>
<tr>
<th>Questions</th>
<th>Example response</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Q1. What’s the situation? Feel free to explain it in as much detail as you’d like.”</td>
<td>“My son has moved away and left no way for me to get in contact with him.”</td>
<td>Provides context for the activity</td>
</tr>
<tr>
<td>“Q2. What part of the situation is the most troubling?”</td>
<td>“The fact that he does not care enough to reach out to me and let me know he is safe.”</td>
<td>Sets an agenda for the rest of the activity</td>
</tr>
<tr>
<td>“Q3. What are you thinking to yourself?”</td>
<td>“I hope he is okay and safe. I wonder why he would do this. I thought we had a good relationship.”</td>
<td>Identifies troubling thoughts</td>
</tr>
<tr>
<td>“Q4. What thought is the most troubling?”</td>
<td>“I don’t know if he is safe.”</td>
<td>Focuses attention on the most troubling thought</td>
</tr>
<tr>
<td>“Q5. What do you feel when you think this?”</td>
<td>“Panicked and worried.”</td>
<td>Reinforces the core CBT principle that thoughts trigger feelings</td>
</tr>
<tr>
<td>“Q6. When you have these feelings, what actions do you take? What actions do you avoid?”</td>
<td>“I try to refocus my thoughts on something else. I try to avoid thinking about what bad things could be happening to him.”</td>
<td>Identifies behaviors that are caused by the cascading effect of thoughts and feelings</td>
</tr>
<tr>
<td>“Q7. Retype the summary of the situation in the following format:</td>
<td>“I am triggered by thoughts of my son taking off and not staying in contact. I think about all the bad things that could happen and why he would do this. I feel panicked and worried. When feeling this way I try to think about other things and not focus on the negative of the situation.”</td>
<td>Synthesizes past reflection by highlighting the connection between the trigger and its manifestations</td>
</tr>
<tr>
<td>Trigger:</td>
<td>“The trigger does justify it. This is my child that I raised. I no longer know where he is, I cannot get in touch with him and I don’t know if he is okay.”</td>
<td>Challenges potentially negative thought patterns</td>
</tr>
<tr>
<td>Thought:</td>
<td>“I raised my child to be independent and he is trying to exercise that independence for the first time in his life. He needs me to take a step back for a while so that he can do this on his own.”</td>
<td>Encourages alternative thoughts that can provoke different feelings and behaviors</td>
</tr>
<tr>
<td>Feeling:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aCBT: cognitive behavioral therapy.
Study 1: User Perspectives After Onetime Use of Our RQA

Overview
Our first study gathered qualitative feedback on the qualities that people saw in the proposed RQA irrespective of other factors (eg, when it was being used and how it compared with other interventions). We used surveys to collect diverse feedback from a broad demographic. Subsequently, we used semistructured interviews to gather deeper insights into some of the salient topics.

Participants
We initially recruited 50 participants from AMT. Participants were required to have a minimum approval rating of 95%. We did not incorporate explicit attention check questions in our surveys but implemented a thorough manual review process to ensure data quality. Two independent members of our research team examined each response, discarding any that were incomplete or contained nonsensical or irrelevant content. Because of data quality issues, we discarded data from 8 (16%) of the 50 participants, leaving us with a final sample of 42 (84%). This cohort of 42 AMT crowdworkers included 35 (83%) men and 7 (17%) women, with an average age of 34.6 (SD 9.99) years. We identified these participants as M1 to M42, and they were compensated CAD $4 (US $2.97) for their time. We also recruited 6 additional people via email and word of mouth from a university campus community to serve as interview participants. This cohort of 6 people included 1 (17%) man and 5 (83%) women with an average age of 19.7 years. We identified these participants as L1 to L6, and they were compensated CAD $15 (US $11.15) for their time. There were no inclusion criteria because we were interested in observing how our RQA would be perceived by the general population.

Study Procedure
All participants were asked to complete the RQA online on the Qualtrics survey platform (Qualtrics International Inc), with all questions being presented on a single page. The data from this survey were saved and made accessible to the research team. After participants finished the RQA, they were requested to provide their feedback on the activity through a separate survey. The questions included, but were not limited to, the following: “How did this activity affect your stress levels?” “How did you feel about answering these questions in this online format?” and “Was any part of the activity not helpful or could be improved?” The university students also answered similar questions, although they participated in semistructured interviews immediately after completing our RQA. The interviews took 45 to 60 minutes and were held either in person or through different videoconferencing platforms.

Data Analysis
The survey responses were analyzed using a thematic analysis approach [44]. After the interviews were transcribed, 2 researchers examined the data together to familiarize themselves with the general sentiments of the participants. The researchers then individually applied the open coding process [45] to a subset of the data to develop their own preliminary codebooks. After sharing their codebooks with one another, the researchers held multiple discussions to consolidate the codes into a shared codebook. Next, they applied this codebook to a different subset of the data and again refined the codebook. Finally, the researchers reached a consensus and applied the final codebook to separate halves of the data.

Study 2: Comparing the RQA With a Baseline

Overview
The observed benefits of reflective questioning may be attributed to the act of discussing a troubling situation rather than the structured questions themselves. Furthermore, clinical psychologists raised concerns that asking participants to answer >2 questions may be overwhelming. To investigate these possibilities, we compared the effects of the RQA with those of simply asking participants to discuss their troubling situation without structured questions. In this baseline activity, participants were required to write about a stressful situation in as much detail as possible in response to a single question. If the structured questions in the RQA provided additional benefits over the baseline activity despite the added time commitment, we posited that the RQA would warrant further exploration as a tool for promoting self-reflection.

Participants
For study 2, we again used AMT and adhered to the same participant recruitment and data quality assurance procedures from study 1. We initially recruited 255 participants for this study. Of the 255 participants, after the data screening process, we excluded 40 (15.7%; n=17, 43% from the baseline group and n=23, 58% from the RQA group) owing to issues related to data quality. This led to a final participant count of 215, with 111 (51.6%) individuals randomly assigned to the baseline group and 104 (48.4%) to the RQA group. Our study included participants of different genders, with 61.9% (133/215) identifying as men, 35.8% (77/215) identifying as women, and 2.3% (5/215) preferring not to disclose their gender. The mean age of the participants was 33.8 (SD 9.51) years. As with study 1, all participants were compensated CAD $4 (US $2.97) for their participation, and there were no inclusion criteria.

Study Procedure
The study had a between-participants design in which participants were randomized into 1 of 2 conditions. The first condition, which we consider the RQA condition, entailed participants completing our 9-question RQA. The second condition, which we call the baseline condition, asked participants to reflect on a troubling situation they were experiencing in as much detail as they wished, answering only a single question.

We expected the RQA to take longer to complete than the baseline condition, given that it involved answering more questions. However, we were also interested in participants’ perceptions of the activity’s length and the value they placed
on the additional time spent. By using a between-participants design, the study aimed to assess whether completing the RQA would lead to differences in outcomes compared with the baseline condition.

**Data Analysis**

We collected data before and after participants completed their respective activities to evaluate the hypotheses outlined in the following subsections.

**Hypothesis 1 (Perceived Benefits)**

We hypothesized that participants in the RQA condition would experience more instantaneous stress relief from completing the activity than those in the baseline condition.

To evaluate this hypothesis, we asked participants to rate how useful they felt the activity was. We call this measure *perceived utility*, and it was measured using a 7-point scale (ranging from −3 for strongly disagree to +3 for strongly agree). We also asked participants to rate the degree to which they were feeling troubled about their selected situation before and after the activity. These ratings were provided using an 11-point scale (ranging from −5 to +5) to increase the resolution with which people could express their stress. We call the difference between the ratings before and after the activity the *perceived stress change*, with positive values indicating a decrease in stress.

Both perceived utility and perceived stress change were compared across conditions using independent samples 1-tailed Welch t tests. For each measure, the null hypothesis (H₀) was that the mean for the RQA condition would be less than, or equal to, the mean for the baseline condition. By contrast, the alternative hypothesis (H₁) was that the mean for the RQA condition would be greater than the mean for the baseline condition.

**Hypothesis 2 (Elapsed Time)**

We hypothesized that participants in the RQA condition would take more time to complete the activity than those in the baseline condition; yet the perceived time commitment would not be significantly different.

To evaluate this hypothesis, we recorded the time it took for participants to complete the activity, the number of words they typed across all questions, and a self-reported rating using a 7-point scale (ranging from −3 to +3) of whether they felt the activity was worth their time. We call these measures completion time, response length, and perceived time commitment, respectively. All 3 measures were compared across conditions using independent samples 1-tailed Welch t tests. For each measure, the null and alternative hypotheses were set in a similar manner as detailed for hypothesis 1.

**Study 3: Observing Repeated Engagement With the RQA**

In our third and final study, we aimed to assess the effectiveness of the RQA in a real-world setting as a periodic intervention and explore the most effective ways to prompt engagement through low-cost asynchronous CMC platforms such as email and SMS text messaging.

**Participants**

We recruited 11 participants (n=8, 73% men and n=3, 27% women) with an average age of 20.6 years. Participants were recruited via email invitations and word of mouth from the same university campus community as study 1 without any inclusion criteria. We refer to these participants as D1 to D11. Participants were not compensated for completing our RQA to avoid undue influence on their level of engagement; however, they were compensated CAD $10 (US $7.43) for completing surveys and CAD $15 (US $11.15) for the interviews.

**Study Procedure**

Participants were recruited to take part in our study for 2 weeks. During the enrollment phase, participants were asked to specify the hours during each day when they would prefer to receive a notification to complete the RQA. They were asked to provide separate preferences for email and SMS text message, and they were allowed to select multiple times during a given day. Participants were then randomized into 1 of 2 groups. One group received emails during the first week and SMS text messages during the second week, whereas the other group experienced the reverse. The notifications prompted participants to complete the RQA and provided them with a link that took them to a web page containing the RQA. We used the same link each time, and participants were aware of this fact.

Participants were prompted to complete the activity once per day for up to 3 days within a given week, similar to what has been done in previous work [46]. Of the 11 participants, 8 (73%) were available for >3 days, and the days on which they received prompts were randomly selected, whereas 3 (27%) were available for <3 days (D2, D8, and D9), and they received a message on every day of their availability.

At the end of the study, participants were asked to complete an exit survey containing questions about their overall experience and their CMC preferences in the context of the RQA. They were then invited to a semistructured interview session to elaborate on their experience. The interviews lasted 15 to 30 minutes, with frequent topics including the barriers people faced while completing the RQA, the applicability of the RQA to their lives, and the trade-offs of being prompted to complete the RQA repeatedly. Of the 11 participants, 7 (64%) took part in the interviews. The interviews were conducted over the Zoom teleconferencing platform (Zoom Video Communications, Inc).

**Data Analysis**

We recorded a variety of data to assess how people engaged with our RQA. We measured how often participants responded to our prompts without a limit on how long they took to respond. In other words, if a participant received a prompt in the morning but waited until the next day to complete our RQA, we still counted this as a response. We calculated the response rate in this way because it is well documented that people respond to emails and SMS text messages at their convenience rather than at the moment of reception [47]. As in study 2, we asked participants to rate their stress using an 11-point scale before and after the activity, and we report the change in this rating.

We also report the time it took for participants to complete the RQA and the word count of their responses as proxies for
engagement. We analyzed the interview responses using the same procedure that was applied to study 1; however, we did so with a new blank codebook.

**Ethical Considerations**

We recognize that conducting research on mental health can raise several ethical issues; for example, our particular set of questions can induce stress or symptoms of depression and anxiety, particularly when participants are asked to recall a troubling situation. To mitigate these negative outcomes, we clearly explained the potential risks in the consent materials and reminded participants that the RQA was part of a research study. We also provided survey participants with the contact information of several mental health helplines. The interviewers were trained to clearly explain the goal of the project and maintain an appropriate level of empathy and support. Interviewers were also trained to run the Columbia-Suicide Severity Rating Scale protocol [48] if interviewees showed any indication of self-harm or suicidal ideation. Furthermore, interviewees had the option to skip any question they did not want to answer or to leave the interview session at any point. All our research activities were approved by the University of Toronto Research Ethics Board (36582).

**Results**

### Study 1: User Perspectives After Onetime Use of Our RQA

During our first study, we elicited 4 major themes related to the benefits and pitfalls of our RQA for first-time users. We provide evidence for each theme in the following subsections.

**Appreciation for Structured Reflection**

Participants were appreciative of the fact that our RQA helped them break down the components of their stressful situation. By deconstructing the situation, participants felt that they were able to become more aware of the causes of their negative emotions, putting their thoughts “in the right order” (L2). Some of the participants also noted that the activity helped them recognize faulty thought patterns:

> The activity helped me pinpoint my maladaptive coping strategy...[it] led me to think more with my brain and less with my immediate emotional reaction. [M17]

**Venting Negative Thoughts Through Writing**

Participants enjoyed expressing their thoughts and feelings through writing because it allowed them to “get out all thoughts and feelings and take that weight off of my shoulders” (L5). Moreover, some of the participants appreciated seeing their thoughts typed out in front of them, commenting that the act of writing helped solidify previously nebulous or disjointed thoughts; for example, L4 thought that the RQA forced them to dissect their feelings that would have otherwise been unorganized.

M11 suggested that writing about their thoughts allowed them to examine their situation “from an outside perspective,” almost as if they were analyzing someone else’s situation instead of their own. This affordance made it easier for them to ignore personal tendencies and instead think more objectively about their thought process.

**Helping Identify Solutions**

Participants also stated that the activity prompted them to adopt a problem-solving approach to improve their situation. They could better identify the root cause of their stress because they were prompted to describe their troubling situation in a structured order, which made it easier for them to find a solution to their problem. As the final question of our RQA prompted users to consider alternative ways of thinking, participants such as L3 felt empowered because they were often able to emerge from the activity with at least 1 prototype solution.

**Incidental Negative Side Effects**

Our RQA did not unilaterally help people become less worried about their troubling situation. L5 noted that as they were considering an alternative line of thinking, they found it confusing to keep track of both their original thought process and the reframed one. L6 felt that this confusion led directly to frustration, whereas others were frustrated because they could not identify a solution by the end of the activity:

> The questions just made me think about how much pain I was in and really didn’t offer any solution whatsoever to the stress. [M19]

Some of the participants also felt at a loss when asked to think of alternative perspectives on their thoughts.

### Study 2: Comparing the RQA With a Baseline

#### Overview

Of the 215 participants, 111 (51.6%) were randomly assigned to the baseline condition and 104 (48.4%) to the RQA condition. The summary statistics for the measures that were collected during study 2 are shown in Table 2.
Table 2. Summary measures and statistics from study 2. Statistical significance between measures in the reflective questioning activity (RQA) and baseline conditions is indicated in the first column according to independent samples 1-tailed Welch t tests. Means are given with the SE within each condition. For each measure, we set our hypotheses as follows: the null hypothesis (H0) was that the mean for the RQA condition would be less than, or equal to, the mean for the baseline condition; by contrast, the alternative hypothesis (Ha) was that the mean for the RQA condition would be greater than the mean for the baseline condition.

<table>
<thead>
<tr>
<th>Measure</th>
<th>RQA condition, mean (SD; SE)</th>
<th>Baseline condition, mean (SD; SE)</th>
<th>t test (df)</th>
<th>P value</th>
<th>Cohen d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived utility</td>
<td>1.2 (2.04; 0.2)</td>
<td>0.5 (2.11; 0.2)</td>
<td>2.82 (213)</td>
<td>.003</td>
<td>0.38</td>
</tr>
<tr>
<td>Perceived stress change</td>
<td>0.7 (2.04; 0.2)</td>
<td>−0.4 (1.05; 0.1)</td>
<td>4.46 (213)</td>
<td>&lt;.001</td>
<td>0.61</td>
</tr>
<tr>
<td>Completion time (min)</td>
<td>8.9 (8.16; 0.8)</td>
<td>1.6 (3.16; 0.3)</td>
<td>9.09 (213)</td>
<td>&lt;.001</td>
<td>1.27</td>
</tr>
<tr>
<td>Response length (words)</td>
<td>87 (118.297; 11.6)</td>
<td>29 (57.95; 5.5)</td>
<td>4.52 (213)</td>
<td>&lt;.001</td>
<td>0.63</td>
</tr>
<tr>
<td>Perceived time commitment</td>
<td>−0.2 (2.04; 0.2)</td>
<td>−0.3 (2.11; 0.2)</td>
<td>0.33 (213)</td>
<td>.37</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Hypothesis 1 (Perceived Benefits)

Participants in the RQA condition saw significantly more utility in completing the activity than those in the baseline condition \((t_{213}=2.82; P=.003; \text{Cohen } d=0.38)\). The average rating for our RQA was 1.2 (SE 0.2), whereas the average rating for the baseline activity was 0.5 (SE 0.2). Although both these averages were near the neutral score of 0, there were many more positive ratings for our RQA. Of the 215 participants who used our RQA, \(n (79\%)\) gave a nonneutral positive score, whereas only \(n (57\%)\) did the same for the single-question activity. Participants also reported a statistically significant change in stress rating in the RQA condition compared with the baseline condition \((t_{213}=4.46; P<.001; \text{Cohen } d=0.61)\). Whereas people who used our RQA experienced a mean decrease of 0.7 (SE 0.2) point in their perceived stress rating, people who used the single-question activity actually experienced a mean increase of 0.4 point. A paired 1-tailed \(t\) test analyzing scores before and after engaging with the RQA condition indicated a statistically significant decrease in stress scores post-RQA, relative to their levels before starting it \((t_{103}=3.59; P<.001; \text{Cohen } d=0.36)\).

These results suggest that the additional questions from our RQA may not only help in potentially mitigating stress but also possibly counteract an initial increase in stress from revisiting the troubling situation.

Hypothesis 2 (Elapsed Time)

Participants in the RQA condition took 8.9 (SE 0.8) minutes on average to complete the activity, whereas those in the baseline condition took only 1.6 (SE 0.3) minutes on average; the difference between the 2 conditions according to this measure was statistically significant \((t_{213}=9.09; P<.001; \text{Cohen } d=1.27)\). We also found that participants wrote significantly longer responses while going through our RQA. Participants in the RQA condition wrote 87 (SE 11.6) words on average, whereas those in the baseline condition wrote 29 (SE 5.5) words on average; this difference was also statistically significant \((t_{213}=4.52; P<.001; \text{Cohen } d=0.63)\). Although the RQA required significantly more effort, there was no statistically significant difference in people’s subjective perceived time commitment \((t_{213}=0.33; P=.37; \text{Cohen } d=0.05)\). We conclude from these results that people found value in the additional time they spent completing the series of questions.

Study 3: Observing Repeated Engagement With Our RQA

Overview

Figure 1 illustrates participants’ response rate to our RQA sent via email and SMS text message over the course of the study period. The figure not only shows the aggregated data across all interactions with our RQA but also splits the results according to the CMC platform through which the prompts were sent. We do not rely on quantitative data to claim that one way of delivering an RQA is better than the other; instead, we look into qualitative data to understand the role that technology plays in supporting long-term engagement with RQA.
**Overall Engagement**

We observed moderate engagement throughout the 2-week period of our study. We sent participants 54 prompts via email (n=27, 50%) and SMS text message (n=27, 50%), and participants completed the RQA in 27 (50%) of these cases. On 3 (11%) of these 27 occasions, participants completed the RQA twice in response to the same prompt; therefore, our RQA was actually completed 30 times during the study.

On average, people spent 18.5 (SE 1.2) minutes, wrote a total of 212 (SE 24.2) words, and experienced a stress level reduction of 1.2 (SE 0.3) points after completing the RQA.

Participants were much more engaged with our RQAs in this study compared with the AMT crowdworkers in study 2 (ie, they spent more time and typed longer responses), even as they completed it multiple times. One explanation for this discrepancy could be the amount of time participants were willing to commit to the study. Participants in study 2 likely completed our RQA in the midst of other crowdsourcing tasks or during their busy workdays. By contrast, participants from study 3 were able to pick a suitable time at their convenience, which in turn gave room for a longer time investment. A participant validated this hypothesis from their experiences:

> Although I initially said that I would be available in the morning, I found the best time to do it in the time between 9 and 11 PM. I used to see the emails and text messages shortly after they came, but I used to only do them at my convenient times in the night. [D3]

Our quantitative and qualitative data show that people could spend as little or as much time as they wanted with the activity without the need for significant scaffolding. In the interviews, participants expressed similar opinions about the benefits of our RQA as they did in our previous studies. Most notably, participants saw benefits to having a structured way of organizing thoughts because it helped them identify triggers and devise an alternate way of thinking.

**Repeated Engagement With the RQA**

A major goal of this study was to observe how participants engaged with the RQA over time. Unsurprisingly, we observed that the response rate decreased over time. Figure 1 shows that the response rate was 64% (7/11) when participants received their first prompt and then 55% (6/11) for the second prompt. By the time they had seen 6 prompts, the response rate went all the way down to 38% (3/8). When we asked participants to explain this trend during our interviews, the main complaint was that doing the same activity in such a short interval was boring and tedious. D3 mentioned that the length of the activity was acceptable for a onetime event, but when they had to do the activity thrice in the same week, it “came across as a chore.” Another participant expressed similar sentiments:

> When it started coming every other day, I felt like I had to do a school homework. So I felt a little bit of pressure to do the activity. [D10]

Participants expressed that they would have preferred to have larger intervals (eg, once a week) between the times they were requested to go through the RQA. This was not only because of the monotony of the task but also because participants struggled to think of new troubling situations to reflect upon:

> By the time I got the last prompt, I could not find a stressful situation in my life. Maybe the frequency should vary depending on the amount of stress a person is going through. [D6]

Participants acknowledged that regularly prompting them to do our activity had value. A few of them noted that they would have forgotten to revisit the RQA had they not been given periodic reminders. D1 also believed that they “got more
comfortable with the activity [over time] and started setting aside a time to do the activity.”

Repeated engagement with our RQA also helped people form habits that yielded benefits outside of the activity itself; for example, D10 informed us that they did the activity multiple times in their mind either to think about how their previous responses could be improved or how they could apply these questions in a new situation. D4 found that doing the activity multiple times was a good mental exercise to prepare themselves for less stressful situations that they may encounter during the day.

**CMC Platforms**

Another goal of this study was to gain insights into the role of technology in deploying RQAs. Figure 1 shows that there was a noticeable difference between the response rates for email versus SMS text message. Even in our exit survey, 8 (73%) of the 11 participants said that they would prefer email over SMS text message for doing this activity, whereas the rest of the participants (3/11, 27%) had no preference. One of the main reasons for the preference was the affordances of desktop and laptop computers when it came to completing the RQA. Most notably, participants commented on how computers are better suited for reading and writing longer passages of text:

> Typing is very difficult in mobile phones. The screen size is small and editing stuff is a nightmare. On the other hand, if you want to write a long answer, you would probably do that on the computer because the process is just much easier. [D1]

Participants also felt that doing the RQA on a computer minimizes the chance for distractions; for example, D7 commented that sitting in front of their computer gave them the “right mindset to do the activity.” With a computer, they felt that they had control over their workspace because they could easily close other tabs and applications. By contrast, when they tried to do the activity on their mobile phone, there were cases when a call or a push notification disrupted their train of thought.

Although email was generally preferred for completing the RQA, many people agreed that mobile phones are a great mechanism for sending notifications and reminders. Some of the participants (eg, D4 and D6) expressed the concern that people may not check their emails as frequently as they check SMS text messages:

> Most of the time, I have my phone in my hand, whereas I check my emails at most once or twice a day. So if you need me to do something immediately, you would probably need me to reach via text messages. I can respond to an email even 2 days later. [D4]

Participants also informed us of instances when they switched between the 2 CMC platforms. When D6 was prompted to do the RQA over SMS text message, they sent the link to themselves over social media and then accessed it on their desktop computer to complete the RQA. Some of the participants posited that the 2 CMC platforms could be integrated into the same system:

> What you can do is you can ask me to answer the questions in the text message, but at the same time you will also send me an email that has the links to the actual page. [D6]

Alternatively, others suggested that the RQA could be advertised over social media platforms such as Facebook or Instagram because people normally access their accounts across multiple devices. In doing so, people could have the option to choose whichever platform they see fit.

**Discussion**

**Principal Findings**

In this work, we aimed to understand the benefits of a brief digital intervention that people could complete on their mobile phone or computer to lessen their concerns about a troubling situation. Our second study showed that doing the RQA could be more effective in reducing instantaneous stress compared with simply reflecting on a troubling situation without structured questions, whereas our first and third studies elicited qualitative findings that we hope will inform the design of future interventions in this space. Most notably, we found that participants appreciated the RQA for its ability to help them undergo a structured analysis of their troubling situation, identify solutions to improve their situation, and vent their negative feelings. Although participants felt that the series of questions was worth the additional time commitment, we also saw some obstacles toward long-term engagement with the RQA: the monotony of doing the same activity several times, the limited affordances of mobile phones, and the importance of having the prompts align with the occurrence of new troubling situations.

Our findings indicate that people from the general population saw value in engaging with a simple lightweight reflection activity without an active conversational partner. Although there has been significant research effort toward making mental health platforms more sophisticated and humanlike [37,49], our work shows that simpler interfaces can also yield benefits. Across all our studies, participants expressed that the structured nature of the RQA played a pivotal role in making them more aware of their troubling emotions. By deconstructing past events, participants were able to view their feelings in an organized manner and from a third-person perspective, enabling them to reevaluate whether their feelings were justified. The writing activity acted as a medium through which they could externalize repressed emotions, a helpful practice that has been noted by past psychology research [50]. People often falsely assume that their problems are a reflection of their own identity or their relationship with others. Failing to separate problems from persons can cause people to identify themselves as different from what society considers normal, eventually leading them to fixate on their negative traits [51]. Our RQA provided people with the opportunity to explore the relationship between their problem and their own self but from a different perspective.

The RQA also offered a general structure that people could adapt to their own life situations. We saw that a few of the participants applied the same line of questioning outside of the activity itself, hinting at longer-lasting benefits. We foresee that

---

https://formative.jmir.org/2024/1/e47360 JMIR Form Res 2024 | vol. 8 | e47360 (page number not for citation purposes)
RQAs could serve as a gateway for people struggling with stress and depression to engage with more complex activities and therapeutic tools. Validating this potential would require a longer study, but our research already demonstrates the hurdles that RQA interventions must overcome to support long-term engagement.

**Improving the Design of RQAs**

The success of the RQA in our work does not mean that future RQAs could not be even better. Although we observed an average decrease in participants’ stress levels after completing the RQA, some of the participants from study 1 remarked that the activity left them confused and without a concrete solution. We hypothesize that such concerns could be remedied by providing users with sample responses to each question as a source of inspiration. These examples could be curated by researchers, or they could be collected from previous users who voluntarily contributed their responses to a database [52]. Topic modeling could be used to tag the examples with keywords related to their subject matter, and an information retrieval system could rank the relevance of these examples [53]. Collaborative filtering could even be used to gradually collect ratings for each example and then tailor examples to individuals’ preferences [54].

Another way that RQAs could be made better is by personalizing the questions. The activity could ask users to rate the perceived benefit of each question, or we could use the average response length as a proxy for estimating the utility of each question. Using this information, we could extend or emphasize questions that individuals find most beneficial. We could also use this information to remove questions that induce stress. However, thought records and behavioral chaining are intentionally designed processes with many critical steps; therefore, removing questions may detract from the activity’s benefits.

Our 9-question RQA took inspiration from CBT principles, but future work could investigate RQA designs based on other psychological principles; for example, encouraging expressions of gratitude or social connections with others can play a key role in stress and depression management [55,56], and RQAs built around these practices can similarly help people manage their well-being. Future work could also explore different activity structures. Many of the participants (8/11, 73%) in study 3 complained about the inconvenience of typing on their smartphones; therefore, an alternative activity could ask people to record and listen to their own voices for reflection. Another activity could encourage peer support by starting conversations among online peers. Finally, researchers could create brief activities centered around other psychological frameworks beyond CBT, with past examples being centered around mindfulness [57], motivational interviews [58], and acceptance and commitment therapy [59].

**Considerations for Long-Term Engagement**

Our 2-week deployment in study 3 enabled us to gain insights into how people would engage with RQAs over a period of time. Although participants were pleased with the fact that they could specify their hours of availability, receiving prompts for the RQA 3 times within the same week was overwhelming for most of them (3/8, 38%). The biggest criticism was that people received multiple prompts without experiencing a new troubling event; therefore, they either had to go through our RQA while analyzing the same event as before or recalling a troubling event from the distant past. Ideally, the frequency of prompts would adapt dynamically according to a person’s needs. A participant suggested that users should have control over how often they receive reminders to complete our RQA, explaining that individuals who experience more stress than others might benefit more from doing these activities in short intervals. Going a step further, future work could integrate physical activity trackers, smartphone sensors, and Internet of Things devices to automatically detect periods of heightened stress [60,61], turning our RQA into a just-in-time adaptive intervention.

Another issue with completing our RQA too often was that answering the same set of questions became boring and tedious; yet, adjusting the prompt frequency alone may not be enough to resolve these concerns. One way to add variety would be to mix an RQA with other microinterventions, as was done by Paredes et al [13] in their PopTherapy work. Brief interventions such as our RQA could also serve as a gateway to more time-consuming exercises or professional therapy. By giving people a preview of the potential improvement in the mood that they can receive from articulating their thoughts and emotions, habits can be formed, and users may become more motivated to build on this momentum [62].

**Limitations**

Rather than developing a mental health intervention for people experiencing clinical depression or other psychological disorders, our intention was to design our RQA for as broad a population as possible. It would be imperative for researchers to conduct further studies specifically with individuals with mental health disorders to understand the benefits and potential risks of digitally delivered RQAs. We suspect that self-reflection could not only serve as a convenient mechanism for people to practice what they learn in psychotherapy but also perpetuate negative thought patterns. We also recognize that our participant cohorts—AMT crowdworkers and university students—do not represent all aspects of the general public. Most of our qualitative findings were not tied to participants’ specific contexts, and we did not find any obvious evidence of substantial differences among the cohorts. Nevertheless, future work could deploy RQAs to more diverse populations.

**Conclusions**

In this work, we used CBT principles to design a brief RQA that helps people articulate, reflect on, and change their thoughts and emotions about a troubling situation. The 3 studies we presented in our paper provide evidence that people are willing to engage with, and find value in, brief self-reflection activities delivered through CMC platforms, even without scaffolding such as training or real-time feedback. We found that providing people with a brief online activity not only helped them reduce their perceived stress levels related to a self-selected situation but also helped them challenge their potentially negative thought patterns and identify alternative ways of thinking. We also found that people were willing to use the RQA more than once, although future work is needed to strike a balance between...
utility and monotony. We hope that our work inspires other researchers to explore new formats for brief interventions that help people with their everyday struggles.

Acknowledgments

The authors extend their gratitude to Dr Norman Farb (University of Toronto) for his valuable input. This work was supported by grants from the Office of Naval Research (N00014-18-1-2755 and N00014-21-1-2576), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2019-06968), and the National Science Foundation (2209819), as well as a gift from the Microsoft AI for Accessibility program. We would also like to thank Linjia Zhou (University of Toronto) and Jai Aggarwal (University of Toronto).

Data Availability

Anonymized data sets will be made available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References


**Abbreviations**

AMT: Amazon Mechanical Turk  
CBT: cognitive behavioral therapy  
CMC: computer-mediated communication  
RQA: reflective questioning activity
Exploring User Perspectives on Brief Reflective Questioning Activities for Stress Management: Mixed Methods Study

Bhattacharjee A, Chen P, Mandal A, Hsu A, O'Leary K, Mariakakis A, Williams JJ

URL: https://formative.jmir.org/2024/1/e47360
doi:10.2196/47360
PMID:38329860

©Ananya Bhattacharjee, Pan Chen, Abhijoy Mandal, Anne Hsu, Katie O'Leary, Alex Mariakakis, Joseph Jay Williams. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Investigation of the Impact of Wellinks on the Quality of Life and Clinical Outcomes in Patients With Chronic Obstructive Pulmonary Disease: Interventional Research Study

Kerri A Pierz1, PhD; Nicholas Locantore2, PhD; Gretchen McCreary2, MA; Robert J Calvey1, MBA; Nickole Hackney1, BS, RRT; Pooja Doshi1, MPH, MBA; John Linnell1, BA; Abirammy Sundaramoorthy1, MD; Carol R Reed1, MD; Julie Yates2, BS

1Wellinks (Convexity Scientific, Inc), New Haven, CT, United States
2COPD Foundation, Miami, FL, United States

Abstract

Background: Wellinks is a remote disease management solution that provides novel chronic obstructive pulmonary disease (COPD) care delivery.

Objective: This study evaluated the satisfaction, engagement, and clinical outcomes of Wellinks participants. This study also investigated the cadence of health coaching for patients with COPD.

Methods: A 24-week interventional study was conducted by Wellinks and the COPD Foundation in 2022. Adults with COPD were recruited by the COPD Foundation in the United States and determined to be eligible if they had phone and internet access, owned a smartphone, and were not currently participating in pulmonary rehabilitation. All study participants provided written informed consent. The Wellinks solution included remote health coaching, pulmonary rehabilitation, and group education; participants were provided the Wellinks app and smart spirometry and pulse oximetry devices. Participants were offered 6 coaching sessions in the first 12 weeks. For the second 12-week period, participants either reduced frequency or discontinued coaching; all other components of the Wellinks solution remained unchanged. The COPD Self-Efficacy Scale, Modified Medical Research Council dyspnea scale, pulmonary function, pulse oximetry, and patient-reported healthcare resource utilization were the clinical outcome measures. Nonclinical outcomes included engagement and satisfaction with Wellinks and net promoter score.

Results: In total, 141 adults consented and completed Wellinks onboarding; 84.4% (n=119) of whom remained engaged throughout the 24-week study. Participants had a mean age of 70 (SD 7.8; range 48-88) years, and 55.7% (n=78) were female. Most participants (n=119, 84.4%) completed all 6 coaching sessions during the first 12-week period. Compliance with spirometer and pulse oximeter use was 82.3% and 89.4%, respectively, at week 1 but waned over the study period to 8.5% and 9.2%, respectively, at the end of the study. Participants indicated a high degree of satisfaction with Wellinks, with 95.5% (n=85) and 91% (n=81) of participants indicating that they agreed or strongly agreed that the educational content and health coaching, respectively, were valuable. At the end of the study, the net promoter score was +64 and +55 in the coaching continuation and discontinuation arms, respectively. A significant improvement from baseline to end of the study was observed in the COPD Self-Efficacy Scale total score (P<.001) and domain scores (P<.001 for each domain). In total, 35.1% (n=27) of participants improved by at least 1 category of change on the 5-point Modified Medical Research Council dyspnea scale from baseline to week 24.

Conclusions: This study confirmed the feasibility of using a remote model of care delivery to support people living with COPD. The insights gained in this study have allowed for further refinement and personalization of the Wellinks care model. Findings related to the combined use of technology and personal care delivery should be considered by others developing remote disease management tools.
Introduction

Chronic medical conditions are highly prevalent among US adults and require long-term management strategies that invoke the need for participatory medicine. Strategies and services that support patient self-management are capable of reducing the impact of chronic disease on the individual and on the health care system [1]. Self-management strategies should optimize and preserve health, reduce symptoms and the impact of disease on daily life, improve quality of life, and build patient and provider relationships [1].

Chronic disease management programs have evolved over time to deliver remote health care and are generally inclusive of both technological (eg, wearables and mobile apps) and personal components, such as coaching or counseling [2]. While society is immersed in the Internet of Things, remote health care delivery needs to be more than biometric monitoring alone and to be effectively integrated with the delivery of care [3]. Although available devices can track mobility, heart rate, oxygen levels, blood pressure, cardiac activity, and body temperature and detect posture and falls (and more), the use of the data is limited if it is not integrated, shared, and applied to the delivery of care. Peyroteo et al [3] cite more than 100,000 apps that have been created to use data from various biometric sensors but also note the lack of integration with care systems as a key limitation to maximizing health outcomes.

Chronic obstructive pulmonary disease (COPD) is a chronic medical condition of the lungs affecting more than 16 million adults in the United States [4]. It is among the top causes of disability worldwide and is projected to become the leading cause of death by disease by 2030 [5]. Annual US health care expenditures on COPD exceed US $49 billion, with employer, federal, and state spending on health care services reaching unsustainable levels [6]. Costs aside, the toll on those struggling with respiratory diseases has been widely reported to lead to the significant presence of comorbidities such as cardiac diseases, diabetes, hypertension, osteoporosis, and mental health disorders [7].

Wellinks goes beyond remote patient monitoring. Wellinks is a COPD disease management solution that pairs technology with personalized health coaching and respiratory therapy services to offer a novel approach to COPD remote disease management. Wellinks is a care partner delivering remotely accessible pulmonary rehabilitation, clinical coaching, a mobile app, and connected devices for home monitoring of pulmonary function.

In a previously published 8-week pilot study of Wellinks, it was demonstrated that patients with COPD with an average age of 79.6 years were able to successfully use the devices provided (ie, Flvip nebulizer [Convexity Scientific, Inc], Smart One spirometer [Medical International Research], and NoninConnect smart pulse oximeter 3230) [Nonin Medical, Inc]) as well as enter data into the Wellinks mobile app [8]. Study participants reported the app to be valuable (13/16, 81%) and easy to use (15/16, 94%). This feasibility study provided preliminary evidence for the willingness and capability of this patient population to successfully use the digital tools provided by Wellinks [8].

Since the original pilot study, the Wellinks solution has expanded to include respiratory therapy and health coaching services, in addition to some modifications of the technological components described earlier. With such iteration, not only was it important to replicate the previously reported feasibility results but also to explore clinical outcome measures and refine the duration and frequency of health coaching. Described herein is the ASPIRE study conducted in partnership with the COPD Foundation that was designed to explore clinical and nonclinical outcomes associated with the use of the updated Wellinks solution inclusive of both personal (health coaching and respiratory therapy) and technological components. The objectives of this study were to determine to what degree study participants would engage with the various components of the Wellinks solution over time and whether any clinical outcomes could be identified to correlate with engagement. This study also sought to collect qualitative feedback on the components of the Wellinks solution and observe any impact of decreased frequency of engagement with the Wellinks team, in service of refining the care delivery model offered by Wellinks.

Methods

Study Design

This 24-week, prospective, interventional research study of the Wellinks COPD solution including the use of Bluetooth-connected devices, patient mobile app, COPD-related education, and health coaching services was conducted from December 2021 through September 2022. This study was designed to gather data on the quality of life and clinical impact of the use of the Wellinks COPD solution, in addition to collecting feedback from patients and investigators to inform further optimization of this intervention. The study was posted to ClinicalTrials.gov (NCT05259280).

Ethical Considerations

The conduct and performance of this study were in accordance with applicable sponsor and investigator responsibilities as described in Title 21 Code of Federal Regulations 812 and other Good Clinical Practice guidance. Institutional review board (IRB) or ethics committee oversight was required as human participants or data from humans were used. IRB approval of
the study protocol and study-related materials was obtained from Western IRB prior to beginning any study-related procedures (IRB protocol 20141136).

Recruitment

Eligible participants were recruited through the COPD Foundation Patient-Powered Research Network, COPD360Social, and various newsletters. Eligible patients were invited to participate in the study. Participants were required to give informed consent before study-specific procedures could proceed. Eligible study participants included adults (≥18 years of age) with a diagnosis of COPD. Participants had to have access to a home telephone (landline or mobile) and the internet and must have had a smartphone (ie, iPhone 6S or later model, running iOS 14.0 or later model, and Android 6 or later model). Individuals must have been proficient in the English language, living or staying in the United States throughout the study duration, willing and able to comply with study requirements, and able to provide written informed consent. Exclusion criteria included current participation in other interventional clinical trials and current participation in a pulmonary rehabilitation program.

Intervention

The Wellinks COPD solution combined personal and technological elements to remotely support enrolled participants living with COPD. The personal elements of the program included one-on-one access to health and wellness coaches or nurse practitioners trained in health coaching methodologies. Health coaches provided support to participants via phone, video, and text interactions throughout the study period. The role of the coach was to support the participant by providing disease-state and treatment-related education, establishing and supporting the attainment of individual health goals, and encouraging adherence to treatment and attendance at clinic visits. Remote pulmonary rehabilitation programs were provided by the health coach, personalized for each participant, and included individual home-based exercise guides or videos and group educational sessions led by Wellinks respiratory therapists that were held remotely.

The technological elements of the program included a mobile app and Bluetooth-enabled medical devices. The Wellinks mobile app downloaded to an iOS (iPhone) or Android device allowed participants to connect with their coach; track goals, medications, and symptoms; and review data from the connected devices provided. The Smart One personal spirometer was used by participants at home to collect peak expiratory flow (PEF) and forced expiratory volume in 1 second (FEV₁). The NoninConnect 3230 pulse oximeter was used at home to measure blood oxygenation (saturation of peripheral oxygen [SpO₂]) and pulse rate. Data from both devices were transmitted via Bluetooth to the participant’s smartphone. Technical support was available to all study participants throughout the duration of the study to answer questions about the technological components or to troubleshoot any issues.

Study Procedures

All baseline assessments were collected via a survey of all consenting participants followed by an onboarding call between a Wellinks coach and the participant to ensure the technical set-up of the app and devices and introduce the coaching process. Participants were instructed to use the connected devices (pulse oximeter and spirometer) at least once a week throughout the duration of the study, use the app to track symptoms and medications, and monitor their own spirometry and pulse oximetry data throughout the study. The frequency of use of each component of Wellinks was recommended to each participant, but in order to best emulate real-world use, the health coaches encouraged but did not mandate the use of all available components.

Participants were sequentially assigned at the time of enrollment in an alternating fashion to arm 1 or arm 2 by the Wellinks head health coach. Participants were not informed of this assignment until the completion of the first 12-week period. In the first 12-week study period, health coaching was offered to all participants in the form of one-on-one 30-minute remote sessions scheduled every other week for a total of 6 sessions over the 12-week period. In addition, participants were instructed to individually perform the remote pulmonary rehabilitation program as directed by their health coach, and they were invited to attend group educational sessions held weekly throughout the study period. In the second 12-week study period, all components of the program remained the same except for the level of personal contact with health coaches. Participants assigned to arm 1 continued with a lower level of engagement with their coach in the form of SMS text messaging or up to 3 brief check-in meetings (15-minute sessions) for an additional 12 weeks. Participants who were assigned to arm 2 discontinued access to the Wellinks health coaches for the second 12 weeks of the study.

Outcomes

The nonclinical objectives of this study were to describe the experience of patients using the Wellinks solution through the assessment of patient engagement as well as by patient-reported satisfaction. Outcome measures included compliance with protocol-recommended device use, compliance with attendance at scheduled coaching sessions, ratings of the degree to which participants valued individual components of the Wellinks COPD solution, and net promoter score (NPS; ie, “How likely is it that you would recommend Wellinks to a friend or colleague?” 0=not at all likely to 10=extremely likely). Spirometer and pulse oximeter data could be synced with the Wellinks app; as such, the use of data from the app provided the necessary data to determine whether participants used these at-home devices. However, the spirometer results could only be viewed by the participants via the app, while the pulse oximeter could be viewed independently of the app. Therefore, the compliance with the pulse oximeter uniquely may be underestimated.

The clinical objectives of this study were to determine whether the use of the Wellinks COPD solution could improve the quality of life for patients with COPD, reduce healthcare resource utilization (HRU) over time, and improve pulmonary function as measured by connected devices. Quality of life was indirectly ascertained by the interpretation of results from the COPD Self-Efficacy Scale (CSES) and Modified Medical
Research Council (mMRC) dyspnea scale, based on known correlations reported in the published literature [9]. Pulmonary function was measured using at-home devices to collect FEV1, PEF, and SpO2. Patient-reported HRU was collected via survey.

The CSES is used to assess the confidence of a patient related to their ability to avoid breathing difficulty based on responses to 34 questions within 5 domains; each question is scored from 1 (not at all confident) to 5 (very confident) [10]. A higher score thus reflects a greater degree of confidence on the part of the respondent. Total scores can range from 34 to 170. The CSES is divided into 5 domains: negative affect, intense emotional arousal, physical exertion, weather or environmental factors, and behavioral risk factors [10].

The mMRC dyspnea scale provides an assessment of a patient’s shortness of breath and its impact on daily activities. At onboarding, data from the mMRC dyspnea scale were combined with patient-reported exercise habits to individualize the remote pulmonary rehabilitation program to be suitable to each study participant’s level of functioning. Participants were assigned to 1 of 6 different exercise programs based on mMRC score (low=0, 1, or 2 or high=3 or 4) and self-reported level of exercise (low, medium, or high) at baseline. mMRC was also assessed at week 12 and week 24 to explore changes over time.

Pulmonary function was measured as a change from baseline to week 12 and week 24 in FEV1, PEF, and SpO2 based on patient use of the Bluetooth-connected spirometer and pulse oximeter provided. When used and connected, these data were captured in the Wellinks app. At the start of the study, participants were asked to use the pulse oximeter and spirometer at least weekly throughout the duration of the study.

Patient-reported HRU was collected through a web-based survey and relied upon individual recall. HRU is a reflection of the patient’s desire or need to seek care and is a measure that can be used to inform the economic impact of an intervention. At baseline, participants were asked to report certain HRU (ie, COPD-related physician visits, emergency department visits, and hospital admissions) in the 3-month and 1-year periods prior to enrollment. Participants were asked the same HRU questions at week 12 and week 24 of the study, each with a 3-month recall period. Outcome measures were assessed at baseline, week 12, and week 24 of the study. Any adverse events or serious adverse events were collected via spontaneous reporting from the study participants.

Statistical Analyses

The planned sample size for this study (n=150) was based on the expected feasibility for recruitment. No formal statistical power calculations were performed to size this study.

Study data were summarized for arm 1, arm 2, and full study cohort. Unless otherwise specified, data were summarized as number and percentage for categorical variables and as mean and SD for continuous variables. All statistical analyses were exploratory in nature. P values for statistical tests are 2-sided tests and not adjusted for multiplicity. Analyses of change from baseline values at week 12 and week 24 were performed for each arm and the full study cohort using 2-tailed t tests. Least squares (LS) mean and LS mean change from baseline at each time point with corresponding SEs for change and P values were produced.

For mMRC, a responder was defined as a participant with an improvement from baseline of 1 category or more. For example, a participant who changes from “3: I have to stop for breath after walking for ~100 yards” at baseline to a postbaseline value of “2: I walk slower than others of my age because I am out of breath, or I have to stop often to catch my breath” would be classified as a responder at that time point. Participants who either remain in the same category or worsen were classified as nonresponders.

Results

Demographics and Baseline Characteristics

A total of 153 individuals were consented in this study, of whom 141 were fully enrolled (ie, consented and completed onboarding of devices and app). Disposition of participants in the study is described in Figure 1.

The demographics of the study population are presented in Table 1. Study participants had a mean age of 70 (SD 7.8; range 48-88) years, 78% (n=110) were 65 years of age or older, and 55.3% (n=78) were female. The population was 90.8% (n=128) White and 97.9% (n=138) non-Hispanic or Latino. There were no statistically significant differences between the 2 treatment arms for any of the demographic variables.

In the study population, 83.7% (n=118) of participants were former smokers (with 77.1% [n=91] of these having quit more than 5 years ago), and 9.2% (n=13) were current smokers with a mean use of 12 (SD 6.1) cigarettes per day at the time of the study.

It was self-reported that 82.3% (n=116) of the population was under the care of a pulmonologist for their COPD, and 39.7% (n=56) reported a primary care physician participating in the management of their COPD alone or together with the pulmonologist. A majority (74.5%, n=105) of participants had been living with COPD for at least 5 years at the time they were enrolled in this study. The severity of disease was self-reported to be moderate (51.1%, n=72) or severe (39.7%, n=56), and 58.2% (n=82) lacked an exercise plan at the study start. Some degree of home oxygen use was reported by 61% (n=86) of study participants (45.4% [n=64] daily use and 15.6% [n=22] as-needed use).

More than half of the study population self-reported emphysema (71.6%, n=101) or bronchitis (53.9%, n=76). High blood pressure (56.7%, n=80) and anxiety (46.1%, n=65) were among the most common nonrespiratory medical conditions reported by the study population.

https://formative.jmir.org/2024/1/e47555
**Figure 1.** Study flow diagram. Flow of participants through the study protocol is described as inclusive of the number of individuals consented to participate (N=153) and enrolled (n=141), followed by completion of certain milestones throughout the 24-week study period. Reasons for withdrawals from the study are reported.

**Table 1.** Summary of participant demographics. Descriptive information about the study cohort at baseline is presented for the combined cohort and for individuals assigned to arm 1 and arm 2 separately.

<table>
<thead>
<tr>
<th>Parameter and statistic or variable</th>
<th>Combined (n=141)</th>
<th>Arm 1 (n=73)</th>
<th>Arm 2 (n=68)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>70 (7.6)</td>
<td>70 (7.8)</td>
<td>70 (7.5)</td>
</tr>
<tr>
<td>Range</td>
<td>48-88</td>
<td>48-87</td>
<td>40-60</td>
</tr>
<tr>
<td>65 years or older, n (%)</td>
<td>140 (86.3)</td>
<td>68 (93.2)</td>
<td>72 (80.9)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>38 (55.3)</td>
<td>40 (52.1)</td>
<td>40 (58.8)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>5 (3.5)</td>
<td>3 (4.1)</td>
<td>2 (2.9)</td>
</tr>
<tr>
<td>White</td>
<td>128 (90.8)</td>
<td>65 (89)</td>
<td>63 (92.6)</td>
</tr>
<tr>
<td>Others</td>
<td>3 (2.1)</td>
<td>1 (1.4)</td>
<td>2 (2.9)</td>
</tr>
<tr>
<td>Unknown or declined</td>
<td>5 (3.5)</td>
<td>4 (5.5)</td>
<td>1 (1.5)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>1 (0.7)</td>
<td>0 (0)</td>
<td>1 (1.5)</td>
</tr>
<tr>
<td>Non-Hispanic or Latino</td>
<td>138 (97.9)</td>
<td>73 (100)</td>
<td>65 (95.6)</td>
</tr>
<tr>
<td>Unknown or declined</td>
<td>2 (1.4)</td>
<td>0 (0)</td>
<td>2 (2.9)</td>
</tr>
<tr>
<td>Height (inches), mean (SD)</td>
<td>67 (4.1)</td>
<td>67 (4.0)</td>
<td>66 (4.2)</td>
</tr>
<tr>
<td>Weight (lb), mean (SD)</td>
<td>180 (46.9)</td>
<td>184 (46.5)</td>
<td>175 (47.1)</td>
</tr>
<tr>
<td>Smoking status at baseline, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>13 (9.2)</td>
<td>8 (10.6)</td>
<td>5 (7.4)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>118 (83.7)</td>
<td>58 (79.5)</td>
<td>60 (88.2)</td>
</tr>
<tr>
<td>Never smoked</td>
<td>10 (7.1)</td>
<td>7 (9.6)</td>
<td>3 (4.4)</td>
</tr>
</tbody>
</table>

**Nonclinical Outcomes**

**Engagement Metrics**

In total, 84.4% (n=119) of all participants completed all 6 coaching sessions in the first 12-week period of the study. Among participants assigned to arm 1 (continued coaching), attendance diminished session-to-session in the second 12-week period of the study, with only 52.1% (n=38) of those assigned to arm 1 completing the third (final) 15-minute coaching session in the second study period.
Participants were advised to use the Bluetooth-connected spirometer and pulse oximeter at least weekly throughout the 24-week duration of the study. Spirometer compliance peaked at the start of the study with 82.3% (n=116) of participants compliant during week 1, but compliance decreased to a smaller proportion of participants at week 12 (n=59, 41.8%) and week 24 (n=12, 8.5%). Similarly, pulse oximeter compliance also peaked at week 1 with 89.4% (n=126) of participants using the pulse oximeter as recommended, and this rate of compliance decreased to 42.6% (n=60) and 9.2% (n=13) at week 12 and week 24, respectively. Compliance with the spirometer or the pulse oximeter use did not differ by treatment arm. For the entire study period, 21.3% (n=30) and 22.6% (n=32) of participants were compliant with spirometer and pulse oximeter use, respectively, for more than 75% of the study period (18 or more of 24 weeks), while 30.5% (n=43) and 29.1% (n=41) were compliant with spirometer and pulse oximeter use, respectively, for 25% or less of the study period (6 or less weeks).

Similar rates of compliance were observed with the use of the Wellinks app; compliance with mobile app use peaked at the start of the study with 94.3% (n=133) compliance in week 1, which declined to 50.4% (n=71) and 22.7% (n=32) at week 12 and week 24, respectively. For the entire study period, 23.4% (n=33) were compliant with app usage for 25% or less of the study period, and 28.4% (n=40) of participants were compliant with app usage for 75% or more of the study period.

One-quarter of study participants attended multiple educational webinar group sessions (3 or more sessions attended). One-half of study participants did not attend any of the educational webinar group sessions.

The web-based week 12 survey was sent electronically to participants after completion of coaching session 6 and was returned by 78.7% (n=111) of participants. The web-based week 24 survey was sent electronically to participants after 24 weeks had elapsed since the start of the study; 73.9% (n=54) of participants in arm 1 and 51.5% (n=35) of participants in arm 2 completed the week 24 survey. The differences in survey completion rates between the 2 treatment arms may be attributable to the difference in level of engagement with Wellinks; specifically, it is possible that arm 2 participants who had less engagement with the Wellinks team between week 12 and week 24 had less interest or motivation in returning the survey.

**Satisfaction Metrics**

Participants were asked via survey at week 12 and week 24 to indicate their level of agreement with various statements aimed to understand whether they valued individual components of the Wellinks solution. Options included strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree. Table 2 presents the proportion of participants who indicated they strongly agreed or agreed with each statement at the end of the study (week 24). Data are shown for the combined cohort not separated by treatment arm due to similar findings across the arms. Only 1 statement appeared to reflect a difference by treatment arm at the end of the study: 92.6% (n=50) of respondents in arm 1 and 68.6% (n=24) of respondents in arm 2 strongly agreed or agreed that “using the Wellinks solution has helped me to learn more about my COPD.” This difference may reflect the higher engagement with health coaches in arm 1 throughout the second half of the study period. There was a low level of disagreement with any of these statements indicating that most study participants find value in the components of the Wellinks solution and the entirety of the offering, regardless of group assignment.

**Table 2.** Participant agreement with statements about intervention components.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Participants who agreed or strongly agreed, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think having access to educational content is valuable.</td>
<td>85 (95)</td>
</tr>
<tr>
<td>Overall, I found the Wellinks solution to be valuable.</td>
<td>84 (94)</td>
</tr>
<tr>
<td>I think having meetings with my health coach is valuable.</td>
<td>81 (91)</td>
</tr>
<tr>
<td>I think being able to message my health coach is valuable.</td>
<td>81 (91)</td>
</tr>
<tr>
<td>I think being able to take and log pulse oximeter measurements at home is valuable.</td>
<td>80 (90)</td>
</tr>
<tr>
<td>I found the Wellinks app easy to use.</td>
<td>79 (89)</td>
</tr>
<tr>
<td>I think being able to take and log spirometry measurements at home is valuable.</td>
<td>75 (84)</td>
</tr>
<tr>
<td>Using the Wellinks solution has helped me to learn more about my COPD.</td>
<td>74 (83)</td>
</tr>
<tr>
<td>Using the Wellinks solution has helped me to manage my COPD better.</td>
<td>72 (81)</td>
</tr>
<tr>
<td>I think being able to track and log my symptoms in the app is valuable.</td>
<td>66 (74)</td>
</tr>
<tr>
<td>I think having my medication schedule in the app is valuable.</td>
<td>55 (62)</td>
</tr>
<tr>
<td>Using the Wellinks solution has helped me to take my COPD medication as needed.</td>
<td>39 (44)</td>
</tr>
</tbody>
</table>

*aSurvey responses are from 89 participants who completed these questions in the end-of-study survey at week 24 (n=54 from arm 1 and n=35 from arm 2). The proportion of participants who selected that they “agreed” or “strongly agreed” with each statement is shown. Statements are listed in rank order from the statement with the highest degree of agreement to the lowest.

bCOPD: chronic obstructive pulmonary disease.
Net Promoter Score

Participants were asked after week 12 and week 24 whether they would recommend Wellinks to friends, family members, or associates who also live with COPD to determine the NPS. Overall, the week 12 NPS was +57, and the week 24 NPS was +60. NPS differed by assigned treatment arm; NPS for arm 1 and arm 2 was +64 and +55, respectively, at week 24.

Clinical Outcomes

COPD Self-Efficacy Scale

CSES scores were collected at baseline, week 12, and week 24 through a web-based survey. At baseline, the mean total score was 103.9 (SD 28.71), with the lowest domain scores on average observed for physical exertion and weather or environmental factors.

The CSES total score significantly improved from baseline to the end of the first 12-week study period during which all participants received the same level of coaching (LS mean change from baseline 11.1, SE 3.10; P<.001; n=96). These improvements were sustained across the entire study cohort at week 24 (LS mean change from baseline 23.6, SE 4.81; P<.001; n=77).

After week 12, participants were split by assignment to arm 1 or arm 2. Significant improvements in total CSES score from week 12 to week 24 were also observed in arm 1 (LS mean change 8.6, SE 4.04; P=.04; n=38) and arm 2 (LS mean change 10.6, SE 4.33; P=.02; n=34). In total, 5 participants did not complete CSES at week 12 but did complete the CSES at week 24.

Scores in all domains were significantly improved from baseline to end of the study in both arms (P<.001 for all domain comparisons except arm 2 for negative affect [P=.006] and intense emotional arousal [P=.002]). The minimally clinically important difference for CSES has not been found in the literature. The greatest differences were observed in the physical exertion and behavioral risk factors domain as shown in Table 3.

Table 3. COPD\textsuperscript{a} Self-Efficacy Scale domain scores.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Week 12 (n=96)</th>
<th>Week 24 (n=77)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score\textsuperscript{b}</td>
<td>LS\textsuperscript{c} mean change (SE)</td>
</tr>
<tr>
<td>Negative affect</td>
<td>3.6</td>
<td>0.3 (0.10)</td>
</tr>
<tr>
<td>Intense emotional arousal</td>
<td>3.6</td>
<td>0.2 (0.09)</td>
</tr>
<tr>
<td>Physical exertion</td>
<td>3.1</td>
<td>0.5 (0.11)</td>
</tr>
<tr>
<td>Weather or environmental factors</td>
<td>3.2</td>
<td>0.4 (0.10)</td>
</tr>
<tr>
<td>Behavioral risk factors</td>
<td>3.4</td>
<td>0.4 (0.11)</td>
</tr>
</tbody>
</table>

\( a \)COPD: chronic obstructive pulmonary disease.

\( b \)Domain scores have a scale of 1-5, calculated as the mean rating of each domain question.

\( c \)LS: least squares.

mMRC Dyspnea Scale

The baseline mMRC scores reflect variability in the study population regarding dyspnea. mMRC scores range from 0 (“I get out of breath only when I engage in strenuous exercise”) to 4 (“I am often too out of breath to leave the house, or I get out of breath even when dressing”). The population mean score was 2.0 (SD 1.26) at baseline. The distribution of baseline scores and the mMRC response rates at week 12 and week 24 can be observed in Table 4. In total, 31.6% (n=30) of participants improved by at least 1 category on mMRC from baseline to week 12, and 35.1% (n=27) improved from baseline to week 24. No differences between treatment arms were observed from week 12 to week 24 (\( P=.77 \), chi-square test).
Table 4. mMRC responder analysis. Baseline mMRC scores are shown for the study population (n=141).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mMRC score at baseline (n=141)</td>
<td></td>
</tr>
<tr>
<td>0: I get out of breath only when I engage in strenuous exercise.</td>
<td>13 (9.2)</td>
</tr>
<tr>
<td>1: I get out of breath when I am hurrying or walking up a slight hill.</td>
<td>47 (33.3)</td>
</tr>
<tr>
<td>2: I walk slower than others of my age because I am out of breath, or I have to stop often to catch my breath.</td>
<td>38 (26.9)</td>
</tr>
<tr>
<td>3: I have to stop for breath after walking for ~100 yards.</td>
<td>16 (11.3)</td>
</tr>
<tr>
<td>4: I am often too out of breath to leave the house, or I get out of breath even when I am getting dressed.</td>
<td>27 (19.1)</td>
</tr>
</tbody>
</table>

Week 12 responder status\(^b\) (n=95)

- Improved\(^c\) 30 (31.6)
- No change 53 (55.8)
- Worsened\(^d\) 12 (12.6)

Week 24 (end of study) responder status\(^e\) (n=77)

- Improved\(^c\) 27 (35.1)
- No change 36 (46.8)
- Worsened\(^d\) 14 (18.2)

\(^a\)mMRC: Modified Medical Research Council.

\(^b\)Week 12 responder status is reported for 95 participants for whom mMRC data were available at baseline and week 12.

\(^c\)Participants were indicated to have “improved” if their score decreased by one or more points from baseline.

\(^d\)Participants were indicated to have “worsened” if their score increased by one or more points from baseline.

\(^e\)Week 24 responder status is reported for 77 participants for whom mMRC data were available at baseline and week 24.

**Pulmonary Function and Pulse Oximetry**

Interpretation of the FEV\(_1\), PEF, and SpO\(_2\) data collected in this study was limited by the small sample size and the declining use of the connected devices throughout the study period. Use of the pulse oximeter waned over the study period: 126 (89.4%) were compliant with pulse oximeter use in study week 1, which fell to 60 (42.6%) in study week 12, and further to 13 (9.2%) in study week 24. Use of the spirometer also waned over the study period: 116 (82.3%) were compliant with spirometer use in study week 1, which fell to 59 (41.8%) at study week 12, and further to 12 (8.5%) at study week 24. Based on the limited data set, these data cannot be reliably analyzed to make any conclusions about changes in pulmonary function or pulse oximetry throughout the study.

**Patient-Reported HRU**

Interpretation of the patient-reported HRU data is limited by the infrequency of events reported. In total, 90% (n=127) and 93.6% (n=132) of study participants reported no COPD-related emergency room visits or hospitalizations, respectively, in the 3 months prior to baseline. Similarly, 89.6% (n=95) and 93.4% (n=99) of study participants reported no COPD-related emergency room visits or hospitalizations, respectively, during the study period. As expected, physician visits were more common than emergency room visits or hospitalizations. However, at all time points, participants most commonly reported none or only 1 COPD-related physician visit for the prior 3-month period. The event rate for physician visits is thus also inadequate for detection of any impact of Wellinks; furthermore, there is inadequate data to analyze any effect by treatment group.

**Safety**

No adverse events were reported by the participants during the study period.

**Discussion**

**Principal Results**

Since the Wellinks clinical model has evolved over time to include the availability of respiratory therapy and health coaching services, questions emerged as to whether the target population—people living with COPD—would engage as they had done in a previously reported pilot study of a more basic program offering [8] and whether data may be collected to better optimize the frequency and duration of the personal health coaching component. Thus, the purpose of this 2-arm study design was to collect clinical and nonclinical data to optimize the appropriate duration and frequency of health coaching with Wellinks.

Study participants displayed a high degree of engagement with the health coaching component of Wellinks. By contrast, the study population had substantial attrition in the use of the mobile app and connected devices throughout the entire study duration despite rating all of them as highly valuable.

Importantly, this study provided evidence for the first time of the clinical value of patient participation in Wellinks. Significant improvements in COPD self-efficacy and breathlessness...
(mMRC) were observed for study participants regardless of assignment to arm 1 or arm 2. The improvement in the CSES observed after engaging in the Wellinks program may reflect the increased belief among study participants in their own ability to do certain activities, despite any potential perceived limitations due to their COPD diagnosis. Improved confidence in physical activities would be predicted to perpetuate a greater level of physical activity, and associated health outcome improvements may result.

Approximately one-third of participants demonstrated an improvement in breathlessness over the course of the study; improvements in shortness of breath as measured by mMRC are also reflected in the improvements in CSES, wherein the greatest degree of improvement was observed in the physical exertion domain. Although this study design does not allow for clear causal relationships to be determined, one hypothesis is that the remote pulmonary rehabilitation and education provided by the Wellinks health coaches may have improved breathlessness, which then also resulted in greater confidence (self-efficacy) on the physical exertion domain of the CSES [9]. Taken together, we can infer that the use of Wellinks improved self-efficacy and breathlessness, which may predict an improvement in the quality of life for these patients.

One difference between the treatment arms was observed in the NPS values. It is hypothesized that the higher NPS value observed for arm 1 compared to arm 2 may be attributable to the higher degree of interaction between health coaches and the study participants assigned to that arm. It has been frequently reported that digital interventions have the greatest value when combined with personal coaching or counseling [2,11,12], and this greater value may be reflected in the NPS.

**Limitations**

The key limitation of this study is missing data for certain outcomes of interest, such as pulmonary function and HRU. Pulmonary function was assessed by way of home use of a Bluetooth-connected spirometer and pulse oximeter. Interpretation of these results is significantly limited by the lack of consistent use of these devices throughout the study period. Very low compliance with the study-directed use of once per week for each device resulted in a very small sample size, from which clear conclusions cannot be drawn. Although data interpretation is thus limited, this design was intentional to best reflect the real-world use of Wellinks; specifically, health coaches did not mandate the use of the devices but did remind participants to use them as appropriate.

Low long-term compliance with the connected devices is not an entirely surprising finding. It has been previously reported that remote monitoring alone with various biometric devices is subject to failure if it is not effectively integrated into the existing health care delivery model [3]. These results suggest that more needs to be done within the Wellinks clinical model to integrate the device data with the remote pulmonary rehabilitation and health coaching components of the program. It would be important to better understand whether the limited use of the devices was due to technical challenges, due to a lack of perceived value, or some other reason. If participants do not recognize the value of the data collected by these devices, it is possible that more can be done to educate patients about the information, provide context for interpreting the results that are recorded, and integrate health care providers into the process. Furthermore, the Wellinks model includes various components that allow for flexibility to meet patient needs; therefore, compliance with certain components may be expected to vary from individual to individual, in part as a reflection of different needs and preferences of each participating person.

HRU was assessed in this study based on participant self-report using a 3-month look-back period. The main limitations to interpreting these data are the low frequency of events reported and the recall bias that can result from this approach. Future studies of Wellinks will rely on verifiable information from electronic medical records or claims databases to inquire about HRU. Missing data were also the result of failure of some participants to complete all surveys per the study protocol. Notably, based on a meta-analysis of 1071 web-based surveys, completion rates average 44% [13], making this survey response rate better than average, although still an important limitation.

Additionally, the enrolled population may reflect a highly motivated subset of people living with COPD, given their existing engagement with the COPD Foundation prior to the study start; those who were recruited from the COPD Foundation Patient-Powered Research Network, by definition, have self-selected to contribute to research activities, which reflects a high degree of motivation. The same attributes may not be present in the general COPD population. Furthermore, the demographics of this study cohort may not fully reflect the demographics of the COPD population in the community. There was slightly more representation of females compared to males, which is consistent with observed trends of increased prevalence of COPD among women, while a decrease in prevalence has been observed in males over recent decades [14]. However, race and ethnicity are known to impact COPD risk but yet are not well represented in this study cohort [15].

**Comparison With Prior Work**

It has been recognized in prior research that engagement with technology among older adults is dependent upon personal support from professionals or peers [16]. The high receptivity of this study population to personal health coaching sessions as compared to the low receptivity to the use of connected devices may also reflect this need for personal connection and support. In designing digital solutions for this population, it will be important to consider the value of the personal connection between the individual and their coach as a means to achieving greater adoption of associated technologies, such as the app and connected devices included in this study.

There are limitations to comparing the previously published pilot study of Wellinks to the study reported here; the populations differ in important ways (ie, in the pilot study, participants were recruited from a single provider’s practice, whereas in the ASPIRE study reported here, participants were recruited nonpersonally through the COPD Foundation), and the intervention differs as well (eg, the pilot study included most of the same technological components but lacked the health coaching component included here).
Conclusions
This study demonstrates the interest and satisfaction of an ambulatory COPD population with the additional support and services provided by Wellinks. Health coaching appeared to be the most valuable component of Wellinks in this study. Signals of clinical outcome improvements in this study are encouraging and would best be further explored in larger cohorts to assess meaningful impact on a population level in terms of clinical improvement and impact on HRU.

Strategies and services to improve chronic disease self-management, such as with what Wellinks offers to patients with COPD, have been shown to reduce the burden of chronic diseases on individuals and the health care system at large. The data reported here are valuable not only to further optimize Wellinks but also to inform novel program design by others.

Future Directions
Future studies will explore the ability of newer and modified versions of Wellinks to reduce hospital readmissions following an acute exacerbation as well as to explore the integration of Wellinks into a large accountable care organization. There remains a significant opportunity to bring remote disease management tools to people living with COPD, and these studies will further build the evidence base and support the long-term scalability of the program.

Acknowledgments
This study was sponsored and funded by Wellinks (Convexity Scientific, Inc) of New Haven, Connecticut, and was executed in partnership with the COPD Foundation. The authors wish to acknowledge and sincerely express gratitude to the people living with COPD who have participated in this study. Without the willingness of research participants, it would not be possible to validate and iterate on the tools and services designed to meet their needs.

Data Availability
The data sets generated during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
KAP receives consulting compensation from Wellinks. NL is an independent contractor for the COPD Foundation and a former employee of GSK plc. CRR is a consultant to and shareholder of Wellinks. RJC, NH, PD, and AS are employees of Wellinks.

References


**Abbreviations**

COPD: chronic obstructive pulmonary disease  
CSES: COPD Self-Efficacy Scale  
FEV₁: forced expiratory volume in 1 second  
HRU: healthcare resource utilization  
IRB: institutional review board  
LS: least squares  
mMRC: Modified Medical Research Council  
NPS: net promoter score  
PEF: peak expiratory flow  
SpO₂: saturation of peripheral oxygen
The Impact of Intervention Design on User Engagement in Digital Therapeutics Research: Factorial Experiment With a Mixed Methods Study

Hyerim Lee, BA; Eung Ho Choi, PhD; Jung U Shin, PhD; Tae-Gyun Kim, PhD; Jooyoung Oh, PhD; Bokyoung Shin, MA; Jung Yeon Sim, MA; Jaeyong Shin, PhD; Meelim Kim, PhD

1Department of Psychology, College of Liberal Arts, Yonsei University, Seoul, Republic of Korea
2Department of Dermatology, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea
3Department of Dermatology, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea
4Department of Dermatology, Severance Hospital, Cutaneous Biology Research Institute, Yonsei University College of Medicine, Seoul, Republic of Korea
5Department of Psychiatry, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea
6Institute of Behavioral Sciences in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
7Department of Integrative Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
8Department of Medical Device Engineering and Management, Yonsei University Graduate School, Seoul, Republic of Korea
9Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
10Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, CA, United States
11The Design Lab, University of California San Diego, San Diego, CA, United States
12Center for Wireless & Population Health Systems, Calit2’s Qualcomm Institute, University of California San Diego, San Diego, CA, United States

*these authors contributed equally

Corresponding Author:
Meelim Kim, PhD
Herbert Wertheim School of Public Health and Human Longevity Science
University of California San Diego
Atkinson Hall, 9500 Gilman Dr, La Jolla
San Diego, CA, 92121
United States
Phone: 1 323 776 5171
Email: mek007@health.ucsd.edu

Abstract

Background: User engagement is crucial for digital therapeutics (DTx) effectiveness; due to variations in the conceptualization of engagement and intervention design, assessment and retention of engagement remain challenging.

Objective: We investigated the influence of the perceived acceptability of experimental intervention components and satisfaction with core intervention components in DTx on user engagement, while also identifying potential barriers and facilitators to user engagement.

Methods: We conducted a mixed methods study with a 2 × 2 factorial design, involving 12 outpatients with atopic dermatitis. Participants were randomized into 4 experimental groups based on push notification (“basic” or “advanced”) and human coach (“on” or “off”) experimental intervention components. All participants engaged in self-monitoring and learning courses as core intervention components within an app-based intervention over 8 weeks. Data were collected through in-app behavioral data, physician- and self-reported questionnaires, and semistructured interviews assessed at baseline, 4 weeks, and 8 weeks. Descriptive statistics and thematic analysis were used to evaluate user engagement, perceived acceptability of experimental intervention components (ie, push notification and human coach), satisfaction with core intervention components (ie, self-monitoring and learning courses), and intervention effectiveness through clinical outcomes.

Results: The primary outcome indicated that group 4, provided with “advanced-level push notifications” and a “human coach,” showed higher completion rates for self-monitoring forms and learning courses compared to the predetermined threshold of clinical significance. Qualitative data analysis revealed three key themes: (1) perceived acceptability of the experimental intervention
components, (2) satisfaction with the core intervention components, and (3) suggestions for improvement in the overall intervention program. Regarding clinical outcomes, the Perceived Stress Scale and Dermatology Life Quality Index scores presented the highest improvement in group 4.

**Conclusions:** These findings will help refine the intervention and inform the design of a subsequent randomized trial to test its effectiveness. Furthermore, this design may serve as a model for broadly examining and optimizing overall engagement in DTx and for future investigation into the complex relationship between engagement and clinical outcomes.

**Trial Registration:** Clinical Research Information Service KCT0007675; http://tinyurl.com/2m8jrmy

(JJMIR Form Res 2024;8:e51225) doi:10.2196/51225

**KEYWORDS**

atopic; dermatitis; experimental design; mobile health; patient engagement; research methodology

**Introduction**

Digital Therapeutics in General

With the rapid advancement of digital technology, digital therapeutics (DTxs) have emerged as a promising approach to either enhance the value of conventional health care delivery systems or have the potential to substantially substitute the existing system [1]. DTx refers to “an evidence-based intervention that is driven by high-quality software programs to prevent, manage, or treat a disease or disorder” [2]. Using technology and data analytics, DTx holds numerous benefits in health care: (1) it can encompass a wide range of physical and mental health conditions (mostly chronic) like diabetes, oncology treatment management, insomnia, attention-deficit/hyperactivity disorder (ADHD), and substance use disorder [3]; (2) it can provide personalized care with data-driven treatment options [4]; and (3) it can reduce health care costs [5]. Given these significant potential benefits, it is crucial to understand how the efficacy of DTxs in therapy can be improved. To achieve such improvement, diverse and comprehensive research regarding the DTx development process should be conducted to successfully implement and optimize these promising interventions.

**User Engagement Issues in DTx**

It is widely acknowledged that user engagement is important for improving the effectiveness of DTx [6]. Engagement in DTx can be defined as “the extent (eg, amount, frequency, duration, and depth) of use and subjective experience characterized by attention, interest, and affect” [7,8]. Although user engagement significantly impacts the effectiveness of DTx, assessing and retaining it is challenging. The possible reasons for this may include (1) a lack of shared awareness regarding the useful perception of engagement, (2) engagement in DTx is not a stationary but a dynamic process, and (3) it is a multifaceted construct, capturing the user’s behavior, cognitive, and emotional states. Several systematic reviews have investigated DTx intervention components (eg, self-monitoring, reminders, and rewards) that are linked with higher engagement [9,10]. However, the findings of these studies do not provide conclusive evidence about the intervention components that help patients become more engaged with the DTx. This occurs due to substantial variation in the definition of engagement and intervention design in DTx. Thus, an in-depth analysis of the intervention components and a concrete definition of user engagement should be established, particularly during the design phase of DTx.

**Methods for Evaluating Intervention Components in Digital Intervention**

For systematically evaluating how intervention design influences user engagement, the optimization methods from the multiphase optimization strategy (MOST) can be used with a couple of representative intervention components from a wide range of possible options. MOST allows for efficient testing through a randomized experiment, including a factorial experiment, which allows for the simultaneous examination of different intervention design factors [11]. Many recent studies, however, used only traditional randomized controlled trials (RCTs) as the primary study design to test the efficacy of the intervention as a package [12-15] and to examine the relationships between engagement level and clinical outcomes through post hoc analysis [6,16,17]. Using only RCTs as an evaluation design may pose some challenges to the effective evaluation of DTxs, as they are considered complex, context-dependent, and individually tailored interventions that purport to maximize its effectiveness [18,19]. Thus, additional evaluation methods for DTxs, such as adaptive study designs (eg, sequential multiple assignment randomized trial and factorial trial from MOST), must be considered to provide robust evidence during its design and development phases.

**Aims of This Study**

Here, we aimed to examine the impact of the perceived acceptability of the experimental intervention components (ie, push notification and human coach) and satisfaction with the core intervention components (ie, self-monitoring and learning courses) in DTx on user engagement (Figure 1). We used “Atomind,” a DTx for patients with atopic dermatitis (AD), developed for clinical trial purposes, with a primary focus on optimization as a refinement process before validating its effectiveness through larger RCTs. This was a proof-of-concept study with an experimental $2 \times 2$ factorial design, using both quantitative (eg, in-app behavioral data) and qualitative (eg, semistructured interviews) assessment methods. We hypothesized that those who received the advanced level of each experimental intervention component would pass the threshold of clinical significance of user engagement metrics in DTxs. Moreover, the qualitative analysis of satisfaction with the core intervention component would identify the potential barriers and facilitators to user engagement. This study could
also inform how to optimize and evaluate other DTx in this field.

Figure 1. Overview of the impact of intervention designs on user engagement in “Atomind”.

Methods

Study Design

This full factorial experiment had 2 experimental intervention components (Figure 2), each of which was implemented at 2 different levels: push notification (“basic” or “advanced”) and human coach (“on” or “off”). Participants were randomly allocated to 1 of the 4 experimental groups in the $2 \times 2$ full factorial design. All participants engaged in self-monitoring and learning courses as core intervention components during the 8-week intervention period. We applied a mixed methods approach by collecting quantitative (eg, surveys) and qualitative (eg, semistructured interviews) data to examine the perceived acceptability of experimental intervention components, satisfaction with the core intervention components, and suggestions for improvement in the overall intervention program. We conducted the interviews after 8 weeks of treatment.

Figure 2. A $2 \times 2$ factorial design exploring the perceived acceptability of experimental intervention components in this digital therapeutics (DTx) study, featuring different combinations of basic- versus advanced-level push notification and “on” versus “off” human coach.

<table>
<thead>
<tr>
<th>Human coach</th>
<th>Push notification</th>
<th>Basic level</th>
<th>Advanced level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>Basic push notification</td>
<td>Group 1</td>
<td>Additional push notification</td>
</tr>
<tr>
<td></td>
<td>Human coach Off</td>
<td></td>
<td>Human coach Off</td>
</tr>
<tr>
<td>On</td>
<td>Basic push notification</td>
<td>Group 3</td>
<td>Additional push notification</td>
</tr>
<tr>
<td></td>
<td>Human coach On</td>
<td></td>
<td>Human coach On</td>
</tr>
</tbody>
</table>
Experimental Intervention Components

Push Notification
Participants randomized to “basic-level push notification” received basic push notifications that encouraged users to log in and complete tasks at time points chosen by users. Participants randomized to “advanced-level push notification” received not only basic push notifications but also additional push notifications when they did not complete in-app self-monitoring forms, weekly classes, or missions after receiving the basic push notifications. Additional push notifications contained emotionally supportive phrases (eg, “It’s a bit annoying, right? But don’t forget that sustained use of the app can help reduce your symptoms.” And, “Malang is waiting for <username>! Haven’t you finished the class yet? Don’t give up and let’s start!”). Push notifications are classified into 4 different categories: self-monitoring, learning course, mission, and personalized feedback report. An overview of the 2 groups’ push notifications is presented in Multimedia Appendix 1.

Human Coach
Participants with this experimental intervention component turned “on” received tailored guidance and assistance from a human coach. The coach sent weekly motivational messages to maintain participants’ engagement through a different digital application called KakaoTalk, which is the most popular instant messaging app in South Korea, with 94% of the entire Korean population as users. The coach spent a total of 6 hours a week—2 hours a day over 3 days—to manage the participants. The coach kept the participants motivated, held them responsible, provided feedback, and monitored their progress to keep them on track. Participants could address difficulties or questions they encountered with the app through 2-way communication. Besides the app’s information, participants could also ask questions about skin health and mental well-being and receive answers from the coach. Conversely, participants with this experimental intervention component turned “off,” received nothing, and conducted self-care.

Participants
All participants were outpatient patients who met the eligibility criteria, including individuals who (1) were aged 19 years or older and had mild to severe AD, (2) were able to understand verbal and written Korean, and (3) had their own smartphone. Participants who met the eligibility criteria were assigned randomly to 4 experimental groups in a 1:1:1:1 ratio using program IDs generated within the Atomind app.

Intervention
Atomind is an app-based intervention program that helps individuals manage skin conditions and AD symptoms. It was developed by Huray Inc, South Korea (Multimedia Appendix 2). The app’s content is based on cognitive behavioral therapy (CBT) and a mindfulness approach to support healthy behavioral habits and regulate negative emotions. The app prompts users to complete in-app self-monitoring forms on a daily, weekly, and monthly basis, focusing on motivation, skin condition, behavioral change, and mental health. Weekly videos demonstrate educational information that can help relieve AD symptoms and CBT strategies for regulating negative thoughts and emotions. After watching the video, users were asked to demonstrate their understanding by passing a postquiz. The overall topic of the weekly video is listed in Multimedia Appendix 3. Moreover, missions are provided to help users apply their newly acquired skills in real life. Users can access personalized graphic feedback based on their self-monitoring.

Outcomes
Outcome measures were collected by using (1) in-app behavioral data, (2) physician- or self-reported questionnaires, and (3) semistructured interviews. At baseline, participants were asked to complete a demographic questionnaire pertaining to their age, gender, educational level, and health-related measures (medical and family health history, health literacy, etc).

The primary outcome was the user engagement of the intervention, measured by in-app behavioral data on core intervention components, including percentages of self-monitoring forms and learning courses completed. We collected qualitative data on the perceived acceptability of experimental intervention components (ie, push notification and human coach), satisfaction with the core intervention components (ie, self-monitoring and learning courses), and suggestions on any improvement for the overall intervention program through semistructured interviews. The interviews were conducted over the telephone by 2 research team members after 8 weeks of intervention. A semistructured interview guide (Multimedia Appendix 4) was used to guide the interviews, lasting 15-20 minutes for each.

Furthermore, other clinical outcome measures were assessed at baseline, 4 weeks, and 8 weeks of intervention. Designated dermatologists assessed the severity of AD using the eczema area and severity index (EASI), including the severity of 4 signs (erythema, edema or papulation, excoriation, and lichenification; range 0-72) [20]. Atopic eczema severity reported by patients was measured with the patient-oriented eczema measure (POEM; range 0-28), a 7-item questionnaire for monitoring the care of patients with atopic eczema [21]. Insomnia severity was measured with the insomnia severity index (ISI; range 0-28), a 7-item questionnaire assessing perceived insomnia severity using a Likert-type scale [22]. Perceived stress level was measured with the perceived stress scale (PSS; range 0-40), a 10-item questionnaire assessing psychological stress [23]. Quality of life was measured by the dermatology life quality index (DLQI; range 0-30), a 10-item questionnaire assessing how much the patients’ skin problems have affected their lives over the past week [24], and fear of negative evaluation was measured using the brief fear of negative evaluation (BFNE; range 12-60) scale, which is a 12-item questionnaire assessing the degree of anxiety about perceived negative evaluation [25]. The assessment methods and assessment period for each measurement are shown in Multimedia Appendix 5.

Statistical Analysis
Descriptive statistics were used to analyze quantitative data, including in-app behavioral data and clinical outcomes. We initially recruited and enrolled 12 participants, with 3 individuals for each group; however, of the initial 12 individuals, 3 participants (1 in group 2, 1 in group 3, and 1 in group 4) were...
excluded from the analysis due to medication changes during the intervention period.

We set the threshold of clinical significance (TCS) for user engagement, considering the period of each assessment. Previous research with larger sample sizes has shown that individuals with high efficacy typically maintain an engagement rate between 50% and 80% [26,27]. However, given the smaller sample size in this proof-of-concept study, a more stringent approach has been applied in setting the TCS for user engagement. For self-monitoring, the TCS is determined if the average completion rate of self-monitoring forms is ≥90%. For learning courses, the TCS is set if the average completion rate of learning courses is ≥80% throughout the intervention period.

The perceived acceptability of each experimental intervention component and the satisfaction of each core intervention component were also examined by semistructured interviews. Qualitative data were analyzed using thematic analysis. The verbatim transcriptions of the interviews were used to extract the responses, which were categorized into items focusing on the perceived acceptability of the experimental intervention components, satisfaction of the core intervention components, and suggestions on any improvement for the overall intervention program.

To measure the interventions’ effectiveness, we assessed the changes in the average of clinical outcomes (eg, EASI, POEM, ISI, PSS, DLQI, and BFNE) before and after the intervention in the 4 groups and the 2 different levels of each experimental intervention component.

**Ethical Considerations**

All study activities were conducted in adherence to ethical standards and received approval from the institutional review boards of the organizing sites, including Severance Hospital (4-2022-0922), Wonju Severance Christian Hospital (CR322035), and Bundang Cha Hospital (CHAMC 2022-05-005-001). The trial was registered on the Clinical Research Information Service (KCT0007675). Participants provided voluntary, written, and informed consent after a thorough explanation of the clinical trial. Privacy measures included data anonymization and secure storage. Participants received US $80 in compensation for their contribution, a detail communicated during the informed consent process.

**Results**

**Sample Characteristics**

A total of 12 adults (mean age 31.1 years; range 20-43 years) were recruited between August and November 2022 (Figure 3). Of the 12 participants, 2 (17%) had mild AD, 7 (58%) had moderate AD, and 3 (25%) had severe AD. More details regarding the sample characteristics are presented in Table 1.
Table 1. Sample characteristics of the participants (N=12).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Male</td>
<td>7 (58)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>5 (42)</td>
</tr>
<tr>
<td>30-39</td>
<td>5 (42)</td>
</tr>
<tr>
<td>40-49</td>
<td>2 (17)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
</tr>
<tr>
<td>High school graduate or less</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Currently enrolled in or graduated from college</td>
<td>7 (58)</td>
</tr>
<tr>
<td>Currently enrolled in or graduated from graduate school</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Severity of atopic dermatitis</strong></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>2 (17)</td>
</tr>
<tr>
<td>Moderate</td>
<td>7 (58)</td>
</tr>
<tr>
<td>Severe</td>
<td>3 (25)</td>
</tr>
<tr>
<td><strong>Duration of disease (years)</strong></td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>1 (8)</td>
</tr>
<tr>
<td>11-20</td>
<td>4 (33)</td>
</tr>
<tr>
<td>21-30</td>
<td>6 (50)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Comorbidity of other allergic diseases</strong></td>
<td></td>
</tr>
<tr>
<td>Atopic dermatitis only</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Comorbid with other allergic diseases</td>
<td>9 (75)</td>
</tr>
<tr>
<td><strong>Family history of allergic diseases</strong></td>
<td></td>
</tr>
<tr>
<td>Atopic dermatitis</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Allergic rhinitis</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Food allergy</td>
<td>1 (8)</td>
</tr>
<tr>
<td>None</td>
<td>3 (25)</td>
</tr>
<tr>
<td><strong>Alcohol consumption frequency over the past year</strong></td>
<td></td>
</tr>
<tr>
<td>Not at all in the past year</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Less than once a month</td>
<td>1 (8)</td>
</tr>
<tr>
<td>About once a month</td>
<td>1 (8)</td>
</tr>
<tr>
<td>2-4 times a month</td>
<td>2 (17)</td>
</tr>
<tr>
<td>2-3 times a week</td>
<td>4 (33)</td>
</tr>
<tr>
<td><strong>Health literacy</strong></td>
<td></td>
</tr>
<tr>
<td>Health literacy (score 15 out of 15)</td>
<td>9 (75)</td>
</tr>
</tbody>
</table>

Primary Outcome

Regarding the user engagement rates among different groups (Figures 4A and 4B), groups 2 (90.9%), 3 (95.5%), and 4 (97%) showed higher completion rates for self-monitoring compared to the predetermined TCS (90%). Additionally, groups 2 (83.3%) and 4 (91.7%) had higher completion rates for learning courses than the TCS (80%). These results indicate that group 4, provided with advanced-level push notifications and a human coach, had the highest user engagement during the intervention.
As shown in Figures 4C and 4D, “advanced-level push notification” (93.9%) and “human coach on” (96.2%) were the experimental intervention components that exceeded the predetermined TCS for self-monitoring (90%). The experimental intervention components that exceeded the TCS for learning courses (80%) were also “advanced-level push notification” (87.5%) and “human coach on” (85.4%). Overall, “advanced-level push notification” and “human coach on” demonstrated the highest user engagement among the experimental intervention components.

**Secondary Outcome**

Qualitative data were organized into three key themes: (1) perceived acceptability of the experimental intervention components, (2) satisfaction of the core intervention components, and (3) suggestions for improvement in the overall intervention program. Table 2 presents all themes and subthemes with corresponding quotes.
Table 2. Key themes, subthemes, and quotes from semistructured interviews.

<table>
<thead>
<tr>
<th>Themes, subthemes, and components</th>
<th>Verbatim examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key theme 1: perceived acceptability of the experimental intervention components</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Push notification component</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Technical aspect</strong></td>
<td></td>
</tr>
<tr>
<td>Basic push notification</td>
<td>It would be better if we could choose the time to receive notifications, and it would be better if we could receive the notification functioned similarly to a wake-up alarm that rings again if not checked...</td>
</tr>
<tr>
<td>Advanced push notification</td>
<td>I lead a busy life, so receiving notifications was helpful. In fact, I think it was better for me to receive notifications frequently.</td>
</tr>
<tr>
<td><strong>Content aspect</strong></td>
<td></td>
</tr>
<tr>
<td>Basic push notification</td>
<td>The (content) of the notifications was all good. It would be nice to have additional features like setting reminders for taking medication. Or maybe some information on whether I've applied moisturizer or taken my medicine today. Something like that would be useful.</td>
</tr>
<tr>
<td>Advanced push notification</td>
<td>The notification content was good enough as it was, with just simple and neat notifications.</td>
</tr>
<tr>
<td><strong>Human coach component</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Technical aspect</strong></td>
<td></td>
</tr>
<tr>
<td>Human coach off</td>
<td>It would be better if there was a feature that allowed patients to send messages to report any technical errors or issues...It might be better to communicate using this feature.</td>
</tr>
<tr>
<td>Human coach on</td>
<td>I wish there was a channel where atopic patients could communicate with each other.</td>
</tr>
<tr>
<td><strong>Content aspect</strong></td>
<td></td>
</tr>
<tr>
<td>Human coach off</td>
<td>It would be great if we could receive feedback for emergency situations.</td>
</tr>
<tr>
<td>Human coach on</td>
<td>For example, it would be more effective to ask direct questions like 'have you reduced your medication dosage?' rather than asking about difficulties or inconveniences...</td>
</tr>
<tr>
<td><strong>Key theme 2: satisfaction with the core intervention components</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Self-monitoring</strong></td>
<td></td>
</tr>
<tr>
<td>Building health habits</td>
<td>I used to forget to take my medicine, but ever since I started using the app to check it, I've been taking it every morning and before bed, and I've been doing it consistently. I've established a routine of recording it separately from the app. By recording the questionnaire every day, I can now monitor the daily improvement or worsening of my condition, which was the best part of the app.</td>
</tr>
<tr>
<td><strong>Learning course</strong></td>
<td></td>
</tr>
<tr>
<td>Acquiring reliable information</td>
<td>It was great to learn about the parts that I used to miss with reliable information.</td>
</tr>
<tr>
<td><strong>Key theme 3: suggestions for improvement on the overall intervention program</strong></td>
<td></td>
</tr>
<tr>
<td>Diversity of daily self-monitoring form questions</td>
<td>I wish, that depending on the symptoms, different questions would be asked to determine whether the symptoms improved or worsened from the previous day.</td>
</tr>
<tr>
<td>Burdensomeness of self-monitoring feature</td>
<td>Taking pictures of my body to check the skin lesion was burdensome. The questionnaire was too lengthy.</td>
</tr>
<tr>
<td>Not tailored contents</td>
<td>There was information that would have been useful, if symptoms hadn't been so severe. It would be better to recommend it to patients with mild symptoms. The quiz following the video was so simple that I didn't even need to watch it and could simply answer the questions correctly. This is why I stopped watching the weekly videos.</td>
</tr>
<tr>
<td>Motivating factors</td>
<td>It would be great to include elements that can boost motivation, such as fun factors or any benefits.</td>
</tr>
<tr>
<td>Technical issues</td>
<td>There were times when I couldn't continue with the survey for a few days because some questionnaire items wouldn't move forward at all. So, it would be great if those issues could be improved.</td>
</tr>
</tbody>
</table>
Key Theme 1: Perceived Acceptability of the Experimental Intervention Components

Perceived acceptability was measured using the components’ technical and content aspects. Regarding the technical aspect of the push notification component, 60% (3/5) participants receiving “basic-level push notifications” responded that they would like the push notification frequency to increase. Moreover, 20% (1/5) responded that it would be better to select the time and frequency of the push notifications and be reminded if they did not complete the task. And 50% (2/4) receiving “advanced-level push notifications” were overall satisfied with the current push notification frequency. Regarding the content aspect of this component, both groups responded that they were satisfied with the provided notification contents. However, 20% (1/5) of participants in the group receiving “basic-level push notifications” suggested that it would be helpful to receive a push notification reminding them to take medication or apply some moisturizer.

Regarding the technical aspect of the human coach component, 40% (2/5) participants assigned to the “human coach off” component requested a 1:1 communication channel within the app, as they could not receive assistance from a human coach. A total of 75% (3/4) of participants assigned to the “human coach on” component preferred to have an in-app communication channel rather than using a different instant messaging app (ie, KakaoTalk) for communication with a human coach. And 25% (1/4) of participants also suggested integrating a community feature for patients to communicate with each other. Regarding the content aspect of the human coach component, 20% (1/5) of participants assigned to the “human coach off” suggested adding a telehealth feature for emergencies. And, 25% (1/4) participants assigned to the “human coach on” preferred the coach asking specific questions related to symptom management, such as “Have you taken your medicine today?” or “Have you visited the hospital?” rather than the questions relevant to the app use, like “Is there anything difficult or uncomfortable while using the app?”

Key Theme 2: Satisfaction With the Core Intervention Components

Satisfaction was measured for each core intervention component, self-monitoring, and learning courses. Regarding the self-monitoring component, 78% (7/9) of participants reported that self-monitoring helped build health habits, including better medication adherence, reduced scratching behavior, and consistent use of moisturizers. Moreover, they could easily track their symptoms through weekly reports, which helped them monitor their symptoms over time. Regarding the learning course component, 56% (5/9) of participants indicated that they acquired reliable information through weekly videos.

Key Theme 3: Suggestions for Improvement on Overall Intervention Program

Suggestions for improvement in the overall intervention program were divided into 5 subthemes. First of all, 33% (3/9) of participants recommended diversifying the questions of the daily self-monitoring form, as they found them to be repetitive and lacking in variation. Second, 22% (2/9) of participants found the self-monitoring burdensome as they had to upload lesion pictures daily. Third, 44% (4/9) of participants felt the learning course was not sufficiently tailored to their needs. They found the video content insufficiently helpful for patients with severe disease; the postquiz questions were unchallenging; and the video was too long. Fourth, 11% (1/9) of participants suggested adding motivating factors to the intervention program to make them more engaged with the app. Lastly, technical issues within the app were mentioned. A total of 33% (3/9) of participants recommended improving its performance, such as fixing bugs in the self-monitoring feature, reducing duplicate push notifications, and improving video sound quality.

Clinical Outcomes

Descriptive statistics, for example, mean (SD), were used to analyze clinical outcomes by group and experimental intervention component. The ISI score showed the greatest improvement in group 2 (mean change –4.50, SD 6.36). The EASI, POEM, and BFNE scores showed the highest improvement in group 3 (mean change –10.20, SD 9.90, mean change –3.00 score, SD 15.56, and mean change –4.50, SD 6.36, respectively). The PSS and DLQI scores presented the greatest improvement in group 4 (mean change –3.50, SD 3.54, and mean change –6.00, SD 11.32, respectively).

Regarding the push notification component, the ISI, PSS, DLQI, and BFNE scores showed the highest improvement in the “advanced-level push notification” component (mean change –0.25, SD 6.34, mean change –1.75, SD 2.99, mean change –4.75, SD 8.06, and mean change –3.00, SD 3.46, respectively). Regarding the human coach component, the EASI, POEM, PSS, and BFNE scores presented the highest improvement in the “human coach on” component (mean change –6.73, SD 7.41, mean change –1.50, SD 11.73, mean change –3.00, SD 2.94, and mean change –4.25, SD 3.69, respectively). More detailed results can be found in Figure 5.
Figure 5. Differences in clinical outcomes between groups (A-F) and components (G-L) from preintervention to postintervention at 8 weeks. BFNE: brief fear of negative evaluation; DLQI: dermatology life quality index; EASI: eczema area and severity index; ISI: insomnia severity index; POEM: patient-oriented eczema measure; PSS: perceived stress scale.

Discussion

Principal Findings

Our primary objective was to investigate how the perceived acceptability of experimental intervention components and satisfaction with core intervention components affect user engagement in DTx. We examined in-app behavioral data on core intervention components (ie, percentages of self-monitoring forms and learning courses completed) as a user engagement metric. As hypothesized, the TCS of user engagement was achieved in group 4, where all 2 experimental factors were advanced simultaneously. Furthermore, clinical outcomes related to the mental health of patients with AD improved in group 4. This study identified potential barriers and facilitators of user engagement through semistructured interviews on the patients’ satisfaction with core intervention components. Overall, our analysis of Atomind data suggests that incorporating advanced-level push notifications with a human coach, tailoring contents with various self-monitoring tools, and implementing some motivational factors (eg, rewards) may improve user engagement.

Comparison With Previous Work

To the best of our knowledge, this is the first study to examine the impact of different levels of push notification, human coach,
and satisfaction with core intervention components on user engagement in DTx using mixed methods. Although there is a proliferation of clinical research on user engagement with mobile health apps, the majority only conducted traditional RCTs [28-39] or optimization trials with a single type of assessment method [40-44]. The findings from these earlier studies with traditional RCTs only explained how the intervention as a package affected user engagement; they could not identify the specific intervention elements that impacted it [29]. Additionally, only assessing quantitative data from optimization trials (eg, factorial experiments) limits the understanding of barriers and facilitators affecting user engagement [45,46]. In contrast, this study clearly showed that advanced-level push notifications and communication with a human coach are the main factors enhancing user engagement. Furthermore, our qualitative analysis showed that advanced-level push notifications were sufficient in frequency to serve as a reminder in busy daily lives, and their content was concise enough to be acceptable. Although communication with a human coach improved user engagement, our qualitative findings suggest that the human coach platform should have been implemented in the internal system of the Atomind app with more diverse questions and detailed responses. Using a mixed methods approach to assess various factors contributing to user engagement in Atomind enabled us to gain insights into the “what, how, and why” of this phenomenon, which is critical to figuring out what steps must be taken to improve an intervention.

Establishing a TCS for user engagement has been applied, as this study is a proof-of-concept study with a small sample size. This approach allows for resource-efficient research with clear go-or-no-go decision-making, lowering the risk of confirmatory bias [19]. Concerning TCS determination, each previous study had its own logic established and multiple metrics to account for user engagement [33,47]. This is because user engagement is a multifaceted concept with no universal consensus on how to perceive it [6,7,34]. Among the various metrics of user engagement from previous research, the completion of specific activities or modules of the intervention was the most commonly used metric for user engagement [29-34,36,39,43,48]. Similarly, we measured user engagement in the app by assessing the completion rate of the core intervention components. In this study, self-monitoring is for daily activity, while learning courses are for weekly activity. Thus, we set up different levels of TCS for each activity to assess user engagement; the completion rate was 90% for self-monitoring forms and 80% for learning courses.

Regarding the clinical outcomes from this study, people who received the advanced level of experimental intervention components saw improvement in the majority of psychological symptoms (eg, stress, quality of life, and fear of negative evaluation), which was more than the physical symptoms related to AD. These findings correspond with previous research that suggests digital interventions should focus mainly on improving mental health conditions to support better physical health conditions [49,50]. This trend is caused by several inherent factors of mental health interventions, including the stigma associated with mental problems and diagnosis-specific barriers to accessing mental health services [51]. Likewise, Atomind is a digital intervention for patients with AD that encourages healthy behaviors and mental health conditions for effective symptom management. Thus, improving psychological measures by engaging with Atomind indicates that it achieved the intended proximal outcome.

Limitations and Future Directions

First, the statistical power of this study is insufficient to determine significant effects before and after the intervention. However, setting reasonable TCS for quantitative data and collecting qualitative data will support our findings on DTx optimization for use in well-powered RCTs. Second, the Atomind app is only available for use on the Android operating system. To overcome this limitation, we provided Android smartphones during the intervention period to those (n=5) who had other operating systems on their smartphones. Despite this effort, the user experience with Atomind, which is closely related to user engagement, may be affected. Lastly, technical issues with the app occurred frequently during the intervention period, which may affect user engagement. As Atomind was in the development phase, these problems could have taken place; however, its technical system should be improved in a later version and used for future clinical research.

Conclusions

This proof-of-concept, mixed methods study with an experimental 2 × 2 factorial design demonstrates the impact that perceived acceptability of experimental intervention components and satisfaction with core intervention components in DTx have on user engagement. The findings will be used to refine the intervention and inform the design of the next RCT to test its effectiveness. Furthermore, this research design may serve as a model for how to examine and optimize overall engagement in DTx in broad terms; it will help future research investigate the complex relationship between engagement and clinical outcomes.

Acknowledgments

This research was supported by the Seoul R&BD Program (grant BT210048; project name: Development and Demonstration of a Digital Therapeutics Platform Service for Atopic Dermatitis Treatment) through the Seoul Business Agency, funded by the Seoul Metropolitan Government.
Data Availability
The data sets generated and/or analyzed during this study are not publicly available due to the need to maintain privacy and confidentiality, but are available from the corresponding author on reasonable request. Requests for access to specific data points or additional information will be considered on a case-by-case basis.

Authors' Contributions
MK and JS conceptualized and developed the study’s design. MK provided the intellectual framework for this research. EHC, JUS, and TGK were in charge of the recruitment and data collection of participants. JO and BS served as human coaches, providing guidance and assistance to the participants. HL and JYS conducted interviews and handled the analysis of qualitative data. HL and MK contributed significantly to the data analysis and interpretation. HL and MK wrote the manuscript and edited its contents. MK and JS conducted a thorough review of the manuscript. All authors approved the final version of the manuscript for submission.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Overview of push notifications between two groups.
[DOCX File, 17 KB - formative_v8i1e51225_app1.docx]

Multimedia Appendix 2
Atomind app sample screen.
[DOCX File, 758 KB - formative_v8i1e51225_app2.docx]

Multimedia Appendix 3
Overall topics of weekly videos.
[DOCX File, 15 KB - formative_v8i1e51225_app3.docx]

Multimedia Appendix 4
Semi-structured interview guideline.
[DOCX File, 17 KB - formative_v8i1e51225_app4.docx]

Multimedia Appendix 5
Assessment methods and assessment period for each measurement.
[DOCX File, 15 KB - formative_v8i1e51225_app5.docx]

References


42. Materia FT, Smyth JM. Acceptability of intervention design factors in mHealth intervention research: experimental factorial study. JIMIR Mhealth Uhealh 2021;9(7):e23303 [FREE Full text] [doi: 10.2196/23303] [Medline: 34309563]


Abbreviations

AD: atopic dermatitis
ADHD: attention-deficit/hyperactivity disorder
BFNE: brief fear of negative evaluation
CBT: cognitive behavioral theory
DLQI: Dermatology Life Quality Index
DTx: digital therapeutics
EASI: eczema area and severity index
ISI: insomnia severity index
MOST: multiphase optimization strategy
POEM: patient-oriented eczema measure
PSS: perceived stress scale
RCT: randomized controlled trial
TCS: threshold of clinical significance

Zahra Zandesh, PhD
Information Technology and Statistics Department, Tehran University of Medical Sciences, Tehran, Iran

Corresponding Author:
Zahra Zandesh, PhD
Information Technology and Statistics Department
Tehran University of Medical Sciences
Keshavarz Blvd, next to the intersection of Qods St.
Tehran, 1417653761
Iran
Phone: 98 2181633102
Email: Zandesh.z@gmail.com

Abstract

Background: Privacy in our digital world is a very complicated topic, especially when meeting cloud computing technological achievements with its multidimensional context. Here, privacy is an extended concept that is sometimes referred to as legal, philosophical, or even technical. Consequently, there is a need to harmonize it with other aspects in health care in order to provide a new ecosystem. This new ecosystem can lead to a paradigm shift involving the reconstruction and redesign of some of the most important and essential requirements like privacy concepts, legal issues, and security services. Cloud computing in the health domain has markedly contributed to other technologies, such as mobile health, health Internet of Things, and wireless body area networks, with their increasing numbers of embedded applications. Other dependent applications, which are usually used in health businesses like social networks, or some newly introduced applications have issues regarding privacy transparency boundaries and privacy-preserving principles, which have made policy making difficult in the field.

Objective: One way to overcome this challenge is to develop a taxonomy to identify all relevant factors. A taxonomy serves to bring conceptual clarity to the set of alternatives in in-person health care delivery. This study aimed to construct a comprehensive taxonomy for privacy in the health cloud, which also provides a prospective landscape for privacy in related technologies.

Methods: A search was performed for relevant published English papers in databases, including Web of Science, IEEE Digital Library, Google Scholar, Scopus, and PubMed. A total of 2042 papers were related to the health cloud privacy concept according to predefined keywords and search strings. Taxonomy designing was performed using the deductive methodology.

Results: This taxonomy has 3 layers. The first layer has 4 main dimensions, including cloud, data, device, and legal. The second layer has 15 components, and the final layer has related subcategories (n=57). This taxonomy covers some related concepts, such as privacy, security, confidentiality, and legal issues, which are categorized here and defined by their expansion and distinctive boundaries. The main merits of this taxonomy are its ability to clarify privacy terms for different scenarios and signalize the privacy multidisciplinary objectification in eHealth.

Conclusions: This taxonomy can cover health industry requirements with its specifications like health data and scenarios, which are considered as the most complicated among businesses and industries. Therefore, the use of this taxonomy could be generalized and customized to other domains and businesses that have less complications. Moreover, this taxonomy has different stockholders, including people, organizations, and systems. If the antecedent effort in the taxonomy is proven, subject matter experts could enhance the extent of privacy in the health cloud by verifying, evaluating, and revising this taxonomy.

(JMIR Form Res 2024;8:e38372) doi: 10.2196/38372

KEYWORDS
taxonomy; privacy; security; legal; cloud computing
**Introduction**

**Background**

Cloud computing is among the hottest core technical topics in the digital world. It has broad-ranging effects across IT, business, software engineering, and data storage. One of the main effects is an increase in capability. According to the National Institute of Standards and Technology (NIST) definition, “cloud computing is a model for enabling convenient, resource pooling, ubiquitous, on-demand access which can be easily delivered with different types of service provider interactions” [1,2]. Cloud technology can meet the requirements of the health care industry. It has some benefits like helping health organizations to reduce their costs by replacing and migrating all IT infrastructure, platforms, and software to the cloud, and providing integrated services across multiple organizations with delivery of better access to IT knowledge, resources, and services in a more technical and economical way.

The cloud in the health care context can increase medical record accessibility and make medical history available for individuals. Moreover, it can enhance cooperation among various stakeholders in the health industry through the integration of electronic medical information from dispersed locations and can reduce medical error complications to achieve patients’ lifesaving goals [3-9]. A health record includes a chronological account of an individual’s tests and treatments, and is a critical part of any health care lawsuit about health care procedures [10-14]. These documents can play an important role in guarding individuals based on medical ethics concerns, patients’ rights, and the bill of rights in each country [15-18]. Therefore, acceptance of any kind of computing technology with the combination of medical informatic applications can change the boundaries of health care organizations [1].

Despite all these benefits, the sharing and storing of sensitive electronic health data and personal health information through the cloud raise various privacy and security concerns [2,3]. An important concern is the probable release of health information to third parties who are not authorized to access the information. The distributed architecture of the cloud causes many difficulties like service accessibility, data reliability, data management, scalability, interoperability, privacy, security, data ownership, regulation and standards, organizational change, business process reengineering, etc [3-8].

The tradeoffs between the pros and cons of this technology depend on the approaches that governments introduce to address the privacy, security, and legal challenges in such a complicated domain like health care.

The challenges are magnified several times when there are no definite implications for some essential and technical concepts. For example, privacy in the digital world is a term with different meanings, which can clearly include a wide range of concepts and can completely differ from its traditional comprehension [3]. Moreover, some interpreters have explained this word as “vague and evanescent” [4]. Therefore, a lack of transparency in the privacy concept has made policy making difficult [19-21].

In these occasions, judges and legislators cannot obviously speak about privacy harms, especially at intersections with other fields like free speech, effective consumer transactions, and security, which are quite controversial. It is completely understandable that privacy and the related implications are complex and multidimensional, and are thus considered legal, philosophical, or even technical.

Furthermore, the involute definitions of privacy and cloud technological risks have stopped governments from adopting cloud technology in the health industry, and if cloud technologies are introduced in the health industry, issues like security, privacy, and legal obstacles play preventive roles. In other words, using cloud capabilities in the health industry without proper setups can lead to disastrous outcomes, such as blackmail and threats. As the relationship between the growth of eHealth and privacy value is quite obvious, it is necessary to create a balance between the pros and cons of these technologies in this new era. Health care stakeholders in different countries have taken many efforts to identify political and legal challenges in this domain and have developed appropriate supplements and technological infrastructure for the health cloud [22-24]. Moreover, the obstacles have led them to revise and redesign required concepts to make them compatible for the new paradigm [13,14,16,17].

A review of previous taxonomies appears necessary to obtain a better overall view. The most popular and famous taxonomies in this domain were analyzed by their features and attributes. The goals, use, and dimensions of each taxonomy in the privacy era are presented in Table 1.

Almost all reports in Table 1 declared that privacy is a multilateral concept that needs analysis from different sides. In addition, the reports indicated that the data value has grown incredibly, which could be the most valuable asset for organizations and individuals, but privacy-preserving concerns were illustrated as nonignorable challenges. Some reports only dealt with security services and presented those as privacy matters, while others only paid attention to legal issues or data features. Obviously, most of them were not specifically designed based on cloud technology features or health care scenarios.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Title</th>
<th>Goals</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barker et al [25]</td>
<td>Data privacy taxonomy</td>
<td>This taxonomy was designed for privacy features and had 4 dimensions, each of which had their own subcategories and demonstrated their relationships in data repositories, such as database management systems, which are used for data mining.</td>
<td>Purpose, Visibility, Granularity, and Retention</td>
</tr>
<tr>
<td>Antón et al [26]</td>
<td>Taxonomy of privacy requirements for websites</td>
<td>The authors analyzed websites to design an internet privacy policy taxonomy for goal mining and extraction of prerequisite goals from postrequirement text artifacts. The goals of privacy in this work are classified as privacy protection and privacy vulnerabilities.</td>
<td>Privacy protection goals: Notice and awareness, Choice and consent, Access and participation, Integrity and security, and Enforcement and redress; Privacy vulnerabilities: Monitoring, Aggregation, Storage, and Transfer of information phases</td>
</tr>
<tr>
<td>Asaddok et al [27]</td>
<td>Usability, security, and privacy taxonomy for mobile health applications</td>
<td>This taxonomy provided a model for mobile health applications, which were identified based on a study on products on the market. It had 3 dimensions, and each of them had their own subcategories (overall 10).</td>
<td>Usability, Security, and Privacy</td>
</tr>
<tr>
<td>Heurix et al [28]</td>
<td>Taxonomy for privacy-enhancing technologies</td>
<td>This taxonomy was designed to provide a classification method owing to the various features of privacy-enhancing technologies. The purpose was to cover various techniques, such as anonymization or encryption, with different application scenarios. Each of its dimensions had its own subsets.</td>
<td>Scenario, Aspect, Aim, Foundation, Data, Trusted third party, and Reversibility</td>
</tr>
<tr>
<td>Kotz [29]</td>
<td>Threat taxonomy for mobile health privacy</td>
<td>This work presented a taxonomy for mobile health privacy and emphasized mobility and networking with many risks. There was a focus on the effects that threats could have, and threats were organized based on their type.</td>
<td>Misuse of patient identities, Unauthorized access or modification of PHI, and Disclosure of PHI</td>
</tr>
<tr>
<td>Skinner et al [30]</td>
<td>Information privacy taxonomy for collaborative environments</td>
<td>This taxonomy had 3 dimensions, and each dimension was interrelated and had different influences over information privacy. These dimensions translated into 3 corresponding views of information privacy within a collaborative environment, like computation view, content view, and structural view.</td>
<td>Time, Matter, and Space</td>
</tr>
<tr>
<td>Stein [31]</td>
<td>Taxonomy of privacy</td>
<td>This work organized all kinds of harms and is one of the most well-known taxonomies in the field. Four different types of harmful activities covered by privacy were identified. Each activity type had its subactivities (n=16).</td>
<td>Information collection, Information processing, Information dissemination, and Invasion</td>
</tr>
<tr>
<td>Vatsalan et al [32]</td>
<td>Taxonomy of privacy-preserving record linkage techniques</td>
<td>Privacy-Preserving Record Linkage taxonomy is another study that provides an overview of techniques that allow linking of databases among organizations. These techniques provide privacy preservation at the same time.</td>
<td>Privacy aspects, Linkage techniques, Theoretical analysis, Evaluation, and Practical aspects</td>
</tr>
<tr>
<td>Zandesh et al [3]</td>
<td>Legal framework for a health cloud</td>
<td>This work was a systematic review that introduced a legal framework for the health cloud with 5 main pillars and 17 subcomponents, and defined the role of legal aspects in the reliability of eHealth.</td>
<td>Compliance, Data protection, Identity credential access management, Ownership, and Quality of service</td>
</tr>
<tr>
<td>Olla et al [33]</td>
<td>Mobile health taxonomy</td>
<td>This taxonomy had 8 categories under 3 main pillars owing to the application’s intended purpose.</td>
<td>Medical use cases, Technical modalities, and Consideration</td>
</tr>
<tr>
<td>Association for Computing Machinery [34]</td>
<td>Computing classification system from the ACM</td>
<td>This taxonomy was developed to organize papers received in the ACM Digital Library or events hosted by the ACM.</td>
<td>Cryptography, Formal methods and theory of security, Security services, Intrusion/anomaly detection and malware mitigation, Security in hardware, System security, Network security, Database and storage security, Human and societal aspects of security and privacy, and Software and application security</td>
</tr>
<tr>
<td>Computer Security Division/NIST [35]</td>
<td>Computer security resource center classification from the NIST</td>
<td>This classification was a significant reference for cybersecurity considerations that provided a comprehensive model for cybersecurity knowledge.</td>
<td>Security and privacy-specific research domains, Technologies, Applications, Laws and regulations, Types of activities, and Business sectors</td>
</tr>
<tr>
<td>IEEE [36]</td>
<td>IEEE taxonomy</td>
<td>IEEE developed a taxonomy to organize papers received in IEEE Xplore Digital Library or events hosted by IEEE.</td>
<td>Access control, Computer security, Cryptography, Data security, Information security, and Terrorism</td>
</tr>
</tbody>
</table>
This study has several implications. It redefines privacy with the concept about privacy. Therefore, a clear and precise taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy. The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.

### Problem Statement

Despite all previous studies, it appears that more efforts are needed to redefine the privacy concept in the health domain, especially in the cloud context. The nomenclature and classification confusion in privacy terminology prevent businesses from finding a comprehensive solution for the domain requirements [22-24]. It is worthwhile to note that taxonomy use is an effective approach. Regarding the research question, our attempts focus on reaching a comprehensive concept about privacy.

The main challenges are related to what we already know and what we need to know. Therefore, a clear and precise taxonomy would be helpful to identify the specifications of privacy in a dynamic environment and would help in conducting future research projects for evaluating its impacts. A taxonomy was developed in this study, and the study contributions are presented below.

**Study Contributions and Objectives**

This study has several implications. It redefines privacy with regard to the health cloud and focuses on identifying the main approaches to deal with the contributed factors and dimensions that rely on taxonomy designing.
legal, and the second layer has 15 components, with each of them having subcomponents (n=57). This taxonomy has some advantages like presenting the hierarchical root of concepts and the inherited features of taxonomies. The specific implementation was performed by selecting published English papers related to the concept of health cloud privacy from several databases and relying on predefined keywords and search strings, followed by a classification design through a qualitative content analysis approach.

Hence, this taxonomy could cover health industry requirements with its specifications like health data and scenarios, which are considered to be the most complicated among businesses and industries. Therefore, this taxonomy could be generalized to other domains and businesses with less complications.

Previous taxonomies in the privacy domain have also been covered in this article, and the designing steps of the new taxonomy are presented in the Methods section.

### Methods

#### Methodology Analysis

One of the main concerns in various disciplines is how to group disciplines based on taxonomies. Such a classification has given taxonomies a pivotal role for researchers and practitioners in investigations and businesses as it has enabled them to comprehend and analyze complex domains [42,43].

Covering both descriptive knowledge and prescriptive knowledge, design science also consists of taxonomies as a type of conceptual knowledge in its epistemology. The research goal at the conceptual level is essentialist: concepts and conceptual frameworks at this level aim at identifying essences in the research territory and their relationships [44].

The term taxonomy is different from other similar words. Compared with classification, in some literature, it refers to groupings that are derived based on empirical studies with involvement of cluster analysis and statistical techniques. This definition is also referred to as numerical taxonomy [45].

Taxonomy is also considered as a classification scheme [46], and it is possible to use the terms of classification scheme, taxonomy, and typology as substitutes of each other. A previous report mentioned 3 approach categories for taxonomy: inductive, deductive, and intuitive [43].

With respect to the inductive approach, empirical cases are taken into account. In the following step, they are analyzed so as to realize dimensions and characteristics in the taxonomy. In this type of analysis, a variety of statistical techniques, such as cluster analysis, or other less rigorous techniques are employed [47].

In the deductive approach, the taxonomy involves theory or conceptualization rather than empirical cases. The method uses a logical process that results from a sound conceptual or theoretical foundation in order to clarify dimensions and characteristics in the taxonomy. It is considered to be similar to the cladistics approach in biology [47]. The method may involve an analysis of empirical cases so that evaluation or even modification of the taxonomy can be performed.

The intuitive approach is considered in the case of necessity. The objects are categorized based on what a researcher comprehends. In this approach, the taxonomy is offered on the basis of the perceptions of a researcher. This technique is not explicitly used [47].

Our proposed privacy taxonomy is derived by the deductive approach. Thematic analysis, which is often called qualitative content analysis, is considered as the methodology for the implementation of the deductive approach and as one of the most favorable methodologies in taxonomy creation [19]. Content analysis, as a research method, is a systematic and objective means of describing and quantifying phenomena. It is also known as a method for analyzing documents. This research method is used for making replicable and valid inferences from the data to their context, with the purpose of providing knowledge, new insights, a representation of facts, and a practical guide to action. In most cases, those concepts or categories are applied to construct models, conceptual systems, conceptual maps, or categories [20].

This type of taxonomy development needs a complete literature review like a systematic or structured review because a systematic review relies on the following: definite time, definite inclusion criteria, definite information sources, and structured study selection according to predefined PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. This method has been described in the Taxonomy Development Characteristics subsection.

#### Ethics Approval

This study did not include human participants or animals, and thus, ethics approval was not required.

#### Eligibility Criteria

Published English papers (inclusion criterion 1) related to privacy aspects in the health cloud (inclusion criterion 2) were used to create a privacy taxonomy for the health cloud.

#### Information Sources

Designated databases, including Web of Science, IEEE Digital Library, Scopus, Google Scholar, and PubMed, were searched from April to June 2020 to identify relevant articles.

#### Study Selection

Study selection involved the following 5 different phases:

1. Health and computer science databases were chosen to cover all related publications. This step was applied to papers after 2010.
2. “Health cloud,” “privacy,” “medical ethics,” “data management,” “compliance management,” and “medical devices” were the keywords considered with divergent MeSH (Medical Subject Headings) terms.
3. Different search strategies on keywords were adopted for each electronic database to obtain more relevant papers.
4. The identified papers were screened based on the eligibility criteria using their titles, abstracts, and keywords.
5. Papers not eliminated in the previous phase were read completely.

**Taxonomy Development Characteristics**

The new taxonomy was developed on the basis of the deductive approach in 6 phases. The initial phase involved reading data intensively and assessing the papers. The second phase involved configuring the main dimensions to align with the research goals. This phase analyzed the results through Excel files. The third phase included data coding in main classes where the results were categorized. In the fourth phase, the main classes were structured and then arranged into components and subcomponents in an inductive manner, and subcomponents were designated to components. In the fifth phase, the results were categorically analyzed and then presented. The final phase involved reporting and documentation.

A total of 2042 papers were identified, of which 585 were discarded because of repetition in different databases (first layer of filtering according to inclusion criterion 1). The remaining 1457 papers were analyzed on the basis of their titles, abstracts, and keywords. Ultimately, the outcome was divided into 3 categories (second layer of filtering according to inclusion criterion 2).

In the second layer of filtering, initially, 150 papers were chosen according to the privacy, security, and legal domains in the health cloud, which were related to the first category of this work (Figure 1). By reading the full texts in this category, it can be judged that different headlines like compliance management, data management, data governance, information security services, medical ethics, patients’ rights, privacy issues, and technology considerations play important roles in privacy management discipline and influence privacy preservation in the health cloud environment. The identified domains provided a new map and road for the construction of the taxonomy of privacy. These domains led to the identification of probable dimensions, components, and subcategories in related contexts.

Subsequently, with the above-mentioned domains and according to the second layer of filtering (inclusion criterion 2), the rest of this work was conducted, which helped to group the 1307 remaining papers. The full texts of the papers were analyzed according to their details. The findings of the analysis phases showed that many related factors can influence privacy-preserving topics in the health cloud. Consequently, the identified factors were coded and grouped into direct and indirect groups for taxonomy creation, and they formed the second and third categories of the PRISMA guidelines. These factors influence privacy preservation in the health cloud. The findings of study selection are shown in a PRISMA flow diagram (Figure 1).

In this study, according to a previous report [43], attempts were made to cover all qualitative attributes, such as conciseness, robustness, comprehensiveness, extendibility, and explanatory ability. The aim was to develop a taxonomy based on a set of dimensions, with each including characteristics describing the objects comprehensively in a specific domain of interest.

Table 2 presents the 6 phases involving the formation and adoption of our taxonomy. The subsequent sections present a detailed introduction with respect to each dimension’s components and subcomponents. The privacy taxonomy can be provided in several different approaches, and hierarchical taxonomy is the most notable method.

![Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. IC: inclusion criterion.](https://www.jmir.org/2024/1/e38372/)
Table 2. Taxonomy development phases.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Thematic analysis method/qualitative content analysis method</th>
<th>Adoption in our work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reading data intensively and assessing papers</td>
<td>A total of 1457 papers were identified from among 2042 papers. The papers were analyzed on the basis of their titles, abstracts, and keywords, and their privacy, security, and legal features were chosen.</td>
</tr>
<tr>
<td>2</td>
<td>Configuring the main dimensions to correspond to the goals of this paper</td>
<td>The full texts of selected results in the previous phase were analyzed and processed by their details in an Excel spreadsheet. The outcome was divided into 3 categories: The first category involves the identification of privacy, security, and legal domains in the health cloud, and 150 related papers were identified. The second and third categories involve the identification of direct and indirect factors that impact privacy preservation in the health cloud. A total of 1307 remaining papers were examined by their contents.</td>
</tr>
<tr>
<td>3</td>
<td>Data coding in main classes</td>
<td>The most frequent and important features were categorized into 76 analytical categories.</td>
</tr>
<tr>
<td>4</td>
<td>Structuring the main classes and configuring components and subcomponents inductively on the material, and assigning subcomponents to components</td>
<td>The analytical categories were then synthesized into the taxonomy. The taxonomy requires a multidimensional and hierarchical structure, and each tier in the hierarchy inherits all attributes of the tier immediately above it. The highest level in the hierarchy has the greatest generality and vice versa. The subcomponents may be used to improve the domain concept under consideration and the relationships between the nodes and leaves in the hierarchy. Iterative processes can lead to taxonomy constructors. The privacy taxonomy provides a heuristic representation of hierarchies with 4 dimensions of privacy and branches in each dimension. This model allows for more specification of independent variables in the model development and with regard to the research objectives.</td>
</tr>
<tr>
<td>5</td>
<td>Performing category-based analyses and presenting the results</td>
<td>The taxonomy has 3 layers, of which the first layer has 4 main dimensions, including cloud, data, device, and legal. The second layer has 15 components, and each of them has subcomponents (n=57). This well-organized taxonomy has some advantages like presenting the hierarchical root of concepts and the inherited features of taxonomies.</td>
</tr>
<tr>
<td>6</td>
<td>Reporting and documentation</td>
<td>Finally, the taxonomy was derived and proposed from the abstraction of each of the dimensions.</td>
</tr>
</tbody>
</table>

Results

After analyzing the identified papers and considering taxonomy development, with respect to studies related to the first category of the method in the digital world, it was found that only documented rules and regulations did not comply with the privacy, security, and legal requirements in the health cloud. To be more precise, compliance alone cannot consider and resolve all the privacy, security, and legal requirements in such a dynamic environment like the cloud, and as mentioned before, some other headings like compliance management, data management, data governance, information security services, medical ethics, patients’ rights, privacy issues, and technology considerations play important roles. To cover all these domains and overcome previous deficiencies, a taxonomy of privacy, security, and legal issues in the health cloud was designed.

As illustrated in Figure 2, this taxonomy has 3 layers. Different features in this context were initially grouped into 4 dimensions, namely the cloud specification, legal aspect, data specification, and device specification in the context of privacy. This classification provided the first or most comprehensive level of generality in the taxonomy of privacy. Other factor identification was related to the next level of taxonomy, and the second and third levels of taxonomy creation and identification led to the introduction of direct and indirect factors for privacy preservation. Then, the basic building blocks or dimensions, components, and subcategories were realized with a qualitative content analysis. The second layer identified 15 components, with each of them having subcomponents (n=57). This model allows for more specification of independent variables in model development and with regard to research objectives.

The findings of this paper helped to process and define privacy by identifying a composite set of variables that represent to the extent possible the true nature of interventions and by incorporating the major dimensions of privacy and their constituent parts. Moreover, the findings led to the creation of a new conceptual diagram, which has been presented in Figure 3. The main outcomes or results of this taxonomy appear in this figure, which provides a definite boundary for each of the ambiguous terms like privacy, legal, and security. This figure displays conceptual coverage and overlapping boundaries of these terms in the digital health world, which are crucial for future research, policy making, and the actual management of privacy.

According to the proposed taxonomy, each circle has its subdomains. In the Discussion section, each dimension’s components and subcomponents are introduced in detail.
Discussion

Principal Findings

The details of each dimension’s components and subcomponents (Figure 2) are provided. The main characteristics included in the taxonomy are described and discussed to answer the research question, and an attempt was made to focus on reaching a comprehensive concept regarding privacy.

The question is as follows: Which dimensions and factors affect privacy taxonomy and should be considered in current health cloud projects or systems for privacy preservation?

As mentioned in the Results section, to provide a clear and precise taxonomy according to the method steps, selected papers were studied and analyzed deeply, which led to 4 new dimensions, namely cloud, legal, data, and device. All these dimensions were related to privacy specifications.

In the below sections, each dimension of the proposed taxonomy, and its components and subcomponents are described extensively to provide better understanding for audiences.

Implications

Dimension 1: Cloud

The first dimension of this taxonomy is the cloud, which incorporates all aspects of cloud computing technology. It is an evolving paradigm that is useful in the health care context and has an indirect impact on privacy. The cloud dimension has 3 main characteristics, each of which has its specialty: architecture, deployment, and communication. According to the NIST definition, the cloud can be defined based on its characteristics as follows: an architecture or service model, which is defined based on its limited taxonomy, and it can also be defined based on its deployment model with service delivery or business operation, which can affect its features [48]. It is worthwhile to mention that each state of these components will
affect the privacy of information in the cloud, which cannot be ignored. Furthermore, several methods of communication can be defined between the cloud providers and the cloud customers in the cloud. Each of them contains characteristics having an indirect effect on privacy. These aspects are grouped into 3 parts in Figure 4, with each containing subcomponents.

**Figure 4.** Components of the cloud dimension. IaaS: infrastructure as a service; PaaS: platform as a service; SaaS: software as a service.

**Architecture or Service Model**

There are several service models defined for the cloud, and their subcomponents constitute the first component of the cloud dimension [48-51].

Software as a service (SaaS) enables the client to receive services from applications where providers use cloud services to provide the services. It is important to note that the client cannot manage and control the cloud infrastructure, including networks, servers, operating systems, and storage, or even individual application capabilities.

Platform as a service (PaaS) enables the client to provide services on the cloud through consumer-created or acquired applications created using some programming languages, libraries, services, and tools in the cloud. The difference is that the client no longer manages and controls the cloud infrastructure, including networks, servers, operating systems, and storage, or even individual application capabilities. Hence, the client only controls the executed application and the configuration settings for the application-hosting environment.

Infrastructure as a service (IaaS) enables the client to provide processing, storage, network, and other fundamental computing resources, where the client can deploy and run arbitrary software including operating systems and applications. The client does not manage or control the underlying cloud infrastructure and has control over operating systems, storage, and deployed applications and possibly limited control of select networking components.

**Deployment Model**

The second subcategory of this component is the cloud deployment model [48,49].

Private cloud is used by a single organization that has different consumers and stakeholders. This infrastructure may be administered or handled by that organization, a third party, or their combination and may exist on or off the premises.

Community cloud is used by a specific community of consumers from organizations with shared concerns. This infrastructure may be administered or handled by one or more organizations in the community, a third party, or their combination and may exist on or off the premises.

Public cloud is provided for open use by the general public. This infrastructure may be administered or handled by a business, academic, or government organization or their combination. It exists on the premises of the cloud provider.

Hybrid cloud is composed of two or more distinct cloud infrastructure (private, community, or public), which remain unique entities. They are bound together by standardized or proprietary technology that enables data and application portability.
Communication

Regarding eHealth, providing health care services depends on several communication technologies. This is because each choice contains its characteristics, for which providing security requirements is very important. This section can be divided into the 3 subcategories of synchronicity, network design, and connectivity in terms of its details [22].

Synchronicity is employed to coordinate scheduling and technology. Depending on the schedule, telemedicine services can be provided in 2 modes. The first mode is “real-time,” and it refers to a situation in which the people involved in the care and the care providers are related at the same time with each other but in different location situations. The second mode is “store and forward,” and it refers to a donating situation in which the people involved in the care and the care providers are not connected at the same time. Both modes include different technological infrastructure, including video conferencing, telemetry, and remote sensing, as well as other modes of interactive health communication.

Network design/configuration contains the 3 modes of virtual private networks, open internet, and social networks, and in all of these, the information is posted and then shared. To effectively protect the confidentiality of the information of these states, different security settings are required.

Connectivity may be divided into wired and wireless, with different levels of bandwidth and the attendant speed and resolution or quality of service.

Dimension 2: Legal

According to the assessed studies, the second dimension of this taxonomy is the legal dimension, which can independently provide a framework of legal issues raised in the health cloud. The identified elements of the legal framework have a direct impact on information privacy, which include the 5 main scopes of compliance, data protection, identity credential access management (ICAM), ownership, and quality of service [3]. It should be mentioned that these scopes have a series of subcategories that have been explained in the below text. According to the research findings, privacy and legal issues are completely related and intertwined issues in terms of eHealth. The legal framework scopes are considered as the main components of this dimension (Figure 5).

Figure 5. Components of the legal dimension. ICAM: identity credential access management.

Compliance

The scope of compliance contains the 3 subscopes of standard, law/act/regulation, and policy/guideline [3].

Standard is a document confirmed through consensus by a recognized body that is provided for repeated and common use, and involves rules, guidelines, or characteristics for products or related processes and production methods in which compliance is not mandatory.

Legislation is comparable with statutory law. Legislation restricts the legal requirements as well as the cost or punishment for breaking the law. Most regulations are issued by governments [52].

Policy or guideline is a formal, brief, high-level report or proposal that indicates an organization’s principles, goals, objectives, and acceptable procedures for a topic [3]. Guideline is related to general instructions in order to achieve policy principles. It provides a framework to implement the required procedures.

Data Protection

The second scope of this dimension encompasses the details of data protection to provide the technical mechanisms of the requirements introduced in the first scope. Data protection is distributed into the 3 main classes of technical, administrative, and physical issues according to the NIST, Health Insurance
Portability and Accountability Act (HIPAA), and Certified Information Systems Security Professional (CISSP) [53].

Technical aims to define supply-related techniques, such as confidentiality, integrity, and nonrepudiation of cloud-based patient data. Confidentiality is the guaranteeing process that makes data property or information available or accessible only for authorized people or processes [54]. Integrity is the property to ensure the prevention of data or information tampering in an unauthorized manner. Nonrepudiation involves service guarantees to make an action taken undeniable.

Administrative involves security infrastructure with a management and development approach, and the implementation and support of systems are discussed [38].

Physical measures policies and procedures to protect the electronic information systems of an entity and the related buildings and equipment from natural and environmental hazards as well as unauthorized intrusion [53].

Identity Credential Access Management

The third scope of the second dimension includes data access management, which is a key factor in patients’ rights and medical ethics. Some pertinent fields like identification and authentication, authorization and access control, auditing and monitoring, and user training issues are also placed in this scope. This is a process in which a unique identity is defined for the person or system [53]. It is known as the first step in the access control process, such that it controls any activity based on the identity or entity of the user.

The process of identification and authentication identifies and authenticates the user, which is possible based on the elements and private data created by the user [53].

Authorization is the process of defining the resources and the level of access for the user [53].

System monitoring or auditing is the last loop of this cycle that plays an important role in recording the log of all the activities, events, and performances of the users who have access. Moreover, it is considered a security check [55], which is very important to identify problems and violations with accounts, access, information disclosure, and system operation.

Data Ownership

The fourth scope of this dimension is related to data ownership, which is responsible for concepts such as information ownership and responsibility. Information control not only speaks about the creation, modification, and other convolutional procedures of data, but also deals with the rights of individuals to grant or revoke their access to others [12].

The ownership of data in the cloud may rely on the nature of the stored data [12]. Data owners must be able to assess, control, and restrict their data during storage, use, and disclosure [56,57]. Nevertheless, the existing shortcomings in the implementation of these statements in the cloud are considered as some of the essential problems for implementing the cloud in the health sector [57]. This scope encompasses some subsopes like data location issues, third party issues, and patient consent.

Data location involves the storage of data. One of the points in the cloud is that data storage can be carried out in any places, even unknown ones.

Patient consent is derived from the ethical and basic principles of human and citizenship rights in terms of the patient’s discretion [58,59]. In this regard, the patient has the freedom to decide whether the tests and surgeries on the organs can be performed before any action [59-61].

Third party is considered as a cloud provider that does not have any role in the patient’s treatment process as a beneficiary. Nevertheless, it has access to all patient information that can cause several legal dilemmas.

Quality of Service

In the fifth scope of this dimension, some issues, such as contract, service availability, and interoperability, are stated, and this has been referred to as quality of service (QoS). It defines guaranteed levels of performance, availability, reliability, interoperability, throughput, performance, response time, etc, all of which are regarded as major factors influencing the quality of service in cloud computing [62].

Contract issues involve a service level agreement (SLA). This is a mutual agreement between cloud service providers (CSPs) and end users. Quality of service management systems monitor resources, storage, networks, virtual machines, service migration, and fault tolerance [63-65].

Availability involves principles ensuring that authorized users at a proper time have access to the data [53].

Interoperability involves the ability of the system to render services using multiple service providers while preserving the integrity of the data. This feature can be used for all kinds of clouds so that if migration to a different system is required, it can be seamlessly carried out [63,64].

Figure 6 illustrates the coverage of information security services by legal dimension elements in privacy taxonomy. It is impossible to preserve privacy without considering information security services in dynamic environments, such as the cloud, as these services can ensure benefits in terms of outsourcing the health records [3].
Figure 6. Compatibility between legal frameworks in a security service. ICAM: identity credential access management; SLA: service level agreements.

Dimension 3: Data

Data structures are critical in various cloud environments, such as data storage features, data processing methods, and data preserving solutions, designed for this dynamic ecosystem. The third major dimension of our proposed privacy taxonomy is related to data characteristics, which have been divided into the 5 subcategories of data type, data life cycle, data usability, data sensitivity, and data acquisition methods. Figure 7 depicts the structure of the data dimension, although the components of this dimension have an indirect effect on privacy.

Figure 7. Components of the data dimension.

Data Type

Any data related to health conditions, reproductive outcomes, causes of death, and quality of life are health data [66].

It is worthwhile to mention that health data can measure several criteria, such as clinical, environmental, and socioeconomic factors, both at the individual and population levels, including information about a person’s behavior related to his or her wellness. The accumulation of collected and utilized health data occurs when interacting with health care organizations. The collected data typically contain the received service types, the
results of those services, and the clinical outputs or information included in those services.

Health data can be classified into 2 structured or unstructured types. The structured type is a standard that can be simply exchanged between health information systems [66]. For example, a patient’s name, date of birth, or blood test result can be recorded in a structured data format. However, unstructured health data are not standard, unlike the structured type. Emails, audio recordings, or physician notes about a patient are examples of unstructured health data.

Advances in the digital world have improved the collection and use of health data and the databases in the health care industry, which have certain complexities. Overall, in terms of health care, the data can be classified based on the data type as follows [21]: alphabetical data/textual data/narrative data, numerical data/measurements/coded data, signal data, images/graphic data/pictures, voice, and videos/film.

**Data Life Cycle**

The second scope of the third dimension in the designed privacy taxonomy is data life cycle, which contains 7 phases [67], each including its requirements for privacy. This cycle encompasses the following phases: data generation, data transmission, data storage, data access, data reuse, data archiving, and data disposal. Data life cycle is comparable with the cloud requirements [68-70].

Data generation involves CSPs receiving requests from their users to generate the related data so that they can assign their access control policies.

Data transmission involves CSPs generating a secure transmission channel to verify user data reliability. Besides, they use encryption methods and the digital certificate mechanism between servers.

Data storage involves the role of CSPs to ensure the conformity of the data in the right place according to the agreements and rules.

Data access involves the CSPs ensuring the validity of users’ identity to protect them from spoofing and verifying the proper execution of the data access policy.

Data reuse can lead to leakage of sensitive or personal data, which is a reason for not providing services in the cloud. In the big data era, data sharing has made this phase quite primitive.

Data archiving involves 3 main operations, including band encryption, long-range storage, and data retrieval.

Data disposal is mainly aimed at placing the data completely and effectively in the cloud and removing unnecessary parts.

**Medical Data Usability**

Medical data have very diverse functions, including personal interests, public health, medical research, and development [21]. The use of the data in applications is categorized into 2 modes of primary and secondary. Primary is a state where the collected medical data are employed to provide medical care. Secondary is a state where the collected medical data are employed for purposes except care.

Here, it is worth noting that digitization and updating based on medical information technology have increased the use of medical data at both primary and secondary levels [21,71]. The data in the patient’s medical file appear in 1 of the following 3 formats based on their origin and applications: demographic data (identification data/date of birth, admission, discharge, biometric identifiers, phone number, and health record number); clinical data (clinical results/images/summaries, medical data, case management, public health data, performance data, and referral management); and administrative data (insurance documents/financial information and nonclinical data focused on record keeping surrounding a service, such as hospital discharge information; it can be part of an electronic health record as well; claims data, which include information regarding insurance claims).

**Data Sensitivity**

One of the important points in privacy preservation is the grading of data regarding their degree of importance. It is performed according to data sensitivity to classify the data based on their sensitivity and the extent of their impact on the patient and the health organization. Accordingly, these importance-based data cannot be disclosed, changed, or destroyed without permission. Classification of the database helps to specify the level of security required by the data. The data are categorized based on their importance level as presented below [72].

Restricted sensitivity of data involves a situation where the data have high sensitivity (restricted sensitivity), and unauthorized access and disclosure of the data may result in significant risks, leading to severe or disastrous adverse effects on the operations and assets of an organization or individual, particularly a patient or health care institution. This level of sensitivity needs the highest level of security controls that must be applied to restricted data.

High sensitivity of data involves a situation where the data have high sensitivity, and unauthorized access and disclosure of the data may alter or destroy the data, leading to serious adverse effects on the operations and assets of an organization or individual, particularly a patient or health care provider. This level of sensitivity needs a reasonable level of security controls that should be applied to private data.

Moderate sensitivity of data involves a situation where the data have moderate sensitivity, such that unauthorized access and disclosure, alteration, or destruction of the data would result in moderate risks for the operations and assets of an organization or individual, especially a patient or health care institution.

Low sensitivity of data involves a situation where the data have less sensitivity, and unauthorized access and disclosure, including alteration or destruction of the data, would lead to a limited risk to the operations and assets of an organization or individual, especially a patient or health care institution, or there will not be any risks.

**Data Acquisition Methods**

When emerging health services arise from the context of modern technologies, such as the cloud, mobiles, wireless multimedia sensor networks (WMSNs), and Internet of Things (IoT), some
new scenarios are raised for health care services. These scenarios consist of patient care in hospitals, patient care at home, and self-care scenarios, with each representing a special type. Hence, the protection of data privacy in each scenario requires its characteristics. The important point in terms of privacy preservation in any of these scenarios is to know how to collect the data. Overall, there are 2 collection methods in all these scenarios [21].

In the *manual* method, data are described subjectively or objectively by the patient and then inferred by health care providers. Then, these data are entered into health information systems manually through personal portals. In the *device base* method, several medical devices (either wired or wireless) collect data. Subsequently, the collected data are sent to applications for processing to be used by health care providers. Evidently, different types of devices will be fully described in the next section since they play substantial roles in ensuring privacy.

Figure 8. Components of the device dimension. PDA: personal digital assistant.

**Dimension 4: Device**

The last dimension identified for the taxonomy of privacy is concerned with devices and their features because, with the advancement of technology, data collection is practically entrusted to devices. Thus, ensuring data privacy is the most important concern of stakeholders in terms of diversity of use.

A medical device is an outfit used to evaluate or diagnose a medical condition [61], for example, electrocardiography machines, ultrasound machines, x-ray machines, different sensors, wireless sensors, and mobile health apps that run on smartphones. Ensuring data privacy on these devices has been an issue in many studies, which makes it challenging in terms of the cloud. As a result, regarding privacy in the cloud, it is essential to consider the features of medical devices. Certainly, the elements defined in this section will have an indirect impact on information privacy in the health cloud. As shown in Figure 8, the device dimension is divided into 2 subcategories: device types and application types.

**Device Types**

WMSNs involve wireless sensors, which are some of the most common devices in the medical world. It is considered as the smallest network and has unique features such as large-scale implementation, portability, and reliability [73]. It should be mentioned that the sensor network encompasses a set of independent nodes with low cost, energy, and memory, and limited computing power [73]. The health care industry has experienced a dramatic transformation with the use of WMSNs [74]. The main aim of WMSNs is to collect and transfer environmental data to central databases or remote locations.

IoT is another popular tool in recent years [65,75], which has created a new technological paradigm in the health care industry. In eHealth, IoT has provided the possibility of interaction and communication between “things” via the internet. In future health care circumstances, IoT will connect subjects and health care professionals seamlessly [76,77].
These technologies can be used for eHealth applications, such as computer-assisted rehabilitation, early detection of medical issues, and emergency notifications. However, there is an issue because several factors limit the use of these technologies. The most important factor is legal issues related to the privacy and security of the data transmitted [78-81].

Smartphones have become an integral part of life. Thus, they can act as a gateway between the wireless body area network (WBAN) and IoT [82-84]. Essentially, the smartphone's sensor data or high-resolution camera images are sampled, processed into medical information, and displayed [84]. Using smartphones for medical purposes can be very useful because millions of people have their own smartphones today and can access medical applications designed for health care [61].

Tablets/personal digital assistants have the same applications as smartphones, acting as a gateway to collect medical data beyond providing accessibility to reference textbooks [85].

Personal computers play a pivotal role in information management. Computers potentially alter the traditional approach that physicians use to communicate with patients [86] and have an essential role in information management. In other words, they can change the traditional ways of providing health services to patients and replace them with novel innovative methods [86].

All of the above-mentioned tools with increasing use in medicine must comply with certain features to ensure the privacy of data since ignoring these features can cause some irreparable damage.

**Application Types**

Care processes across virtually all basic medical specialties and subspecializations associated with disease entities, sites of care, and treatment modalities are included. The vast array of these applications and the complexity of the medical practice and medical specialization are listed separately [22]. The second device subcategory is related to application types.

Basic specialties include content areas around specific diseases, including diabetes, stroke, and posttraumatic stress disorder, and such applications have been developed. Moreover, programs may differ by the site of care, including the intensive care unit, outpatient psychiatry unit, emergency department, and home. Some programs were organized around specific treatment modalities such as rehabilitation and pharmacy. Over 40,000 health applications have been used on smartphones [61]. The World Health Organization has classified mobile health applications as follows [70]: toll-free emergency, health call centers, public health emergencies, mobile telemedicine, information initiatives, appointment reminders, community mobilization, treatment compliance, patient records, surveillance, health surveys, patient monitoring, decision support systems, and awareness raising [20]. Depending on the site of care, these applications have several privacy requirements that must be identified and met. In other words, the privacy of a user's data in the devices depends on the security of the designed computer programs.

**Comparisons to Existing Literature**

From these dimensions, it is understandable that the legal dimension and its subcomponents have direct influence on privacy and other dimensions like data, device, and cloud along with their subcomponents, as well as an impact on privacy preservation concerns in the cloud environment.

In contrast with other taxonomies, this taxonomy sides with health data specification and cloud considerations, which appear critical. Therefore, this article first tries to adopt the privacy taxonomy in the cloud context, especially in the health cloud, and the remainder is dedicated to redefining privacy terms with new details.

The health care domain has the most complicated scenarios and most varied data among businesses. Thus, when a taxonomy fits with its requirements, the taxonomy might be appropriate for other domains, businesses, and scenarios that are complex. In fact, the user of the model should exercise judgment as to the appropriate level of detail necessary to test the target hypothesis.

**Usability and Experimental Use of This Taxonomy**

This well-organized taxonomy has some advantages like presenting the hierarchical root of concepts and inherited features of taxonomies. It provides a heuristic representation of hierarchies with 4 dimensions of privacy and the branches of each dimension. This model allows for more specification of independent variables in model development and with regard to research objectives. Experimental use of this taxonomy depends on the following stages: scenario clarification stage, device and system specification stage, data specification stage, and privacy mapping stage.

In the first stage, the specification of cloud-based scenarios should be clarified. For example, which service model and cloud deployment have been chosen for health care delivery and which communication method has been chosen to connect the stakeholders individually or with each other (synchronized or unsynchronized; wired or wireless).

In the second stage, the use of medical devices and application types for data collection should be prominent and transparent to users because each device has its specific privacy requirements.

In the third stage, data specifications collected in each scenario should be explicated because the veracity in data specifications can lead to variations in privacy strategies. For instance, in one scenario, electrocardiography data detected by the WMSN and transferred via a designated mobile health app to the cloud for storage, processing, and use will have special privacy requirements. In another self-care scenario, subjective data that are just entered through a cloud-based personalized portal need a different set of privacy requirements.

In the fourth stage, to ensure privacy preservation in all means, the identified features in other stages should match with legal components from the proposed taxonomy. For example, proper corresponding security services like authentication, authorization, auditing, confidentiality methods, integrity, and nonrepudiation methods should be chosen for each type of health
care scenario in the digital world. Through these approaches, stakeholders can trust eHealth.

This taxonomy generally has 2 layers of stakeholders (people and organizations, and applications and systems).

The first layer involves people and organizations, including patients; health cloud and general cloud providers; health care providers (eg, physicians and nurses); health care organizations (eg, hospitals, laboratories, drug stores, and physicians’ offices); cloud app developers and vendors; health domain stakeholders (eg, insurance companies and financial organizations); researchers and practitioners working in areas like health, cloud, data management, security, and privacy; medical ethics authorities; organizations planning to design and deploy cloud services and migrate to cloud platforms and services; governments and legislation bodies; and national or international standardization bodies. These groups, according to the scenario clarification stage, device and system specification stage, and data specification stage, map their privacy preferences with respect to the proposed privacy taxonomy.

The second layer involves applications and systems that are affected by this taxonomy, including patient assessment systems; telemedicine systems; medical imaging systems; public health systems; hospital information systems; clinical information systems; health data secondary use systems; teleconsultation systems; self-care systems; and medical device and wireless system producers (WMSN, IoT, etc). These systems by their provisions can meet privacy requirements according to the proposed privacy taxonomy.

Considering the above-mentioned stakeholders, among the main approaches to deal with privacy challenges, identifying the contributing factors and dimensions can be helpful to manage this domain.

Limitations of the Study and Future Work

This study has some limitations. The interchangeable use of some related terms like “security,” “privacy,” and “legal” made the close assessment of articles difficult, and it was challenging to obtain findings from related comprehensive articles with regard to health industry scenarios.

An attempt was made to include English papers; therefore, the results must be considered within the scope of the English literature and studies in a specific interval. Any papers published before or after the search interval were not included; however, there is always the possibility of missing some relevant information or bias.

Future studies can be conducted to identify or propose definite standards and requirements for privacy preservation in each subcategory of known dimensions. It is hoped that the proposed taxonomy will not only clarify nomenclature proliferation in privacy for the health cloud or eHealth, but also provide a useful guide for research and policy making.

This taxonomy is not a finished product and needs more attention with regard to development and improvement. The process has been initiated with the hope that others in the field will be interested in it and complement the privacy taxonomy in the health cloud. Furthermore, this taxonomy can be considered as the subject matter for experts in various domains of privacy for assessment, testing, revision, and verification.

Conclusion

This research was conducted to identify the factors affecting privacy in the health cloud and classify them to provide a unique and comprehensive taxonomy through the investigation of related papers. It redefines the health cloud privacy term by using a deductive approach.

The proposed taxonomy tries to provide the true and full perspectives of the intervention, management, and handling of other variables, as well as itemize the expected outcomes and determine how best to assess them, thus clarifying the units of analysis in health cloud privacy research.

The subscribed elements have been classified into the 4 main dimensions of cloud, legal, data, and device. Moreover, since taxonomy designing is an iterative process, 15 components and 57 elements were added to these 4 main dimensions in 3 layers.

Among all these elements, those classified in the legal dimension had a direct impact on data privacy in the cloud. However, other elements will have an indirect impact on ensuring data privacy in the cloud.

In the second step, this taxonomy tried to clarify the privacy concept in eHealth, which is a multidisciplinary context, and tried to remove the ambiguities between existing definitions in the field of security and define a clear boundary for the words. This led to the distinction and clarification of the overlapping and vague structure of related concepts, and privacy was defined by identifying the discrete sets of variables representing specific privacy configurations and definitive boundaries for “security,” “privacy,” and “legal” terms, which are crucial for future research, policy making, and the actual management of privacy. Therefore, users can have a more accurate definition of the concepts in this field in the future.

This taxonomy is designed to satisfy the needs of emerging technologies, such as mobile health, health IoT, telemedicine, etc, which use cloud devices in their infrastructure. Moreover, it can be considered as supplementary classification and a reference for current privacy, security, or technological taxonomies.

Hence, this taxonomy can cover health industry requirements with its specifications like health data and scenarios, which are considered as the most complicated among businesses and industries. Therefore, the use of this taxonomy could be generalized and customized to other domains and businesses that have less complications.

This paper has also reviewed the most popular previous taxonomies in the privacy domain.
Acknowledgments
I thank the Health Information Department of Allied Medical Sciences for providing very close support, and useful feedback and advice.

Conflicts of Interest
None declared.

References


27. Asaddok N, Ghazali M. Exploring the usability, security and privacy taxonomy for mobile health applications. 2017 Presented at: International Conference on Research and Innovation in Information Systems (ICRIIS); July 16-17, 2017; Langkawi, Malaysia. [doi: 10.1109/ICRIIS.2017.8002472]


29. Kotz D. A threat taxonomy for mHealth privacy. 2011 Presented at: Third International Conference on Communication Systems and Networks; January 04-08, 2011; Bangalore, India. [doi: 10.1109/COMSNETS.2011.5716518]


34. ACM Computing Classification System. Association for Computing Machinery. URL: https://dl.acm.org/ccs [accessed 2024-01-04]


38. International Federation for Information Processing. URL: https://ifip.org/ [accessed 2024-01-04]


57. Bernsmed K, Cruzes D, Jaatun M, Haugset B, Gjere E. Healthcare services in the cloud -- obstacles to adoption, and a way forward. 2014 Presented at: Ninth International Conference on Availability, Reliability and Security; September 8-12, 2014; Fribourg, Switzerland. [doi: 10.1109/ares.2014.28]


68. Bernsmed K, Cruzes D, Jaatun M, Haugset B, Gjere E. Healthcare services in the cloud -- obstacles to adoption, and a way forward. 2014 Presented at: Ninth International Conference on Availability, Reliability and Security; September 8-12, 2014; Fribourg, Switzerland. [doi: 10.1109/ares.2014.28]


80. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. URL: https://iris.who.int/handle/10665/44607 [accessed 2024-01-04]


Abbreviations

- CSP: cloud service provider
- IoT: Internet of Things
- NIST: National Institute of Standards and Technology
- PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- WMSN: wireless multimedia sensor network

©Zahra Zandesh. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Evaluating the Intensity of Exposure to MTV Shuga, an Edutainment Program for HIV Prevention: Cross-Sectional Study in Eastern Cape, South Africa

Sarah Mulwa¹, MSc; Venetia Baker¹, MSc; Cherie Cawood², PGDip; David Khanyile², BA; Dominique O'Donnell², BSocSci; Sophie Sarrassat¹, PhD; Simon Cousens¹, DipMathStat; Isolde Birdthistle¹, PhD

¹Faculty of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom
²Epicentre Health Research, Durban, South Africa

Corresponding Author:
Sarah Mulwa, MSc
Faculty of Epidemiology & Population Health
London School of Hygiene & Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 20 7636 863
Email: sarah.mulwa@lshtm.ac.uk

Abstract

Background: MTV Shuga is an edutainment campaign designed to equip young people with knowledge, motivation, and informed choices to protect themselves from HIV infection. From 2019 to 2020, a total of 10 episodes of a new dramatic series, MTV Shuga “Down South 2” (DS2), were broadcast via television and the internet, alongside complementary media activities.

Objective: This study aims to investigate whether the intensity of DS2 exposure was linked with positive HIV prevention outcomes in a setting with high HIV prevalence and relatively low levels of HIV testing.

Methods: We analyzed data from a web-based survey of participants aged 15 to 24 years in South Africa in 2020. The survey was promoted via social media platforms of schools, universities, and communities in Eastern Cape, South Africa. The primary exposure of interest was the intensity of exposure to DS2, measured by the number of episodes of DS2 watched on the television or the internet or listened to on the radio (out of 10 episodes). Individuals who had not watched or listened to any DS2 episode were classified according to other MTV Shuga content that they had accessed. We estimated associations between the intensity of DS2 exposure and HIV-related outcomes, including knowledge of HIV status, awareness of HIV self-testing (HIVST) and pre-exposure prophylaxis (PrEP), uptake of HIVST, and demand for HIVST and PrEP, adjusting for potential confounders using multivariable logistic regression.

Results: Among the 3431 survey participants, 827 (24.1%) were exposed to DS2. Specifically, 18.1% (622/3431) watched or listened to only 1 DS2 episode, and 2.4% (82/3431), 1.7% (58/3431), and 1.8% (62/3431) watched or listened to 2 to 4, 5 to 7, and 8 to 10 DS2 episodes, respectively. Increasing the exposure to DS2 was associated with improvements in most outcomes. Exposure to multiple episodes (eg, 2-4, 5-7, and 8-10) was associated with successively higher odds of knowing one’s HIV status, awareness of PrEP and HIVST, and uptake of HIVST compared with no MTV Shuga exposure, albeit with statistical uncertainty around some estimates. The interest in using HIVST or PrEP was high overall (>80%), with no measurable differences by DS2 intensity.

Conclusions: We found evidence consistent with a dose-response relationship between MTV Shuga DS2 exposure and outcomes, including knowledge of HIV status, awareness and uptake of HIVST, and awareness of PrEP among young people in Eastern Cape. This indicates that greater engagement with a youth-focused edutainment campaign can improve HIV testing and prevention options in a setting and population with high need. However, only a few participants accessed multiple DS2 episodes despite its availability on multiple media platforms. We conclude that there is potential to benefit more young people by increasing access to and interest in the show.

(JMIR Form Res 2024;8:e44111) doi:10.2196/44111
KEYWORDS
young people; media; evaluation; dose-response; edutainment; HIV prevention; mobile phone

Introduction

Background

Over the course of the HIV/AIDS epidemic, edutainment and mass media campaigns have played an important role in health promotion, with some successfully raising awareness of HIV transmission, prevention, and treatment as well as contributing to increases in condom use [1-10]. To reach their target audiences, campaigns have used different modes of delivery such as print (eg, comic books), audio (eg, radio), video campaigns (eg, television), celebrity influencers, the use of peers or community workers, or a combination of different strategies [1,10]. As technology and the internet have become more accessible, edutainment campaigns have expanded to incorporate digital solutions, such as social media, mobile phone apps, websites, and on-demand streaming channels, to achieve broader reach [3,11].

Progress in reducing new HIV infections has been reported in many countries [12]. However, gaps within the HIV prevention cascade persist among young people [13,14], a key target population for HIV prevention in many African countries. In South Africa, for instance, the analysis of survey data indicates that knowledge of HIV status among individuals aged 15 to 24 years has steadily increased from 19% in 2005 to 59% in 2017, but this proportion is still low compared with older age groups [15]. HIV testing is a crucial step in HIV prevention and treatment programming [12]. As a supplement to provider- or client-initiated HIV testing and counseling, HIV self-testing (HIVST) can help reduce the testing gap among younger individuals who may prefer convenience and privacy and who may forego or be unable to access health care facilities [12,16]. In addition, awareness and uptake of pre-exposure prophylaxis (PrEP), which is now recommended as part of combination prevention for individuals at increased risk of HIV acquisition, remains low among young people [17]. As new HIV prevention options become available, edutainment campaigns can be adopted to enhance awareness through messaging that resonates with young people. This can, in turn, create demand for, facilitate uptake, and support the effective use of such options among young people.

Evidence of the impacts of edutainment campaigns that rely on dramatized series among young people has grown in recent years but remains limited [1,6,8,9]. MTV Shuga is an edutainment campaign designed to equip young people with knowledge and motivation to protect themselves from HIV infection while navigating healthy relationships. Shuga uses compelling characters and storylines to disseminate messages on HIV prevention through a series of parallel but interlinked storylines, which are offered over multiple episodes and media channels. From 2019 to 2020, a new series of MTV Shuga called Down South 2 (DS2), comprising 10 episodes, was produced in South Africa. DS2 episodes were broadcast via television and the internet alongside complementary media and offline activities, which included a documentary, peer-led discussions, DS2 graphic novels, and community-based viewing events [6,18]. The producers of DS2 conducted formative research through focus group discussions with young people to develop and validate the content, storylines, character development, and scripts, as they did for all previous MTV Shuga series [6,19,20]. Briefly, DS2 storylines largely revolve around characters who have recently left high school and are navigating life and the challenges that many young people face, including financial hardship, family conflict, and sexual relationships. Detailed descriptions of example storylines for DS2 have been documented previously [19]. In a recent evaluation of the DS2 series among young people aged 15 to 24 years in South Africa, we found that young people who engaged with the DS2 campaign (any component) were more likely to know their HIV status, use HIVST, and be aware of PrEP compared with those who did not engage with DS2 [6]. However, this analysis did not explore the intensity of DS2 exposure and whether greater exposure is linked with greater impacts.

In the absence of randomization, exploring whether the impact of an intervention differs across different intensities of exposure can support the plausibility of a causal association between the intervention and the outcomes [21,22]. Although little information exists regarding associations between the amount of exposure to edutainment campaigns and HIV-related outcomes among young people [1,8], published evidence from mass media campaigns in general populations suggests that high exposure intensity is associated with better health outcomes. In an evaluation of a weekly television soap opera on HIV/AIDS in Côte d’Ivoire by Shapiro et al [23], women who had watched ≥10 episodes (out of 20) were more likely to use condoms compared with nonviewers, whereas no effect was seen among women who watched <10 episodes. The authors concluded that repeated exposure to relevant content through continuous engagement partly contributed to the observed effects [23]. In an evaluation of an intervention to address domestic violence in South Africa (Soul City campaign, “SC4”), participants in rural residences with high exposure to SC4 television media (accessed ≥9 out of 13 episodes) were more likely to “do something to stop domestic violence” compared with those with no or low (<5 episodes) exposure, with little effect seen among those with moderate exposure (5-8 episodes) [24].

Objectives

In this study, we investigated whether the intensity of DS2 exposure was linked with differential impacts on HIV prevention outcomes. Understanding and documenting such effects can offer insights to implementers and program designers on ways to maximize engagement and benefits for adolescents and young adults.

Methods

Study Setting, Sample, and Data Collection

The analysis for this paper was based on quantitative data from a mixed methods evaluation study conducted in 2020. This
study aimed to evaluate the impacts of engagement with DS2 on HIV-related outcomes among young people in Eastern Cape, South Africa. The quantitative component used a self-administered web-based survey hosted on a website free of internet data charge for users and captured sociodemographic characteristics; exposure to MTV Shuga content (eg, how young people engage with DS2 and how many episodes they watched or listened to); and outcomes including knowledge of HIV status, awareness and uptake of HIVST, and awareness of PrEP. The survey was promoted via social media platforms; through targeted advertisements; and via social media accounts of schools, universities, and communities in the Eastern Cape. To avoid the risk of SARS-CoV-2 transmission, all study activities were conducted remotely. Additional details of the mixed methods evaluation are documented elsewhere [6].

**Measures**

**Exposure Variables**

In this study, the primary exposure of interest was the intensity of exposure to DS2 dramatic series, measured by the number of episodes of DS2 watched on television or the internet or listened to on the radio (out of 10 episodes). Some individuals had not watched or listened to any DS2 episodes but accessed other formats of Shuga, including Down South 1 (DS1, the first series of Down South), which preceded DS2. Rather than grouping these individuals with those not exposed to any Shuga content, we classified them based on the specific MTV Shuga content they had accessed. These additional categories captured exposure to other MTV Shuga content, for which the intensity could not be inferred. On the basis of this definition, seven exposure categories are analyzed in this paper: (1) no exposure to MTV Shuga content, (2) exposure to any MTV Shuga format other than the DS2 dramatic series, (3) exposure to DS1 series, (4) exposure to only 1 DS2 episode, (5) exposure to 2 to 4 DS2 episodes, (6) exposure to 5 to 7 DS2 episodes, and (7) exposure to 8 to 10 DS2 episodes.

To understand with whom young people watched or listened to DS2 and whether this influenced the intensity of DS2 exposure and the impacts of exposure, we generated a secondary exposure variable with four categories as follows: watched or listened to DS2 (1) alone, (2) with peers only (eg, friends or partners), (3) with parents (eg, mother, father, grandparents and siblings or peers), and (4) with siblings (eg, siblings only or siblings and peers).

**Outcome Variables**

Outcomes included knowledge of HIV status, awareness and willingness to use HIVST and PrEP, and the uptake of HIVST. Furthermore, we analyzed 3 sexual behavior outcomes (ever had sex, had sex in the past 12 months, and condom use during last sex with current or last partner [in the past 12 months]). The definition of each of the outcomes is summarized in Multimedia Appendix 1. All measures were self-reported.

**Analysis**

We summarized the frequencies and proportions of respondents reporting each of the above mentioned outcome measures based on the intensity of DS2 exposure. We estimated the associations between the intensity of DS2 exposure and each outcome using multivariable logistic regression models, adjusting for potential confounding variables. To limit the number of confounding variables adjusted for in the regression models, the selection of covariates in this study was informed by an earlier analysis that assessed the effects of binary exposure to DS2 (yes or no) on the abovementioned outcomes [6]. The initial set of confounding variables was identified using a directed acyclic graph to represent the hypothesized causal relationship between exposure to DS2, study outcomes, and other sociodemographic characteristics [6]. Only variables associated with each outcome at P≤.10 in the previous analysis were included in this analysis. From these regression models, we present the unadjusted and fully adjusted odds ratios (aORs) with their respective 95% CIs. To assess evidence of a dose-response relationship between DS2 intensity and the odds of each outcome, we compared 2 multivariable models: in the main model, the intensity of DS2 exposure was included as a categorical variable (model 1), whereas the other model assumed a linear relationship (model 0). The results from model 0 indicated whether the data were consistent with a linear trend, whereas a likelihood ratio test comparing model 0 with model 1 provided additional information regarding whether the relationship was more complex than linear [25]. All multivariable logistic regression models were restricted to individuals with nonmissing responses for each outcome. This decision was informed by the earlier analysis, which showed similar findings between complete case analysis and imputation methods [6]. In a secondary analysis, we assessed (1) whether the distribution of intensity of DS2 exposure varied by who young people watched or listened to DS2 with using χ² tests and (2) whether these variations had differential effects on outcomes of interest following the regression approach described previously. This secondary analysis was conducted only among individuals who said they had watched or listened to DS2.

**Ethical Considerations**

Ethics approvals were obtained from the Biomedical Research Ethics Committee at the University of KwaZulu-Natal (Ref: BREC/00000477/2019), the London School of Hygiene and Tropical Medicine (Ref: 17996), and the World Health Organization (Ref: ERC.0003283). All participants aged ≥18 years provided web-based consent, and participants aged <18 years provided their informed assent, with their parents or guardians providing informed consent [6]. To ensure participants’ confidentiality, we did not collect any identifying information other than mobile phone numbers for those who completed the survey to facilitate transfer of mobile data credit of ZAR50 (approximately US$5). We de-identified the data prior to conducting the analysis.

**Results**

**Characteristics of the Study Sample**

A total of 4145 records from the web-based survey were created by the users. In total, 82.8% (3431/4145) of the records were taken forward for analysis after removing records without full consent (407/4145, 9.8%) or sex information (144/4145, 3.5%) and likely duplicates (163/4145, 3.9%). We identified potential
duplicates using a combination of date of birth and mobile phone numbers. Respondents were predominantly female (2020/3431, 58.9%), aged 20 to 24 years (2352/3431, 68.6% vs 1079/3431, 31.5% aged 15 to 19 years), enrolled in education (2857/3431, 83.3%), and resided in urban settings (2923/3431, 85.2%; Table 1). Household ownership of media assets was high, with proportions ranging from 56.8% (1949/3431) for computers or other digital devices to 82.5% (2832/3431) for televisions. Most respondents had their own smartphones (2922/3431, 85.2%), and 50.8% (1744/3431) had their own computer. Digital media engagement was high in the study population: 86.1% (2953/3431) reported using the internet and social media platforms at least once a week, and 74.1% (2542/3431) and 62.4% (2141/3431) watched television or listened to the radio at least weekly, respectively. A detailed summary of the study population has been described elsewhere [6].
Table 1. Sociodemographic characteristics of the survey participants overall and by intensity of DS2 exposure (N=3431).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Exposure categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No MTV Shuga exposure</td>
</tr>
<tr>
<td>Total (N=3431), n (%)</td>
<td>1944 (56.7)</td>
</tr>
<tr>
<td>Age group (y), n (%)</td>
<td></td>
</tr>
<tr>
<td>15-19 (n=1079)</td>
<td>540 (50.1)</td>
</tr>
<tr>
<td>20-24 (n=2352)</td>
<td>1404 (59.7)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male (n=1317)</td>
<td>851 (64.6)</td>
</tr>
<tr>
<td>Female (n=2020)</td>
<td>1018 (50.4)</td>
</tr>
<tr>
<td>Other (n=94)</td>
<td>75 (79.8)</td>
</tr>
<tr>
<td>Schooling and employment, n (%)</td>
<td></td>
</tr>
<tr>
<td>In school† (n=726)</td>
<td>326 (44.9)</td>
</tr>
<tr>
<td>TVET‡ (n=967)</td>
<td>764 (79)</td>
</tr>
<tr>
<td>University (n=1164)</td>
<td>513 (44.1)</td>
</tr>
<tr>
<td>Employed (n=106)</td>
<td>57 (53.8)</td>
</tr>
<tr>
<td>Unemployed (n=357)</td>
<td>212 (59.4)</td>
</tr>
<tr>
<td>Unknown (n=111)</td>
<td>72 (64.9)</td>
</tr>
<tr>
<td>Language spoken at home, n (%)</td>
<td></td>
</tr>
<tr>
<td>English (n=259)</td>
<td>123 (47.5)</td>
</tr>
<tr>
<td>isiXhosa (n=2745)</td>
<td>1592 (58)</td>
</tr>
<tr>
<td>Zulu (n=224)</td>
<td>106 (47.3)</td>
</tr>
<tr>
<td>Other§ (n=203)</td>
<td>123 (60.6)</td>
</tr>
<tr>
<td>Province, n (%)</td>
<td></td>
</tr>
<tr>
<td>Eastern Cape-Mthatha (n=2462)</td>
<td>1472 (59.8)</td>
</tr>
<tr>
<td>Eastern Cape-O.R. Tambo or other parts of Eastern Cape (n=364)</td>
<td>169 (46.4)</td>
</tr>
<tr>
<td>Other provinces (n=536)</td>
<td>269 (50.2)</td>
</tr>
<tr>
<td>Unknown (n=69)</td>
<td>34 (49.3)</td>
</tr>
<tr>
<td>Residence, n (%)</td>
<td></td>
</tr>
<tr>
<td>Urban setting (n=2923)</td>
<td>1665 (57)</td>
</tr>
<tr>
<td>Rural setting (n=287)</td>
<td>133 (46.3)</td>
</tr>
<tr>
<td>Unknown (n=221)</td>
<td>146 (66.1)</td>
</tr>
<tr>
<td>Food insecurity, n (%)</td>
<td></td>
</tr>
<tr>
<td>Never or rarely (n=1878)</td>
<td>1131 (60.2)</td>
</tr>
<tr>
<td>Sometimes (n=1139)</td>
<td>576 (50.6)</td>
</tr>
<tr>
<td>Often or always (n=218)</td>
<td>94 (43.1)</td>
</tr>
<tr>
<td>Unknown (n=196)</td>
<td>143 (73)</td>
</tr>
<tr>
<td>Relationship status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Not in a relationship (n=1453)</td>
<td>925 (63.7)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Exposure categories</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>No MTV Shuga exposure</td>
</tr>
<tr>
<td>Ever in a relationship (n=1317)</td>
<td>668 (50.7)</td>
</tr>
<tr>
<td>Unknown (n=661)</td>
<td>351 (53.1)</td>
</tr>
<tr>
<td>Ever had sex, n (%)</td>
<td>No (n=1291)</td>
</tr>
<tr>
<td></td>
<td>Yes (n=1330)</td>
</tr>
<tr>
<td></td>
<td>Unknown (n=810)</td>
</tr>
<tr>
<td>Called a helpline or searched for information on HIV on the internet, n (%)</td>
<td>No (n=1667)</td>
</tr>
<tr>
<td></td>
<td>Yes (n=1352)</td>
</tr>
<tr>
<td></td>
<td>Unknown (n=412)</td>
</tr>
</tbody>
</table>

aDS2: Down South 2.
bFormats other than the Down South dramatic series (eg, Alone Together miniseries on COVID-19).
cDS1: Down South 1.
dIndividuals who had accessed DS2 formats not offered via radio, television, or the internet were classified as having watched 1 episode of DS2.
eRespondents in either primary or secondary school.
fTVET: technical and vocational education and training.
gIncludes respondents whose language is unknown.
hThis measure includes websites or helplines such as B-wise, Loveline, and Childline but excludes the MTV Shuga website.

**Intensity of Exposure to MTV Shuga DS2**

The components and proportions comprising each exposure categories are summarized in Table 2. Of 3431 respondents, 1487 (43.3%) had been exposed to some form of MTV Shuga content. This comprised 9.9% (n=338) of the respondents exposed to MTV Shuga formats other than the DS2 dramatic series, 9.4% (n=322) of the respondents exposed to DS1 but not DS2, and 18.1% (n=622) of the respondents exposed to 1 DS2 episode. The proportion of respondents exposed to multiple DS2 episodes was low at approximately 2% within each category of 2 to 4 episodes (84/3431, 2.4%), 5 to 7 episodes (59/3431, 1.7%), or 8 to 10 episodes (62/3431, 1.8%). Repeated exposure to DS2 episodes was mainly through television or the internet (Table 2).
<table>
<thead>
<tr>
<th>Exposure category and components</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No MTV Shuga exposure</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>1944 (56.7)</td>
</tr>
<tr>
<td><strong>Exposure to other MTV Shuga formats</strong></td>
<td></td>
</tr>
<tr>
<td>Watched the MTV preview of the show called 16 and Pregnant</td>
<td>856 (24.9)</td>
</tr>
<tr>
<td>Watched MTV public service announcements with short MTV videos with health messages</td>
<td>596 (17.4)</td>
</tr>
<tr>
<td>Watched any MTV Shuga Alone Together episodes on YouTube (Google LLC) or the MTV Shuga website</td>
<td>280 (8.2)</td>
</tr>
<tr>
<td>Searched for information on HIV on the MTV Shuga website</td>
<td>210 (6.1)</td>
</tr>
<tr>
<td>Answered a polling question about an MTV Shuga DS\textsuperscript{b} episode</td>
<td>267 (7.8)</td>
</tr>
<tr>
<td>Ever posted any comments about an episode of MTV Shuga DS\textsuperscript{c}</td>
<td>294 (8.6)</td>
</tr>
<tr>
<td>Exposed to other Shuga formats only: said yes to at least one of the above (but not exposed to DS1\textsuperscript{d} or DS2)</td>
<td>338 (9.9)</td>
</tr>
<tr>
<td><strong>Exposure to DS1 episodes</strong></td>
<td></td>
</tr>
<tr>
<td>Ever watched MTV Shuga DS1 on television, MTV Shuga website, or YouTube</td>
<td>763 (22.2)</td>
</tr>
<tr>
<td>Ever listened to MTV Shuga DS1 on the radio</td>
<td>245 (7.1)</td>
</tr>
<tr>
<td>Exposed to DS1: said yes to at least one of the above (but not exposed to DS2)</td>
<td>322 (9.4)</td>
</tr>
<tr>
<td><strong>Exposure to 1 DS2 episode</strong></td>
<td></td>
</tr>
<tr>
<td>Ever watched 1 MTV Shuga DS2 episode on television, MTV Shuga website, or YouTube</td>
<td>238 (6.9)</td>
</tr>
<tr>
<td>Ever listened to 1 MTV Shuga DS2 episode on the radio</td>
<td>71 (2.1)</td>
</tr>
<tr>
<td>Read the MTV Shuga DS2 graphic novel\textsuperscript{g}</td>
<td>344 (10)</td>
</tr>
<tr>
<td>Watched the documentary called MTV Shuga in real life that was broadcast at the end of DS2\textsuperscript{g}</td>
<td>386 (11.3)</td>
</tr>
<tr>
<td>Attended small group discussion facilitated by a peer educator on DS2 (at a clinic, school, university, TVET\textsuperscript{h}, or somewhere else)\textsuperscript{g}</td>
<td>513 (14.9)</td>
</tr>
<tr>
<td>Attended a community event on DS2 anywhere\textsuperscript{g}</td>
<td>292 (8.5)</td>
</tr>
<tr>
<td>Exposed to 1 DS2 episode (exposure to any of the listed components)</td>
<td>622 (18.1)</td>
</tr>
<tr>
<td><strong>Exposure to 2-4 DS2 episodes</strong></td>
<td></td>
</tr>
<tr>
<td>Ever watched 2-4 MTV Shuga DS2 episodes on television, MTV Shuga website, or YouTube</td>
<td>75 (2.2)</td>
</tr>
<tr>
<td>Ever listened to 2-4 MTV Shuga DS2 episodes on the radio</td>
<td>30 (0.9)</td>
</tr>
<tr>
<td>Exposed to 2-4 DS2 episodes either on television, MTV Shuga website, YouTube, or radio</td>
<td>84 (2.5)</td>
</tr>
<tr>
<td><strong>Exposure to 5-7 DS2 episodes</strong></td>
<td></td>
</tr>
<tr>
<td>Ever watched 5-7 MTV Shuga DS2 episodes on television, MTV Shuga website, or YouTube</td>
<td>53 (1.5)</td>
</tr>
<tr>
<td>Ever listened to 5-7 MTV Shuga DS2 episodes on the radio</td>
<td>17 (0.5)</td>
</tr>
<tr>
<td>Exposed to 5-7 DS2 episodes either on television, MTV Shuga website, YouTube, or radio</td>
<td>59 (1.7)</td>
</tr>
<tr>
<td><strong>Exposure to 8-10 DS2 episodes</strong></td>
<td></td>
</tr>
<tr>
<td>Ever watched 8-10 MTV Shuga DS2 episodes on television, MTV Shuga website, or YouTube</td>
<td>59 (1.7)</td>
</tr>
<tr>
<td>Ever listened to 8-10 MTV Shuga DS2 episodes on the radio</td>
<td>7 (0.2)</td>
</tr>
<tr>
<td>Exposed to 8-10 DS2 episodes either on television, MTV Shuga website, YouTube, or radio</td>
<td>62 (1.8)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}People exposed to other MTV Shuga formats only were not exposed to Down South 1 or Down South 2.

\textsuperscript{b}DS: Down South series.

\textsuperscript{c}DS2: Down South 2.

\textsuperscript{d}DS1: Down South 1.

\textsuperscript{e}Participants exposed to DS1 may have been exposed to other MTV Shuga formats (but not DS2).

\textsuperscript{f}Participants exposed to DS2 may also have been exposed to DS1 or other MTV Shuga formats.
Participants were classified as having watched 1 episode of DS2.

TVET: technical and vocational education and training.

Of the 238 participants who had watched DS2 on television, MTV Shuga website, or YouTube, 162 (68.1%) had started watching DS2 the year before our evaluation. The intensity of exposure to DS2 was similar across most categories of participant characteristics, although there were some relatively small differences of approximately 5% to 8% in absolute terms for schooling or employment status, sexual history, and food insecurity; 9.7% (129/1330) of those who had ever had sex accessed ≥2 DS2 episodes compared with 3.2% (41/1291) of those who had never had sex, the corresponding proportions were 11.9% (26/218) among those who reported experiencing food insecurity often versus 6.2% (70/1139) among those who experienced moderate food insecurity, and approximately 9.5% (69/726) among those in school or university versus 2% (19/967) among those in technical and vocational education and training institutions (Table 1).

Regression Results for DS2 Intensity

The proportion of respondents who knew their HIV status, were aware of PrEP and HIVST, and had used HIVST (ever and in the past year) increased with increasing DS2 exposure intensity. There was evidence of a nonlinear association between the intensity of DS2 exposure and knowledge of HIV status (nonlinear $P=.06$), awareness of PrEP ($P<.001$), and awareness of HIVST ($P=.003$). For uptake of HIVST and willingness to use HIVST or PrEP, the data were consistent with a linear trend ($P=.38$ for lifetime use of HIVST; $P=.33$ for use of HIVST in the past 12 months; $P=.36$ for willingness to test self with HIVST kit; $P=.76$ for willingness to give kit to partner; and $P=.46$ for willingness to take PrEP every day). The proportion of respondents reporting each of these outcomes was always the lowest among those not exposed to any MTV Shuga content (Multimedia Appendix 2).

Knowledge of HIV status was 28.1% (431/1535) among those not exposed to any MTV Shuga content and highest among those who had watched or listened to 8 to 10 DS2 episodes (45/53, 84.9%; Multimedia Appendix 2). In the adjusted analysis, increasing DS2 exposure was associated with increasing knowledge of HIV status. Compared with individuals not exposed to any MTV Shuga content, the aORs for those who had watched or listened to 1 DS2 episode, 2 to 4 DS2 episodes, 5 to 7 DS2 episodes, and 8 to 10 DS2 episodes were 2.65 (95% CI 2.01-3.49), 3.92 (95% CI 2.05-7.48), 3.82 (95% CI 1.84-7.91), and 5.72 (95% CI 2.46-13.32), respectively (Figure 1; Multimedia Appendix 2).

**Figure 1.** Forest plots summarizing the associations between different MTV Shuga Down South 2 (DS2) exposure intensities and HIV testing and pre-exposure prophylaxis (PrEP) outcomes. HIVST: HIV self-testing.
Similarly, increasing DS2 exposure was associated with increased awareness of PrEP in the adjusted analysis; for example, 47.2% (215/456) of those who had watched or listened to only 1 DS2 episode versus 17.1% (251/1469) of those not exposed to any MTV Shuga content (aOR 2.38, 95% CI 1.83-3.09) and 76% (38/50) of those who had watched or listened to 8 to 10 episodes versus 17.1% (251/1469) of those who were not exposed to any MTV Shuga content (aOR 7.25, 95% CI 3.56-14.79; Figure 1; Multimedia Appendix 2). Exposure to DS2 was not associated with willingness to take PrEP (which was high overall at 1851/2284, 81%; Figure 1; Multimedia Appendix 2).

Regarding lifetime awareness of HIVST (ever heard of HIVST), proportions ranged from 18.8% (284/1509) among those who were not exposed to any MTV Shuga content to 56.4% (270/479) among those who had watched or listened to only 1 DS2 episode and to 75% (39/52) among those who had watched or listened to 8 to 10 episodes. In the adjusted analysis, those exposed to 2 to 4, 5 to 7, or 8 to 10 episodes of DS2 had >5 times higher odds of being aware of HIVST compared with those not exposed to any MTV Shuga content (Figure 1; Multimedia Appendix 2). Lifetime use of HIVST was 26.4% (125/474) among those who had watched or listened to only 1 DS2 episode versus 7.8% (115/1483) among those not exposed to any MTV Shuga content (aOR 2.49, 95% CI 1.83-3.38), 32.8% (21/64) among those who had watched or listened to 2 to 4 DS2 episodes (aOR 3.93, 95% CI 2.16-7.15), 35.2% (19/54) among those who had watched or listened to 5 to 7 episodes (aOR 4.74, 95% CI 2.50-9.02), and 40% (21/53) among those who watched or listened to 8 to 10 episodes (aOR 4.57, 95% CI 2.46-8.50).

Likewise, in the adjusted analysis, watching or listening to an increasing number of DS2 episodes was associated with increased odds of using HIVST in the past year (Figure 1; Multimedia Appendix 2). Among those who had never used an HIVST before, interest in using HIVST on oneself or interest in giving an HIVST kit to a partner was high overall (83% for both), with no differences in the number of DS2 episodes accessed (Figure 1; Multimedia Appendix 2).

Compared with respondents who were not exposed to any MTV Shuga content, DS2 audiences who had watched or listened to multiple episodes of DS2 were more likely to ever have sex in their lifetime and had sex in the past 12 months (Figure 2; Multimedia Appendix 3). Among those who reported having sex within the past 12 months, there was no evidence of a departure from a linear trend between DS2 exposure intensity and condom use ($P = .23$; Figure 2; Multimedia Appendix 3). There was evidence of a nonlinear association between DS2 exposure intensity and sexual history (ever or in the past 12 months; nonlinear $P < .001$).

**Figure 2.** Forest plots summarizing the associations between different MTV Shuga Down South 2 (DS2) exposure intensities and sexual behavior outcomes.
Effects of Other MTV Shuga Content or Formats

Regarding knowledge of HIV status, awareness of PrEP and HIVST, and lifetime uptake of HIVST, there was evidence that exposure to other forms of MTV Shuga content (other than the DS2 dramatic series) was beneficial, compared with no exposure to MTV Shuga content. For instance, knowledge of HIV status was 74.4% (215/289; aOR 2.52, 95% CI 1.78-3.56) among those who had watched DS1 and 63% (182/289; aOR 1.84, 95% CI 1.33-2.54) among those who had accessed other MTV formats compared with 28.1% (431/1535) among those with no exposure to MTV Shuga content (Figure 1; Multimedia Appendix 2). Furthermore, those exposed to these other forms of MTV Shuga content were more likely to have had sex (ever or within the past 12 months) compared with those with no MTV Shuga exposure (Figure 2; Multimedia Appendix 3).

How Young People Watched or Listened to DS2

Among those who had been exposed to at least 1 DS2 episode and provided information on how they accessed the series, 50.6% (119/235) had watched or listened to DS2 with someone, 33.6% (79/235) accessed alone, and 15.7% (37/235) reported a combination of the 2 (Table 3). Among those who had watched or listened to DS2 with someone, 45.5% (71/156) did so with peers only, 25% (39/156) with parents, and 29.5% (46/156) with siblings. The findings were similar by age group (P=.45; Table 3). The intensity of DS2 exposure did not differ significantly by how young people accessed DS2 (P=.46), although higher proportions of repeat viewers watched or listened to DS2 with a parent than those who watched or listened to 1 DS2 episode. For example, 19.3% (16/83) of the participants who accessed 2 to 4 DS2 episodes did so with a parent compared with 2.9% (1/35) among those who accessed 1 DS2 episode (Figure 3). We were unable to establish whether our outcomes of interest differed by the intensity of DS2 exposure, given how young people accessed DS2, because of insufficient data to conduct the regression analyses.

Table 3. Descriptive summaries of how young people watched or listened to Down South 2.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Age group (y), n (%)</th>
<th>Total (n=235), n (%)a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15-19 (n=104)a</td>
<td>20-24 (n=131)a</td>
</tr>
<tr>
<td>Did you usually watch or listen to MTV Shuga Down South season 2 on your own or with someone?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On my own</td>
<td>31 (29.8)</td>
<td>48 (36.6)</td>
</tr>
<tr>
<td>With someone</td>
<td>54 (51.9)</td>
<td>65 (49.6)</td>
</tr>
<tr>
<td>Both</td>
<td>19 (18.3)</td>
<td>18 (13.7)</td>
</tr>
<tr>
<td>With whom did you usually watch or listen to MTV Shuga Down South season 2?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>31 (29.8)</td>
<td>48 (36.6)</td>
</tr>
<tr>
<td>With peers only</td>
<td>27 (26)</td>
<td>44 (33.6)</td>
</tr>
<tr>
<td>Friends</td>
<td>24 (32.9)</td>
<td>30 (36.1)</td>
</tr>
<tr>
<td>Friends and partners</td>
<td>1 (1.4)</td>
<td>9 (10.8)</td>
</tr>
<tr>
<td>Partners</td>
<td>2 (2.7)</td>
<td>5 (6)</td>
</tr>
<tr>
<td>With parents plus</td>
<td>22 (21.2)</td>
<td>17 (13)</td>
</tr>
<tr>
<td>Parents</td>
<td>6 (8.2)</td>
<td>3 (3.6)</td>
</tr>
<tr>
<td>Parents and friends</td>
<td>4 (5.5)</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>Parents, friends, and partners</td>
<td>1 (1.4)</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>Parents and siblings</td>
<td>5 (6.8)</td>
<td>3 (3.6)</td>
</tr>
<tr>
<td>Parents, friends, and siblings</td>
<td>4 (5.5)</td>
<td>2 (2.4)</td>
</tr>
<tr>
<td>Parents, friends, partners, and siblings</td>
<td>2 (2.7)</td>
<td>6 (7.2)</td>
</tr>
<tr>
<td>Parents, partners, and siblings</td>
<td>0 (0)</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>With siblings plus</td>
<td>24 (23.1)</td>
<td>22 (16.8)</td>
</tr>
<tr>
<td>Siblings</td>
<td>15 (20.5)</td>
<td>11 (13.3)</td>
</tr>
<tr>
<td>Siblings and friends</td>
<td>7 (9.6)</td>
<td>6 (7.2)</td>
</tr>
<tr>
<td>Siblings, friends, and partners</td>
<td>2 (2.7)</td>
<td>2 (2.4)</td>
</tr>
<tr>
<td>Siblings and partners</td>
<td>0 (0)</td>
<td>3 (3.6)</td>
</tr>
</tbody>
</table>

aAmong those who had watched or listened to at least 1 Down South 2 episode and provided information on how they accessed the episodes.
Discussion

Key Findings

This study aimed to understand whether the intensity of exposure to the MTV Shuga DS2 series had differential impacts on the outcomes of interest. We found evidence that the levels of knowledge of HIV status, awareness of PrEP, and use of HIVST (ever or in the past year) increased with increased exposure to DS2, consistent with a dose-response effect. We found no evidence of an effect of DS2 exposure intensity on the interest in using HIVST or PrEP (a proxy for demand), which was already high in the study population. In our study sample, repeated exposure to DS2 episodes was mainly through television or the internet.

Regarding how young people access DS2, findings indicate a mix of preferences, with the majority watching or listening to DS2 alone or with peers (eg, friends or partners) and less so with siblings and parents. Although there was little influence of these preferences on the intensity of DS2 exposure, we found higher proportions of repeat viewers watched or listened to DS2 with a parent, compared with those who accessed only 1 episode. This suggests that accessing DS2 with parents may have helped some young people to watch more episodes. However, it is worth noting that some of our study participants avoided watching with parents because they feared that it would be awkward or that the parents would judge or lecture them on sex and relationship matters, as documented in our related qualitative research [19]. Others may have preferred watching or listening to DS2 alone for privacy reasons, as documented in another study [11]. These findings point toward the need to ensure privacy and nonjudgmental spaces in the delivery of campaigns targeted to young populations. Our related qualitative research showed that watching or listening to DS2 spurred interpersonal conversations and discussions with peers and sometimes with parents, as reported elsewhere [3,19]. Discussions where peers engaged in debates about DS2 characters and storylines led to shared learning and support systems where people felt safe to discuss sexual health topics and disclose personal information [19]. This could deepen the influence of the show for the viewer (beyond just watching), for example, through greater internalization, and future research could consider these events as mediators along a pathway between exposure (eg, to MTV Shuga content) and HIV and health outcomes.

The observed dose-response effects on knowledge of HIV status, awareness of PrEP, and use of HIVST (ever or in the past year) might reflect that more exposure facilitates increased levels of narrative engagement and, in turn, helps the viewers or listeners to retain a greater amount of relevant content compared with someone with little or no exposure. This might particularly be true when the audience is interested in finding out what eventually happens to the characters, often facilitated by the immersive nature of the DS2 series. As documented in our qualitative research, the emotional and dramatic storylines kept participants engaged as they wanted to know what would happen next [19]. In addition, the re-emphasis of content across multiple storylines allows participants to access content that they may have missed previously. Previous research has documented that participants in dramatized series tend to actively reflect on the plotlines and often compare the characters’ ways of confronting dilemmas across storylines, which presents a learning opportunity for the audience [3,5]. In a cluster randomized trial of MTV Shuga in Nigeria, the effects of MTV Shuga were stronger for viewers who were more immersed or highly identified with the characters while watching the show [9]. Although we did not measure identification with characters in this analysis, we drew on other evidence from the larger mixed methods study, which indicates that young people found the show relatable. In a miniseries called MTV Shuga: Alone Together (which aimed to disseminate timely and accurate information on COVID-19), developed in the same manner as
DS2, ≥90% of young people (in the same study sample as this paper) who had watched the miniseries indicated that they found it entertaining, informative, memorable, and realistic (I Birdthistle, PhD, unpublished data, February 2021). One of the key findings from the nested qualitative study conducted as part of our research to understand why young people engaged with DS2 was that young people found the show relatable as the storylines reflected real-life issues that they or people they knew experienced [19]. Younger participants appreciated how the show embraced the uncertainty that often surrounds such decision-making (as opposed to simplistic or moralistic messaging). In addition, some viewers reported being introduced to HIVST and PrEP for the first time through the show, and less experienced viewers felt more prepared for future sexual relationships and decision-making based on the scenarios explored in DS2 [19]. The high identification with the DS2 characters and storylines likely enhanced the effect of watching DS2. The larger effects among those who watched or listened to multiple DS2 episodes suggest the potential benefits of viewing dramas as a whole rather than as a series of parallel storylines, as documented elsewhere [5]. However, it is worth noting that a viewer or listener may miss relevant content completely when scenes are relatively few and short [6]. Although there is uncertainty around our effect estimates, given the low proportion of participants who accessed multiple episodes of DS2, there remains a clear overall pattern for these outcomes (except those on demand) that more DS2 exposure is better.

We did not find evidence of a dose-response relationship between DS2 exposure intensity and demand outcomes (ie, interest in using HIVST and PrEP), which were already high in the study population. Similarly, higher DS2 exposure intensity did not result in greater condom use. Although DS2 influenced awareness and motivation to use HIV prevention products and services, including HIV testing and PrEP, the actual provision of these products was not part of the DS2 campaign [6]. Qualitative research findings indicate that many participants were unsure of the availability of these services and products in their own setting, which could partly explain the limited effect on demand for these products [6]. This finding highlights the crucial role that actual provision and access to these HIV prevention tools play in influencing behavior, in addition to awareness and motivation to use these methods [26]. It is possible that our analysis is underpowered to detect smaller differences when assessing the dose-response effect for demand outcomes, given the high absolute proportions of these outcomes and the fact that a high proportion of our participants were randomized to either a control group or to 1 of the 4 experimental arms (based on different storytelling formats), only 6% of those assigned to experimental arms went on to watch the full episodes after exposure to different formats [3]. These estimates are much lower than those from multimedia studies targeting general populations, where exposure to multiple episodes (≥25) is as high as 30% [4,23].

The low exposure to multiple episodes in our study may reflect various issues inherent in engaging a relatively young population (aged <25 years). First, young viewers or listeners often have no control over when episodes are aired on television or radio channels and given that they might actively avoid watching or listening to MTV Shuga episodes with parents or older siblings for reasons summarized previously, this can limit their ability to access complete or multiple episodes. Limited affordability and availability of the internet and data plans may prohibit access to content offered via streaming platforms such as YouTube or the MTV Shuga website [28]. Even with extensive efforts to create relatable storylines, we acknowledge that it can be challenging to capture the wide range of complexities in the lives of adolescents and young adults in 1 show, and there may be a proportion of participants who are not interested in the show entirely. In addition, it is possible that some young people may have decided to stop watching or listening after 1 episode because they did not find the episode interesting enough. Future research can aim to understand the complete range of reasons for “disengagement” in different contexts and age groups.

Given our study findings that more DS2 exposure is better for many of the outcomes and the fact that a high proportion of our study sample engaged with “offline” DS2 content, it might be worth finding ways to facilitate and increase offline viewing of DS2 episodes. In particular, peer education and school education programs are good complementary options because they (1) do not require constant internet connectivity (which may prohibit access via YouTube or MTV Shuga website) and (2) can create a safe space for young people (who prefer to watch alone or with peers) to watch without having to worry about the reactions and judgment of older people. In addition, short clips or extensions can be used to deliver and highlight critical components (eg, HIV testing and PrEP) so that people do not miss them. These could be offered through social media platforms, which are popular among young people. We note here that the specific combinations of activities that program designers and implementers use will depend on the scope and
goals of the campaign. For instance, if promoting high intensity of exposure (ie, access to multiple episodes) is the goal, then complementary activities may need to incorporate and deliver full episodes rather than sharing shorter formats (eg, on social media), with the latter more suited for raising awareness and enhancing interest in general. Many web-based platforms also allow viewers to download episodes and watch or listen later, and sensitizing young people about this option can potentially facilitate more access. The recent inclusion of MTV Shuga on Netflix is likely to widen its reach, although it may reach those who are already connected and able to pay for the service. Furthermore, for some young people, getting their parents interested in watching Shuga might improve their exposure intensity. As we learn from the COVID-19 pandemic, public health players need to be innovative and adaptive, and using a combination of strategies to reach the target populations is worthwhile [29].

**Strengths and Limitations**

One of the study limitations is that we relied on self-reported measures that are subject to recall bias. Some of the behavioral questions (eg, on sexual history) may also be subject to social desirability bias. To ensure participants' confidentiality and to help minimize social desirability bias, we did not collect any identifying information (although mobile phone numbers for those who completed the survey were obtained to facilitate the transfer of mobile data credit). Furthermore, the survey was self-administered, and we anticipated that participants would complete the interviews in private, which could increase the accuracy of the self-reported information. Regarding recall bias, we feel that most of our behavioral outcome measures are related to events that people are likely to remember (eg, testing for HIV in the last year and taking PrEP). Many of our questions included options such as “don’t know” and “prefer not to answer” to accommodate participants who may not have an opinion regarding the question at hand. Approximately 6 (68%) out of 10 participants had started watching DS2 the year before we conducted our evaluation, and this may have influenced their ability to accurately recall the exact number of DS2 episodes accessed. Participants who were exposed to DS2 but did not know how many episodes they had watched or listened to were assumed to have been exposed to only 1 DS2 episode. Furthermore, we did not collect information on how frequently offline components such as graphic novels were accessed, and thus, individuals exposed to DS2 content not offered through radio, television, or the internet were also assumed to have been exposed to only 1 DS2 episode. We may have misclassified these individuals if they had indeed been exposed to multiple episodes. Although we adjusted for various potential confounders in all our analyses, we cannot rule out residual confounding and other possible explanations for the observed associations. Moreover, it is possible that those with a higher awareness of HIV in general might be more likely interested in accessing DS2, resulting in reverse causality. The strengths of the study include the assessment of multiple outcomes, the rich data on exposure to multiple MTV Shuga content, and information on how young people accessed DS2. Among those not exposed to DS2, we were able to capture exposure to other MTV Shuga content. Although this resulted in “nonnatural” categories of DS2 exposure intensity, it allowed us to identify whether individuals exposed to other MTV Shuga content had better outcomes than those not exposed to any MTV Shuga content at all.

**Conclusions**

Several studies have examined the effects of edutainment campaigns on sexual and HIV-related health outcomes among young people; however, few have examined the intensity of exposure and whether increased engagement resulted in greater benefits. In this study, we found that increasing DS2 exposure was associated with increasing knowledge of HIV status, awareness of PrEP, and use of HIVST. This is consistent with a dose-response effect and supports the plausibility of a causal association between DS2 and HIV prevention outcomes among young audiences. Overall, relatively few participants viewed multiple episodes of DS2, and supporting young people to view or listen to more episodes of the DS2 campaign can yield benefits for more young people. If promoting high intensity of DS2 exposure (ie, access to multiple episodes) is the goal for a given campaign, incorporating complementary activities that support the delivery of full episodes rather than sharing shorter formats may be useful. A more complete and immersive experience can be offered through better and more equal digital access and through school programs and peer education programs, taking into account young people’s preferences when designing and delivering these campaigns.

**Acknowledgments**

The authors would like to thank all the young people who contributed their time to this study during the COVID-19 restrictions in 2020. The authors are grateful to Antonio Duran-Aparicio for the support in making this project possible. The authors acknowledge the dedication of MTV Shuga peer educators and coordinators (Yvonne Diogo and Lesedi Thwala) in South Africa; the resilience of the team at Epicentre Health Research; and the commitment of the MTV Staying Alive Foundation, especially Georgia Arnold and Sara Piot.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest**

None declared.
Multimedia Appendix 1
Definitions of outcome measures included in the analysis.
[XLSX File (Microsoft Excel File), 11 KB - formative_v81e44111_app1.xlsx ]

Multimedia Appendix 2
Associations between different MTV Shuga Down South 2 exposure intensities and HIV testing and pre-exposure prophylaxis outcomes.
[XLSX File (Microsoft Excel File), 16 KB - formative_v81e44111_app2.xlsx ]

Multimedia Appendix 3
Associations between different MTV Shuga Down South 2 exposure intensities and sexual behavior outcomes.
[XLSX File (Microsoft Excel File), 14 KB - formative_v81e44111_app3.xlsx ]

References


18. MTV staying alive foundation. MTV Shuga Down South. URL: https://www.mtvshugacom/downsouth/episodes/ [accessed 2022-10-17]


Abbreviations

- aOR: adjusted odds ratio
- DS1: Down South 1
- DS2: Down South 2
- HIVST: HIV self-testing
- PrEP: pre-exposure prophylaxis
Nonuse of Blended Web-Based and Face-To-Face Cognitive Behavioral Therapy for Alcohol Use Disorder: Qualitative Study

Kristine Tarp1,2,3, MA, PhD; Regina Christiansen4,5, MA, PhD; Randi Bilberg4,5,6, MSc, PhD; Simone Borkner1, BSc; Caroline Dalsgaard1, BSc; Marie Paldam Folker1, MA; Anette Søgaard Nielsen4,5,7,8, MA, PhD

1Research Unit of Digital Psychiatry, Centre for Digital Psychiatry, Mental Health Services in the Region of Southern Denmark, Odense, Denmark
2Research Unit of Digital Psychiatry, Department of Clinical Research, University of Southern Denmark, Odense, Denmark
3The National Research Centre for the Working Environment, Copenhagen, Denmark
4Unit of Clinical Alcohol Research, Department of Clinical Research, University of Southern Denmark, Odense, Denmark
5Psychiatric University Hospital – University Function, Mental Health Services in the Region of Southern Denmark, Odense, Denmark
6Department for Finance and Planning, Lillebaelt Hospital, University Hospital of Southern Denmark, Vejle, Denmark
7Brain Research Inter-Disciplinary Guided Excellence (BRIDGE), Department of Clinical Research, University of Southern Denmark, Odense, Denmark
8Open Patient data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark

Corresponding Author:
Kristine Tarp, MA, PhD
The National Research Centre for the Working Environment
Lersø Parkallé 105
Copenhagen, 2100
Denmark
Phone: 45 21292332
Email: ket@nfa.dk

Abstract

Background: The use of digital technologies for health care has been the focus of social studies, which have concentrated on the digital divide between individuals who use technology and those who do not—with the latter often being considered as individuals with shortcomings. In Denmark, 91% of the population have computers and 97 out of 100 families have internet access, indicating that lack of access to technology is not the primary reason for nonuse. Although previous studies have primarily focused on participants’ perspectives of using internet-based treatment for alcohol use disorder (AUD), no study has investigated individuals’ reasons to prefer face-to-face treatment over blended face-to-face and internet-based cognitive behavioral therapy (bCBT) for AUD among treatment-seeking populations.

Objective: The aim of this qualitative study was to investigate the nonuse of bCBT among patients with AUD. Specifically, this study aims to explore patients' reasons for choosing not to receive treatment via this format.

Methods: This study was conducted among Danish patients with AUD who were enrolled in the study “Blending internet treatment into conventional face-to-face treatment for alcohol use disorder (Blend-A)” but had not used bCBT. The participant group consisted of 11 patients with AUD: 3 women and 8 men. The age range of the participants was 29-78 years (mean 59 years). Individual semistructured interviews were conducted using cell phones to gather participants’ reasons for not choosing bCBT. The interviews were recorded, transcribed, and analyzed using thematic analysis. Five authors performed the analysis in 3 steps: (1) two authors read the transcripts and coded themes from their immediate impression of the material, (2) one author provided feedback, which was used to group overlapping themes together or create new themes that better reflected the content, and (3) the remaining two authors provided feedback on the analysis to improve its structure, readability, and relevance to the research aim.

Results: We found that the participants had various reasons for choosing face-to-face treatment over bCBT; these reasons were more related to personal matters and lesser to digital health literacy. We identified 4 themes related to personal matters for choosing face-to-face treatment over bCBT: (1) patients’ need for attending sessions in person, (2) preference for verbal communication, (3) desire for immediate feedback, and (4) feeling more empowered and motivated with face-to-face sessions.

Conclusions: This study provides valuable insight into participants’ perspectives on blended therapy for AUD and highlights the importance of considering personal factors when designing digital health interventions. Our study indicates that most of the
participants choose not to use bCBT for AUD because they perceive such treatment formats as impersonal. Instead, they prefer direct communication with the therapist, including the ability to express and comprehend facial expressions and body language.

**Introduction**

**Background**

Social studies of the use of digital technologies for health care have focused on the digital divide—the division between individuals using technology and those who do not—the latter viewed as merely individuals with deficits [1,2]. These deficits may cover a range of difficulties and barriers, which can occur when engaging with technology [3]. In general, social groups with higher education and higher income seem to have more knowledge, motivation, and competency in initiating steps toward a healthy lifestyle [4], in addition to having fewer barriers toward digital solutions [5]. For example, Heponiemi et al [6] described how individuals who do not use computers have lesser education, higher unemployment, lower income, and poorer health, and found a risk of digital exclusion among those who have lower socioeconomic status, poorer health, or are more socially isolated. Nonetheless, in Denmark in 2020, 97 of 100 families had internet access, 96% had a mobile phone, and 91% had a computer [7], which indicates that people in Denmark are regular users of digital technologies. Therefore, it may be anticipated that digital treatment interventions targeting individuals with alcohol use disorder (AUD), like guided internet-based cognitive behavioral therapy (iCBT), may be an appreciated intervention among the Danish population, although there may be barriers toward seeking treatments for AUD in general and specifically toward digital solutions.

A review [8] has shown the general barriers toward seeking traditional AUD treatments. For example, Wallhed Finn et al [9] conducted a study among nontreatment seekers with alcohol dependence on their perceptions of alcohol consumption, dependence, and barriers toward seeking face-to-face (FtF) treatment for AUD. They found participants to be generally negative toward FtF treatment, for example, due to stigma and shame. For this group of nontreatment seekers, an internet-based intervention like iCBT may be perceived as a potential first step toward entering treatment—both to assess one’s alcohol use and to receive guidance for suitable treatment.

Barriers toward engaging in iCBT for AUD have not been investigated much. In a study on attrition during a web-based treatment for problem drinkers, Postel et al [10] found the second most common reason for noncompletion of an internet-based intervention to be dissatisfaction with the intervention itself, for example, that it was too time consuming or demanding and did not meet personal needs. In another study of user experiences of internet-based treatment for problematic alcohol use, Ekström and Johansson [11] identified the following barriers toward engaging in internet-based AUD treatment: lack of recognition in the content of the intervention, too much text and repetition, too little (meaningful) support or feedback, lack of contact with a therapist, and lack of guidance.

Combining iCBT with FtF CBT is referred to as blended CBT (bCBT). bCBT for AUD may propose a treatment solution that combines a high level of discretion and flexibility in addition to being guided and person-centered [12]. It might, so to speak, offer the best from both the aforementioned treatment modes. Participants in that study [12] were offered bCBT, but they opted out of using iCBT as they preferred solely FtF CBT. We found this intriguing and important since we anticipated that bCBT, in particular, would be perceived as an attractive offer due to the high familiarity of the Danish population with digital technology. Thus, in this study, we wished to explore participants’ reasons for deciding against and opting out of using bCBT. To our knowledge, no study has previously investigated patients’ reasons to prefer FtF over bCBT or iCBT for AUD among treatment-seeking populations.

**Aim**

In this study, we sought to understand individuals’ perceptions of bCBT and iCBT for AUD when they are introduced to this type of treatment format by the therapists. In particular, we aimed to explore participants’ reasons for choosing a treatment strategy that solely consists of FtF treatment and not digital solutions, when offered the possibility of a flexible combination of FtF and iCBT.

**Methods**

**Settings**

This study is a substudy under the study “Blending internet treatment into conventional face-to-face treatment for AUD (Blend-A)” [12]. At the beginning of the overall study, 18 Danish municipal treatment institutions participated, but only 14 clinics remained throughout the whole study period. The clinics are quite similar in structure and treatment content offers. In Denmark, the municipalities offer AUD treatment free of charge to the individual patient. The treatment is based on treatment manuals stemming from evidence-based treatment methods such as CBT and motivational interviewing (MI) [13]. A typical treatment course entails acute treatment for withdrawal symptoms, followed by a series of either individual or group-based sessions. The duration of the treatment courses depends on the patients’ needs.

A treatment layout for AUD that consists of a combination of FtF treatment and internet-based modules, which was developed in The Netherlands [14], was translated and adjusted to fit Danish language and culture [15]. During the Blend-A study,
all patients who entered AUD treatment in the participating clinics were offered to receive all or part of their treatment course in Blend-A. Blend-A is operated as bCBT, where patients can use iCBT on a web-based platform, hosted by the Dutch company Minddistrict, in combination with attending FtF CBT sessions at the clinic. The degree of blending is agreed upon among specific clinics, therapists, and patients. One example is that, when blending, the patient would attend FtF CBT every sixth instead of fourth week. The platform entails 21 sessions with written material, visual resources, and assignments following CBT and MI. The therapist can offer a short paragraph of written feedback on some of the solved assignments for further elaboration during the FtF CBT. The format is flexible and the patients can access the web-based platform when it suits them. The platform can be accessed anonymously, if needed. The patient can go back and look at the earlier solved assignments, if needed. The implementation of the study commenced in June 2020 and ran until ultimo December 2022.

Participant Recruitment

Participants in this study were recruited among the participants in the Blend-A study who did not engage in bCBT. In total, 1033 participants were enrolled in the Blend-A study; of these, 606 (58.6%) did not register for an iCBT profile on the web-based platform, and thus, did not make use of the bCBT offer. All Blend-A participants filled out a baseline survey and were invited to fill out a 6-month follow-up survey, no matter to which degree they had made use of bCBT for AUD, if at all. The 6-month follow-up survey was collected electronically or on the telephone by researchers, not knowing until the last questionnaire, whether the Blend-A participants had actually used bCBT for AUD. A random sample of 60 participants participating in the Blend-A study without bCBT were telephoned by author SB and invited to participate in telephone-based individual interviews for this study about their reasons for not wanting to use bCBT for AUD. Some did not answer the telephone, some did not feel that they could contribute as they could not remember having been offered bCBT, and some did not wish to participate and gave no reason for this. Twelve participants agreed to participate and scheduled an appointment for an interview; in 1 case, however, we failed to reach the participant.

Data Collection

Data were collected using semistructured individual interviews with an interview guide, available in Multimedia Appendix 1. The interview guide was not pilot tested, and no repeat interviews were performed. The questions were inspired by relevant subjects found in the literature, asking about the participants’ background, experiences with using digital technology in their everyday lives, and their reasons for not choosing the offered bCBT for AUD. Furthermore, the questions were open-ended, leaving room for pursuing any given direction set by the participant. The interviews were conducted by a psychology student intern, Jakob Godsk Nielsen, and the first author KT. Neither were involved in the clinical treatment in this study. No relationship between the interviewer and interviewee was established prior to study commencement. The interviewees had no prior knowledge about the researchers, and no characteristics about the interviewees were reported to the interviewees other than that the interviewers were researchers. The interviews were conducted over telephone, lasted between 30 minutes and 45 minutes, and were audiotaped and transcribed in NVivo (QSR International) in full length by authors SB and CD. The transcribed interviews were not returned to the interviewees for commenting and corrections, and no field notes were made during the interviews. We used COREQ (Consolidated Criteria for Reporting Qualitative Studies) [16] as a checklist for reporting on the interviews. Data were anonymized and securely stored.

Data Analysis

The transcribed interviews were analyzed in the qualitative software support system NVivo by using thematic analysis [17]. First, all transcripts were read to obtain an overall immediate impression of the material. Along reading, the material was coded by themes that came to the mind of the authors (KT and RC, both female postdocs, who holds MA degrees in anthropology and philosophy, PhD degrees within health sciences, and approximately 10 years of experience within the field). Second, another author (ASN) commented on the transcripts with the coded themes. Based on these comments, the authors (KT and RC) recoded the material with focus on overlapping themes grouped together or recoded with new themes that more accurately specified the content. Lastly, the 2 remaining authors (RB and MPF) gave their feedback on the themes, structure, and readability of the analysis, leading to the final themes as expressed by the participants who had chosen not to use bCBT. The participants did not provide feedback on the findings. The collected themes are further described in the forthcoming results section.

Ethics Approval

This study was conducted according to the current ethical standards. The protocol for the Blend-A study was approved by the scientific research ethics committee of the Region of Southern Denmark (project identification S-20190166G). The Danish Data Protection Agency gave the permission to collect and store data (record 20/12692). After receiving both oral and written information about the study, the participants signed a consent form. Further, participants were informed about their rights to withdraw their consent at any time, without any consequences on their treatment course.

Results

Participant Sample Description

Ultimately, 11 participants participated in this nonuse study. The baseline characteristics of the participants are shown in Table 1. The participant group consisted of 3 women and 8 men, with a mean age of 59 (SD 16) years. The youngest participant was 29 years, and the oldest participant was 78 years. Five of the participants were married or in a relationship, 5 were single, and 1 was widowed. The participants were all educated—either within craftsmanship or had a short, intermediate, or long higher education. Five were currently employed, 3 were unemployed, and 3 had retired. They all described having an everyday schedule, wherein they got up and went about their daily
activities. Besides, 2 of them had additional mental illness, and 4 had somatic illness. The participants had used alcohol problematically for 3-30 years (mean 11 years). Their reasons for seeking treatment were to find someone to talk to and to receive help, support, and advice. They expressed that they needed tools to reduce their alcohol use. During the interviews, the participants self-assessed themselves to be super users of technology (n=2), intermediate users (n=7), and having limited digital competencies (n=2). Compared to the profile of 44,516 patients, who had a total of 88,057 treatment courses in Danish alcohol treatment institutions between 2006 and 2014 [18], our study sample was somewhat older.

Table 1. Baseline characteristics of our study participants compared to those of patients seeking treatment in Danish alcohol treatment institutions in 2006-2014.

<table>
<thead>
<tr>
<th></th>
<th>Nonuse sample in this study (N=11)</th>
<th>National Danish profile [18]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD), range</td>
<td>59 (16), 29-78</td>
<td>46-49&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Excessive alcohol use (years), mean (SD), range</td>
<td>11.19 (7.41), 3-30</td>
<td>12-13&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>3 (27)</td>
<td>41&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Married/in a relationship (yes), n (%)</td>
<td>5 (45)</td>
<td>16&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Education (low), n (%)</td>
<td>6 (55)</td>
<td>8&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Education (intermediate), n (%)</td>
<td>2 (18)</td>
<td>38&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Education (high), n (%)</td>
<td>3 (27)</td>
<td>4&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Employment (yes), n (%)</td>
<td>5 (45)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Additional mental illness (yes), n (%)</td>
<td>2 (18)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Somatic illness (yes), n (%)</td>
<td>4 (36)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Technology user (low level), n (%)</td>
<td>2 (18)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Technology user (intermediate level), n (%)</td>
<td>7 (63)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Technology user (high level), n (%)</td>
<td>2 (18)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>These data represent ranges in percentages as mentioned in [18].
<sup>b</sup>Values are presented in percentage, as the exact n values are not provided in [18] and cannot be calculated.
<sup>c</sup>N/A: not applicable.

Description of Themes

Two participants considered themselves to be technology super users, 7 felt that they were intermediate users, and 2 felt that they were inadequate digital users. The 2 latter participants felt that they had insufficient digital competencies for receiving treatment via the internet, as they did not understand it and felt unacquainted and terrible at it. One participant elaborated as follows:

...I am really bad at internet and all such technical stuff. I am also old. I did not grow up with it. But during that time where I had to be able to use it at work, I learned the basics to manage. Besides that, I have never done more about it. And when I stopped working, I also got rid of the internet. I simply don’t use it. [Participant ID 1002]

It was unclear if the participant was reluctant about bCBT because of insufficient digital competencies or if the participant merely did not find technology use engrossing. The participant did not believe that he or she had the digital competencies to make use of bCBT, even though the participant had been a former internet user and had used technologies at work. However, in general, neither this nor the other participants reported feeling insecure or unsafe about using the internet as such. Nine out of 11 participants who chose to receive FtF CBT instead of bCBT for AUD explained that their choice had rather to do with them feeling a need for attending the sessions in person. This need consisted of multiple facets. The participants seemingly felt being exposed by having to enter a treatment center, thereby accepting that someone might recognize them. They described that when having reached this far in surrendering to the fact that they had an alcohol problem, they could not risk that treatment could fail. The participants had a wish to gain as much as possible from the treatment course, considering FtF CBT to be the safe route to success.

The participants considered that when feelings were involved, there was a risk that they would become emotionally upset during the process. When being upset, they anticipated that there might be a difference between being at the computer alone and being in a room with a person who might tell if you were, for example, anxious. One participant described how being in a vulnerable position demands a level of maintenance, which may succeed through verbal communication upheld by the therapist, as this could offer dialogue, nuance, and reflection, asking more deeply into and seeing behind difficult issues. The participants seemed to link physical appearance with the ability to move the therapy forward. Some of the participants explained that videoconferencing and telephone calls would also be okay
if it was a synchronous conversation but not as a substitute for physical attendance during treatment.

Accordingly, an often mentioned factor that influenced the participants’ choice was that they considered internet-based treatment to be impersonal and that they preferred to be physically present in the same room as the therapist. The perceived benefit of FiF CBT was, according to the participants, the possibility of instantaneous communication. The participants believed being physically near to the therapist would enable a more trustful relationship, as described below:

...here people in question need help, and they need a pat on the back when things go well and help when things go bad. I do not think you can do that over mail (… with an email, you just become a number in a line, instead of a person who needs a shoulder to cry on and an ear that minds to listen….). When the matter is alcohol, then I do not think it is something that can take place over an email. Because then I think that I would feel pissed on. [Participant ID 957]

The above quote shows how participants perceive internet-delivered therapy to be impersonal—a concept often used in digital technologies although not specified. The quote above shows that what constitutes not perceiving the FiF treatment as impersonal is the ability to feel the presence of the therapist and having the feeling of being understood and respected—something that the participant considered was difficult to be accomplished over an email from the therapist. A participant considered that to receive an email, even if it is a part of the internet-delivered treatment program, was equal to being as a number in a line.

Another factor that had an impact on the participants preferring attending sessions in person was the ability to see the therapists' body language and look at them into their eyes during the sessions. This ability led the participants to believe that they could better comprehend the therapists and their responses. Being able to have questions elaborated on and clarified immediately was of importance to 5 of the participants. One example concerned personally sensitive subjects, where the participants found it easier to receive a response if they articulated the matters in a conversation compared to an email exchange, wherein the therapist might not have the time to answer right away. The participants’ wish was to have such issues settled instantly. One participant elaborated on the importance of receiving a quick clarification on outstanding matters:

...If I am in front of the therapist I am talking to, I would be able to get a response right here and right now. Then I can park it and it does not have to live inside my body anymore. Then it is out of the body. Away, fine, it is gone, finished (…) then I can move on. Then it doesn’t live inside my brain, fill up space, or spend resources anymore. [Participant ID 1091]

The immediate feedback was of importance to the participants in situations where they felt alone and in doubt about how to understand a question. The physical presence of the therapist enabled them to receive a quick clarification and thus be able to move on. The verbal communication made sense to the participants as they felt safer and assessed it to be more giving. They felt that they could gain more from verbal communication because they could tell, explain, and inform, which enabled them to instantly see reactions or signals from the therapist, which they needed to act on, or give the therapist the possibility to ask questions allowing the participant to elaborate. It was their experience that a message is better understood when you look into each other’s eyes while communicating, as it is easier to reflect on what has been said and let it sink in before one answers and then give a more precise answer based on the discussions and the reflections based on the discussion. Below, is the transcript of one participant as a voice for all:

...It is because there is some communication that you cannot always see, and something happens when you talk to people that does not happen when you write. What happens is that you reflect differently when you have a conversation and a dialogue. It is also easy to write that “all is well.” In a person, you can see if it is and maybe say “well, are you sure about that?” (Laughs). [Participant ID 1128]

Another advantage of FiF mentioned by the participants was that sitting in front of the therapist made them feel more obliged to adhere to the treatment or more compliant in relation to not drinking. The participants expressed how they felt dutiful and believed that if they had agreed upon attending an FiF CBT session, they would not cancel it. They imagined that it would be easier to cancel an internet-based session, thereby giving them an opportunity to choose the easy way out.

The participants also believed that it was beneficial that FiF therapy enabled synchronous sessions with the therapist, while on the web-based platform, the correspondence is asynchronous and the participants can use it when they want, which they believed to be risky for them and thereby the treatment. Sessions held in person make it more difficult to cheat the therapist with regard to drinking compared to web-based sessions, where they considered that it would be easier to continue their drinking. One participant unfolded this drawback as follows:

...When you make an agreement, I think it is nice that you can look each other into the eyes. Especially when it is about alcohol, then I cannot just say “I promise.” There is just something about the human contact. [Participant ID 957]

In other words, the participant considered that the commitment is stronger if expressed FiF to the therapist compared to in writing during a web-based session. Thus, choosing solely FiF CBT rather than also making use of internet-based solutions meant deciding on taking responsibility for and committing strongly to their own treatment. Finally, the participants reported that FiF CBT sessions enabled them to concentrate on their situation and focus on the treatment—a dimension that is perceived as necessary to maintain the consistency in their rehabilitation. Thus, we suggest that participants gauge their need for treatment and choose the treatment that best suits them and would be beneficial for them. Participants in this study are aware that they will not continue their treatment if it fails in their life and find that FiF CBT is a better option than just “keeping up appearances” through internet-delivered therapies.
Discussion

This study aims to investigate the perceptions that are prevalent among participants who decide to opt out of the possibility of using bCBT for AUD and instead continue with FtF CBT without merging it with internet-based modules. We found that the participants had various reasons for preferring FtF CBT over bCBT, and these reasons were mainly related to personal choices.

Participants’ Assumed Need for Attending Sessions In Person

Being physically in front of the therapist was considered to strengthen a more personal connection between the participant and the therapist and thus the central reason for preferring a synchronous FtF verbal dialogue. Participants considered that FtF allows for all aspects of communication with the therapist to come into play, including nonverbal communication, eye contact, and body language. In a systematic review on women’s expectations and experiences regarding eHealth treatment, Verhoeks et al. [19] found 3 studies that showed women’s negative expectations with regard to receiving eHealth treatment. Those studies showed that the eHealth treatment was perceived as rather impersonal treatment and that the participants valued immediate and empathic responses in their dialogues with the therapist and stressed the importance of nonverbal communication through eye contact and bodily expressions. In the study by Verhoeks et al. [19], the women expressed an intuitive preference for FtF CBT. They feared that the absence of personal contact would make their treatment course more impersonal and impact negatively on their alliance with the therapist, their motivation, and consequently on their treatment outcome.

In this study, we found that the participants had similar feelings as they took their rehabilitation process seriously. They chose the treatment form that they believed they could gain the most from—to them, it was FtF CBT. In general, we found that participants choose FtF CBT rather than bCBT because they had been in a vulnerable position in their rehabilitation process. In particular, it was of importance to them to have a sense of privacy and having a person in front of them when the subject is a personal and vulnerable matter. Since some of the participants had recently stopped drinking, they felt vulnerable in the situation because they realized how alcohol had got hold of them and now were dependent on a therapist’s assistance and guidance to help reduce their alcohol use.

Participants Preferring Verbal Communication

In particular, the participants in our study considered verbal communication to be better than digital written communication to express dialogue, nuances, and reflections and to allow clinicians to ask deeper questions and grasp difficult issues. Further, the participants considered that they would gain more from verbal communication than from digital communication, as they assumed that it enabled clinicians to immediately react to participant’s signals and ask questions about their understanding of the said. The review of Verhoeks et al. [19] also reported difficulties in explaining complex situations and feelings in written text compared to communicating through verbal FtF sessions. In their study [19], participants commented that they were afraid that the therapist would misunderstand their issues given in writing. In this study, the participants stressed the importance of a conversation, in which they would be able to ask questions and discuss problems with the therapist because they needed a more in-depth dialogue about their problems. These findings are corroborated by Runz-Jørgensen et al. [20]. In their study, the participants perceived web-based treatment as undesirable because the therapist would just be waiting for the next person in line and they felt neglected. It should, however, be noted that not all participants agreed with the above interpretation. Other studies report that participants have used text-based interventions for AUD and have found them to be a positive experience in their treatment course [21,22].

The Meaning of Receiving Immediate Feedback on Outstanding Matters

We found that the participants emphasized a need for immediate feedback from their therapist on outstanding matters—a need that they felt that asynchronous digital communication could not fulfil. In situations where they felt alone and in doubt about how to understand a question, it was of importance for them to receive instant elaboration on the matter. Here, the physical presence of the therapist enabled them to receive a quick clarification and thus be able to move on in their rehabilitation. This finding is congruent with findings from the review by Verhoeks et al. [19], where the women wished to be able to ask questions and receive feedback from their therapist during their treatment course. The women stressed the importance of physical presence as they otherwise doubted the quality of the feedback. In continuation of this, participants in the study by Runz-Jørgensen et al. [20] had a wish for even longer FtF consultations with the therapist as they felt that there was not enough time to ask questions or address concerns that were of importance to them.

It can be hypothesized that integrating videoconference-based conversations with therapists in digital treatment solutions for AUD might acknowledge and apprehend participants’ preferences in terms of being able to communicate at a distance without the loss of sensorial stimulation. Further research is thus needed in order to secure that digital solutions become attractive and preferred by participants. We anticipate that our findings may be used for developing information material addressed to therapists regarding participants’ concerns toward bCBT for the therapist to accommodate the potential participant barriers beforehand. Moreover, our findings may be used to inform participants prior to treatment about their possibilities of combining treatment forms in accordance with their specific needs at specific times.

Strengths and Limitations

This study has limitations. The relatively small sample size (N=11) may be a limitation in this study. However, in an experiment with data saturation and variability, Guest et al. [23] found the first 6 interviews to be crucial for the emerging of meta themes. Based on this finding, they recommended a minimum of 6 interviews for developing meaningful themes.
during an inductive analysis. Further, Crouch and McKenzie [24] found that a small number of participants is usable for facilitating the interviewer-interviewee alliance, thereby increasing the validity of semistructured interviews. This study is strengthened by the use of independent parallel coding and code check, which increase internal validity and reliability [25] and thus enhance the credibility of the analysis [26,27]. However, it may be a limitation that we have not used stakeholder check [27]. We saw that, compared to the profiles of the patients in Danish alcohol treatment institutions, our participant population was older. This difference in sample population characteristics may have had an influence on our study results. The relatively higher mean age in our sample may have had an impact, as participants may have less experience with and thereby less interest in using digital interventions [28]. It is also a possibility that the subjects covered in the interview guide and the way of asking by the interviewers may have affected the answers given by the interviewees, but it does not change the fact that they initially did not make use of the bCBT offer.

**Conclusion**

We found multiple reasons for participants choosing FtF CBT over bCBT. Participants expressed a preference for FtF, in particular, due to positive expectations in the various dimensions of FtF, which they felt were important. The participants were worried that they would not feel as motivated, empowered, and obliged to complete treatment if it partly consisted of iCBT, as they would if it purely consisted of FtF sessions with a therapist.

**Acknowledgments**

This study was financially supported by TrygFonden (grant 127727). The authors thank the patients and staff at the municipal alcohol use disorder treatment institutions participating in the “Blending internet treatment into conventional face-to-face treatment for alcohol use disorder (Blend-A)” study. The authors also thank all the colleagues working with Blend-A at the Center for Digital Psychiatry and the Research Unit for Clinical Alcohol Research, especially Ramlo Abdulkadir Mohamed and Jeppe Tryggedsson for retrieving data and Jakob Godsk Nielsen for conducting interviews.

**Data Availability**

The data sets on which this study is based will be made available by the authors on request.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Interview guide for patients who opted out of using blended internet-based and face-to-face treatment for alcohol use disorder. [DOCX File, 17 KB - formative_v81e45471_app1.docx ]

**References**


Abbreviations

AUD: alcohol use disorder
bCBT: blended cognitive behavioral therapy
Blend-A: Blending internet treatment into conventional face-to-face treatment for alcohol use disorder
COREQ: Consolidated Criteria for Reporting Qualitative Studies
CBT: cognitive behavioral therapy

https://formative.jmir.org/2024/1/e45471
FTF: face-to-face
iCBT: internet-based cognitive behavioral therapy
MI: motivational interviewing

Edited by T Leung; submitted 03.01.23; peer-reviewed by M Johansson, L Boß; comments to author 20.03.23; revised version received 05.04.23; accepted 12.11.23; published 13.02.24.

Please cite as:
Tarp K, Christiansen R, Bilberg R, Borkner S, Dalsgaard C, Folker MP, Nielsen AS
Nonuse of Blended Web-Based and Face-To-Face Cognitive Behavioral Therapy for Alcohol Use Disorder: Qualitative Study
JMIR Form Res 2024;8:e45471
URL: https://formative.jmir.org/2024/1/e45471
doi: 10.2196/45471
PMID: 38349711

©Kristine Tarp, Regina Christiansen, Randi Bilberg, Simone Borkner, Caroline Dalsgaard, Marie Paldam Folker, Anette Søgaard Nielsen. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
The Effect of Web-Based Culinary Medicine to Enhance Protein Intake on Muscle Quality in Older Adults: Randomized Controlled Trial

Emily Salas-Groves1, PhD; Michelle Alcorn2, PhD; Allison Childress1, PhD; Shannon Galyean1, PhD

1Nutritional Sciences, Texas Tech University, Lubbock, TX, United States
2Hospitality and Retail Management, Texas Tech University, Lubbock, TX, United States

Corresponding Author:
Shannon Galyean, PhD
Nutritional Sciences
Texas Tech University
1301 Akron Avenue
Lubbock, TX, 79409
United States
Phone: 1 806 834 2286
Email: shannon.galyean@ttu.edu

Abstract

Background: The most common age-related musculoskeletal disorder is sarcopenia. Sarcopenia is the progressive and generalized loss of muscle mass, strength, and function. The causes of sarcopenia can include insufficient nutritional status, which may be due to protein-energy malnutrition, anorexia, limited food access and eating ability, or malabsorption. In the United States, 15.51% of older adults have been diagnosed with sarcopenia. Culinary medicine (CM) is a novel evidence-based medical field that combines the science of medicine with food and cooking to prevent and treat potential chronic diseases. CM helps individuals learn and practice culinary skills while tasting new recipes. Therefore, this program could successfully reduce barriers to protein intake, enabling older adults to enhance their diet and muscle quality.

Objective: This study aimed to examine how a web-based CM intervention, emphasizing convenient ways to increase lean red meat intake, could improve protein intake with the promotion of physical activity to see how this intervention could affect older adults’ muscle strength and mass.

Methods: A 16-week, single-center, parallel-group, randomized controlled trial was conducted to compare a web-based CM intervention group (CMG) with a control group (CG) while monitoring each group’s muscle strength, muscle mass, and physical activity for muscle quality. The CMG received weekly web-based cooking demonstrations and biweekly nutrition education videos about enhancing protein intake, whereas the CG just received the recipe handout. Anthropometrics, muscle mass, muscle strength, dietary habits, physical activity, and cooking effectiveness were established at baseline and measured after the intervention. The final number of participants for the data analysis was 24 in the CMG and 23 in the CG.

Results: No between-group difference in muscle mass ($P=0.88$) and strength (dominant $P=0.92$ and nondominant $P=0.72$) change from the prestudy visit was detected. No statistically significant difference in protein intake was seen between the groups ($P=0.50$). No nonsignificant time-by-intervention interaction was observed for daily protein intake ($P=0.08$). However, a statistically significant time effect was observed ($P≤0.001$). Post hoc testing showed that daily protein intake was significantly higher at weeks 1 to 16 versus week 0 ($P<0.05$). At week 16, the intake was 16.9 (95% CI 5.77-27.97) g higher than that at the prestudy visit.

Conclusions: This study did not affect protein intake and muscle quality. Insufficient consistent protein intake, low physical activity, intervention adherence, and questionnaire accuracy could explain the results. These studies could include an interdisciplinary staff, different recruitment strategies, and different muscle mass measurements. Future research is needed to determine if this intervention is sustainable in the long term and should incorporate a follow-up to determine program efficacy on several long-term behavioral and health outcomes, including if the participants can sustain their heightened protein intake and how their cooking skills have changed.

Trial Registration: ClinicalTrials.gov NCT05593978; https://clinicaltrials.gov/ct2/show/NCT05593978
Introduction

The guidance of the National Institute on Aging classifies older adults as those aged 65 years and older [1]. As adults age, several age-related diseases can occur, the most common being cardiovascular disease, cancer, Alzheimer disease, Parkinson disease, osteoporosis, and sarcopenia [2]. A Global Burden of Disease study in 2017 [3] revealed that 31.4% of all diseases were age related. These age-related diseases, combined with the body and life changes that occur with aging, could contribute to compromised nutritional status. These body and life changes can be physiological, psychosocial, and economic [4]. All these factors play a significant role in nutrition and food choices, which are barriers to appetite and diet quality. Therefore, current research strategies aim to acquire healthy aging and prevent age-related diseases.

Aging can lead to age-related musculoskeletal disorders [5] caused by an imbalance between muscle protein’s anabolic and catabolic pathways, leading to overall skeletal muscle mass (SMM) loss [6]. The most common age-related musculoskeletal disorder is sarcopenia. Sarcopenia is the progressive and generalized loss of muscle mass, strength, and function [2,7,8]. Muscles affected include skeletal [9], smooth [10], and cardiac [11]. Consequently, sarcopenia increases fall and fracture risk [12], impairs daily living activities performance [13], increases cognitive impairment [14], decreases the quality of life [15], and leads to death [16].

In research, the general sarcopenia prevalence ranges from 0.2% to 86.5%, with prevalence in women ranging from 0.3% to 91.2% and prevalence in men ranging from 0.4% to 87.7% [17]. In the United States, 15.51% of older adults have been diagnosed with sarcopenia, demonstrating its magnitude of being a public health burden [18]. Therefore, early identification and intervention are the key factors for achieving improved sarcopenia outcomes. According to the European Working Group on Sarcopenia in Older People (EWGSOP), a sarcopenia diagnosis requires the measurements of muscle mass, strength, and function [6].

Although many factors lead to sarcopenia, the 2 crucial factors that can be controlled in older adults are inadequate nutritional intake and physical inactivity [19,20]. Older adults tend to have anabolic resistance, defined as “a blunted stimulation of muscle protein synthesis (MPS) to common anabolic stimuli in SMM” [21]. Therefore, increasing protein-dense food ingestion and habitual physical activity are frontline strategies to support muscle mass, performance, and health [21]. The Society for Sarcopenia, Cachexia, and Wasting Disease provided protein recommendations for treating and preventing sarcopenia at a minimum of 1.0 to 1.5 g/kg body weight per day with exercise [22]. The protein quality is also critical in age-related SMM anabolism. Research on how protein-rich whole foods (eg, lean red meat) can enhance MPS over supplementation in older adults is rising [23]. Recent data suggest that a moderate 113 g (30 g of protein) serving of animal protein (eg, lean beef) can increase MPS by approximately 50% [24]. Therefore, the per-meal anabolic threshold recommendation is 25 to 30 g of protein [23-25]. Unfortunately, older adults’ protein needs are usually not met. Independent older adults answered the 2005-2014 National Health and Nutrition Examination Survey (NHANES) [26], revealing that up to 46% are not meeting the protein intake recommendation.

Physical activity directly impacts muscle quality and quantity [27]. Inactivity in older adults can promote sarcopenia development [28,29], whereas physical activity increases muscle strength [30,31] and mass [32,33]. Therefore, physical activity is vital to lower sarcopenia prevalence [34-36]. Specifically, resistance training and balance exercises are considered the best for sarcopenia prevention [27,37-41]. Steps through activity trackers can help determine one’s physical activity [42]. Accomplishing 10,000 daily steps is suggested to positively influence body composition (eg, weight and body fat) and improve health parameters (eg, quality of life) [43]. Therefore, nutrition and physical activity have been seen to be essential in countering sarcopenia [44].

More interventions focusing on nutrition and lifestyle changes are essential in decreasing chronic disease and health care costs [45]. Educating and empowering individuals to change their lifestyles can be less costly than medications and invasive procedures [45]. Culinary medicine (CM) is a novel evidence-based medical field defined by combining the science of medicine with food and cooking [46]. CM differs from traditional lifestyle and nutrition interventions by attempting to empower the patient to care for herself or himself safely, effectively, and happily with food and beverages as a primary care technique [47]. It helps people access and eat nutrient-dense meals to prevent and treat potential chronic diseases [46]. Individuals learn and practice culinary skills while tasting new recipes [45]. Also, they can incorporate their favorite foods into their eating plan while learning how to enhance diet quality through new foods (eg, different types of vegetables) and meal preparation tips (eg, defrosting techniques) [47,48]. If executed appropriately, CM can be taught to all populations regardless of culinary skill, educational level, or socioeconomic background [45]. A CM curriculum typically includes practical applications in supermarkets and home kitchens [49]. These practical applications include basic nutrition knowledge and instruction on how to apply that knowledge to diet therapies [49]. However, limited studies report whether a web-based CM curriculum could be as effective as in-person.

Multiple randomized controlled trials report that CM significantly improved individuals’ culinary knowledge, healthy dietary patterns, and self-efficacy for healthier cooking [50-54]. Thus, highlighting CM’s potential as a nutrition intervention could lower the risk of diet-related chronic disease among older adults. However, few studies in this area include older adult participants; none exclusively focused on an older adult population, and only 6% of CM programs were taught by a qualified health professional [55]. Additionally, CM
interventions have been very heterogeneous, indicating a lack of variety in how the intervention is conducted compared with others [55]. Therefore, this study could advance our knowledge of CM and sarcopenia prevention in older adults. A web-based CM program might be an innovative strategy to improve protein intake in independent older adults at home. In addition, this program could successfully reduce barriers to protein intake, enabling older adults to enhance their diet and muscle quality. This factor could be vital because research surrounding CM within older adults is in its infancy. Therefore, our study aimed to examine how a web-based CM intervention, emphasizing convenient ways to increase lean red meat intake, could improve protein intake with the promotion of physical activity to see how this intervention could affect older adults’ muscle strength and mass.

Methods

Study Design

A 16-week, single-center, parallel-group, randomized controlled trial compared a web-based CM intervention group (CMG) with a control group (CG) on their protein intake, cooking effectiveness, muscle strength, muscle mass, and physical activity. The study was conducted at Texas Tech University Nutrition and Metabolic Health Initiative (NMHI), Lubbock, Texas. Participants were permitted to remove themselves from the trial at any time.

Ethical Considerations

A human study compliance review was submitted to the institutional review board at Texas Tech University, Lubbock, Texas. The study was expedited for review and received approval (IRB2021-693). Once participants were recruited and eligibility was determined, an initial appointment was set up at Texas Tech University NMHI. A research team member described the study in detail, and participants were asked to sign a consent form stating willingness to participate. The participants’ information collected for the study was deidentified, given a code number, and kept on the researchers’ computer at Texas Tech University NMHI. The research team offered the participants the vívofit 4 watch (Garmin) as compensation, which they used to complete the study.

Recruitment, Screening, and Participants

Flyers, newsletters, and word of mouth were essential for recruitment. When participants agreed to enroll in the study, they filled out an initial screening questionnaire to help determine whether they met the eligibility criteria. The inclusion criteria involved individuals who are aged 65 years or older, able to cook for themselves, physically active (eg, no need for equipment for assistance), and able to use a computer and mobile device. The exclusion criteria included individuals aged <65 years; those with limited mobility (eg, need for equipment for assistance), cognitive dysfunction (eg, dementia), a heart pacemaker, or type 1 or type 2 diabetes with insulin use; current smokers; those with some form of amputation; those who unable to use a computer and mobile device or unable or unwilling to wear the vívofit 4 watch (Garmin) for the duration of the study; and those undergoing or had recently undergone a severe medical procedure or diagnosis.

Participants were recruited and enrolled from June 2022 to August 2022, with data collection completed in December 2022. If a participant dropped out of the study, a new participant would replace and be allotted to the same group as the participant they replaced. A total of 52 older adults, including both men and women, met the study’s eligibility criteria. Assessments were conducted at the prestudy, weekly, and poststudy time points.

Intervention Design and Study Procedures

Prestudy Visit

Before their visit, participants were told to refrain from exercising for 48 hours, taking alcohol for 12 hours, and wearing clothes with any metals. Informed consent was obtained before starting the assessments. The assessments included completing 4 questionnaires: Community Healthy Activities Model Program for Seniors (CHAMPS), Dietary Screener Questionnaire (DSQ), protein questionnaire, and cooking effectiveness questionnaire. Afterward, grip strength, height, and weight were measured. Then, the participants were scanned by dual-energy x-ray absorptiometry (DXA). After completing their scan, they were given a vívofit 4 watch (Garmin). Lastly, the participants were randomized to either CMG or CG and provided their study’s subject code (eg, Beef Study 01), grip strength and DXA results, and exercise handouts. Both groups were advised to consume 25 to 30 g of protein during every meal, and all questions were answered. A follow-up email was sent providing a sample of a 2-week workout plan based on the exercise recommendation handouts and reminders of the study protocol.

Weekly Interventions

The CMG received weekly web-based cooking demonstrations with a recipe handout and biweekly nutrition education video on general nutrition information based on the Nutrition Care Manual content from the Academy of Nutrition and Dietetics [56], all provided by email at the beginning of each week. Meanwhile, the CG just received the recipe handout by email. Therefore, this intervention was developed to show how effective the hands-on and visual intervention provided to the CMG is compared with just general reading of a recipe with no further education provided to the CG. In addition, at the end of each week, both groups received their weekly protein and cooking effectiveness questionnaires.

A total of 20 recipes focusing on lean ground beef were provided for this study. Before starting the study, the research team tested each recipe and adjusted it as needed based on visual, flavor, and dish size. Then, the cooking demonstration was recorded once the recipe was approved for the study. For weeks 1 and 2, three recipes were sent to the participants. For the remainder of the study, 1 recipe was sent weekly. In addition, educational videos on a specific nutrition topic were sent every 2 weeks. These topics provided the participants with further nutrition education, which is essential regarding their diet outside of protein.
Poststudy Visit

After their 16th week, the participants had their final data collected. At the end of the visit, the primary researcher shared the pre- and poststudy DXA and grip strength results with the participant and answered any questions.

Outcome Measurements

Questionnaires

The following outcomes were measured: weekly activity level through CHAMPS, the diet through the DSQ, protein intake through a protein questionnaire, cooking confidence and attitude using a pre- and poststudy cooking effectiveness questionnaire, and intervention compliance through weekly cooking effectiveness.

CHAMPS is a 41-item questionnaire [57] that assesses the weekly frequency and duration of various lifestyle physical activities that are appropriate for older adults. The DSQ was developed for the 2009-2010 NHANES [58]. It is a 30-item questionnaire that assesses the frequency of consumption of selected foods and drinks in the past month, such as intakes of fruits and vegetables, red and processed meat, dairy or calcium products, added sugars, and whole grains or fiber. The protein questionnaire is a modified version of the rapid self-administered dietary protein food frequency questionnaire, which contains 37 items evaluating the weekly intake of different types of meat, dairy products, eggs, and beans [59].

Lastly, the pre- and poststudy cooking effectiveness questionnaires measured participants’ cooking confidence, attitudes, and challenges or barriers. In addition, the weekly cooking effectiveness reported each group’s compliance toward their intervention. The prestudy cooking effectiveness questionnaire includes 14 items, the weekly cooking effectiveness questionnaire includes 5 items, and the poststudy cooking effectiveness questionnaire includes 33 items.

Anthropometrics

Height was measured using a Charder HM: 200P stadiometer (Charder Electronic Co Ltd) to the nearest half inch. Body weight was measured by a Brecknell MS-1000 wheelchair scale (Brecknell) to the nearest 0.5 lbs.

Muscle Quality

Lean body and fat mass were measured using a Norland XR-800 DXA (Swissray International, Inc) to the nearest gram. Muscle strength was measured by a Camry Digital Hand Dynamometer (Camry Scale) to the nearest kilogram for dominant and nondominant hands. Steps were measured by the vívofit 4 watch (Garmin).

Statistical Analysis

The study was powered to identify pre- to poststudy changes between the groups. A similar study [60] was used to develop the necessary sample and effect size using the G^*power software (version 3.1.9.6; Heinrich Heine University Düsseldorf). Calculations were made for a total sample of 52 participants (26 participants per group) to obtain a statistical difference in muscle strength and mass between the groups, assuming an α of 5%, effect size of 0.72, power of 80%, and 10% inflation for dropouts. Data were imported to SPSS (version 29; IBM Corp) for analysis. DXA measuring muscle and fat mass was the study’s primary outcome measure. Secondary outcomes included protein intake in grams, muscle strength in kilogram, average daily steps, frequency of physical activity in minutes per week, height in inches, and weight in kilograms.

Participants were randomized to the CMG or the CG by block randomization using 2 blocks with 26 codes. On the basis of the assigned participant’s study code, the primary researcher enrolled the participants into their group at the end of their initial visit. Therefore, the allocation was not concealed. The analysis assessed the effect of the intervention with the completers. Any missing data were replaced with the last observation carried forward before analyses of all measurements via single imputation. Participants were excluded from data analysis if they did not complete over 50% of their weekly questionnaires or, after enrollment, met an exclusion criterion.

Results are presented as mean (SD), mean (95% CI), ranges, or frequencies. P<.05 was considered statistically significant. Linear mixed models were used to assess the differences in protein intake between the groups at the end of the intervention. The model included the fixed effects of time, intervention, and time-by-intervention interaction. Participants were modeled as a random effect to account for the repeated measures design. When a significant main effect was observed, post hoc analyses were conducted and the Tukey-Kramer method was used to adjust for multiple comparisons. Within-group muscle mass and strength differences, as well as physical activity and diet quality differences, were estimated using an independent samples (1-tailed) t test for variables measured before and after the study.

Study Population

In total, 64 participants expressed interest in the study. Of these, 8 (13%) were excluded during web-based or telephone screening due to failing to meet the inclusion criteria or losing contact. A total of 56 participants were eligible for inclusion and were randomized: 29 to the CMG and 27 to the CG. A total of 25 participants in the CMG, compared with 24 in the CG, completed the 16-week weekly questionnaires and both study visits. Of the eligible 56 participants, 7 (13%) withdrew or dropped out before the completion of the study. Of the 7 participants, 6 (86%) dropped out due to medical reasons unrelated to the study, and 1 participant (14%) dropped out due to family reasons. Of the 56 participants, 2 (4%) participants had to be excluded from the data analysis because 1 participant had bariatric surgery during the study and the other completed less than 50% of their weekly questionnaires. Therefore, a total of 49 participants were included for the data analysis (CMG: 24/29, 83%; CG: 23/27, 85%). See the CONSORT (Consolidated Standards of Reporting Trials) study flow diagram (Figure 1) for the study details.
The prestudy characteristics of the groups are presented in Table 1. The study included a greater proportion of female (38/47, 81%) and White (44/47, 94%) participants. The mean age, weight, and BMI of the participants in the CMG were 71.4 (SD 5.2) years, 76.6 (SD 17.4) kg, and 28.0 (SD 6.0) kg/m², respectively. In the CG, they were slightly older (mean 73.2, SD 5.8 years) but had lower weight (mean 69.4, SD 15.0 kg) and BMI (mean 26.1, SD 5.0 kg/m²). The CG was found to be more physically active than the CMG. Regarding diet, the CG consumed more fiber, calcium, dairy, vegetables, and fruit than the CMG. Meanwhile, the CMG consumed more daily added sugar than the CG. However, both groups consumed the same amount of daily whole grains.
### Table 1. Prestudy participant characteristics in body composition, physical activity, and diet components (n=47).

<table>
<thead>
<tr>
<th>Variable</th>
<th>CMG (n=24)</th>
<th>CG (n=23)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4 (17)</td>
<td>5 (22)</td>
</tr>
<tr>
<td>Female</td>
<td>20 (83)</td>
<td>18 (78)</td>
</tr>
<tr>
<td><strong>Age, mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71.4 (5.2)</td>
<td>73.2 (5.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Age group (years) n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>18 (75)</td>
<td>15 (65)</td>
</tr>
<tr>
<td>75-84</td>
<td>6 (25)</td>
<td>7 (30)</td>
</tr>
<tr>
<td>≥85</td>
<td>N/A</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>1 (4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Hispanic or Mexican American</td>
<td>2 (8)</td>
<td>N/A</td>
</tr>
<tr>
<td>White</td>
<td>21 (88)</td>
<td>23 (100)</td>
</tr>
<tr>
<td><strong>Body composition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>76.6 (17.4)</td>
<td>69.4 (15.0)</td>
</tr>
<tr>
<td>Height (inches), mean (SD)</td>
<td>65 (3.6)</td>
<td>64.1 (3.7)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>28.0 (6.0)</td>
<td>26.1 (5.0)</td>
</tr>
<tr>
<td><strong>BMI (kg/m²) n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤18.5, underweight</td>
<td>N/A</td>
<td>2 (9)</td>
</tr>
<tr>
<td>18.6-24.9, normal</td>
<td>8 (33)</td>
<td>8 (35)</td>
</tr>
<tr>
<td>25-29.9, overweight</td>
<td>9 (38)</td>
<td>8 (35)</td>
</tr>
<tr>
<td>30-34.9, class I obesity</td>
<td>4 (17)</td>
<td>5 (22)</td>
</tr>
<tr>
<td>35-39.9, class II obesity</td>
<td>2 (8)</td>
<td>N/A</td>
</tr>
<tr>
<td>≥40, class III obesity</td>
<td>1 (4)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Physical activity (min/wk), mean (SD)</strong></td>
<td>838.8 (545.9)</td>
<td>930.0 (649.1)</td>
</tr>
<tr>
<td><strong>Diet components, mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fiber (g)</td>
<td>15.9 (2.8)</td>
<td>16.7 (2.9)</td>
</tr>
<tr>
<td>Calcium (mg)</td>
<td>905.2 (180.6)</td>
<td>932.1 (167.3)</td>
</tr>
<tr>
<td>Whole grain (ounce)</td>
<td>0.8 (0.3)</td>
<td>0.8 (0.3)</td>
</tr>
<tr>
<td>Total added sugar (teaspoons)</td>
<td>15.3 (4.9)</td>
<td>13.8 (2.9)</td>
</tr>
<tr>
<td>Dairy (cup)</td>
<td>1.4 (0.4)</td>
<td>1.6 (0.5)</td>
</tr>
<tr>
<td>Vegetables (cup)</td>
<td>1.5 (0.3)</td>
<td>1.6 (0.4)</td>
</tr>
<tr>
<td>Fruit (cup)</td>
<td>0.8 (0.3)</td>
<td>1.0 (0.4)</td>
</tr>
</tbody>
</table>

---

**Muscle Mass and Strength Outcomes**

There was no between-group difference in the muscle mass change from the prestudy visit \((P=0.88; \text{Table 2})\). Using the EWGSOP sarcopenia diagnosis [61], 21% (5/24) of the CMG and 26% (6/23) of the CG had low muscle mass at the prestudy visit. At the poststudy visit, 21% (5/24) of the CMG and 22% (5/23) of the CG had low muscle mass.

---

\(^{a}\)Randomized controlled trial (June 2022 to August 2022; Texas Tech University Nutrition and Metabolic Health Initiative) evaluating the effect of a web-based culinary medicine intervention on protein intake, cooking effectiveness, muscle strength, muscle mass, and physical activity in an older adult population aged 65 years and older.

\(^{b}\)CMG: culinary medicine intervention group.

\(^{c}\)CG: control group.

\(^{d}\)N/A: not applicable.
Table 2. Mean muscle mass and strength of participants at the pre- and poststudy visits (n=47).

<table>
<thead>
<tr>
<th>Variablea</th>
<th>CMGb (n=24), mean (SD)</th>
<th>CGc (n=23), mean (SD)</th>
<th>Poststudy between-group differences, mean (95% CI)</th>
<th>P valueb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Body composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle mass (g)</td>
<td>40,424.3 (9891.6)</td>
<td>41,042.4 (9857.0)</td>
<td>39,816.9 (7496.0)</td>
<td>39,974.3 (7581.4)</td>
</tr>
<tr>
<td>Fat mass (g)</td>
<td>33,043.5 (12,416.6)</td>
<td>33,324.4 (12,654.0)</td>
<td>26,942.5 (9499.2)</td>
<td>26,124.9 (1744.3)</td>
</tr>
<tr>
<td>Muscle mass dominant (kg)</td>
<td>20.1 (6.4)</td>
<td>22.0 (6.5)</td>
<td>21.1 (7.5)</td>
<td>22.7 (6.9)</td>
</tr>
<tr>
<td>Muscle mass nondominant (kg)</td>
<td>18.5 (6.4)</td>
<td>20.3 (6.3)</td>
<td>19.2 (7.5)</td>
<td>20.8 (7.9)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>76.6 (17.4)</td>
<td>76.9 (17.5)</td>
<td>69.4 (15.0)</td>
<td>68.6 (14.3)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>28.0 (6.0)</td>
<td>28.0 (6.0)</td>
<td>26.1 (5.0)</td>
<td>25.8 (4.7)</td>
</tr>
<tr>
<td>Physical activity (min/wk)</td>
<td>838.8 (545.9)</td>
<td>968.75 (619.9)</td>
<td>930.0 (649.1)</td>
<td>948.3 (602.7)</td>
</tr>
<tr>
<td>Steps</td>
<td>_e</td>
<td>5921.5 (2887.8)</td>
<td>—</td>
<td>6049.9 (3597.5)</td>
</tr>
</tbody>
</table>

aThe independent samples t test was used to compare between-group differences at the poststudy visit.
bCMG: culinary medicine intervention group.
cCG: control group.
dP value refers to between-group differences by the independent samples t test.
eNot available.

Similar results were seen for muscle strength. There was no between-group difference in the muscle strength change from the prestudy visit (dominant: P=.92 and nondominant: P=.72). When comparing the classification of muscle strength for the dominant hand, the CMG was considered 29% (7/24) weak, 67% (16/24) normal, and 4% (1/24) strong at the prestudy visit. At the poststudy visit, the CMG was considered 33% (8/24) weak, 46% (11/24) normal, and 21% (5/24) strong. The CG was considered 13% (3/24) weak, 83% (19/24) normal, and 4% (1/24) strong at the prestudy visit. At the poststudy visit, the CG was considered 13% (3/23) weak, 74% (17/23) normal, and 13% (3/23) strong.

When comparing the classification of muscle strength for the nondominant hand, the CMG was considered 42% (10/24) weak, 54% (13/24) normal, and 4% (1/24) strong at the prestudy visit. At the poststudy visit, the CMG was considered 38% (9/24) weak, 50% (12/24) normal, and 13% (3/24) strong. On the other hand, the CG was considered 30% (7/23) weak, 65% (15/23) normal, and 4% (1/23) strong at the prestudy visit. At the poststudy visit, the CG was considered 30% (7/23) weak, 57% (13/23) normal, and 13% (3/23) strong.

Per the EWGSOP sarcopenia diagnosis [61], 38% (9/24) of the CMG and 30% (7/23) of the CG could be diagnosed with probable sarcopenia. In comparison, 8% (2/24) of the CMG and 9% (2/23) of the CG could be diagnosed with sarcopenia at the prestudy visit. At the poststudy visit, 33% (8/24) of the CMG and 17% (4/23) of the CG could be diagnosed with probable sarcopenia, whereas 8% (2/24) of the CMG and 9% (2/23) of the CG could be diagnosed with sarcopenia at the poststudy visit.

Protein Intake and Diet Quality

Figure 2 reveals the mean (SD) daily protein intake in grams for each week of the study for each group. A nonsignificant time-by-intervention interaction was observed for daily protein intake (Figure 2 and Table 3; P=.08). There was also no statistically significant difference in protein intake between the interventions (P=.50). However, a statistically significant time effect was observed (P<.001). Post hoc testing showed that daily protein intake was significantly higher at weeks 1 to 16 versus week 0 (P<.05) in the cohort. At week 16, protein intake was 16.9 (95% CI 5.77-27.97) g higher than that at the prestudy visit.
Figure 2. Mean (SD) daily protein intake in grams for each week of the study for each group. CM: culinary medicine; CN: control.

Table 3. Dietary intake of participants at the pre- and poststudy visits (N=47).

<table>
<thead>
<tr>
<th>Variable</th>
<th>CMG (n=24), mean (SD)</th>
<th>CG (n=23), mean (SD)</th>
<th>Poststudy between-group differences, mean (95% CI)</th>
<th>P value&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein (g)</td>
<td>60.6 (5.1)</td>
<td>73.3 (5.4)</td>
<td>50.5 (5.2)</td>
<td>71.7 (5.4)</td>
</tr>
<tr>
<td>Fiber (g)</td>
<td>15.9 (2.8)</td>
<td>16.3 (2.5)</td>
<td>16.7 (2.9)</td>
<td>16.1 (2.8)</td>
</tr>
<tr>
<td>Calcium (mg)</td>
<td>905.2 (180.6)</td>
<td>904.4 (164.6)</td>
<td>932.1 (167.3)</td>
<td>930.5 (251.3)</td>
</tr>
<tr>
<td>Whole grain (ounce)</td>
<td>0.8 (0.3)</td>
<td>0.7 (0.3)</td>
<td>0.8 (0.3)</td>
<td>0.7 (0.3)</td>
</tr>
<tr>
<td>Total added sugar (teaspoons)</td>
<td>15.3 (4.9)</td>
<td>12.5 (2.7)</td>
<td>13.8 (2.9)</td>
<td>13.2 (2.3)</td>
</tr>
<tr>
<td>Dairy (cup)</td>
<td>1.4 (0.4)</td>
<td>1.4 (0.4)</td>
<td>1.6 (0.5)</td>
<td>1.6 (0.7)</td>
</tr>
<tr>
<td>Vegetables (cup)</td>
<td>1.5 (0.3)</td>
<td>1.6 (0.4)</td>
<td>1.6 (0.4)</td>
<td>1.6 (0.4)</td>
</tr>
<tr>
<td>Fruit (cup)</td>
<td>0.8 (0.3)</td>
<td>0.8 (0.3)</td>
<td>1.0 (0.4)</td>
<td>0.8 (0.2)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Linear mixed-effects model analysis was used to compare between-group differences after the study for protein, whereas an independent samples t-test was used for the remaining variables.

<sup>b</sup>CMG: culinary medicine intervention group.

<sup>c</sup>CG: control group.

<sup>d</sup>P value refers to linear mixed-effects model analysis of between-group differences over time (time×treatment interaction).

Each group was evaluated to see how many participants met their protein needs (1.0-1.2 g/kg body mass per day). In the CG, 39% (9/23) participants did not meet their needs, 26% (6/23) did meet their needs, and 35% (8/23) exceeded their needs during the study. In the CMG, 58% (14/24) participants did not meet their needs, 8% (2/24) did meet their needs, and 33% (8/23) exceeded their needs during the study. Additionally, in all the completed protein questionnaires, the CMG and the CG had blank answers for 15.4% (63/408) and 12.5% (49/391) of their questions, respectively. When evaluating the daily intake for each dietary component from the DSQ (Table 3), the components stayed close to the same when comparing pre- with poststudy results.

Cooking Effectiveness

For the CMG, participants reported watching 82.8% (318/384) of the intervention videos. The primary reason reported on why they did not watch the videos was “not interested in watching” (21/56, 38%). Additional reasons included personal reasons, traveling or vacation, or they did not receive the video. For the CG, participants reported that they read 94.8% (349/368) of the recipes sent to them. The primary reason why the participants did not read the recipe was “busy” (5/13, 39%). Additional reasons included personal and medical reasons, laziness, uninterest, not receiving the video, and having their spouse read it.
When examining whether both groups cooked the recipe learned through web-based videos or just by reading the recipe, the CMG cooked more recipes than the CG (64.8%, 249/384, vs 62.5%, 230/368). Based on the questionnaires with responses outside of “N/A,” the CMG and CG did not cook primarily because of “holiday, traveling, or vacation” (CMG: 20%, 25/125, and CG: 26.5%, 35/132). See Table 4 for the remaining reasons. Barriers or complications that were reported from both groups when either watching the videos or preparing the recipe included borrowing ingredients from a neighbor; recipe serving size being too big; confusion toward either the ingredients or methods; changing or not including ingredients to meet taste or diet preference; finding certain ingredients at the store; too much spice or ingredient in the recipe for their palette; standing for an extended period was challenging; difficulties in scheduling time and energy to shop, prepare, or cook; or taking more initiative to prepare recipe themselves.

<table>
<thead>
<tr>
<th>Reasons</th>
<th>CMG(^a) (n=125), n (%)</th>
<th>CG(^b) (n=132), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holiday, traveling, or vacation</td>
<td>25 (20)</td>
<td>35 (26.5)</td>
</tr>
<tr>
<td>Busy</td>
<td>16 (12.8)</td>
<td>26 (19.7)</td>
</tr>
<tr>
<td>Spouse prepared it</td>
<td>1 (0.8)</td>
<td>17 (12.9)</td>
</tr>
<tr>
<td>Not interested in cooking</td>
<td>18 (14.4)</td>
<td>14 (10.6)</td>
</tr>
<tr>
<td>Ate leftovers</td>
<td>1 (0.8)</td>
<td>11 (8.3)</td>
</tr>
<tr>
<td>Medical reason</td>
<td>9 (7.2)</td>
<td>7 (5.3)</td>
</tr>
<tr>
<td>Fixed other recipe</td>
<td>17 (13.6)</td>
<td>6 (4.5)</td>
</tr>
<tr>
<td>Did not go to the store</td>
<td>5 (4)</td>
<td>4 (3)</td>
</tr>
<tr>
<td>Recipe too large</td>
<td>3 (2.4)</td>
<td>4 (3)</td>
</tr>
<tr>
<td>Food preference</td>
<td>19 (15.2)</td>
<td>3 (2.3)</td>
</tr>
<tr>
<td>Ate out</td>
<td>1 (0.8)</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>Did not have the recipe</td>
<td>0 (0)</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Financial reason</td>
<td>0 (0)</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>No reason provided</td>
<td>2 (1.6)</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Personal reason</td>
<td>6 (4.6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Confusion toward ingredients</td>
<td>2 (1.6)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\(^a\)CMG: culinary medicine intervention group.  
\(^b\)CG: control group.

At the end of the study, both groups were asked about the main challenges or barriers to maintaining their protein intake (Tables 5 and 6). Meanwhile, the CMG participants were asked how the CM videos specifically helped clarify managing their protein intake (Table 7) and what the most memorable thing they recalled after watching the video or what their favorite part of the CM videos was. All CMG participants were reported having no technical difficulties accessing and watching the videos.

<table>
<thead>
<tr>
<th>Challenge or barrier</th>
<th>Value (n=24), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding low-fat protein options</td>
<td>1 (4)</td>
</tr>
<tr>
<td>I was eating all the time</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Keeping track of protein intake</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Knowing which protein is healthiest or easiest</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Limiting protein intake</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Not consuming enough daily</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Price of protein</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Deciding the correct protein amount to eat</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Protein variety</td>
<td>3 (13)</td>
</tr>
<tr>
<td>No answer provided</td>
<td>3 (13)</td>
</tr>
<tr>
<td>No issues</td>
<td>9 (38)</td>
</tr>
</tbody>
</table>
Table 6. Control group’s main challenges or barriers to maintaining their protein intake.

<table>
<thead>
<tr>
<th>Challenge or barrier</th>
<th>Value (n=23), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High calories in cheese or red meat</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Paying attention when shopping</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Time to prepare</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Eating 25-30 g of protein was too much for me</td>
<td>1 (4)</td>
</tr>
<tr>
<td>No answer provided</td>
<td>3 (13)</td>
</tr>
<tr>
<td>No issue</td>
<td>16 (70)</td>
</tr>
</tbody>
</table>

Table 7. How the culinary medicine videos specifically helped clarify managing their protein intake.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Value (n=24), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding serving or portion size</td>
<td>7 (29)</td>
</tr>
<tr>
<td>New cooking ideas</td>
<td>2 (8)</td>
</tr>
<tr>
<td>How important protein is to our health</td>
<td>2 (8)</td>
</tr>
<tr>
<td>I am visual learner, so helped my confidence</td>
<td>2 (8)</td>
</tr>
<tr>
<td>I realized that I do not eat enough protein</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Introduce more protein into my own recipes</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Learning new skills in cooking</td>
<td>1 (4)</td>
</tr>
<tr>
<td>How easy it is to manage protein intake by cooking myself</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Hard to tell how much protein I got from eating a serving size</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Helped but I get busy and forget to eat during the day</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Waste of time</td>
<td>1 (4)</td>
</tr>
<tr>
<td>No answers provided</td>
<td>3 (13)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

To the authors’ knowledge, a study has yet to be performed with CM explicitly targeting the older adult population to enhance their protein intake. However, a statistically significant time effect was observed ($P \leq 0.001$). Furthermore, post hoc testing showed that daily protein intake was significantly higher at weeks 1 to 16 versus week 0 ($P < 0.05$). At week 16, protein intake was 16.9 (95% CI 5.77-27.97) g higher than that at the prestudy visit. This result indicates that protein intake increased in the cohort with the information provided to both groups.

Nevertheless, there was no additive effect of the CMG over what the CG received because no between-group differences were observed for any primary or secondary outcomes. Insufficient consistent protein intake, low physical activity, adherence to the intervention, and accuracy of the questionnaires could explain the results. Also, participants’ ethnicity, average age, gender, and BMI were similar in both groups and affected the diversity of the study’s population; therefore, the outcomes were not tested against them because there was no vast difference to indicate a relationship. Given the limited representation of men in the cohort, the results cannot be generalized to men, Hispanic participants, and African American participants.

The accuracy of each group’s protein questionnaire could play a factor because they were self-administered. Self-administered questionnaires are more susceptible to item nonresponse [62]. The CMG and the CG had blank answers for 15.4% (63/408) and 12.5% (49/391) of their questions, respectively, suggesting that their intake could have been higher and explained how their muscle mass was overall maintained. Additionally, the participants were not asked to change their diet outside their protein intake. The DSQ reported that participants’ diets did remain the same.

Comparison With Prior Work

Before the study started, both groups were recommended to consume 25 to 30 g of protein per meal in addition to daily physical activity. These recommendations are similar to Paddon-Jones and Rasmussen’s [63] findings, reporting that approximately 25 to 30 g of protein per meal is a valuable strategy for maintaining muscle mass in older adults. This strategy would mean that the participants would have to eat approximately 75 to 90 g of protein daily. The CMG met this range from weeks 6 to 15, but the CG met this range during weeks 4, 11, and 12. Specifically, 39% (9/23) of the CG and 58% (14/24) of the CMG did not meet their needs (1.0-1.2 g/kg body weight per day). The 2005-2014 NHANES [26] reported that 31% to 50% of older adults did not meet their protein recommendations. Our population was in this range. Therefore, these results could also explain why muscle mass did not...
significantly increase between the groups. However, the estimated average requirement for 51 to 70 years is 0.66 g/kg/d, and the recommended dietary intake is 0.8 g/kg/d for all adults over 18 years old, including older adults [64]. Therefore, in the context of adequate energy intake, muscle mass was maintained in this cohort if their protein intake was consistent with these levels.

Grip strength has been used in research to determine overall body strength [65,66]. However, there were no between-group differences in muscle strength change from the prestudy visit. Kim et al [67] found no association between the amount and change (increase or decrease) in daily total protein intake with the incidence or prevention of low muscle strength, which was similar to our results. Additionally, a longitudinal study [68] indicated that 25 to 30 g of protein per meal is associated with greater muscle strength in older adults. However, this recommended intake did not consistently happen in our study, and participants did not meet their calculated needs, which could affect their muscle strength. Physical activity also did not impact muscle strength. Similar results were seen with Ramsey et al [69], who also saw no association between the number of steps and handgrip strength.

When looking at their steps, current evidence suggests that healthy older adults should meet approximately 7000 to 10,000 steps per day [70]. However, our study showed that 67% (16/24) of the CMG and 57% (13/23) of the CG did not meet this range. Also, Park et al [34] reported that individuals who walked at least 7000 to 8000 steps daily likely have muscle mass above the sarcopenia threshold. Because only 33% of the CMG and 44% of the CG met this threshold, it is unsurprising that their lack of steps may have impacted our results.

Lastly, the dropout rates were similar, 14% (4/29) in the CMG and 11% (3/27) in the CG. This rate is lower than the reported average of 20% to 49%, which is commonly seen in dietary clinical trials [71]. In the CMG, 10% (3/29) of participants dropped out due to medical reasons, whereas 3% (1/29) dropped out due to family reasons. In the CG, all the participants dropped out due to medical reasons. These are all common reasons for dropouts in clinical trials [72]. The dropouts were not related to the study, and no adverse effects were reported throughout the study.

Strength and Limitations

Strength

This study is the first to evaluate CM’s effect on enhancing protein intake and muscle quality in older adults, which brings a new aspect to existing CM research. Furthermore, this study allowed us to see if the intervention program improved their knowledge, awareness, and attitude toward protein intake within 4 months. In addition, the feedback from the participants can be applied to future studies.

A registered dietitian (RD), fully trained and qualified with years of experience, developed the whole program with assistance from those with expertise in food service and kinesiology. In addition, an RD implemented the intervention and provided advice if participants needed clarification about their intervention.

Conclusions and Future Direction

To the authors’ knowledge, this study is the first to examine the outcomes of CM in the form of web-based cooking demonstrations and nutrition education to enhance protein intake and muscle quality in older adults. The results reveal insufficient evidence because no between-group differences were observed for primary or secondary outcomes. However, most of the intervention group reported that the cooking demonstrations helped them prepare and cook recipes at home, providing more confidence in the kitchen, and its learning was feasible for them. Our study had an overall dropout rate and data exclusion of 16% (9/56), limiting attrition bias. Additionally, there was a high response rate to the weekly questionnaires, with 84.6% (345/408) for the CMG versus 87.5% (342/391) for the CG, and the response rate goal for most research was approximately 60% [73]. This high response rate was credited to weekly adherence checks and effective accountability in recording their weekly questionnaires. Lastly, this intervention was low-cost and could be easily replicated and enhanced for future research.

Limitations

Although exercise recommendation handouts were given in this study, the main intervention has limitations with a focus on diet and nutrition education. A more comprehensive approach including digital CM education, exercise training sessions, and dietary supplementation would have allowed for a more adequate comparison and expectation of significant differences in muscle quality outcomes. Additionally, the result of this study may not be representative of the general population because the majority were female (38/47, 81%) and White (44/47, 94%), and their ages were similar. Therefore, this study would benefit from seeing its effect on those who lack cooking confidence and skills in addition to a more diverse population setting. In addition, there may be recall and social desirability biases as the questionnaires were self-reported, and the participants knew that the research team was reading the responses. This factor could be lessened through the interview-administered questionnaires. Finally, the protein questionnaire results may not be accurate because of the blank questions.

Some participants reported that they could not cook a recipe because they were on Weight Watchers or had self-proclaimed dietary restrictions (eg, no bread or pasta). This situation was seen in 15.2% (19/125) of the CMG and 2.3% (3/132) of the CG. Also, participants reported that some recipes could have been better for a different season (eg, chili in the winter instead of during the summer). They also voiced concern about some recipes needing smaller portions because they live alone. Additionally, because this intervention was performed in summer, fall, and the beginning of winter, the seasonal changes can explain why participants did not partake in some weeks of the study. For example, the participants did not cook their recipes because of holidays, traveling, or vacations (CMG: 20%, 25/125, and CG 26.5%, 35/132). Another example is that the colder weather and traveling could have impacted the results of the steps because most of the questions asked were about outdoor and in-house activities.
In the future, it would be valuable to further investigate the factors that could have affected this study. In developing and implementing this study, exercise training sessions and a dietary supplement could be included. Additionally, the research study design could include RDs, chefs, exercise physiologists, health coaches, or psychologists. The staff would be essential in creating the study protocol, kitchen equipment checklist, consent forms, scripts, and questions. During recruitment, it would be ideal to obtain a broad age range with an equal gender and ethnicity ratio to help reciprocate the general population. The recipes should consider the season, 1-person portion size, time, cost, and mild flavors. A protein food diary could help keep track of protein intake during the week and help answer the protein questionnaire accurately.

It could be interesting to incorporate muscle biopsy and biomarkers, such as vitamin B₁₂, folate, and creatinine, to evaluate muscle mass further and see if this intervention impacts or could explain why muscle mass outcomes were nonsignificant due to predispositions. However, there are challenges in successfully performing a muscle biopsy in older men and women who are frail or have low body mass [74], so that would be a concern to consider. For biomarkers, no specific recommendations, references, or cutoff values are available to assess muscle mass or quality. Therefore, the biomarkers could be used to notice any significant change within a short time duration. Overall, given the current concern of sarcopenia, these concepts could enhance this intervention further with the information gathered in this study to impact public health.

Acknowledgments
The authors want to acknowledge Texas Tech University and Nutrition and Metabolic Health Initiative (NMHI) for the support of their facilities.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT eHEALTH Checklist (V 1.6.2).
[PDF File (Adobe PDF File), 113 KB - formative_v8i1e49322_app1.pdf ]

References


Abbreviations

CG: control group
CHAMPS: Community Healthy Activities Model Program for Seniors
CM: culinary medicine
CMG: culinary medicine intervention group
CONSORT: Consolidated Standards of Reporting Trials
DSQ: Dietary Screener Questionnaire
DXA: dual-energy x-ray absorptiometry
EWGSOP: European Working Group on Sarcopenia in Older People
MPS: muscle protein synthesis
NHANES: National Health and Nutrition Examination Survey
NMHI: Nutrition and Metabolic Health Initiative
RD: registered dietitian
SMM: skeletal muscle mass

©Emily Salas-Groves, Michelle Alcorn, Allison Childress, Shannon Galyean. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Using #ActuallyAutistic on Twitter for Precision Diagnosis of Autism Spectrum Disorder: Machine Learning Study

Aditi Jaiswal¹, MS; Peter Washington¹, PhD

Department of Information and Computer Sciences, University of Hawaii at Manoa, Honolulu, HI, United States

Corresponding Author:
Aditi Jaiswal, MS
Department of Information and Computer Sciences
University of Hawaii at Manoa
Room 312C, Pacific Ocean Science and Technology
1680 East-West Road
Honolulu, HI, 96822
United States
Phone: 1 8088296359
Email: ajaiswal@hawaii.edu

Abstract

Background: The increasing use of social media platforms has given rise to an unprecedented surge in user-generated content, with millions of individuals publicly sharing their thoughts, experiences, and health-related information. Social media can serve as a useful means to study and understand public health. Twitter (subsequently rebranded as “X”) is one such social media platform that has proven to be a valuable source of rich information for both the general public and health officials. We conducted the first study applying Twitter data mining to autism screening.

Objective: This study used Twitter as the primary source of data to study the behavioral characteristics and real-time emotional projections of individuals identifying with autism spectrum disorder (ASD). We aimed to improve the rigor of ASD analytics research by using the digital footprint of an individual to study the linguistic patterns of individuals with ASD.

Methods: We developed a machine learning model to distinguish individuals with autism from their neurotypical peers based on the textual patterns from their public communications on Twitter. We collected 6,515,470 tweets from users’ self-identification with autism using “#ActuallyAutistic” and a separate control group to identify linguistic markers associated with ASD traits. To construct the data set, we targeted English-language tweets using the search query “#ActuallyAutistic” posted from January 1, 2014, to December 31, 2022. From these tweets, we identified unique users who used keywords such as “autism” OR “autistic” OR “neurodiverse” in their profile description and collected all the tweets from their timeline. To build the control group data set, we formulated a search query excluding the hashtag, “-#ActuallyAutistic,” and collected 1000 tweets per day during the same time period. We trained a word2vec model and an attention-based, bidirectional long short-term memory model to validate the performance of per-tweet and per-profile classification models. We also illustrate the utility of the data set through common natural language processing tasks such as sentiment analysis and topic modeling.

Results: Our tweet classifier reached a 73% accuracy, a 0.728 area under the receiver operating characteristic curve score, and an 0.71 F₁-score using word2vec representations fed into a logistic regression model, while the user profile classifier achieved an 0.78 area under the receiver operating characteristic curve score and an F₁-score of 0.805 using an attention-based, bidirectional long short-term memory model. This is a promising start, demonstrating the potential for effective digital phenotyping studies and large-scale intervention using text data mined from social media.

Conclusions: Textual differences in social media communications can help researchers and clinicians conduct symptomatology studies in natural settings.

(JMIR Form Res 2024;8:e52660) doi:10.2196/52660

KEYWORDS

ASD: autism spectrum disorder; machine learning; natural language processing; public health; sentiment analysis; social media analysis; Twitter
Introduction

Autism spectrum disorder (ASD) is a developmental disability causing physical, cognitive, and behavioral changes and affecting millions of individuals. A core complexity of ASD lies in its dynamic symptom profile that changes with age, often leading to the misattribution of behavioral characteristics to other conditions such as anxiety, obsessive-compulsive disorder, and attention-deficit/hyperactivity disorder [1,2]. Yet there are limitations on the availability of standard tests [3], leading to misdiagnosis or delayed treatments [4], putting patients at risk of developing depression or suicidal tendencies [5]. Social media has become a useful means for real-time public health monitoring, offering insights into individuals’ thoughts, emotions, behaviors, and daily struggles and symptomatology related to health issues. Such nonclinical data hold considerable potential for clinicians and researchers to extract meaningful insights through a less intrusive approach. This digital footprint can be analyzed to study the behavioral symptoms of ASD and other mental health disorders [6].

In recent years, social media has emerged as a promising tool for mining behavioral and observational data. The collection of digital data from social media, wearable devices, and smartphones holds potential for improving health care. Research in mental health, such as identifying depression and mood changes [7-13] and real-time mapping of natural disasters [14,15] or infectious disease spread and its effect on emotional health [16-23] has greatly benefited from such “digital phenotyping” studies. Among social media platforms, Twitter (subsequently rebranded as “X”), known for its concise microblogging nature with tweets limited to 280 characters, has emerged as a valuable source of personalized data, boasting an active monthly user base of around 450 million individuals [24].

ASD has been the subject of multiple clinical trials, reviews, and epidemiological studies conducted using behavioral features such as eye gaze [25], prosody [26], asynchronous body movement [27], facial expressions [28,29], mobile phone data [30-33], or even electroencephalograms [34]. However, only a handful of studies have used social analytical tools [35-38], especially Twitter [39-41], for investigating ASD. In addition, other social networking sites such as Reddit [42-45], Facebook [46], Instagram [47,48], Flickr [49], and Sina Weibo [50] have also provided a valuable source of data for detecting and studying mental health conditions, substance abuse, and risky behaviors. Using these previous works as inspiration, we curated a novel, extensive Twitter data set to study various aspects of social communication that differentiate people with autism from their neurotypical peers on a larger scale than previous work.

Our goal was to examine communication patterns and social cues indicative of emotional states to identify distinctive textual features associated with ASD. The methods outlined in this study could potentially aid researchers and clinicians to understand and analyze linguistic features associated with ASD, enabling the research community to build precision health tools for identifying and monitoring early symptoms, understanding specific behavioral traits, uncovering hidden patterns, proposing a tailored clinical treatment plan or personalized interventions, and offering support within communities. However, it should be noted that this research serves as a supplementary resource for clinicians, aiming to showcase how social media can aid in developing risk assessments, custom treatment plans, and targeted interventions based on the patient’s individual traits, communication style, and lifestyle. Additionally, the methods that we explore in this feasibility study could assist in designing more accessible and user-friendly technologies tailored to the sensory needs of individuals with ASD.

Methods

Overview

Here, we describe the data set curation process (Figure 1), preprocessing steps, and a series of analyses on the curated data. We started by analyzing the sentiments and topics within the data set to discover some qualitative insights. We then performed per-tweet and per-user classifications of ASD to understand the linguistic differences between the users in the ASD and control groups.
Figure 1. Pipeline for the creation of the novel Twitter (subsequently rebranded as “X”) autism data set. API: application programming interface; VADER: Valence Aware Dictionary for Sentiment Reasoning.

Data Collection
In recent years, hashtags such as #MeToo, #BlackLivesMatter, and #StopAsianHate have played significant roles in promoting social movements and campaigns, including those aimed at raising awareness about specific societal issues. Within the ASD community, popular hashtags such as #AutismMom and #AutismParent have represented the perspectives of neurotypical parents, significantly influencing research and policies in this domain. However, these advocacy groups often overshadow adults with autism, creating a gap in their representation within decision-making processes. To address this issue, a paradigm shift occurred in the autism rights movement through the hashtag “#ActuallyAutistic” [51,52]. This movement has emphasized understanding the experiences, challenges, and perspectives of individuals on the autism spectrum, redirecting attention toward them rather than solely focusing on caregivers.

Using the hashtag as the criteria for our corpus selection, we extracted Twitter conversations of users self-identifying with ASD to study the differences in their linguistic patterns. Our data extraction involved using `snscrape` (JustAnotherArchivist) [53], a Python-based library allowing social media scraping without requiring personal Twitter application programming interface keys and providing powerful search functionality to help filter tweets based on various conditions, such as date-time, language, or location. We targeted English-language tweets using “#ActuallyAutistic” posted from January 1, 2014, to December 31, 2022. To identify users self-identifying with ASD, we searched for keywords such as “autism,” “autistic,”
or “neurodiverse” within their profile descriptions (bios). Additionally, we considered usernames and tweet contents for users who used these keywords solely in their usernames. Finally, we extracted all the tweets from the timelines of these users to construct the autism data set, which consists of 3,137,952 tweets from 17,323 individuals. Associated metadata such as username, account created, friend count, and date of tweets posted were also extracted and could be used for statistical or network analysis.

To build a tweet classifier for individuals with ASD and their neurotypical peers, we collected a sample of random tweets as part of the control group. To achieve this, we formulated a search query excluding the hashtag, “-#ActuallyAutistic,” using the advanced query searching operators and methods provided by Dr. Igor Brigadir [54]. However, this approach carries the risk of data leakage, whereby users who have not posted any autism-related content may possess autism-related keywords in their profile description or name. To avoid this, we screened users who had any such keywords in their profile description or usernames, or who were also present in the autism data set, and subsequently removed them from the sample. We collected 1000 control tweets per day during the same time period to obtain a total of 3,377,518 tweets across 171,273 individual users.

Data Labeling

To train a supervised machine learning model effectively, labeled data that associate each data point with a respective class are crucial. We automatically labeled the tweets from the autism data set as belonging to the class “autism,” assigned label 1. All other tweets from the control group data set were labeled as belonging to the class “control group,” assigned label 0. However, it is important to clarify that these tweet labels were used temporarily for classification purposes and were not permanently stored in the data set. It is important to note that obtaining ground-truth labels can be a costly and time-consuming process, and the performance of machine learning models is often found to decrease with a decrease in labeled data set size. Weak supervision approaches leverage partially accurate or noisy sources for annotations, which can be more efficient than manual labeling.

Data Preprocessing

Working with raw, unstructured Twitter data is challenging because the conversational text contains too many noisy elements, such as punctuation, abbreviations, emojis, and other stray characters. Thus, before using such data for model training, it is necessary to clean and preprocess the data, which is an essential step for any natural language processing task. We started by removing the usage of any profane language in the tweets, such as cursing or swear words, using a Python library called better-profanity [55], which is designed to flag inappropriate words using string comparison and mask them using special characters (the default setting uses “*”). While profane language can sometimes be highly emotive and help in understanding the sentiments of a text, we chose to censor any such words while classifying the tweets, as such words can be used by any individual and might not help in classification tasks. However, we considered the contribution of profane language through sentiment analysis and observed that the polarity of the sentiments was almost similar when using clean and uncensored tweets.

We then tokenized the text into words; removed any nonalphanumeric characters, hyperlinks, user mentions, and HTML tags; and converted the word tokens into lower case to avoid any confusion and data redundancy. We removed stop words to avoid adding noise and complexity to the features with no meaningful information. To further simplify the input space and normalize the vocabulary, we applied stemming and lemmatization. We also removed any hashtags or a list of keywords related to ASD such as “actuallyautistic,” “autism,” “autistic,” “autismacceptance,” “autismawareness,” “askingautistics,” “askingautistic,” “neurodiversity,” “neurodivergent,” “allautistics,” “adhd,” “mentalhealth,” “asd,” “diagnosis,” “autistics,” “autismspeaks,” and “autismpride,” which could introduce bias and lead to model overfitting.

Sentiment Labeling

We compared the sentiments of tweets posted by individuals with ASD against those from the control group in order to understand the subjective characteristics and emotional polarity around the topic. Initially, we conducted sentiment analysis on the original data set, which contained profanity. Additionally, we wanted to explore how profane language can affect the sentiments of the tweets, and thus we also conducted sentiment analysis on a pseudoclean data set after removing any profane words. Sentiment analysis commonly involves 2 approaches: machine learning and lexical. We used the Valence Aware Dictionary for Sentiment Reasoning (VADER) [56], a lexical approach specifically attuned to sentiments expressed in social media or microblogs like context, to analyze the sentiments of the curated data set. VADER has been explicitly trained on social media data sets (such as social media posts or New York Times editorials) and requires no training data. VADER applies a set of rules and heuristics to the sentiment scores of the individual words to determine the overall sentiment of the sentence and returns a dictionary of negative, neutral, positive, and overall (normalized) sentiment scores for the sentence.

Topic Modeling

The objective of our topic modeling analysis was to investigate whether there exist specific themes and semantic patterns that are frequently discussed in relation to ASD and can offer insights beneficial for clinicians and policymakers. Topic modeling is an unsupervised learning technique used to uncover concealed topics and coherent themes within textual data. We used the Top2Vec [57] algorithm, which offers a dynamic approach to discovering topics within a corpus of text data by making use of the spatial proximity of the words.

Tweet-Level Classification

Our initial focus involved training a model specifically designed to predict ASD based on the content within individual tweets. To build this tweet classifier, we identified unique users from both the ASD and control data sets, allocating an 85:15 split for training and testing purposes. Data splitting by user rather than by tweet avoids data leakage, where a user’s tweets might scatter across both training and testing sets, potentially leading to a better-profanity.
to overfitting by the model due to learning user-specific patterns. The tweets, with no profanity, were preprocessed as defined in the previous section and formed the training and test sets. The categorical labels, representing whether a tweet belonged to a user in the ASD or control group, were used as the basis for model training and evaluation. Additionally, the training data set underwent an 85:15 split, separating it into training and validation subsets, which was used to fine-tune the model and adjust hyperparameters.

For text-to-numeric vectorization, we used 2 approaches: a bag-of-words term frequency–inverse document frequency (TF-IDF) method and word2vec embeddings. We started by training TF-IDF feature representation using various classical machine learning algorithms: support vector machines, naive Bayes, logistic regression, and XGBoost (extreme gradient boosting), using 5-fold cross-validation and accuracy as the primary evaluation metric to identify the best classification method. We then trained the word2vec model using the best-identified algorithm for better feature representation. This approach captures both semantic and syntactic similarities among words, and we assessed its efficacy using a more comprehensive array of evaluation metrics.

User Profile Classification

Our subsequent task involved training a model to predict ASD by considering all tweets from an individual user’s timeline. To ensure a more representative data set and prevent potential model overfitting, we isolated unique users who had shared a minimum of 5 tweets and split them into an 80:20 ratio for training and testing. The preprocessed tweets from each user were then grouped together to form an individual document. For model training, we used an attention-based, bidirectional long short-term memory (Bi-LSTM) model vectorized with a randomly initialized, self-trained embedding layer. As the tweets vary in their lengths and raw text cannot be directly represented as dense vectors in the way that images can, we used padding and an extra “unknown” token during tokenization to achieve the fixed length input and represent any unseen tokens.

Ethical Considerations

While social media data can help with public health analysis by offering a less intrusive and real-time monitoring approach for disease symptomatology and public sentiments, it also poses ethical challenges by exposing the users to harm or the potential leaking of personally identifiable information. First, this study was approved by all ethics-related regulatory bodies at the University of Hawaii. The study has been approved by the University of Hawaii Institutional Review Board (2023-00248) under an expedited review procedure, and the user information was deidentified. We also ran the request through University of Hawaii institutional data governance to approve this study, where it was determined that the study is exempt from further data governance review due to the inherently public nature of the study data. We also took additional measures not required by the Institutional Review Board. Specifically, we encrypted user IDs, reducing the chances of user reidentification. We also anonymized any user mentions or personal information, such as email addresses, contained within the tweets. These steps were aligned with the ethical considerations outlined in various research studies on social media analysis [58-60].

The public nature of such data can often overshadow the participants’ consent, leaving them unaware or unsure of the inclusion of their data in the research. Williams et al [61] observed that 84% of respondents were not at all or only slightly concerned when using the Twitter posts for university research. However, this leaves a considerable portion of the population who remains concerned. In most cases, it is impractical to obtain consent for large-scale social media analytics research, leaving the responsibility to researchers to safeguard participant data.

The purpose of analyzing social media data is not to provide an immediate intervention but rather to uncover patterns to refine accuracy and help clinicians comprehend the needs of the specific population being studied. With these considerations in mind and to promote interdisciplinary research, the fully anonymized data set can be made available to researchers upon request following a set of protocols to ensure ethicality: we require researchers who request the data set to sign a data use agreement that forbids the researchers from sharing our data set with others and to attest that the data will remain confidential. The Data Use Agreement also forbids attempting to reidentify users represented in the data set.

Results

Data Records

The autism subset, collected from 17,323 self-reported individuals with autism, contains 3,137,952 tweets. The control subset, collected from 171,273 users, consists of 3,377,518 tweets. The combined data set contains the following columns: user ID (a unique value assigned to each Twitter account), profile description (a short summary of the account posted by the user), account created (date-time when the account was created), friends count (number of accounts the user follows), followers count (number of accounts the user is being followed by), tweet date (date-time when the tweet was posted), tweet ID (a unique ID assigned to each tweet), tweet text (original tweet), a list of hashtags present in each tweet, number of replies (number of times the tweet has been replied to), number of retweets (number of times the tweet was retweeted), number of likes the tweet got, and source from where the tweet was posted (web, mobile device, or app). While we focused on using tweet text as the primary source of data, other supporting metadata could be used in the future for network analysis or statistical studies.

Exploratory Data Analysis

The data set’s columns for hashtags and locations were found to contain the highest number of missing values during our analysis. While not all tweets are accompanied by hashtags or location details, users possess the liberty to input any desired location on their profiles. Our analysis revealed that a large portion of users either did not provide their actual location or had inconsistencies in their location entries. Among the top 20 location values identified, most were variations of “United Kingdom,” such as “UK,” “London, England,” “England, United Kingdom,” and “South East, England.” However, other entries
were less informative and included phrases such as “Picnic party” and “My parent’s basement.” Due to the majority of the missing data and to safeguard users’ personal information, we opted to exclude the location column from the data set before using it to train any machine learning algorithms.

Further analysis of the yearly distribution of tweets revealed a rising trend in discussions related to ASD across the years. This trend suggests that individuals on the autism spectrum are increasingly embracing social media platforms, potentially opening up numerous employment prospects and serving as an effective channel to educate the public about developmental delays. Additionally, sharing behavioral symptoms through social engagement could be beneficial to others to build better community support. This increased social involvement may hold significance not only in social science [62] but also in human-computer interaction research [63], offering insights to design more inclusive and efficient digital environments.

**Sentiment Analysis**

The VADER sentiments of most of the ASD and control group tweets were found to be positive and neutral, respectively, as shown in Table 1.

This was supported by another interesting observation: tweets from individuals with ASD comprised a higher character count compared to those from the control group (Figure 2). The histograms depicting the word counts in tweets from both groups follow similar distributions but with a substantial difference in their means. This disparity strongly suggests varying linguistic patterns between these 2 groups.

**Table 1.** Distribution of sentiments in the autism spectrum disorder (ASD) and control group data sets.

<table>
<thead>
<tr>
<th>Data set and VADER sentiments</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In original autism tweets (n=3,137,952)</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>1,528,183 (48.7)</td>
</tr>
<tr>
<td>Negative</td>
<td>812,730 (25.9)</td>
</tr>
<tr>
<td>Neutral</td>
<td>797,039 (25.4)</td>
</tr>
<tr>
<td><strong>In clean autism tweets (n=3,137,952)</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>1,562,700 (49.8)</td>
</tr>
<tr>
<td>Negative</td>
<td>756,247 (24.1)</td>
</tr>
<tr>
<td>Neutral</td>
<td>819,005 (26.1)</td>
</tr>
<tr>
<td><strong>In original control group tweets (n=3,377,518)</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>1,280,080 (37.9)</td>
</tr>
<tr>
<td>Negative</td>
<td>938,950 (27.8)</td>
</tr>
<tr>
<td>Neutral</td>
<td>1,158,488 (34.3)</td>
</tr>
<tr>
<td><strong>In clean control group tweets (n=3,377,518)</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>1,323,987 (39.2)</td>
</tr>
<tr>
<td>Negative</td>
<td>719,411 (21.3)</td>
</tr>
<tr>
<td>Neutral</td>
<td>1,334,120 (39.5)</td>
</tr>
</tbody>
</table>

*aVADER: Valence Aware Dictionary for Sentiment Reasoning.*
Topic Modeling

Using just the ASD data set, multiple topics were discovered, and the word clouds of a few topics are shown in Multimedia Appendix 1.

As it can be seen, the majority of topics were related to behavioral and emotional symptoms such as “hyperactivity,” “fidgeting,” “depressed,” “anxiety,” “trembling,” and “overwhelmed.” Interestingly, a considerable number of documents also focused on terms such as “vaccine,” “therapy,” “misdiagnosis,” and “cats.” These findings may be attributed to the frequent misdiagnosis or delayed diagnosis of ASD, prompting individuals to seek therapy, support, and guidance. The presence of vaccine-related discussions likely stems from misinformation and its negative impact on individuals affected by ASD. However, given the time frame in which the data set was collected, it is also possible that these tweets are related to COVID-19 vaccines. Last, multiple studies [64,65] have found that children with autism are more at ease with cats due to their nonintrusive nature, lack of prolonged eye contact, and their ability to alleviate stress and interpret emotional cues.

Deriving specific topics from the control group’s Twitter conversations was challenging given their scattered and diverse nature. Most of these discussions centered around internet personalities, random conversations, specific days of the week, or special occasions such as birthdays and anniversaries. Interestingly, some broader topics related to animals surfaced in these conversations, but not as specifically focused as observed in autistic user conversations—specifically mentioning cats. Some of these posts also displayed the use of emotional words, suggesting that pets or animals may provide therapeutic benefits.

Technical Validation

The performance metrics for tweet classification are shown in Tables 2 and 3. Table 2 displays the results from TF-IDF feature representations across several classical machine learning models, while Table 3 displays the results using word2vec feature vectors trained with logistic regression. While the TF-IDF vectorization yielded similar accuracy using different machine learning algorithms for tweet classification, logistic regression was chosen as the best predictor due to its superior performance and shorter training time. The results of the word2vec model were found to be consistent with the semantic similarities of the words. For instance, “autism” exhibited higher cosine similarity to terms such as “Aspergers,” “neuroatypical,” and “autism spectrum condition,” indicating the model’s proficiency in capturing semantic relationships between words.

Table 4 displays the results for user classification. Although there is a class imbalance in the number of users with ASD versus controls, the attention-based LSTM model still seems to yield better measures, with $F_1$-scores of 0.7 and 0.9 on the “autism” and “control group” classes, respectively, and an AUC score of 0.78.
Table 2. Summary of results obtained for tweet classification from term frequency–inverse document frequency vectorization to identify the best algorithm based on accuracy.

<table>
<thead>
<tr>
<th>Model</th>
<th>Validation set accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support vector machine</td>
<td>0.615</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.598</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.63</td>
</tr>
<tr>
<td>XGBoost(^a)</td>
<td>0.624</td>
</tr>
</tbody>
</table>

\(^a\)XGBoost: extreme gradient boosting.

Table 3. Summary of results obtained for tweet classification from the word2vec model using the highest performing model, logistic regression.

<table>
<thead>
<tr>
<th>Metric performance on test set</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.73</td>
</tr>
<tr>
<td>(F_1)-score</td>
<td>0.71</td>
</tr>
<tr>
<td>AUC(^a) score</td>
<td>0.728</td>
</tr>
</tbody>
</table>

\(^a\)AUC: area under the receiver operating characteristic curve.

Table 4. Summary of results obtained for user classification from Keras embedding using the attention+Bi-LSTM\(^b\) model.

<table>
<thead>
<tr>
<th>Metric performance on test set</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F_1)-score</td>
<td>0.805</td>
</tr>
<tr>
<td>AUC(^b) score</td>
<td>0.78</td>
</tr>
</tbody>
</table>

\(^b\)AUC: area under the receiver operating characteristic curve.

\(^a\)Bi-LSTM: bidirectional long-short-term memory.

Discussion

Overview

The profound shift in society’s reliance on social media for information, in contrast to traditional news sources, along with the immense volume of generated data, has resulted in an increased focus on the use of natural language processing for text analytics. While research tools using facial expressions [6,66-75] and eye gazing for phenotyping ASD [76,77] are consistently reliable, there exists a current deficiency in standardizing precise methods for assessing deficits in social interaction. Therefore, linguistic and behavioral markers extracted from Twitter conversations can serve as valuable resources to investigate textual variations and social dynamics in naturalistic settings. In this study, we demonstrated the potential of leveraging data mining techniques to learn about ASD and related topics from social media platforms such as Twitter. The \(F_1\)-scores of 0.71 in tweet classification and 0.80 in user classification signify substantial semantic distinctions in messages posted by individuals with and without ASD. Tweets by individuals with autism showed a higher frequency of emotional language, corroborated by the word2vec model’s stronger semantic associations among such words, reinforcing the model’s predictive capability. This finding, coupled with previous studies using computer vision models [76,78], suggests that social phenotypical behavior could be used to support effective ASD screening strategies and facilitate early detection. We also want to emphasize that the National Institutes of Health is actively funding research works [79,80] using data from electronic health records, social media, and mobile devices with novel artificial intelligence–based tools to improve public health surveillance and precision diagnostics, keeping in mind ethical and other societal considerations. We would also like to highlight that any social media analytics research should always be supported by ethical considerations and user privacy.

Limitations

There are certain limitations to consider in this study. While we focused on individuals who self-identified as autistic, there is no clinical validation for their diagnosis. Annotations from clinical experts or crowdsourcing can help. Furthermore, there is a possibility of data leakage, where the identified users may not be autistic but instead could be family members, parents, caregivers, or advocacy organizations belonging to a different study population and still using the hashtags. However, the frequency of this type of leakage is predicted to be rare due to the negative social connotations of using #ActuallyAutistic without a diagnosis. There might also be a possibility of some data leakage of an individual with autism falling into the control group, but with ASD having a prevalence rate of <3%, the model performance should not degrade by more than 3% if an individual who chose not to self-identify themselves with ASD falls into the control group cohort. The predictive power of social media is not to be used at an individual level but at a broader cross-sectional level, possibly combined with self-reported questionnaires for enhanced accuracy in neurological studies.
In addition, the sentiment polarity obtained through VADER may lack accuracy compared to human-labeled sentiments, as human sentiments are influenced by various factors such as surroundings and politics, making reliable labeling challenging. Moreover, this study only considered the English language, potentially missing out on information from other countries or languages that could aid the model in making better predictions. This also raises concerns about the lack of diversity in the data, where only English-speaking users from higher socioeconomic groups or younger adults are represented in the data set, as they comprise a larger portion of Twitter users.

**Future Work**

This study presents several opportunities for future research, including using pretrained large language models such as Bidirectional Encoder Representations from Transformers and Generative Pre-trained Transformers for text classification, topic modeling, and feature extraction. Another interesting avenue is the integration of text data with additional data modalities such as audio and video, which could also be mined from social media. In addition, incorporating auxiliary information into textual features may further improve the effectiveness of machine learning models. Lastly, as the Centers for Disease Control and Prevention have reported that boys are 4 times more likely to receive an ASD diagnosis than girls, gender-stratified analysis using crowdsourcing or other metadata analysis techniques may also hold promise for more precise screening practices.

**Acknowledgments**

The technical support and advanced computing resources from University of Hawaii Information Technology Services—Cyberinfrastructure, funded in part by the National Science Foundation Campus Cyberinfrastructure awards #2201428 and #2232862, are gratefully acknowledged. We used the generative artificial intelligence tool ChatGPT by OpenAI only to edit the grammar of the manuscript.

**Authors’ Contributions**

AJ was responsible for data collection, data analysis, and manuscript writing—the original draft. PW conceptualization, supervision, and manuscript reviewing and editing.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Topics observed in autism spectrum disorder data set using the Top2Vec algorithm.

[DOCX File, 221 KB - formative_v8i1e52660_app1.docx]

**References**


10. Mowery D, Bryan C, Conway M. Feature studies to inform the classification of depressive symptoms from Twitter data for population health. arXiv Preprint posted online on January 28 2017. [FREE Full text]


47. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017;6(1):15 [FREE Full text] [doi: 10.1140/epjds/s13688-017-0100-0]


83. Creating safe AGI that benefits all of humanity. OpenAI. URL: https://openai.com/ [accessed 2024-01-12]

Abbreviations

ASD: autism spectrum disorder
AUC: area under the receiver operating characteristic curve
Bi-LSTM: bidirectional long short-term memory
TF-IDF: term frequency–inverse document frequency
VADER: Valence Aware Dictionary for Sentiment Reasoning
XGBoost: extreme gradient boosting

©Aditi Jaiswal, Peter Washington. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Determining the Right Levels of Health Coaching and Heart Rate Variability Biofeedback in a Workplace Behavior Change Intervention: Multiphase Optimization Strategy Preparation Study

Sean Locke, MA, PhD; Jenna Osborne, MSc

Department of Kinesiology, Faculty of Applied Health Sciences, Brock University, St Catharines, ON, Canada

*all authors contributed equally

Corresponding Author:
Sean Locke, MA, PhD
Department of Kinesiology
Faculty of Applied Health Sciences
Brock University
1812 Sir Isaac Brock Way
St Catharines, ON, L2S 3A1
Canada
Phone: 1 905 668 5550 ext 4958
Email: slocke@brocku.ca

Abstract

Background: Work-related stress is associated with poor job performance and negative health outcomes. Changing health behaviors through corporate wellness programs can improve physical and mental health and help employees manage stress. This project sought to pilot the potential addition of brief coaching and biofeedback to an 8-week web-based self-help program to improve employee stress using the multiphase optimization strategy.

Objective: This study aims to determine which candidate components will be tested in a later optimization phase and at what dose they will be tested, examine the feasibility and acceptability of delivering the different components, investigate whether the outcomes can be feasibly measured, and review evidence to build a conceptual model before the optimization phase.

Methods: The study was positioned within the preparation phase of the multiphase optimization strategy. It is a 2×2×2×2 design with 4 components: 2 types of health coaching and 2 types of biofeedback. All components were tested by turning them on or off. A total of 16 adult office workers (mean age 40, SD 14.3 years; n=15 women) completed an 8-week self-paced web-based stress management and health behavior change program and were randomly assigned to 1 of the 16 conditions, created from a combination of the 4 candidate components. Assessments included web analytics, surveys, and interviews regarding program recommendations, likes, and dislikes.

Results: Findings from the interviews provided suggestions to improve the intervention (eg, separating wellness from stress content) and trial conduct (eg, streamlining the onboarding process). On average, participants logged into the wellness program 83 times (range 36-291), with 75% (12/16) participant retention and 67% (8/12) survey completion. There were no reported problems with coaching or obtaining data from interviews or apps. The interview findings suggested potential mediators to include and assess in a future conceptual model.

Conclusions: The results provided areas to improve the intervention content and trial methods. Instead of progressing to the next scheduled large-scale optimization phase, our plan to iterate through a second preparation phase after making changes to the protocol, apps, and corporate coaching partner.

(JMIR Form Res 2024;8:e47181) doi:10.2196/47181

KEYWORDS

mobile health; mHealth; behavior change; stress management; intervention; pilot study; heart rate variability; health coaching; coach; coaching; coaches; work-related stress; stress; wellness; burnout; behavioral intervention; work; worker; workers; employee; employees; occupational health; job; satisfaction; web-based; remote; corporate; web analytics; biofeedback; survey; surveys; interview; interviews; experience; experiences; attitude; attitudes; opinion; opinion; perception; perceptions; perspective; perspectives; acceptance
**Introduction**

**Background**

Work is one of the major contributors to chronic stress. Work-related stress is experienced by most employees and is related to poor job performance and lower job satisfaction [1,2]. Chronic stress, a common outcome of workplace stress, is a critical risk factor for a variety of negative health outcomes [1,3] and is linked to the development of cardiovascular disease and depression [4-6]. Meta-analytic findings from a study conducted among >80,000 employees demonstrated an increase in coronary heart disease risk for those experiencing work stress (risk ratio 1.16) [7]. There are considerable costs to the employer from unmanaged high stress, resulting in performance decreases and absenteeism [8]. These costs can be ameliorated through employer-funded health and wellness programs [9,10], which can also serve as feasible strategies for chronic disease prevention [11].

There are numerous psychological, social, and physiological benefits of engaging in positive behaviors such as physical activity and healthy eating [12-14]. It is well established that regular engagement in positive health behaviors (eg, exercise and diet) is associated with lower stress [15-17]. Despite the potential benefits, only a few individuals meet recommendations, missing out on these benefits [18,19]. Individuals spend as much as half of their waking hours working, and the workplace offers an opportunity to promote health, including stress management. Behavior change interventions target exercise and diet in the workplace can improve health behaviors [20-24]. Evidence from a review of 23 studies indicates that workplace interventions can improve physical and mental health outcomes [25].

Internet-based interventions represent a scalable approach to health promotion. High rates of internet use and the convenience of accessing content 24 hours a day allow intervention to reach participants, whether they are at work or home [26]. Web-based self-help interventions are structured such that the participants can complete these at their own pace and do not require direct contact with an interventionist. A review of 23 web-based self-help interventions reported significant weight loss across studies [27]. Workplace health and wellness programs are often multifaceted [28]; web-based self-help approaches may be particularly useful for delivering program components such as health information or behavior change strategies designed to help individuals learn about new behaviors and plan their behavior change efforts.

Behavior change techniques (BCTs) represent the active ingredients of an intervention [29]. BCTs are used to target the hypothesized mechanism of action to help individuals change their behaviors. For example, goal setting and self-monitoring are among the most common BCTs in physical activity and diet interventions [30]. These BCTs target and seek to improve individuals’ confidence and ability to self-regulate their behavior. These types of BCTs are suitable for internet-based delivery, given the ease of use and accessibility of websites and apps.

Individual health and wellness coaching can be an effective way to promote behavior change for a variety of health behaviors [31-33]. Motivational interviewing is a person-centered coaching approach to enhance clients’ intrinsic motivation for change [34]. Motivational interviewing is a heart- and mindset approach to coaching that aims to embody compassion, acceptance, partnership, and evocation with the client [34]. There are 4 processes and 4 communication skills to support the spirit of motivational interviewing. The 4 processes (ie, engaging, focusing, evoking, and planning) are designed to guide clients from a state of ambivalence toward a state of willingness to make a healthy change. The 4 communication skills (ie, open-ended questions, affirmations, reflections, and summaries) are fundamental to reflective listening and are ways to demonstrate that spirit. Motivational interviewing–based interventions have consistently been shown to be effective across numerous health behaviors (eg, diet, exercise, and oral health) [35]. Even brief motivational interviewing–based interventions can be effective in improving diet and exercise [36]. Brief motivational interviewing coaching may be a scalable approach for providing health coaching in workplace interventions. Furthermore, individual-level coaching represents a common strategy that multifaceted health promotion programs use in the workplace [28]. A recent meta-analysis revealed that physical activity interventions highlighted biofeedback as 1 of the 4 BCTs related to the effective initiation of physical activity [37]. Heart rate variability (HRV) is a reliable means of acutely and indirectly measuring stress through the autonomic nervous system [38,39]. It has also been used as an index of physical and emotional health [40-42]. HRV biofeedback (HRV-BF) is a noninvasive method for delivering HRV data in real time to the desired user [43,44]. HRV-BF has been shown to be effective across different settings for improving HRV and other psychological and physiological outcomes [45,46]. HRV-BF has been shown to be effective in the workplace [47,48] and may help promote self-awareness and allow individuals to better regulate their psychological and physiological states [49]. HRV-BF may be a beneficial tool alongside behavioral intervention components to promote awareness and self-regulation.

We partnered with CoreHealth, a company that supplies corporate health and wellness coaches with an web platform to deliver a variety of wellness tools and interventions, including coaching, to their clients. This project sought to pilot the potential addition of brief coaching and HRV-BF to an 8-week web-based self-help program using the multiphase optimization strategy (MOST [50]).

**MOST: Preparation Phase**

MOST will be used to pilot this study, to optimize, and evaluate the efficacy of the stress management and health behavior change intervention to improve workers’ stress and well-being. MOST is a 3-phase framework that aims to develop interventions that are effective, efficient, economical, and scalable [50]. The 3 phases include preparation (ie, selecting which intervention and at what level components are feasible to examine), optimization (ie, determining the optimized package of components via an optimization randomized controlled trial), and evaluation (ie, evaluating the efficacy of the optimized
package through an evaluation randomized controlled trial). The factorial design of MOST in the optimization phase will allow us to examine the effect of adding intervention components to a standard 8-week behavior change program. This paper reports the first phase of MOST, the preparation phase, using the reporting recommendations by Landoll et al [51], with the following four study objectives:

1. Determine which candidate components will be tested in a later optimization phase and at what dose they will be tested
2. Examine the feasibility and acceptability of delivering the different components
3. Investigate whether the outcomes can be feasibly measured
4. Review evidence to build a conceptual model before the optimization phase.

Methods

Design Overview
This study is registered with ClinicalTrials.gov (NCT05150574). In this phase, the components are first tested with a few participants to identify implementation problems and refine them to ensure acceptability and feasibility. The proposed future optimization phase will use a $2^4 (2\times2\times2\times2)$ factorial experiment to test all combinations of the 4 candidate components by turning them “on” or “off,” resulting in 16 different conditions in total (Table 1).

Table 1. Combination of intervention components comprising the 16 intervention conditions.

<table>
<thead>
<tr>
<th>Number</th>
<th>Componentsa</th>
<th>8-week program</th>
<th>Daily resting HRVb feedback</th>
<th>Momentary HRV feedback</th>
<th>Behavioral initiation coaching</th>
<th>Practice with feedback coaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>On</td>
<td>Off</td>
<td>Off</td>
<td>Off</td>
<td>Off</td>
<td>Off</td>
</tr>
<tr>
<td>2</td>
<td>On</td>
<td>Off</td>
<td>Off</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
</tr>
<tr>
<td>3</td>
<td>On</td>
<td>Off</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>4</td>
<td>On</td>
<td>Off</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>5</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>6</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
<td>Off</td>
<td>Off</td>
</tr>
<tr>
<td>7</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>8</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>On</td>
</tr>
<tr>
<td>9</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>Off</td>
<td>On</td>
<td>On</td>
</tr>
<tr>
<td>10</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>11</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>12</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>13</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>On</td>
</tr>
<tr>
<td>14</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
<td>On</td>
</tr>
<tr>
<td>15</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>16</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>Off</td>
<td>On</td>
</tr>
</tbody>
</table>

aAll participants will have access to the content on CoreHealth, and they will be randomized to receive 1 of the 16 conditions.
bHRV: heart rate variability.

For this study, we recruited 16 participants, 1 per condition. There are four candidate coaching components being delivered, with a fifth untested component that all participants will receive the following:

1. Daily resting HRV biofeedback: the participants will be asked to record regular HRV measurements in the morning before eating or drinking.
2. Momentary HRV biofeedback: throughout the workday, the participants will be asked to take in-the-moment HRV measurements, with the option of performing a mindfulness activity to manage elevated stress.
3. Health coaching and behavior change initiation: the participants will receive one 15-minute wellness coaching session at the start of week 1 of their program.
4. Health coaching and practice with feedback coaching: the participants will receive one 15-minute wellness coaching follow-up session in either week 2 or week 4.
5. Web-based behavior change program (all participants receive): this is an 8-week web-based program that provides the participants with access to 8 weekly behavior change and stress management modules. The platform also contains information on a variety of health behaviors.

Ethical Considerations
This study received ethics approval from the Brock University Research Ethics Board (REB #20-367) before recruitment. Prospective participants were emailed information about the study, and those interested were directed to a link to complete the eligibility questionnaire via Qualtrics (Qualtrics International).
Inc). Eligible participants provided informed consent on Qualtrics before scheduling an onboarding meeting with study staff. The health coach also provided informed consent to be interviewed. All data were electronic, deidentified, and stored on the institutional servers of the corresponding authors. There was no compensation for participating in this study. Generative artificial intelligence was not used in any part of the manuscript writing.

**Participants and Procedure**

A total of 16 participants were recruited for the study. The eligibility criteria included being an adult aged ≥18 years, being able to read and comprehend English, having an Apple iPhone 8 or newer version (iOS 14 or newer) or an Android (operating 7.0 or newer), having a self-reported BMI <40 kg/m², and reporting at least a moderate amount of daily job stress. We recruited potential participants through email blasts and word of mouth from companies primarily comprising office workers. Interested participants were directed to a web link to determine their eligibility and to provide informed consent. The participants were contacted to schedule a 30-minute onboarding phone call to discuss study procedures and help them set up and learn about the study apps. Following the onboarding call, the participants were asked to complete the baseline survey measures and record baseline HRV for 7 days before starting the 8-week study. For the first 3 days, HRV was measured when the participants woke up in the morning, to be averaged as before and after the program to assess change, whereas for the following 4 days, they were to get accustomed to taking different HRV readings. If requested, the participants were sent reminders to record their baseline HRV readings.

At the end of 8 weeks, the participants were interviewed to discuss likes and dislikes and suggested improvements to the study. At the end of the interview, they were asked to complete a poststudy survey and record 3 more poststudy HRV readings. Furthermore, the wellness coach was interviewed at the end of the study. All interviews were conducted via Microsoft Teams (Microsoft Corp). Interviews were recorded, and written transcripts of the interviews were created. In response to ongoing participant feedback, we made a deviation to the study protocol, which received research ethics approval, and conducted a 15-minute interview approximately halfway through the 8-week program. This was sparked by a few participant emails regarding the onboarding process and the early technical issues. We chose to interview the participants about these challenges while the issues were fresh in their mind rather than waiting for another 4 weeks to ask questions at the poststudy interview.

**Candidate Intervention Components**

There are 5 components being delivered. Four of the components will be tested and examined as part of a future factorial experiment: (1) daily resting HRV biofeedback, (2) momentary HRV biofeedback, (3) behavioral initiation coaching, and (4) practice with feedback coaching. The fifth is a constant component with all the participants receiving an 8-week behavior change and stress management program.

**HRV Biofeedback**

Light Heart is a mobile phone app (CoreHealth) that was recently designed and developed by CoreHealth to be used as a supplementary tool in their web-based platform to provide HRV biofeedback (BCT 2.6 biofeedback). Research assessing the concurrent validity of the app in assessing pulse intervals against a gold-standard electrocardiogram (BioAmp FE132, ADInstruments) is currently being reviewed. The results demonstrated a strong positive linear correlation ($r=0.99; P<.001$) between the Light Heart app and the electrocardiogram. The app was designed to measure and provide HRV biofeedback through photoplethysmography to detect changes in blood flow through the skin. All the participants were asked to download the app to measure their HRV. The participants were then asked to connect to this app via a mobile access code created through the CoreHealth platform. The participants could obtain an HRV reading by opening the app, selecting “take reading,” placing their finger over their rear phone camera, and watching for 60 seconds. The participants were given a visual prompt if they were moving their fingers too much to obtain a clear reading.

Because HRV metrics are complex to understand (eg, SD of the normal sinus beats [sdNN; ms]), the participants view a user-friendly HRV index. The proprietary index uses a 0 to 100 scale to graphically represent an individual’s HRV, with higher scores indicating poorer HRV. To assess their baseline HRV, the participants were instructed to find a quiet, distraction-free place to sit and measure their HRV when they first wake up in the morning and before they eat or drink. There are two forms of HRV biofeedback being tested: (1) momentary and (2) daily resting.

Those receiving momentary HRV biofeedback were instructed to use the app to record in-the-moment HRV measurements throughout the day when they experienced elevated stress. Higher HRV scores indicated elevated stress. In these instances, the participants were recommended to take a few minutes to perform a mindfulness or meditation activity, following 1 of the videos provided on the CoreHealth platform. Following the activity, the participants were prompted to take another HRV measurement to determine whether their HRV improved after performing the stress management activities.

Those receiving daily resting HRV biofeedback were asked to regularly obtain HRV measurements each morning, similar to recording their baseline HRV. In this way, a regular morning measurement may be an indicator of improved HRV over time as a result of making healthy changes known to impact HRV (eg, improved cardiorespiratory fitness, sleep quality, and mindfulness).

**Health Coaching: Behavioral Initiation and Practice With Feedback (BCT 3.1, Social Support-Unspecified)**

All coaching for this study was conducted through CoreHealth’s videoconferencing tool or via telephone. The wellness coach was trained in a communication style called motivational interviewing. Motivational interviewing is a person-centered collaborative counseling style that helps clients strengthen their autonomous motivation and commitment to change [34].
Coaching at 2 different points in the intervention was tested: behavioral initiation coaching and practice with feedback coaching. In this study, those who received coaching components received motivational interviewing–based health and wellness coaching. Consistent with the spirit of motivational interviewing and standard practice for many of CoreHealth’s clients, the participants were given the autonomy to choose the behavior they wished to change. Those receiving behavioral initiation coaching received one 15-minute behavioral coaching session to start week 1 of their program. In this session, the participants established their commitment to change, decided on a health behavior change goal, and were asked to create an action plan and schedule and track their progress using CoreHealth. Those receiving practice with feedback coaching received a follow-up 15-minute session. In this session, clients discussed their level of success in achieving their goals, revised their goals as necessary, and discussed potential barriers to achieving their goals. For this preparation phase, we investigated the 2 different time points to deliver the follow-up coaching at either 2 or 4 weeks after the initiation of counseling to determine which was more acceptable to the coaches to facilitate behavior change.

**Web-Based Behavior Change Program**

The CoreHealth wellness platform was used to deliver the 8-week web-based behavior change program. In consultation with CoreHealth, BCTs used in the current 8-week period consisted of prompts (eg, email reminders to check-in; BCT 7.1 [29]); goal setting, action planning, and self-monitoring tasks (eg, prompting to create a goal and then providing a calendar to schedule and track participants’ behavior; BCT 1.1, 1.3, 1.4, and 2.3); instruction on how to perform a behavior (eg, videos on how to meditate and examples of different types of exercise; BCT 4.1 and 6.1); reframing (eg, suggesting new perspectives to view barriers; BCT 13.2); providing health-related information (eg, text, video, infographics, or others such as recipes and guidelines; BCT 5.1, 5.3, and 5.6); and confidence ruler (motivational interviewing technique 11 [52]). Refer to Table 2 for a breakdown of the intervention content by week. Furthermore, the participants had access to modular health information and videos on the platform. Health information covered a variety of health and wellness behaviors. A total of 5 wellness areas were included on the study platform in consultation with CoreHealth, chosen based on their association to HRV (physical activity, sleep, meditation and mindfulness, diet, and stress management [53]). The intervention content was based on existing CoreHealth behavioral and stress management programming. The participants were asked to download the CoreHealth app, which provided the same features as the web platform but in an app form.
<table>
<thead>
<tr>
<th>Week</th>
<th>Behavior change content</th>
<th>Stress management content</th>
</tr>
</thead>
</table>
| 1    | • Set SMART® long-term goal  
       • Prompt reflection of values underlying the behavior change  
       • Set a 1-week action plan for health and wellness behavior  
       • Confidence ruler                                                                                                                                                                                                                                                                      | • None                                                                                                                                                                                                                                      |
| 2    | • Review previous action plan and adjust or set a new 1-week plan  
       • Overcoming barriers  
       • Confidence ruler                                                                                                                                                                                                                                                                                                                                 | • None                                                                                                                                                                                                                                      |
| 3    | • Review previous action plan and adjust or set a new 1-week plan  
       • Set a stress management goal  
       • Confidence ruler                                                                                                                                                                                                                                                                                                                                 | • Providing information for defining stress  
       • Introduction to platform stress resources  
       • Introduction to mindfulness meditation                                                                                                                                                                                                                                                     |
| 4    | • Review previous action plan and adjust or set a new 1-week plan  
       • Confidence ruler                                                                                                                                                                                                                                                                                                                                 | • Reframing your stress response  
       • Being kind to yourself in stressful situations                                                                                                                                                                                                                                                |
| 5    | • Review previous action plan and adjust or set a new 1-week plan  
       • Confidence ruler                                                                                                                                                                                                                                                                                                                                 | • Cognitive defenses  
       • Disputing negative self-talk  
       • Keeping stress in perspective  
       • Shifting focus forward  
       • Avoiding blame game  
       • Being grateful                                                                                                                                                                                                                                                                            |
| 6    | • Review previous action plan and adjust or set a new 1-week plan  
       • Confidence ruler                                                                                                                                                                                                                                                                                                                                 | • Ways to prevent stress  
       • Identifying and eliminating sources of stress  
       • Replacing negative coping mechanisms with positive ones  
       • Placing stress-relieving habits into easily accessible forms  
       • Boosting stress immunity with regular physical and mental exercise  
       • Making a little plan (1-week plan)  
       • Making a big plan (long-term goal)                                                                                                                                                                                                                                                        |
| 7    | • Review previous action plan and adjust or set a new 1-week plan  
       • Confidence ruler                                                                                                                                                                                                                                                                                                                                 | • Steps to avoid unnecessary stress  
       • Planning the day  
       • Organizing the surrounding and tasks  
       • Building time management skill  
       • Problem-solving and strategizing to manage daily life stress                                                                                                                                                                                                                                  |
| 8    | • Reflect on the past content and set a plan to maintain positive health habits  
       • Review the previous action plan and adjust or set a new 1-week plan  
       • Confidence ruler                                                                                                                                                                                                                                                                                                                                 | • Reflect on the past content and set a plan to maintain positive stress management habits                                                                                                                                                     |

aSMART: specific, measurable, attainable, realistic, and timeframe.

**Measures**

Refer to Tables 3 and 4 for a list of the study measures and participant interview outline.
Table 3. List of the study measures and time of assessment.

<table>
<thead>
<tr>
<th>Measure name</th>
<th>Study measures</th>
<th>Number of items</th>
<th>Example item</th>
<th>Time points (before and after program or after program only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital Anxiety and Depression Scale [54]</td>
<td>Self-measured instances of anxiety and depression</td>
<td>14 items and scale range of 1 (not at all) to 4 (yes, all the time)</td>
<td>“I feel tense or wound.”</td>
<td>Before and after</td>
</tr>
<tr>
<td>Job Stress Scale [55]</td>
<td>Describes possible feelings and emotions felt during time spent at work or because of time spent at work</td>
<td>5 items and scale range of 1 (strongly disagree) to 5 (strongly agree)</td>
<td>“A lot of time my job makes me very frustrated or angry.”</td>
<td>Before and after</td>
</tr>
<tr>
<td>Psychological Well-Being Scale [56]</td>
<td>Assesses participant’s current level of well-being</td>
<td>9 items and scale range of 1 (strongly disagree) to 7 (strongly agree)</td>
<td>“I judge myself by what I think is important, not by the values of what others think is important.”</td>
<td>Before and after</td>
</tr>
<tr>
<td>Self-rated health scale [57,58]</td>
<td>Self-reported health scale</td>
<td>1 item and scale range of 1 (excellent) to 5 (poor)</td>
<td>“In general, over the past four weeks would you say your overall health is”</td>
<td>Before and after</td>
</tr>
<tr>
<td>Self-regulatory efficacy for health behavior goals [59,60]</td>
<td>Rating confidence levels to make health behavior change</td>
<td>8 items and scale range of 0% (not confident at all) to 100% (completely confident)</td>
<td>“How confident are you that you will develop solutions to cope with unexpected barriers that can interfere with achieving your health goals?”</td>
<td>Before and after</td>
</tr>
<tr>
<td>Wellness Behaviors Inventory [61]</td>
<td>Self-reported questions on how regularly participants engage in different health and wellness behaviors</td>
<td>17 items and scale range of 1 (&lt;1 time a week or none) to 5 (every day of the week)</td>
<td>“I walk as much as possible, ie, taking stairs instead of elevator etc.”</td>
<td>Before and after</td>
</tr>
<tr>
<td>Motivation Scale</td>
<td>The extent to which the participant believes the program received motivated change to their stressor health behaviors</td>
<td>5 items and scale range of 1 (not at all motivating) to 5 (extremely motivating)</td>
<td>“To what extent did receiving biofeedback help motivate your healthy changes? (if applicable)”</td>
<td>After only</td>
</tr>
<tr>
<td>Acceptability measures [62]</td>
<td>Acceptability and feasibility of intervention measures</td>
<td>8 items, including 4 for acceptability and 4 for feasibility, and scale range of 1 (completely disagree) to 5 (completely agree)</td>
<td>“This program seems easy to use.”</td>
<td>After only</td>
</tr>
</tbody>
</table>

https://formative.jmir.org/2024/1/e47181  JMIR Form Res 2024 | vol. 8 | e47181 | p.924 (page number not for citation purposes)
Table 4. Semistructured interview outline.

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Interview questions</th>
<th>Time points (before and after program or after program only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>Onboarding process</td>
<td>After only</td>
</tr>
<tr>
<td></td>
<td>How did you find the process of getting enrolled and starting the study?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What was helpful about getting started in the study?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What was challenging about getting started?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What do you think would improve the process of starting the study?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How did you find the process of downloading the apps and logging in?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What do you think would improve the process of downloading and logging into the apps?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CoreHealth platform and content</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What did you like about the platform?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What did you not like about the platform?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Which modules did you use on the CoreHealth platform?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What did you think of the information that you were given?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>To what extent did the content help you make changes to your health or stress management?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Was there anything you would change about the platform? How can we improve it?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Light Heart Content (if applicable)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What did you like about Light Heart?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What did you dislike about Light Heart?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Did you encounter any barriers to measuring your HRV(^a)? Please describe them.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Was there anything you would change about the app? How can we improve it?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Health coaching (if applicable)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tell me about your experience working with your coach? What did you like and dislike about it?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Did the coaching help you make healthy changes? If so, how?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What could be improved?</td>
<td></td>
</tr>
<tr>
<td>Health coach</td>
<td>Tell us about your experience delivering health coaching to our study participants?</td>
<td>After only</td>
</tr>
<tr>
<td></td>
<td>What did you like about the format of the study’s health coaching?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What did you dislike about the format of the study’s health coaching?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What did you think about the different conditions of the research study?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tell us what do you think needs to change about the study format to help participants elicit more health positive change?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Do you have any general feedback about coaching and study format?</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)HRV: heart rate variability.

Analysis

Descriptive statistics were computed using SPSS software (IBM Corp). The 6-phase method proposed by Braun and Clarke [63] was used to inductively analyze the interview data. Interview data were transcribed verbatim. A total of 2 raters read through the responses to become familiar with the data; main codes and subcodes were identified, using reflexive thematic analysis and then exemplar quotations for each code were selected.

Results

Sample Characteristics

A total of 16 adults consented to participate and were randomized to each of the 16 study conditions. The sample consisted of 15 women and 1 man (mean age 40.9, SD 14.3 years). All the participants worked full-time and had a mean BMI of 24.03 (SD 4.75) and mean job stress of 2.66 (SD 0.5). Regarding race, of the 16 participants, 15 (94%) identified as being White and 1 (6%) as being Southeast Asian. Of the 16 participants, 15 (94%) were married or living as married and 1 (6%) did not respond. Of the 16 participants, 11 (69%) reported having a bachelor degree or college diploma, 3 (19%) reported having postgraduate degrees, and 2 (12%) did not respond.

Participant Retention

A total of 4 (25%) of the 16 participants dropped out of the study before completing the 8-week program. Reasons for study dropout included pregnancy (1/16, 6%), life stress unrelated to the study (2/16, 12%), and no stated reason (1/16, 6%). Of the 12 participants who completed the program, 8 (75%) completed the poststudy follow-up assessment. Regarding item-level missing data from complete surveys, there was 8.3% missing survey data in the prestudy measures and 15.6% missing survey data in the poststudy measures.

Participant Engagement and Treatment Fidelity

A total of 12 (75%) out of the 16 participants completed the entire 8-week program, 1 (6%) completed 7 weeks, and 3 (19%) completed ≤3 weeks. Participants logged into the web-based CoreHealth platform an average of 83 times during the study (SD 63.8; range 36-291 log-ins). The CoreHealth program log-ins by group were as follows: daily resting HRV (mean 110.1), momentary HRV feedback (mean 70.7), initiation
coaching (mean 64.4), and feedback coaching (mean 90.6). Among the 12 participants initially randomized to a condition that required the use of Light Heart throughout the program, the average number of HRV readings during the 8-week program was 12.55 (range 0-39), suggesting that the participants recorded an HRV reading with an average of 1.80 readings per week. Of the 4 (25%) participants not randomized to use the Light Heart app during their 8-week program duration, 3 (75%) did not use Light Heart and 1 (25%) used it once during the program. All the participants received the coaching they were randomized to. The average acceptability score was 3.9 (SD 0.66), and the average feasibility score was 3.78 (SD 0.91).

**Trial Conduct Interview Findings**

Although more participants found the trial onboarding process helpful, there were few suggestions for improving the trial methods (Table 5). The most frequently reported code within the trial conduct was the usefulness of the onboarding process (9/16, 56%). However, 4 of the participants reported that the onboarding process could be improved. For example, providing videos in addition to showing how to use the apps in the onboarding call would help participants proceed with the trial when they went on their own. In total, 2 of the participants found that the program length of 8 weeks was sufficient, whereas 4 of the participants mentioned that a shorter program would better fit their behavior change efforts. Informal feedback from CoreHealth staff suggested that it might be difficult to manage the backend for 16 conditions and suggested fewer conditions moving forward.

**Table 5.** Participant interview findings related to trial conduct.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Illustrative quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likes, 2 codes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onboarding process, 9</td>
<td>Perceived like or dislike of the study’s enrollment process</td>
<td>“Using the Light Heart app was challenging but it was helpful to have the onboarding call and having someone to reach out to.” [HRT101]4</td>
</tr>
<tr>
<td>Program length, 2</td>
<td>Perceived like or dislike of the 8-week program duration</td>
<td>“But I think 8 weeks, that’s a good time to actually make changes and then assess those changes.” [HRT101]</td>
</tr>
<tr>
<td>Dislikes, 2 codes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onboarding process, 4</td>
<td>Perceived like or dislike of the study’s enrollment process</td>
<td>“I found it a little bit difficult is that some of the instruction were in a PDF and others were in that email. So you know flip flipping from different pieces of information. And I was like if it was like you know all into the same document so that I know one was for my code.” [HRT108]</td>
</tr>
<tr>
<td>Program length, 4</td>
<td>Perceived like or dislike of the 8-week program duration</td>
<td>“Personally, I think a four to six weeks. I understand 8 weeks kind of helps you get in the habit of doing it, but after six weeks I’ve almost kind of forgot about it.” [HRT111]</td>
</tr>
</tbody>
</table>

4HRT denotes participant number.

**CoreHealth Interview Findings**

There were 8 positive and 4 negative codes associated with the use of the CoreHealth platform (Table 6). A total of 6 different BCTs were coded 17 times as being positive for the program, which included prompts, self-monitoring and planning, instruction on how to perform a behavior, reframing, health-related information, and a confidence ruler. Participants noted that they enjoyed the platform features such as a weekly outlook of the self-paced modules. They also liked the health-related information on mindfulness and stress management, particularly the 10-minute videos (although 3 suggested having shorter video options available). However, the action planning feature of the weekly program drew critical feedback (n=10). Although some did not find any benefit in setting an action plan, others found that it required too much writing and that completing the action plan became repetitious as there was no function to copy the plan details from the previous week. The overall usability of the platform interface received both positive n=5 and negative reactions n=11. Most usability issues surrounded updating the user interface, including streamlining the action planning function. Additional instructions for navigating the platform were suggested 3 times for improving the user experience (eg, how-to-use videos).
Table 6. Participant interview findings related to the 8-week program and the Core Health platform.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Illustrative quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likes, 8 codes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCTs*: 17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subcodes:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence ruler: 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prompts: 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-monitoring: 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions: 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reframing: 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health-related information: 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module length: 2</td>
<td>Perceived like or dislike of the amount of time to complete the action plan</td>
<td>“I thought it was good, I thought it was concise and you know I don’t like things that go on too long and that are too wordy.” [HRT111]</td>
</tr>
<tr>
<td>Usability: 5</td>
<td>Perceptions of the capacity of the system to perform the tasks safely, effectively, or efficiently while enjoying the experience</td>
<td>“I liked the content, it was user friendly like on mobile devices.” [HRT101]</td>
</tr>
<tr>
<td>Dislikes, 4 codes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of stress manage-</td>
<td>Perceived like or dislike of the video duration</td>
<td>“The only issue I had with them. I think there were a little too long so if you could just do like a you know instead of doing the full 10 minutes let me just do it for three or four minutes.” [HRT106]</td>
</tr>
<tr>
<td>ment videos: 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCT Action Planning: 10</td>
<td>Action planning: participants found that weekly action planning required too much writing and commitment</td>
<td>“I found sometimes I end up saying a lot of the same things or like it was something that was out of my control and I couldn’t change....” [HRT107]</td>
</tr>
<tr>
<td>Usability: 11</td>
<td>Perceptions of the capacity of the system to perform tasks safely, effectively, or efficiently while enjoying the experience</td>
<td>“The I guess it’s not the cleanest looking app, but it was functional...” [HRT106]</td>
</tr>
<tr>
<td>Quality of instructions: 3</td>
<td>Perceived quality or clarity of the instructions</td>
<td>“Step by step document or personalized calendar would be nice so that participants can see what I have to do on each day would be helpful” [HRT129]</td>
</tr>
</tbody>
</table>

*BCT: behavior change technique.

**Light Heart Interview Findings**

There were 8 possible codes regarding Light Heart from the interviews (2 likes and 6 dislikes; Table 7). There were both positive n=6 and negative n=8 feedback about the usability of the app. Participants across both the daily resting and momentary HRV components generally liked the look and feel of the app. Feedback about technical issues did not differ across participants receiving either HRV component as the issues were not component specific. Technical issues of poor signal quality, failed access codes early in the program, and elevated light temperature on some Android phones muted enthusiasm for the app in a few participants. Some participants noted that additional instructions about using and troubleshooting the app would enhance app usability. Suggestions included providing more detailed in-app explanations for failed readings owing to poor

https://formative.jmir.org/2024/1/e47181
signal quality and providing more demo videos on how to use the app (e.g., finger placement based on the phone and number of phone cameras).

The participants generally enjoyed recording their HRV readings to obtain biofeedback (11/16, 69%), which primed them to think about managing their stress or health and wellness goals. One participant remarked about obtaining their HRV as follows:

There was a day I felt pretty stressed out so I tested using the HRV and then I did a breathing exercise. I tested again and it had improved. It’s validating to see those numbers. It makes you feel better to measure again and see that it worked. [HRT104]

One participant who was randomized to using both types of HRV-BF remarked, “the best usage for HRV is really [pre-post] biofeedback,” suggesting a preference for momentary HRV (HRT108). All participants had access to both types of readings: pre- and postfeedback readings used in the momentary HRV condition and daily resting readings. This confused participants, regardless of the condition, who were uncertain about which type of reading they should record and when.

Table 7. Participant interview findings related to obtaining heart rate variability (HRV) biofeedback.

<table>
<thead>
<tr>
<th>Interview findings and code</th>
<th>Definition</th>
<th>Example quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Likes, 2 codes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Usability: 6</td>
<td>Perceptions of the capacity of the system to perform tasks safely, effectively, or efficiently while enjoying the experience</td>
<td>“It’s nice not having to keep up with different Instruments.” [HRT101]</td>
</tr>
<tr>
<td>• Biofeedback: 11</td>
<td>Biofeedback: provide beneficial feedback about the body (e.g., physiological or biochemical state) using an external monitoring device as part of a behavior change strategy</td>
<td>“The best usage for HRV is really biofeedback; part of managing stress and finding ways to calm down.” [HRT108]</td>
</tr>
<tr>
<td><strong>Dislikes, 6 codes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Reading length: 2</td>
<td>Perceived like or dislike of the amount of time required to take an HRV reading</td>
<td>“If the readings didn’t take a whole minute” in response to how they would use Light Heart differently [HRT104]</td>
</tr>
<tr>
<td>• Quality of instructions: 5</td>
<td>Perceived quality or clarity of the instructions</td>
<td>“I don’t think there was enough instructions in the app to use it properly. Wasn’t sure how to stop the readings from quitting so maybe a demo video and a little trigger.” [HRT113]</td>
</tr>
<tr>
<td>• Technical problem: 21</td>
<td>Signal quality: capability of the HRV measure to obtain a successful reading</td>
<td>Signal quality: “It was it was difficult to get a reading, and many mornings I would have to try multiple times” [HRT104]</td>
</tr>
<tr>
<td>• Subcodes:</td>
<td></td>
<td>Access codes: numeric code used to link the study’s mobile apps to the CoreHealth platform</td>
</tr>
<tr>
<td>• Signal quality: 10</td>
<td>Access codes: “It was difficult to connect the apps” [HRT101]</td>
<td></td>
</tr>
<tr>
<td>• Access code: 8</td>
<td>Phone temperature: perceived temperature of the flash while taking an HRV reading</td>
<td>Phone temperature: “Phone gets really hot and they find it difficult to get a reading: kind of frustrating using the HRV” [HRT104]</td>
</tr>
<tr>
<td>• Temperature: 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Usability: 8</td>
<td>Perceptions of the capacity of the system to perform tasks safely, effectively, or efficiently while enjoying the experience</td>
<td>“I think it’s complicated with having to choose between baseline and feedback, I would like it to be simplified.” [HRT110]</td>
</tr>
</tbody>
</table>

Coaching Interview Findings

Participant feedback on both the initiation and feedback coaching components were similar. The participants overwhelmingly liked the coaching style n=11 and reported benefiting from the coach using different behavior change strategies (i.e., n=17: goal setting, social support, and providing information; Table 8). One participant felt that the supportive coach helped with their ability to cope with stress, whereas another felt that the added accountability helped keep them on track with their goals. There were no negative responses regarding the BCTs used by the counselors. However, 1 participant did not like the motivational interviewing–based counseling style, suggesting that they wanted the coach to be more directive in providing exercise during the session. A total of 4 respondents suggested that 15-minute sessions were not long enough for coaching to facilitate health behavior changes. Coaching was conducted web-based through the CoreHealth platform, and there were 4 mentions of difficulty in knowing where and how to access videoconferencing.
Table 8. Participant interview findings related to coaching conditions.

<table>
<thead>
<tr>
<th>Interview findings and code</th>
<th>Definition</th>
<th>Illustrative quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Likes, 5 codes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• BCT(^a): 17</td>
<td>Goal setting: setting or agreeing on a goal defined in terms of the behavior to be achieved</td>
<td>Goal setting: “Setting up, setting up the objective, making sure you’re on the right track, accountability.” [HRT108]</td>
</tr>
<tr>
<td>• Subcodes:</td>
<td>Unspecified social support: providing resources (eg, psychological) intended to benefit an individual’s ability to cope with stress [64]</td>
<td>Social support: “She applied all the great practices right, the not, no judgment, supporting... Uh, so look, you know a potential solutions and make you know, just raising ideas, bringing motivation” [HRT108]</td>
</tr>
<tr>
<td>• Goal setting: 5</td>
<td>Health-related information: providing health-related information to the participant</td>
<td>Health-related information: “She understood activity and exercise and all of these things that we talked about it. I mean it was just being familiar with the content on her part was really good.” [HRT106]</td>
</tr>
<tr>
<td>• Social support: 11</td>
<td>Perceived like or dislike of the coaching style</td>
<td>“The thing I liked was it was really kind of nice to talk to a live friendly voice. She was very friendly, very positive, very upbeat.” [HRT111]</td>
</tr>
<tr>
<td>• Health-related information: 1</td>
<td>Perceived like or dislike of the coaching style</td>
<td><strong>Dislikes, 3 codes</strong></td>
</tr>
<tr>
<td>• Counseling length: 4</td>
<td>Perceived like or dislike of the counseling session duration</td>
<td>“[I] dislike the length of it. If you’re going to put the effort in, probably to have someone help you, I think 15 minutes is personally too short.” [HRT115]</td>
</tr>
<tr>
<td>• Technical problems: 4</td>
<td>Perceived challenges associated with using the CoreHealth platform for the web-based coaching session</td>
<td>“When I was trying to log in the second time, I’m like how did I do it the first time and I just like totally forgot... Uh, so we got my gosh I’m going to be late.” [HRT107]</td>
</tr>
<tr>
<td>• Coaching style: 2</td>
<td>Perceived like or dislike of the coaching style</td>
<td>“So it was just her agreeing with me which I don’t know if you like or do not like I think I was expecting it to act like her don’t like a strategy that I had heard over like something that I didn’t think of...” [HRT115]</td>
</tr>
</tbody>
</table>

\(^a\)BCT: behavior change technique.

**Interview With the Wellness Coach**

Overall, the coach was satisfied with the coaching and found that using the spirit of motivational interviewing was beneficial to bridge the gap and allowed the participants to connect to them as a coach. The coach stated that “Motivational interviewing is definitely one that I use a lot.” Although 15 minutes is the standard length of counseling for most low-risk clients, they mentioned that “15 minutes is a tight time frame to be able to really utilize a lot of motivational interviewing tools.” They also suggested that stress and health behavior change content should be separated and introduced at different time points. The coach mentioned that the study’s setup and participants receiving different conditions was a bit confusing. They mentioned the following:

> The communication was probably the hardest part is being able to get in touch with the participants and make sure that they’re on track...I’m used to being that person that goes and holds their hand and helps them along the entire way.

This seemed particularly the case for those receiving feedback counseling but not initiation counseling, where the first communication between the coach and participant occurred part-way through the program. The coach also echoed the challenges that were heard from participants in connecting Light Heart to CoreHealth to view their HRV-BF in CoreHealth as follows:

> I think that’s the main thing is making sure people are utilizing it [Light Heart] properly. And then because, I mean, when people did use it, I received the information in CoreHealth. So, if the information is there, it’s really easy to be able to coach on. It’s just making sure that they see their end.

**Discussion**

**Principal Findings**

We conducted this study to assess participants’ perceptions of different levels of the intervention components to be tested in a future optimization stage. There were many positive perceptions about the intervention components; however, interviews with participants and the coach revealed areas of strength and improvement.

In general, participants expressed more positive than negative reactions toward the onboarding process, consistent with the quantitative acceptability scores. Although some participants found it easy to learn the CoreHealth site and study apps, others suggested that additional resources were needed to help individuals engage with the technology as they started the program (eg, instructional videos). There were also mixed
responses regarding the length of the 8-week program. Although some felt that it was right, others found it too long. Previous literature has suggested that sequentially initiating changing to multiple behaviors (ie, 4 weeks of health behavior change followed by 4 weeks of stress management) may be more effective than having participants concurrently initiate changes in multiple behaviors at the same time [65]. This, combined with the feedback from participants and the coach to separate the stress management content from the health behavior change content, will result in a disaggregation of the content into 2 separate 4-week modules. Notably, there were no differences between delivering feedback coaching at either 2 or 4 weeks. Moving forward, coaching will be delivered at the 4-week time point to align with the updated structure of delivering two 4-week modules.

In reporting what they found helpful about the 8-week program, participants frequently mentioned 6 different BCTs. These responses are an indication of treatment receipt fidelity—that participants received the BCTs on the CoreHealth platform as intended [66]. One exception was the action planning content, which participants found cumbersome to use. The quality of how interventions were implemented using technology could impact the amount of time it took to complete a task [67]. Asking participants to set weekly action plans that were time-consuming may have negatively affected their experience. We plan to work with developers to simplify the action planning script to improve the user experience because health information technology systems that are easy to navigate are more likely to be used [68,69]. We also plan to work with developers to improve the process of connecting the Light Heart and CoreHealth apps to the CoreHealth Web platform through their mobile access codes.

Participants enjoyed the look and feel of the Light Heart app and reported important benefits in obtaining HRV. There were 3 noted areas for improvement related to our integration of the app within this study. First, incorporating the HRV function from Light Heart into the broader CoreHealth app would enhance usability and ease the burden of having to switch between using 2 different apps. Combining the 2 apps would also alleviate the second challenge, which is the difficulty of connecting the apps through mobile phone codes. Third, providing additional instructional resources may help alleviate other challenges that participants experienced when first learning to use the app (eg, finger placement). We believe that making these adjustments will improve the usability and acceptability of the study’s apps.

Despite these challenges, participants reported the benefits of using Light Heart to obtain HRV-BF. Biofeedback has been shown to reduce stress and improve well-being [70]. Consistent with these findings, 1 participant in the momentary HRV-BF group explicitly noted that the biofeedback improved their mindfulness and physical feeling. Biofeedback is a form of self-monitoring. Regular self-monitoring can prime individuals to think about their behavior change goals, particularly for those who do not regularly self-monitor [71]. One participant in the daily resting HRV-BF group suggested that taking regular readings served as a regular reminder of their behavior change goals. These self-reported benefits from the 2 HRV-BF conditions provide an anecdotal indication that the participants received their HRV-BF condition and obtained the hypothesized effects as intended.

One of the key activities of the preparation phase is to pilot the candidate components and gather evidence to build a conceptual model before the optimization phase [50,51]. The initial intervention components and levels were determined in consultation with CoreHealth staff based on their affordability and scalability because they align with their current standard practices. Participants’ feedback on the candidate components provided insight into potential mediators to include and assess in a future conceptual model. For example, feedback about HRV-BF did not substantially differ between the daily resting and momentary components, and the findings pointed to increased awareness of current HRV as a potential mediator. The participants suggested that using different BCTs during coaching helped them track and achieve their wellness goals. This may suggest self-regulatory efficacy in managing one’s health behaviors as a potential mediator of coaching components. We plan to review the candidate components and mediators to develop a conceptual model before the next phase of this project.

**Strengths**

A notable strength of the study was using a multimethod approach used in this preparation phase and adhering to the MOST framework. For example, obtaining user feedback answers calls to include participants in developing and refining health technology [68]. MOST is a progressive framework for the development and optimization of behavioral treatments. Framing this study using MOST provided clear benchmarks for this study’s objectives and a clear path to progress to the next stage of research. Research can be more impactful and have better patient outcomes when key stakeholders are meaningfully engaged [72,73]. One final strength was the partnered approach used to execute this study. Throughout this project, we had a close working relationship with the CoreHealth staff, which allowed for the quick resolution of technological challenges as they arose. For example, we were able to find a quick solution to the challenge of mobile phone codes not working properly at the beginning of the study. Partnered research also increases the likelihood that the findings will be put into practice [74]. The findings from this study will result in changes that are implemented by our research partner. Two such examples include adjusting the action plan widget and adjusting how users navigate between the 2 types of HRV-BF readings in the Light Heart app.

**Limitations**

One possible limitation of conducting research with multiple partners was the continued reliance on all partners to perform their roles to ensure the ongoing success of a project. We had exceptional buy-ins from both CoreHealth and our corporate coaching partner throughout this study. However, owing to the economic impact of the COVID-19 pandemic, our coaching partner is no longer able to deliver health coaching. This possibility was anticipated at the onset of our research with CoreHealth, which indicated that they work with other health coaching companies that can be approached to run the next
phase of the study. Another limitation was that the interview questions may not have been sensitive enough to distinguish between similar HRV-BFs or coaching components. One aim of preparation studies is to gain insight into the delivery of candidate components and different conditions. Future preparation phase research should seek more nuanced feedback about the receptivity of the different component combinations in the individual conditions. Another limitation of this study was the dropout rate, which was higher than anticipated, yielding fewer participant responses to our poststudy assessments. These adherence rates are consistent with other lifestyle intervention trials reporting up to 70% attrition, depending on the length of the follow-up [75-77]. Multiple participants received each of the 4 candidate components; however, 4 (25%) of the 16 conditions did not receive participant feedback and went unpiloted because of dropout.

MOST is a structured framework for developing efficient behavior change interventions. This preparation study will result in improved onboarding and app usability. Given that there were unpiloted conditions, proposed changes based on user feedback, and changes in corporate wellness providers, we plan to reiterate through the preparation phase before proceeding to the optimization phase. Strong partnerships with a commitment to make user-centered adjustments will allow us to keep progressing through the different phases of MOST.

Data Availability

Deidentified data may be made available by reasonable request to the corresponding author.

Conflicts of Interest

None declared.

References


37. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review


Abbreviations

BCT: behavior change technique
HRV: heart rate variability
HRV-BF: heart rate variability biofeedback
MOST: multiphase optimization strategy
sdNN: SD of the normal sinus beats
Determining the Right Levels of Health Coaching and Heart Rate Variability Biofeedback in a Workplace Behavior Change Intervention: Multiphase Optimization Strategy Preparation Study

Please cite as:
Locke S, Osborne J
Determining the Right Levels of Health Coaching and Heart Rate Variability Biofeedback in a Workplace Behavior Change Intervention: Multiphase Optimization Strategy Preparation Study
JMIR Form Res 2024;8:e47181
URL: https://formative.jmir.org/2024/1/e47181
doi:10.2196/47181
PMID:38354036

©Sean Locke, Jenna Osborne. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Measurement of Head Circumference Using a Smartphone: Feasibility Cohort Study

Stefan Yordanov¹*, MD, MRCS; Kalsoom Akhter¹, CPsychol; Jye Quan Teh¹*, BCHIR, MA; Jawad Naushahi¹, LLM, PhD; Ibrahim Jalloh¹*, MA, PhD

Academic Division of Neurosurgery, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, University of Cambridge, Cambridge, United Kingdom

*these authors contributed equally

Corresponding Author:
Stefan Yordanov, MD, MRCS
Academic Division of Neurosurgery
Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust
University of Cambridge
Hills Rd
Cambridge, CB2 0QQ
United Kingdom
Phone: 44 01223 805000 ext 348134
Email: yordanov.stefan@yahoo.com

Abstract

Background: Accurate head circumference (HC) measurement is essential when assessing neonates and infants. Tape measure HC measurements are prone to errors, particularly when performed by parents/guardians, due to individual differences in head shape, hair style and texture, subject cooperation, and examiner techniques, including tape measure placement and tautness. There is, therefore, the need for a more reliable method.

Objective: The primary objective of this study was to evaluate the validity, reliability, and consistency of HC app measurement compared to the current standard of practice, serving as a proof-of-concept for use by health care professionals.

Methods: We recruited infants attending the neurosurgery clinic, and parents/guardians were approached and consented to participate in the study. Along with the standard head circumference measurement, measurements were taken with the head circumference app (HC app) developed in-house, and we also collected baseline medical history and characteristics. For the statistical analysis, we used RStudio (version 4.1.1). In summary, we analyzed covariance and intraclass correlation coefficient (ICC) to compare the measurement's within-rater and interrater reliability. The $F$ test was used to analyze the variance between measurements and the Bland-Altman agreement, $t$ test, and correlation coefficients were used to compare the tape measurement to the measures taken by the HC app. We also used nonvalidated questionnaires to explore parental or guardians’ experiences, assess their views on app utility, and collect feedback.

Results: The total number of recruited patients was 37. Comparison between the app measurements and the measurements with a tape measure showed poor reliability (ICC=0.177) and wide within-app variations (ICC=0.341). The agreement between the measurements done by parents/guardians and the tape measurements done by the researcher was good (ICC=0.901). Parental/guardian feedback was overall very positive, with most of the parents/guardians reporting that the app was easy to use (n=31, 84%) and that they are happy to use the app in an unsupervised setting, provided that they are assured of the measurement quality.

Conclusions: We developed this project as a proof-of-concept study, and as such, the app has shown great potential to be used both in a clinical setting and by parents/guardians in their own homes.

(JMIR Form Res 2024;8:e54194) doi:10.2196/54194

KEYWORDS

head circumference; HC; hydrocephalus; neurosurgery; pediatric neurosurgery; paediatric neurosurgery; neurology; neuro; neurosurgeon; neurologist; mobile health; mHealth; app; apps; application; applications; digital health; smartphone; smartphones;
pediatric; pediatrics; paediatric; paediatrics; infant; infants; infancy; baby; babies; neonate; neonates; neonatal; toddler; toddlers; child; children

**Introduction**

Accurate head circumference (HC) measurement is essential when assessing neonates and infants. HC outside the normal range may indicate a brain development disorder, hydrocephalus, or an intracranial mass lesion. The growth pattern of HC, determined from serial measurements, provides valuable clinical information; for example, changes in HC are used to help determine whether hydrocephalus needs treatment and is predictive of neurodevelopmental outcomes [1].

The importance of HC measurement is recognized worldwide. The World Health Organization advises HC measurements just after birth (although measurements taken within the first 24 hours can be unreliable due to moulding), at the 8-week check, and any time thereafter if there are concerns about the child’s head growth, weight gain, development, or general health. It is also advisable for HC measurement to be performed at any pediatric review in the first 2 years of life. An accurate HC measurement is essential when clinically evaluating a neonate or infant. It monitors slow or excessive growth, assesses the impact of illness and treatment, and identifies those at higher risk of neurodevelopmental disorders [2]. HC should be measured at extremes of body weight (either below the 0.4th centile or above the 99.6th centile) or if there is rapid weight gain [3].

HC is the widest circumference of the head measured using a tape measure. Typically, this is performed by health care workers. With the increasing use of telephone and remote review clinics as a substitute for face-to-face appointments, community measurement of HC is more frequently used to help guide clinical management in various clinical specialties, such as neurosurgery, neonatology, and pediatrics. Parents and caregivers can be taught to measure the HC using tape, but its accuracy and reliability are sometimes insufficient for making clinical decisions. Many caregivers need more confidence to perform HC measurements themselves because of the technical challenges in performing the measurement and the potentially severe implications of erroneous measurements. Often, HC measurements must be checked by a health care professional, necessitating additional health visitor home visits or trips to the general practitioner.

HC measured with a tape measure is also prone to errors due to individual differences in head shape, hairstyles and texture, patient cooperation, and examiner techniques, including tape measure placement and tautness [4,5]. Measuring HC with a tape measure is often challenging due to poor cooperation in infants, particularly in a health care setting. Neonates in the neonatal intensive care unit pose significant challenges for HC measurements due to the risks associated with handling or removing them from the incubator. Many neonates require daily HC measurements.

The challenges of performing HC measurement with a tape measure, both in the hospital and the community, mean many missed opportunities to capture HC. Only 3% of infants presenting to Accident and Emergency had their head circumference measured within 1 year in an Australian hospital [6]. A study from a UK hospital found that HC measurement was performed sporadically in only 1 of 7 infants [7]. In summary, HC assessment is frequently missed due to difficulties measuring with a tape measure and the errors associated with this method. There is, therefore, an urgent need to improve our ability to monitor HC easily, accurately, and reliably, which would facilitate greater dependence on caregivers to perform this assessment independently.

We developed a smartphone app that measures HC using the smartphone camera and automated measurement to improve the accuracy and reliability of HC measurement. Our study’s main objective was to validate the accuracy of the HC app and prove its utility and feasibility when used by both health care workers and parents/guardians.

**Methods**

**Patient Recruitment**

Treating clinicians identified patients eligible for this study (infants <18 months of age) in a tertiary neurosurgical referral center and notified research team members with the consent of parents/guardians. Once identified, research team members approached patients’ parents/guardians to seek informed consent before their clinic appointment or at the bedside on the ward. An information sheet was used. If they chose to participate, we measured the HC during the clinic appointment or on the ward. Sample size estimation was calculated with 90% statistical power to detect a change of means by 1 and an SD of 1. This, along with 10% iteration, equaled 33 participants.

**App Development**

To successfully develop the app, we built the back end and front end of the app separately. Languages used were Python and JavaScript, and the app was built in Expo [8] using React Native [9]. The app’s back end used an algorithm to recognize a reference object and provide a measurement according to the object. Technically, the algorithm consists of a couple of parts. One part includes the foreground and contour recognition code, which recognizes the oval object and then measures the pixel points around its contour. Another part of the algorithm detects the reference object and recognizes the scale in comparison. They both fuse using Python to measure according to the identified contour points of both objects and the introduced scale of the known object sizes, providing a measurement within the scale. Points were manually inputted to clarify the top, bottom, and outermost lateral edges of the baby’s head. This was introduced as an additional measure to align with the already recognized contour and provide a more accurate measurement.

The algorithm was tested to match an SD of 0.5 cm between measurements. This SD target was selected to mirror the accepted variation in tape measurements, aiming for consistency.
and reliability while minimizing bias and variability associated with traditional methods [10].

We used a phantom baby model to test the accuracy of our measurement model (Figure 1). Instructions for parents/guardians are shown in Figure 2.

**Figure 1.** Head circumference app measurement sequence.
Ethical Considerations
Our regional research ethics committee (East of England-Cambridge East Research Ethics Committee) reviewed our study, and Health Regulatory Authority approval was obtained on June 6, 2022 (22/EE/0109). All parents/guardians of patients eligible to participate were approached with information about the study and given sufficient time for consideration of their participation. Informed consent in English was signed before any study-related procedures were undertaken. The data collected as part of the study were anonymized and deidentified. No patient identifiable data were collected as part of this study. Study participants and parents/guardians did not receive any compensation for their participation.

Study Objectives
After the successful trial on a life-size baby model (3B Scientific W17001 Baby Care Model), we developed a trial protocol and a study. The rationale of our study was to validate the HC app in a clinical setting and to review its usability by parents/guardians in the community.

The primary objective was to assess the technical validity (ie, accuracy and precision) and user reliability (ie, consistency of measurements across different raters) of the HC app compared to standard tape measure methods. We also set multiple secondary objectives; these include evaluating user satisfaction with the HC app and comparing the reliability of HC measurements between health care professionals and parents/guardians using both tape and the app.

Baseline and Study Data Collection
All participants had a medical history, clinical examination, and routine investigation details taken from their medical notes. The study phone did not retain any patient information. A nonvalidated questionnaire was used to capture parental or guardian feedback. Public and patient involvement feedback was sought in the design of the questionnaires for parental or guardian feedback. A complete list of questions can be found in Multimedia Appendix 1. HC measurements, as part of the study, were performed using a tape measure along the widest circumference of the baby’s head. Parents/guardians were instructed by a health care professional competent in HC measurement before the measurements were taken. They were also supervised during the procedure by a health care professional who was part of the study team. A specific set of instructions for using the app was prepared to be shown to the parents before they attempted to use the application. The instructions were developed with the help of a patient and a public representative group during the study development phase.

Figure 2. Parent/guardian instructions.
Statistical Analysis

In this study, we introduce a novel measurement method, which requires comprehensive evaluation to ascertain its properties. Reliability, validity, and reproducibility are key criteria in measurement science. We adopt the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) initiative standards, a widely recognized authority in clinical measurement quality [11] (Figure 2), to guide our assessment. In recent years, COSMIN has become a well-recognized organization [12], and consensus definitions can convey a unified message. COSMIN has also been dedicated to improving clinical measurement, including creating guidance and resources for measurement quality. That is why we decided to be guided by their definitions in exploring our novel measurement method’s reliability, consistency, and validity.

Guided by the COSMIN initiative, reliability is defined as the degree to which the measurement is free from measurement errors and includes a subsection of internal consistency (the degree of interrelatedness), reliability (the proportion of total variance due to true variance), and measurement error (systemic random error) [11,13]. To address this domain, we analyzed covariance, intraclass correlation coefficient (ICC), and Cohen Kappa and generated the interrater and within-rater reliability and variance. The ICC is between 0 and 1, where values below 0.5 indicate poor reliability, those between 0.5 and 0.75 indicate moderate reliability, those between 0.75 and 0.9 indicate good reliability, and any value above 0.9 indicates excellent reliability [14].

The COSMIN initiative also defines validity as the degree to which a measurement instrument measures the construct and its purpose to measure. The main subsections of this include content validity, criterion validity, and construct validity. Content validity ensures that the content of a measuring tool adequately reflects the given facts. Construct validity is defined as the degree to which the scores of a measurement instrument are consistent and includes structural validity, which is the degree of reflection of dimensionality, and hypotheses testing, which is synonymous with construct validity. Criterion validity shows how adequate the measurement is based on a “gold standard” [11,13]. For comparisons of agreements between the measures in this study, we used the correlation coefficients and the Bland-Altman method. The Bland-Altman plot analysis is a simple way to evaluate a bias between the mean differences and estimate an agreement interval. This interval encompasses 95% of the differences of the second method compared to the first one [15].

For the statistical analysis, we used RStudio (version 4.1.1). In summary, we analyzed covariance and ICC to compare the measurement’s within-rater and interrater reliability. The $F$ test was used to analyze the variance between measurements and the Bland-Altman agreement, $t$ test, and correlation coefficients were used to compare the tape measurement to the measures taken by the HC app.

Results

Demographics

We recruited 37 patients (23 male and 14 female), with a median age of 9 (IQR 4-16) months.

Normal means distribution was indicated visually with a histogram (Figure 3); however, the Shapiro-Wilk normality test with a $W=0.777$ and $P<.001$ showed a deviation from the normal distribution of means.
Comparison of HC App Measurements Versus Tape Measurements

A 2-sample t test was used to compare the mean values of all the tape measurements with HC app-based measurements. There was a statistically significant difference between the measurements ($P<.001$). The variance in the population means was found to be equal ($F_{5,30}=0.92134; P=.81$).

We then calculated the limits of agreement between the 2 measures using the Bland-Altman plot. The upper limits of agreement were determined to be within a range of 0.706 to 21.472, revealing significant variability in the data points (Figure 4). These were scattered across the field away from the 0 lines, indicating a weak agreement between the 2 methods. We then assessed the ICC for the measurements, which showed a value of 0.177, which is less than 0.5, indicating inferior reliability of the HC app measurements in comparison to the measurements taken by a tape measure.
Comparison of Measurements Within the App (Interrater Reliability)

We compared the HC app’s performance between different raters (researchers and parents/guardians) by calculating the ICC, which yielded a value of 0.341. The limits of agreement, determined using Bland-Altman plots, were also high (lower limit −16.696 and upper limit 20.252), with a mean difference of 11.089 and with a significant scatter of the values. These results show poor agreement between the rater evaluations. (Figure 5). We then performed an $F$ test to assess whether the variance between the 2 groups was equal. This resulted in $F_{35,71}=0.302$ and $P<.001$, indicating a difference in the variance of the measurements between the groups. The same analysis was also performed comparing the consequential researcher measurements, showing similar results ($F_{3,33}=1.994; P=.04$). The standard error of the mean (SEM) between the measurements done by parents/guardians and the researcher measurements was 0.979 and 0.921, respectively. Cohen Kappa was calculated at 0.435, showing a fair agreement between the measures.

Figure 4. Bland-Altman plot; limits of agreement and variability assessment between app measurements and tape measurements.

Figure 5. Bland-Altman plot; limits of agreement and variability assessment between different app measurements.
Comparison Between Parent/Guardian and Health Care Professional Tape Measurement

We also compared the interrater reliability between the measurements taken with a tape measure. The parents/guardians measured the HC under direct supervision and guidance. The ICC was 0.901, indicating good method reliability with very few lines above the 0 line in the plot and a mean difference of 0.5, comparable to other evidence in the literature [16]. The SEM for the medical professional measurements was 0.88, while the SEM for the parent/guardian measurements was 0.97. Cohen Kappa was set at 0.93, showing near-perfect agreement (Figure 6).

Figure 6. Bland-Altman plot; limits of agreement and variability assessment between parent/guardian and health care professional app measurements.

User Feedback

The survey results showed favorable opinions from parents/guardians, with the majority of them being happy to use the HC app in a nonsupervised setting. Most parents/guardians (n=31, 84%) answered that the HC app was easy or extremely easy to use, and 33 (89%) responded that they were very confident in using the HC app after reviewing the instructions (Figures S1 and S2 in Multimedia Appendix 1). The app instructions, created using public and patient involvement feedback, were also valued by the parents (Figure S4 in Multimedia Appendix 1). Once again, the vast majority (n=32, 87%) were either satisfied (n=14) or very satisfied (n=18) with the HC app (Figure S3 in Multimedia Appendix 1). Free-text comments from several parents/guardians described that they found the concept of using the HC app appealing and were happy to be presented with the opportunity to use it. Most parents were also happy to use the app even at home if available (Figure S5 in Multimedia Appendix 1). Furthermore, they indicated that they would be very likely to use the HC app provided that the app’s measurements and reliability improve.

Discussion

Principal Findings

We developed a smartphone app that measures HC using the smartphone camera and automated measurement. Our HC app is less accurate in its current iteration than the standard tape measure. Interrater reliability using the app was poor, but there was no significant difference in the variability between the operators. Parents/guardians also valued the convenience of using the app and the ease of performing the measurements, highlighting the potential of this technology once modified to improve accuracy.

The idea for our HC app was conceived during the COVID-19 pandemic lockdown in 2020, when disruption to health care services significantly reduced face-to-face appointments between patients and health care providers. Parents/guardians can measure HC using tape, but this requires some teaching and is prone to errors. Many parents/guardians express anxiety about assessments typically carried out by health care professionals being delegated to them and used in making clinical decisions. We, therefore, recognized the opportunity to create an automated method for HC measurement that is simple and easy to use and eliminates the errors associated with tape measure placement and patient cooperation.

Smartphone (and wearable) technology plays an ever-increasing role in public health and health care delivery, increasing the ability of health care workers to monitor patients remotely and empowering patients to track their health care metrics. The potential of technology to benefit health care is recognized by the World Health Organization, which introduced the term “mHealth” [16] to denote “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices” [16]. Mobile communication allows remote assessments with results that can be transferred distantly via mobile devices. In this landscape, we feel that replacing the tape measure with a smartphone equivalent is inevitable.

The current version of our HC app requires a reference object and manually placed input points to indicate the top, bottom,
and lateral edges of the baby’s head. The high reproducibility of measurements taken by the same person indicates the ease and reliability of this method. The relatively poor accuracy of our HC app is likely due to confounders such as the photo’s angle and distance and the type of reference object used. Using a bright-colored reference object (eg, a bottle cap) gave us a more accurate reading. However, we recognize that this choice of reference object may also relate to some inaccuracies, making the app less user-friendly and contributing to the difficulties in obtaining the perfect conditions for the photo. Thus, we intend to continue to develop the HC app using newer smartphone technology, such as light detection and ranging (LIDAR), which will improve the accuracy of the measurements, eliminate the need for a reference object, and make the app more user-friendly. This technology is highly advanced, and mobile LIDAR is already incorporated into most modern phones. Since 2020, Apple Inc has introduced a LIDAR sensor in some iPhone and iPad models, namely iPhone 12 Pro, iPhone 12 Pro Max, iPhone 13 Pro, iPhone 13 Pro Max, iPad Pro 2020, and iPad Pro 2021 (as of March 2022) [17]. Dimension can easily be measured between 2 points in the LIDAR 3D point cloud, providing accurate measurements of various objects. This technology is already used in other fields [18,19]. Thus, we are confident that this technology can be reliably used to improve the measurement accuracy and usability of the app.

Feedback from parents/guardians about the HC app revealed that they are confident in using the HC app and value the ability to track HC themselves in the convenience of their homes. They appreciated the ease of performing the measurements and performing them without fear of waking a sleeping infant. Overall, we received positive feedback from both the public and patient panel groups we consulted in preparation for our proof-of-concept study as well as during the trial itself. Many parents gave both written and verbal feedback with ideas and suggestions for how to improve the app’s utility. Most parents suggested features they would like to see, which shows good community engagement, reassuring that the concept can easily reach the target auditorium. Most of the parents were very welcoming to the technological solution we proposed. They expressed their interest in using it in the privacy of their homes after improvements to the app. However, reassurance is needed that the measurements are accurate and reproducible, which is something to be aware of when the final version of the app is distributed. We will keep that in mind in the development of future studies following the development of the final app to increase the adoption in the community.

Importantly, our study confirms that parent/guardian measurements using a tape measure are reliable with an excellent correlation (ICC=0.901) and agreement (kappa=0.93) between health care professionals and parents/guardians. Although our findings are similar to those previously reported by Sullivan et al [4] and support empowering parents/guardians to perform these measurements, these measurements were done under the direct supervision of a health care professional with an additional helping hand during the measurements. Having this in mind, translating these results may provide false reassurance regarding the reliability of parental measurements done at home without supervision. The only thing we can do to manage erroneous measures is to provide sufficient training and guidance to parents/guardians. This does not eliminate errors, but it offers a certain degree of reassurance. The gold standard measure, nevertheless, will remain the one measured by a health care professional. In contrast, the HC app, once improved, has the potential to introduce a sustainable, uniform, and reproducible means of measurement, which can be applied consistently across settings, providing equal reliability.

Limitations
To facilitate recruitment and combined measurements by both health care professionals and parents/guardians within a reasonable time frame, measurements were performed in the clinic rather than at home. A researcher supervised measurements; however, once instructions were given, no further help was provided, as our intention was for the HC app to be reliably used by parents/guardians independently at home. Our study design, therefore, only partially validates the HC app use in the home environment independent of health care professionals. We plan to explore this with future iterations of our HC app. Another important limitation to report is associated with the Shapiro-Wilk normality test. As the data set increases in size, the test can pick up very small variations, which can result in a higher likelihood of rejecting the null hypothesis [20].

Conclusions
Our HC app has demonstrated proof-of-concept for parent/guardian HC measurement using smartphone technology. The feedback collected from parents/guardians confirmed that the technology is easy to use, giving them the confidence to perform the measurement independently. Overall, parents/guardians were interested in this technological solution and were eager to give both written and verbal feedback during the study. This, along with the clinical proof-of-concept, reassured us that the technology is feasible, prompting us to initiate plans to improve the versions of our HC app.

Acknowledgments
Funding was received from Health Enterprise East via the innovation voucher awards in 2021.

Data Availability
The data sets generated during this study are available from the corresponding author upon reasonable request and after approval from the sponsor.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Questionnaire and additional statistics.

References
11. COSMIN. URL: https://www.cosmin.nl/ [accessed 2024-01-30]
19. Assessing the assumption of normality. UC Business Analytics R Programming Guide. URL: https://uc-r.github.io/assumptions_normality#:~:text=Shapiro%2DWilk%20Test%20for%20Normality&text=II%20the%20test%20is%20non,different%20from%20a%20normal%20distribution [accessed 2024-01-28]

Abbreviations
COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments
HC: head circumference
ICC: intraclass correlation coefficient
LIDAR: light detection and ranging

https://formative.jmir.org/2024/1/e54194
JMIR Form Res 2024 | vol. 8 | e54194 | p.945
(page number not for citation purposes)
**PDA:** personal digital assistant

**SEM:** standard error of the mean

© Stefan Yordanov, Kalsoom Akhter, Jye Quan Teh, Jawad Naushahi, Ibrahim Jalloh. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Use Intention and User Expectations of Human-Supported and Self-Help eHealth Interventions: Internet-Based Randomized Controlled Trial

Talia R Cohen Rodrigues1, MSc; Thomas Reijnders1, PhD; Linda D Breeman1, PhD; Veronica R Janssen1,2, PhD; Roderik A Kraaijenhagen1,3, PhD; Douwe E Atsma2, Prof Dr; Andrea WM Evers1,5,6, Prof Dr

1Health, Medical, and Neuropsychology Unit, Leiden University, Leiden, Netherlands
2Department of Cardiology, Leiden University Medical Center, Leiden, Netherlands
3NDDO Institute for Prevention and Early Diagnostics (NIPED), Amsterdam, Netherlands
4Vital10, Amsterdam, Netherlands
5Department of Psychiatry, Leiden University Medical Center, Leiden, Netherlands
6Medical Delta, Leiden University, Technical University of Delft, Erasmus University Rotterdam, Leiden, Delft, Rotterdam, Netherlands

†the BENEFIT consortium

Corresponding Author:
Talia R Cohen Rodrigues, MSc
Health, Medical, and Neuropsychology Unit
Leiden University
Wassenaarseweg 52
Leiden, 2333 AK
Netherlands
Phone: 31 71 527 3627
Email: t.r.cohen.rodrigues@fsw.leidenuniv.nl

Abstract

Background: Self-help eHealth interventions provide automated support to change health behaviors without any further human assistance. The main advantage of self-help eHealth interventions is that they have the potential to lower the workload of health care professionals. However, one disadvantage is that they generally have a lower uptake. Possibly, the absence of a relationship with a health care professional (referred to as the working alliance) could lead to negative expectations that hinder the uptake of self-help interventions. The Unified Theory of Acceptance and Use of Technology (UTAUT) identifies which expectations predict use intention. As there has been no previous research exploring how expectations affect the adoption of both self-help and human-supported eHealth interventions, this study is the first to investigate the impact of expectations on the uptake of both kinds of eHealth interventions.

Objective: This study investigated the intention to use a self-help eHealth intervention compared to a human-supported eHealth intervention and the expectations that moderate this relationship.

Methods: A total of 146 participants were randomly assigned to 1 of 2 conditions (human-supported or self-help eHealth interventions). Participants evaluated screenshots of a human-supported or self-help app–based stress intervention. We measured intention to use the intervention-expected working alliance and the UTAUT constructs: performance expectancy, effort expectancy, and social influence.

Results: Use intention did not differ significantly between the 2 conditions ($t_{142}=-1.133; P=.26$). Performance expectancy ($F_{1,140}=69.269; P<.001$), effort expectancy ($F_{1,140}=3.961; P=.049$), social influence ($F_{1,140}=90.025; P<.001$), and expected working alliance ($F_{1,140}=26.435; P<.001$) were positively related to use intention regardless of condition. The interaction analysis showed that performance expectancy ($F_{1,140}=4.363; P=.04$) and effort expectancy ($F_{1,140}=4.102; P=.045$) more strongly influenced use intention in the self-help condition compared to the human-supported condition.

Conclusions: As we found no difference in use intention, our results suggest that we could expect an equal uptake of self-help eHealth interventions and human-supported ones. However, attention should be paid to people who have doubts about the intervention’s helpfulness or ease of use. For those people, providing additional human support would be beneficial to ensure uptake. Screening user expectations could help health care professionals optimize self-help eHealth intervention uptake in practice.
Introduction

EHealth provides the opportunity to provide remote or automated health care support through digital tools [1]. EHealth is becoming increasingly relevant, for example, because of the physical restrictions during the recent COVID-19 outbreak [2]. During this pandemic, the demand for health care support increased too. Especially vulnerable groups experienced increased mental health difficulties [3,4], which require professional support. However, health care professionals already have a high workload and pressure [5] and, in some cases, even experience an additional workload from using eHealth [6]. Self-help eHealth interventions might provide a potential solution to these problems. Self-help eHealth interventions are defined as interventions in which automated support instead of human assistance is provided [1]. As this means that no human professionals are involved, self-help eHealth interventions are easier and cheaper to widely implement [1].

Despite these advantages, self-help interventions generally deal with low levels of adherence [7-10] and low uptake [11,12]. People generally show a higher intention to start with lifestyle changes using an intervention with additional human assistance compared to a self-help intervention [13]. While there has been extensive research on the factors contributing to nonadherence, there is a notable gap in our understanding when it comes to expectations that influence whether individuals will choose to use an intervention before starting. This information is important, as a growing number of eHealth tools are being developed and proven to be effective but hardly used [14,15]. Therefore, the aim of this study is to investigate whether there is a difference in use intentions between self-help and human-supported eHealth interventions and if user expectations influence the intention to use the intervention. If we know what expectations drive people’s intention to either use self-help or human-supported eHealth interventions, we could predict and even influence their actual uptake [16].

A possible explanation for the low use intention of self-help interventions could be the lack of a relationship with a health care professional [17]. This so-called working alliance, the degree to which a health care professional and patient is involved in a useful and collaborative working relationship [18], is an important predictor of intervention adherence and effectiveness [19,20]. People are more engaged with the intervention and motivated to work on their goals when they feel supported. This effect is not exclusive to face-to-face settings; it is also evident when internet-based human assistance is involved in the use of eHealth interventions [21,22]. It is even shown to be present in self-help eHealth interventions with automated support, using, for example, a human avatar [23-25]. Thus, people can form relationships not only with other people but also with technology [26]. Therefore, we predict that people’s expectations toward a potential future working alliance when using an eHealth intervention will influence their intention to use that intervention.

Other important expectations that may influence the use intention of human-supported and self-help eHealth interventions can be found within the Unified Theory of Acceptance and Use of Technology (UTAUT) [16]. According to this model, 3 different types of expectations explain people’s intention to start with an eHealth intervention. These UTAUT expectations are (1) performance expectancy: the extent to which someone expects that the eHealth intervention will be helpful in reaching their goals; (2) effort expectancy: the extent to which someone expects that the eHealth intervention will be easy to use; and (3) social influence: the extent to which someone expects that important others believe one should use the eHealth intervention [16]. Although the UTAUT model has been used to explain people’s intention to use eHealth in general [27,28], to our knowledge, no studies have used this model to investigate differences in people’s intention to use either human-supported or self-help eHealth interventions.

In this study, we aim to investigate (1) whether there is a difference in use intention between human-supported and self-help eHealth interventions, (2) whether the expected working alliance predicts the use intention of human-supported and self-help eHealth interventions, and (3) what UTAUT constructs predict the use intention of human-supported and self-help eHealth interventions.

Methods

Design and Sample

In an experiment, people were presented with a sham stress management app. In this app, people would either be supported by a human coach or by an automated coach. We decided to use a student sample, as they experience high levels of stress and could therefore benefit from an eHealth stress intervention [29], especially given their increased need for support during the COVID-19 pandemic [3,4]. They were asked to evaluate the screenshots of the app and measure their use intention, the 3 UTAUT constructs (performance expectancy, effort expectancy, and social influence), and their expected working alliance. We used a randomized between-participants design with 2 experimental conditions (human-supported or self-help eHealth interventions). Healthy participants aged 18 years or older, who had a sufficient level of grasp in English, were recruited on the campus of Leiden University with internet-based and offline flyers. Power calculations [30] identified a minimum sample size of 119 to detect a medium effect ($\alpha = 0.15$) with an $\alpha$ of .05, based on a linear multiple regression with 3 predictors.
Procedure and Manipulation

Interested participants could open the internet-based questionnaire and would be offered the internet-based consent form. After reading and agreeing to the informed consent, participants were automatically randomized into 1 of 2 experimental conditions (human-supported or self-help eHealth interventions). In both conditions, participants were instructed to evaluate a nonexistent stress management app for students called “Bye Bye Stress.” They were asked to carefully assess the screenshots of the app and give feedback to help the researchers make the app fit the needs of students. Although the design of the app and the content of the intervention were identical in both conditions, the conditions differed in the type of support that would be offered in the app. In the human-supported condition, the description of the app explained how a human coach would support the participants and provide them feedback. The screenshots of the app showed a picture of a human coach and messages with a human tone of voice (Figure 1). In the self-help condition, the description of the app explained how participants would receive automated feedback. In the screenshots, there was no picture of a human being, and the messages had a neutral tone of voice (Figure 1). All screenshots used in both conditions can be found in Multimedia Appendix 1. After this, participants were asked to complete the questionnaire.

Figure 1. Example screenshot of the app for human-supported (left) and self-help conditions (right).

Measures

Use Intention

The behavioral intention subscale of the UTAUT questionnaire [16] was used to assess use intention. The subscale consists of 3 items (eg, “I would intend to use ‘Bye Bye Stress’ in the next 6 months.”) measured on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A higher score indicates a higher intention to use the app. The scale showed a high internal consistency (Cronbach α= .953).
**Expected Working Alliance**

The expected working alliance was measured with an adjusted version of the Working Alliance Inventory—Short Revised form (WAI-SR) [31], which consists of 12 items measured on a 5-point Likert-type scale ranging from 1 (seldom) to 5 (always). Questions were adjusted to fit the context of the study by using the words “coach,” “lifestyle,” and “intervention” and being written in the future tense (e.g., “The coach and I will collaborate on setting lifestyle goals.”). A higher score indicates a stronger working alliance. The adjusted version had a high internal consistency (Cronbach α=.917).

**Performance Expectancy, Effort Expectancy, and Social Influence**

The constructs predicting behavioral intention according to the UTAUT model—performance expectancy, effort expectancy, and social influence—were measured with the corresponding UTAUT subscales [16]. Each subscale consisted of 4 items (e.g., “I find 'Bye Bye Stress' useful.”), measured with a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A higher score indicates a higher expectation of the app’s efficacy in helping the participant, a higher expectation toward the ease of use of the app, and a higher expectation that important others will approve the use of the app. The performance expectancy, effort expectancy, and social influence subscales all had sufficient internal consistency (Cronbach α of .764, .730, .792, respectively).

**Manipulation Check**

To assess whether participants carefully read the information and whether the manipulation had worked, they were asked to complete a manipulation check question (“During the intervention, I would be supported by...” followed by several options, such as “doctor” or “chatbot”).

**Analyses**

To test whether there was a difference in use intention between conditions, we ran a 2-tailed independent-sample t test with use intention as the dependent variable and condition (human-supported vs self-help eHealth interventions) as the independent variable. To test whether the association between condition and use intention differed for different levels of the UTAUT constructs (performance expectancy, effort expectancy, or social influence) as a covariate in 3 separate analyses. We analyzed both the main effects of condition and the UTAUT construct, as well as their interaction effect on use intention. To further investigate the interaction patterns found in the data, we conducted a simple slopes analysis. To formulate the simple slope equations for both the human-supported condition and the self-help condition, the intercept and the slope were obtained from the parameter estimates of the GLM analysis testing the association between expected working alliance and use intention.

To test whether the association between condition and use intention differed for different levels of the UTAUT constructs of performance expectancy, effort expectancy, and social influence, we conducted 3 univariate GLM analyses with interactions. We added use intention as dependent variable, condition as fixed factor, and each of the UTAUT constructs (performance expectancy, effort expectancy, or social influence) as a covariate in 3 separate analyses. We analyzed both the main effects of condition and the UTAUT construct, as well as their interaction effect on use intention. To further investigate the interaction patterns found in the data, we again conducted 3 simple slopes analyses: the intercept and the slope were obtained for both conditions from the parameter estimates of the GLM analyses testing the association between the UTAUT construct and use intention.

Statistical analyses were conducted with SPSS (version 26; IBM Corp) with a significance level set at \( p \leq 0.05 \).

**Ethical Considerations**

The study was approved by the Psychology Research Ethics Committee of Leiden University (CEP19-1125/557). Furthermore, the study was preregistered through the Center for Open Science [32]. Before the start of the study, participants were asked to sign an informed consent form. After completing all the questionnaires, they were debriefed and provided with a few examples of real internet-based stress management interventions in case they needed one. As compensation, participants received course credits.

**Results**

**Demographic Characteristics**

A total of 146 students participated in our study and completed the questionnaire. Their mean age was 21.8 (SD 4.51) years, 103 (70.5%) were female, and 104 (71.2%) were of Dutch nationality (Table 1). There were no significant differences in demographic characteristics between the 2 groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total sample (n=146)</th>
<th>Human-supported condition (n=73)</th>
<th>Self-help condition (n=73)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>21.8 (4.5)</td>
<td>22.0 (4.6)</td>
<td>21.6 (4.4)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>103 (70.5)</td>
<td>47 (66.2)</td>
<td>56 (76.7)</td>
</tr>
<tr>
<td>Nationality, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>104 (71.2)</td>
<td>49 (67.1)</td>
<td>55 (75.3)</td>
</tr>
<tr>
<td>European (non-Dutch)</td>
<td>37 (25.3)</td>
<td>20 (27.4)</td>
<td>17 (23.3)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (3.4)</td>
<td>4 (5.5)</td>
<td>1 (1.4)</td>
</tr>
</tbody>
</table>

Table 1. Baseline demographic characteristics.
Use Intention Per Condition
We found no significant difference in use intention between the human-supported condition and self-help condition ($t_{142}=-1.133; P=.26$; Table 2). Furthermore, we found no differences between the 2 conditions in any of the other constructs (Table 2).

Table 2. Mean scores and SDs of use intention and its predictors.

<table>
<thead>
<tr>
<th>Variable (scoring range)</th>
<th>Human-supported condition (n=73), mean (SD)</th>
<th>Self-help condition (n=73), mean (SD)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use intention (3-15)</td>
<td>7.5 (3.6)</td>
<td>8.2 (3.6)</td>
<td>.26</td>
</tr>
<tr>
<td>Expected working alliance (12-60)</td>
<td>42.3 (8.2)</td>
<td>40.3 (8.7)</td>
<td>.16</td>
</tr>
<tr>
<td>Performance expectancy (4-20)</td>
<td>13.8 (2.9)</td>
<td>14.0 (2.5)</td>
<td>.69</td>
</tr>
<tr>
<td>Effort expectancy (4-20)</td>
<td>16.9 (2.4)</td>
<td>16.8 (2.3)</td>
<td>.94</td>
</tr>
<tr>
<td>Social influence (4-20)</td>
<td>12.5 (2.9)</td>
<td>12.9 (3.1)</td>
<td>.66</td>
</tr>
</tbody>
</table>

Working Alliance and Use Intention
The GLM showed no significant association between condition and expected working alliance ($F_{1,140}=0.051; P=.82; \eta^2=0$). However, we did find a significant positive association between expected working alliance and use intention ($F_{1,140}=26.435; P<.001; \eta^2=0.159$). We found no significant interaction effect of condition and expected working alliance on use intention ($F_{1,140}=0.367; P=.55; \eta^2=0.003$; Figure 2).

Figure 2. Simple slopes of the effects of expected working alliance, performance expectancy, effort expectancy, and social influence on use intention.

UTAUT Constructs and Use Intention
The GLM showed no significant association between condition and performance expectancy ($F_{1,140}=3.34; P=.07; \eta^2=0.024$). We did, however, find a positive association between performance expectancy and use intention ($F_{1,140}=69.269; P<.001; \eta^2=0.331$) and a significant interaction effect of condition and performance expectancy on use intention ($F_{1,140}=4.363; P=.04; \eta^2=0.030$). An increase in performance expectancy was related to a greater increase in use intention in the self-help condition compared to the human-supported condition (Figure 2).

We also found no significant association between condition and effort expectancy ($F_{1,140}=3.4086; P=.07; \eta^2=0.024$). However, again, we did find a significant positive association between effort expectancy and use intention ($F_{1,140}=3.961; P=.049; \eta^2=0.028$) and a significant interaction effect of condition and effort expectancy on use intention ($F_{1,140}=4.102; P=.045; \eta^2=0.028$). An increase in effort expectancy was related to a greater increase in use intention in the self-help condition but not in the human-supported condition (Figure 2).

Again, we found no significant association between condition and social influence ($F_{1,140}=0.003; P=.96; \eta^2=0$). We did find a significant positive association between social influence and use intention ($F_{1,140}=90.025; P<.001; \eta^2=0.391$) but this time...
we found no significant interaction effect of condition and social influence on use intention ($F_{1,140}=0.020; P=.89; \eta^2=0$; Figure 2).

**Discussion**

**Overview**

In our study, we asked university students to evaluate a sham stress management app. We aimed to investigate whether there is a difference in use intention for self-help eHealth interventions compared to human-supported ones and what user expectations may influence this. We found that people were as likely to start using a self-help eHealth intervention as an eHealth intervention with human support. More than with human-supported interventions, the perception that the intervention might be ineffective or difficult to use limits the intention to start using self-help interventions. See Figure 3 for an overview of the findings.

![Figure 3. Overview of study findings.](https://formative.jmir.org/2024/1/e38803)

Although previous studies show a relatively low uptake and use intention of self-help eHealth interventions [11-13], we did not find differences in use intentions between the self-help and human-supported interventions. Possibly, the health beliefs, perceptions, and skills of our student sample might have played a role in this [33]. Not only do perceptions about the effectiveness or ease of use of an eHealth tool affect the start of an intervention but also perceptions about the risks of getting health-related problems and actually performing the health-promoting behavior [34]. Furthermore, a younger age and higher educational level are related to a higher intention to start eHealth interventions in general [13]. Our sample might therefore have been more open to using eHealth interventions and were less influenced by the presence, or lack thereof, of human support. Future research could focus on investigating the role of age and educational level on use intentions of self-help and human-supported eHealth interventions. Another explanation for the differences in findings between our and previous studies [11,12] could be the use of different outcome measures. Although the UTAUT model predicts that use intention can predict actual use, studies do show that people have difficulties translating their intentions into actual behavior [35]. The objective measure of uptake might therefore have led to different results compared to the more subjective measure of use intention we used, which would be interesting to additionally take into account. Finally, the study that did find a difference in use intention between self-help and human-supported interventions focused on interventions focused on interventions for mental health, such as depression [13]. It would be interesting to test if the need for social support during eHealth interventions depends on the goal of the intervention (eg, psychological vs lifestyle improvements).

Interestingly, we found that an expected working alliance has an equally strong effect on the intention to use either a human-supported or self-help intervention. This result is in line with previous studies showing a positive effect of working alliance on intervention effectiveness and adherence, both within human-supported [21,22] and self-help eHealth interventions with automated support [23-25]. Our findings show that working alliance is not important only during an intervention but even before the intervention has started in the form of expectations. The similar effect of the expected working alliance in both conditions suggests that people not only are able to actually have relationships with technology [26] but also seem to expect building one with the technology they are about to interact with. These results would also mean that improving the expected working alliance before the start of an intervention (eg, by designing a digital character that would welcome the user) would be a way to possibly increase the uptake of self-help eHealth interventions.
Finally, we found that performance and effort expectancy had a stronger effect on the use intention of self-help interventions compared to human-supported interventions. Not only the UTAUT model but also models such as the Health Belief model show that perceived benefits and perceived barriers affect whether people start with a health-promoting behavior, such as stress management [33]. What is new, though, is that the perceived effectiveness and ease of use of the intervention have a more pronounced impact on intention to use an intervention for interventions with an absence of human support compared to interventions where human support is available. This suggests that the perception that the intervention might be ineffective or difficult to use diminishes the intention to start using a self-help intervention but not the intention to start using a human-supported intervention. Meta-analyses show that the mere presence of a human being (even a nonprofessional) is a key ingredient in intervention effectiveness and the prevention of dropout [36-38]. Just the option of having someone available to provide procedural support (related to performance expectancy) or technical support (related to effort expectancy) seems to be enough for people to be motivated to start something new. The presence of a human coach could act as a buffer against negative expectations, which would make it easier for these people to adhere to the intervention [39]. Possibly, the mere presence of social support in the human-supported intervention could compensate for a lack of self-efficacy (the extent to which one believes in his or her own capabilities [40]) that people may feel when using a new intervention [41,42]. This could lower the perceived barriers and increase willingness to start using the intervention [33]. Exploring this further is crucial in a clinical context because individuals with limited social support tend to experience reduced adherence to health interventions and demonstrate less favorable intervention outcomes [39,43]. Even despite the relatively high use intention of self-help eHealth interventions, these results indicate that it is important to take the user’s needs and wishes into account when deciding on the level of human support to provide during an intervention.

**Self-help eHealth interventions will become more and more important in health care practice. To ensure uptake of new eHealth interventions, professionals could screen the user’s expectations toward the intervention’s helpfulness and ease of use beforehand (Table 3). If the user’s expectations turn out to be low, it would be useful to incorporate some level of human support into the eHealth intervention to prevent people from dropping out even before the start of the intervention. Additionally, designers of self-help eHealth interventions could pay extra attention toward its perceived helpfulness and ease of use. Preventing negative user expectations toward the intervention’s performance or effort expectancy could help increase the uptake of self-help eHealth interventions.**

**Table 3.** Items of the Unified Theory of Acceptance and Use of Technology subscales: performance expectancy (PE) and effort expectancy (EE).

<table>
<thead>
<tr>
<th>Item</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE1</td>
<td>I find [name eHealth technology] useful.</td>
</tr>
<tr>
<td>PE2</td>
<td>Using [name eHealth technology] enables me to [target behavior].</td>
</tr>
<tr>
<td>PE3</td>
<td>Using [name eHealth technology] will [target behavior].</td>
</tr>
<tr>
<td>PE4</td>
<td>If I use [name eHealth technology] I will know how to [target behavior].</td>
</tr>
<tr>
<td>EE1</td>
<td>My interaction with [name eHealth technology] is clear and understandable.</td>
</tr>
<tr>
<td>EE2</td>
<td>It would be easy for me to develop the skills needed to use [name eHealth technology]</td>
</tr>
<tr>
<td>EE3</td>
<td>I think [name eHealth technology] would be easy to use.</td>
</tr>
<tr>
<td>EE4</td>
<td>It would be easy to learn how to operate [name eHealth technology].</td>
</tr>
</tbody>
</table>

**Strengths and Limitations**

Our study was not without limitations. For example, although the screenshots of the app were adjusted to the experiences and interests of our sample, it is plausible that the topic of stress management was not equally relevant for all students, which could also have affected use intentions. For future studies, it would be better to tailor the goal of the eHealth intervention (eg, decreasing stress or improving physical activity) to the actual interests of the individual participants to investigate if and how this affects a participant’s use intention. Second, we used a university student population to test our hypotheses. People with a younger age and higher educational level have a more favorable attitude toward eHealth interventions in general [13]. To be able to generalize our findings, future research should investigate whether the same effects are found in other populations. It would be interesting to replicate this study with a target population who would benefit the most from eHealth interventions, for example, older patients with a chronic disease, to see if their expectations toward either human or automated support have similar effects on their intention to start with such interventions.

**Conclusions**

In our study, we investigated what expectations drive the intention to start using self-help and human-supported eHealth interventions. The results suggest that expectations toward the intervention’s helpfulness and ease of use are especially relevant regarding the use of self-help interventions. This means that people who have doubts about the intervention’s usefulness or usability would benefit the most from additional human support. The question, however, remains whether such expectations are also relevant for actual uptake. Our study provides a basis to further investigate user expectations within a clinical sample, which will provide health care practitioners with the tools to influence the uptake of eHealth interventions.
Acknowledgments
This work was supported by the Netherlands Cardiovascular Research Initiative, an initiative with the support of the Dutch Heart Foundation, CVON2016-12 BENEFIT, ZonMw (the Netherlands Organization for Health Research and Development), and the members of the BENEFIT consortium. Furthermore, we would like to thank Yalou Schoot (master student of psychology, Leiden University) for her assistance in the practical implementation of the study procedure and for data acquisition.

Authors' Contributions
TRCR and TR contributed to study design. TRCR was responsible for data acquisition. TRCR, TR, and AWME were involved in data analysis and interpretation and drafting the manuscript. TRCR, TR, LDB, VRJ, RAK, DEA, and AWME contributed to manuscript revision. All authors gave final approval and agreed to be accountable for all aspects of the work ensuring integrity and accuracy.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Screenshots shown in human-supported and self-help conditions.

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).

References


Abbreviations

GLM: general linear model
UTAUT: Unified Theory of Acceptance and Use of Technology
WAI-SR: Working Alliance Inventory–Short Revised form

©Talia R Cohen Rodrigues, Thomas Reijnders, Linda D Breeman, Veronica R Janssen, Roderik A Kraaijenhagen, Douwe E Atsma, Andrea WM Evers. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Effects of Embodiment in Virtual Reality for Treatment of Chronic Pain: Pilot Open-Label Study

Adam Saby¹, MD; Anthony Alvarez², BSc; David Smolins³, MD; James Petros⁴, MBA, MD; Lincoln Nguyen², BSc; Michael Trujillo², PhD; Oytun Aygün⁵, PhD

¹Department of Emergency Medicine, Occupational Health Division, University of California Los Angeles, Los Angeles, CA, United States
²Karuna Labs, San Francisco, CA, United States
³Remedy Medical Group, San Mateo, CA, United States
⁴Allied Pain and Spine, San Jose, CA, United States
⁵Karuna Labs, New York, NY, United States

Corresponding Author:
Oytun Aygün, PhD
Karuna Labs
185 Wythe Avenue Brooklyn
New York, NY, 11249
United States
Phone: 1 641824514
Email: oytunsultanaygun@gmail.com

Abstract

Background: Chronic pain has long been a major health burden that has been addressed through numerous forms of pharmacological and nonpharmacological treatment. One of the tenets of modern medicine is to minimize risk while providing efficacy. Further, because of its noninvasive nature, virtual reality (VR) provides an attractive platform for potentially developing novel therapeutic modalities.

Objective: The purpose of this study was to determine the feasibility of a novel VR-based digital therapy for the treatment of chronic pain.

Methods: An open-label study assessed the feasibility of using virtual embodiment in VR to treat chronic pain. In total, 24 patients with chronic pain were recruited from local pain clinics and completed 8 sessions of a novel digital therapeutic that combines virtual embodiment with graded motor imagery to deliver functional rehabilitation exercises over the course of 4 weeks. Pain intensity as measured by a visual analog scale before and after each virtual embodiment training session was used as the primary outcome measure. Additionally, a battery of patient-reported pain questionnaires (Fear-Avoidance Beliefs Questionnaire, Oswestry Low Back Pain Disability Questionnaire, Pain Catastrophizing Scale, and Patient Health Questionnaire) were administered before and after 8 sessions of virtual embodiment training as exploratory outcome measures to assess if the measures are appropriate and warrant a larger randomized controlled trial.

Results: A 2-way ANOVA on session x pre- versus postvirtual embodiment training revealed that individual virtual embodiment training sessions significantly reduced the intensity of pain as measured by the visual analog scale ($P<.001$). Perceived disability due to lower back pain as measured by the Oswestry Low Back Pain Disability Questionnaire significantly improved ($P=.003$) over the 4-week course of virtual embodiment regimen. Improvement was also observed on the helplessness subscale of the Pain Catastrophizing Scale ($P=.02$).

Conclusions: This study provides evidence that functional rehabilitation exercises delivered in VR are safe and may have positive effects on alleviating the symptoms of chronic pain. Additionally, the virtual embodiment intervention may improve perceived disability and helplessness of patients with chronic pain after 8 sessions. The results support the justification for a larger randomized controlled trial to assess the extent to which virtual embodiment training can exert an effect on symptoms associated with chronic pain.

Trial Registration: ClinicalTrials.gov NCT04060875; https://clinicaltrials.gov/ct2/show/NCT04060875

(JMIR Form Res 2024;8:e34162) doi:10.2196/34162
KEYWORDS
centralized pain; dicentralized pain; digital therapeutics; visual analog scale; Fear-Avoidance Beliefs Questionnaire; Oswestry; Oswestry Low Back Pain Disability Questionnaire; Pain Catastrophizing Scale; Patient Health Questionnaire; sensorimotor; virtual reality; chronic pain; pain; rehabilitation

Introduction

Chronic pain is a major health care problem worldwide. About 20% of the US [1-3] and European [4-6] populations have been reported to experience chronic pain. Chronic pain is associated with anxiety, depression, and other psychological disorders [7-13], as well as impairment in executive functions [14] and decision-making [15]. Opioids are one of the most common treatment strategies for chronic pain despite the potential for abuse, dependence, misuse, and accidental overdose [16,17]. Chronic pain is a consequential burden on society due to impairment of personal well-being, loss of productivity, and opioid use and dependence. Opioid addiction has been declared a public health emergency by the US Health and Human Services due to the potential misuse by more than 2 million people and more than 47,000 deaths annually [18].

While the specific neural mechanisms of chronic pain are variable and elusive, it has been proposed that a maladaptive neuroplasticity in the anterior cingulate cortex (ACC) occurs, leading to the long-term potentiation in neural circuits associated with the ACC [19]. The ACC has been shown to elicit rate, spatial, and temporal coding that is specific to the anticipation of pain [20]. The ACC shares connections with the periaqueductal gray (PAG), which is the primary control center for descending pain neumodulation. The disruption of ACC and PAG circuit activation may impair the ability of the descending pain suppression pathway to modulate pain in patients with chronic pain [21]. Long-term potentiation in the ACC may alter nociceptive processing, resulting in hyperalgesia and hyperpathic anticipation of painful events. The ACC and PAG also interact with the amygdala of the limbic system, which has been suggested to modulate pain pathways [22-27]. Through the involvement of the amygdala in pain pathways, the emotions such as stress and fear modulate pain [24,27-29].

Fear-avoidance models of pain suggest that attitudes and beliefs about pain are predictors of pain-related disability, as pain catastrophizing beliefs cause behavioral avoidance [30-37]. Pain catastrophizing has been defined as an exaggerated negative attitude toward pain experience [38]. Pain catastrophizing has been linked to the transition of acute pain into chronic pain [30,39], as well as a negative attitude toward medical procedures in patients with chronic pain [40,41]. Interventions aiming to develop adaptive psychological attitudes and strategies are suggested to reduce the level of pain catastrophizing and thus having a positive impact on patient outcomes [32,42].

Virtual reality (VR) has emerged as a novel technology for the treatment of pain, showing promise as a treatment strategy for chronic pain, burn pain, acute pain, and reducing the intensity of experimentally induced pain [43-50]. There are 2 strategies of VR that may promote analgesic effects: distraction therapy and immersiveness [47,48,51]. Distraction therapy temporarily diverts the attention of a user, which leads to reduced intensity of pain. Immersiveness, also known as embodiment, is the phenomenon by which a person identifies with and develops a sense of ownership over a body part that is not their own. Embodiment allows a user to interact with an artificial environment and receive altered visual feedback associated with the movement of a virtual avatar, which is controlled by the user’s manipulation of the VR hardware. Immersiveness in VR may activate premotor and somatosensory circuitry associated with the body parts that are embodied [52]. Thus, there may be a reorganization of the sensorimotor representations within the central nervous system in a way that the perception of a painful limb is modified, resulting in people perceiving the limb as less painful [48,53-55]. Furthermore, virtual embodiment has been shown to influence pain-free range of motion in patients with unilateral chronic shoulder pain [49].

Digital rehabilitation interventions could have multidimensional, biopsychosocial effects for managing pain [56,57]. In patients with chronic neck pain using digital rehabilitation therapy, disability and range of motion was improved [58,59], suggesting that VR interventions are promising for treatment of chronic pain. Multiple sessions of Fear Avoidance Beliefs Training through a rehabilitation medical device have shown improvements in the Pain Catastrophizing Scale (PCS), along with reductions in pain intensity and mobility impairment and disability [49]. Active coping strategies, such as exercise in chronic pain, were shown to have better outcomes in pain-related disabilities compared to passive strategies such as taking medication or resting [60]. Immersiveness in VR could potentially provide the patients with a more active participation in their treatment for chronic pain.

Chronic pain is a health care problem worldwide. There is a compulsory need for alternative, noninvasive, and nonaddictive therapeutics for treating chronic pain [61]. The purpose of this study was to determine the feasibility of a novel VR-based digital therapy for the treatment of chronic pain. The safety and feasibility of an intervention through embodiment in VR to treat chronic pain was assessed. In this study, 24 patients with chronic pain received a novel VR-based functional rehabilitation program (Karuna Virtual Embodiment Training [KVET]; Karuna Labs, Inc) over the course of 4 weeks, with 2 sessions per week. Self-perceived pain intensity rating, pain catastrophizing, and self-perceived disability were measured. The measurements were anonymized for analysis. We hypothesized that gradual exposure of patients to functional rehabilitation in immersive therapy in VR can be used to help patients overcome pain-related fears and catastrophizing beliefs. We hypothesized that an immersive VR therapy would have a positive impact on patient outcomes.
Methods

Ethical Considerations
This study’s protocol was approved by Advarra’s institutional review board (reference number: Pro00026459) and conducted in accordance with the ethical standards of the Declaration of Helsinki.

Study Procedure
This trial was registered at ClinicalTrials.gov (NCT04060875). In total, 24 adult participants that were diagnosed with chronic (for 3 months or longer) lower back or chronic upper extremity pain were recruited from local pain clinics over the normal course of business. Individuals with a history of severe mental illness, including schizophrenia, bipolar disorder I or II, and posttraumatic stress disorder; a history of susceptibility to seizures per the participant’s reporting; and pregnant women were not included in this study. Interested participants received information regarding virtual embodiment training and were scheduled for an intake visit. During the intake visit, participants completed a written informed consent and then were administered initial outcome measures (patient-reported pain questionnaires). To assess potential adverse events associated with VR, the Simulator Sickness Questionnaire (SSQ) was administered to all potential participants. The SSQ helps identify the risk of nausea and dizziness in VR [62]. In this study, no patients exceeded the exclusion threshold on the SSQ, and none were excluded from participation on this basis. Qualified participants received information regarding VR therapy and were scheduled for an intake visit. During the intake visit, all recruited candidates completed a written informed consent, and baseline instruments for outcome measures were administered. All recruited participants completed this study.

Intervention
Virtual embodiment sessions were administered on an HTC Vive (HTC Corporation) VR head-mounted display (110° field of view, 1080 x 1200 pixels/eye, and 90 Hz refresh). The HTC Vive hand controllers and Vive trackers were used to provide an immersive VR experience. The HTC Vive accurately tracks limb and trunk position and movement and delivers an immersive experience where a virtual avatar is controlled by manipulating the position of the hand controllers, Vive tracker, and head-mounted display.

Patients received biweekly sessions of virtual embodiment training over the course of 4 weeks. Visual analog scale (VAS) was administered before each session to obtain an indication of baseline pain intensity levels. Patients were then administered virtual embodiment training. Virtual embodiment training sessions progressed from 20 minutes in the initial session to 45 minutes in later sessions. Following each virtual embodiment training session, the VAS was readministered to determine whether individual sessions provide reductions in self-perceived pain intensity. KVET consists of exercises in an embodied VR experience that are designed based on the principles of graded motor imagery (GMI). In this study, patients with upper-extremity chronic pain engaged in the 4 virtual embodiment training modules: laterality, motor imagery, “mirroring,” and predictive coding. Laterality training consisted of movements mirrored from one side to the other by an avatar in VR. In the motor imagery module, the participant experienced a first-person perspective view of virtual avatar limbs moving while being instructed to imagine the movement is happening in his or her own body. In the mirroring module, a participant’s healthy limb is used to produce movements that appear to be occurring in the painful limb. In the predictive coding module, participants were surrounded by floating orbs in a virtual environment (Figure 1). The orbs were placed at random locations within arm’s reach of the participant. Participants were instructed to reach and grab the orbs promoting upper-extremity movement. A description of rehabilitation exercises administered in KVET for patients with chronic low back pain has been previously published [49].
Figure 1. An example of a virtual avatar. In this exercise, the patient is participating in an immersive functional movement exercise. The figure shows a third-person view of what a patient would be doing in an exercise. In the exercise, participants grasp floating orbs in an immersive virtual reality experience. Shoulder flexion, scaption, and internal and external rotation are measured in the experience.

Measures
Self-perceived pain intensity served as the primary outcome measure and was evaluated before and after each virtual embodiment training session using a VAS for pain. Participants were asked to denote pain using a 10-cm line with the left end representing no pain and the right indicating the worst possible pain. Participants would then check off the region of the line they believed accurately represented their current pain. The distance between the left most part of the 10-cm line and the participants check mark was then measured using a standard ruler. This instrument is standard for tracking pain [63]. A 2-way ANOVA on session (1-8) was used to assess differences in VAS score as a function of session and clinical intervention.

To assess the appropriateness of outcome measures on a larger scale, exploratory outcome measures were administered for pain, physical functioning, mental functioning, and disability. Patient-reported pain questionnaires were administered before the first session of virtual embodiment training and after the last session. The validated instruments included the Fear-Avoidance Beliefs Questionnaire (FABQ), Oswestry Low Back Pain Disability Questionnaire (ODI), PCS, and the Patient Health Questionnaire. The FABQ is a questionnaire measuring the effect that pain has on daily life activities [64]. The ODI, frequently used in research of physical therapy efficacy for treatment of lower back pain [65,66], was applied to the subset of participants whose diagnoses included low back pain. The PCS includes 3 subset scales: rumination, magnification, and helplessness [67]. Due to the ordinal nature of subjective pain questionnaires and the likeliness of nonnormal distribution, a Wilcoxon signed rank test was used to assess the difference in the subjective pain questionnaire score before and after clinical intervention.

Results

Self-Perceived Pain Intensity
Self-perceived pain intensity as measured by the VAS improved as a function of individual virtual embodiment training sessions. A 2-way ANOVA on session (1-8) × pre- versus posttraining on the VAS revealed a significant main effect of KVET therapy sessions ($F_{1,8}=14.246; P<.001; \text{Cohen} \ d=0.504$). There was no significant main effect of the number of sessions ($P=0.85$).

Patient-Reported Pain Questionnaires
Pre- and posttraining FABQ work and physical activity scores were recorded for 14 participants. A Wilcoxon signed rank test revealed no significant difference between pre- and post-KVET scores for FABQ-work ($Z=-0.629; P=0.57$; mean pretest score 26.14, SD 13.04; mean posttest score 24.86, SD 13.07) nor FABQ-physical activity ($Z=-1.480; P=0.15$; mean pretest score 15.07, SD 7.08; mean posttest score 13.21, SD 7.16).

Pre- and posttraining Patient Health Questionnaire scores of the 4-week course were recorded for 10 participants. A Wilcoxon signed rank test revealed no significant difference between pre- and post-KVET scores ($Z=-1.262; P=0.23$). Pre- and posttraining ODI scores for the 4-week course were recorded for the population of participants experiencing low back pain. A Wilcoxon signed rank test revealed a significant improvement on how patients rated their disability associated with low back pain after KVET ($Z=-2.819; P=0.003$; mean pretest score 23.07, SD 9.42; mean posttest score 19.73, SD 6.99). A post hoc measurement of effect size revealed a medium treatment effect of KVET (Cohen $d=0.401$).
Pre- and posttraining PCS scores for the 4-week course were recorded for 15 participants. A Wilcoxon signed rank test revealed a significant improvement of pain catastrophizing in the helplessness category ($Z=-2.254; P=0.02$; mean pretest score 11.60, SD 6.62; mean posttest score 9.13, SD 5.10). A post hoc measurement of effect size revealed a medium treatment effect of KVET ($d=0.418$). No significant difference was observed for PCS ruminating ($Z=0.605; P=0.59$; mean pretest score 8.80, SD 4.75; mean posttest score 9.33, SD 3.83), magnification ($Z=-0.247; P=0.84$; mean pretest score 5.27, SD 2.02; mean posttest score 5.13, SD 3.09), or total ($Z=-1.224; P=0.24$; mean pretest score 25.60, SD 11.48; mean posttest score 23.60, SD 11.02). These results suggest that 8 sessions of KVET are an effective treatment for improving the feeling of helplessness associated with chronic pain and allows a person to experience that they can influence pain and movement.

We analyzed the magnitude of change in scores before and after training to assess the degree to which patients with chronic pain improved on pain intensity before and after each session, low back pain disability rating before and after 8 KVET sessions, and the sense of helplessness in dealing with pain before and after 8 KVET sessions. The magnitude was defined as the amount of reduction between pre- and posttraining scores. Figure 2 displays scatter plots of the magnitude of change of the pre-(x-axis) and posttraining (y-axis) scores. A Pearson product correlation analysis revealed a significant correlation for VAS ($r=-0.41; P<0.001$), PCS helplessness ($r=-0.64; P=0.01$), and ODI ($r=-0.69; P=0.003$).

**Figure 2.** The magnitude of improvement on measurements that showed a significant improvement. (A) Magnitude of change for MVAS; (B) magnitude of change for Oswestry; and (C) magnitude of change for PCS helplessness. PCS: Pain Catastrophizing Scale; KVET: Karuna Virtual Embodiment Training; MVAS: mechanical visual analog scale.

Discussion

**Principal Findings**

This study aimed to assess the feasibility and safety of a novel digital therapy that delivers functional rehabilitation through virtual embodiment in VR. Patient-reported exploratory measures were administered to assess the effectiveness of virtual embodiment training as a viable nonpharmacological, noninvasive treatment of pain. Pain intensity, as measured by the VAS, showed statistically significant reductions after each individual KVET session but no improvement session over session. Disability, as measured by the ODI, and pain catastrophizing all improved as a function of a 4-week virtual embodiment training.

In this experiment, pain intensity was reduced after the intervention using embodiment in VR. The effect size was similar to that reported in magnetic neural stimulation for chronic pain in spinal cord injury [68] and acceptance and commitment therapy to treat chronic pain [69], suggesting that this protocol is a promising intervention for patients with chronic pain. The use of VR in interventions has been previously proposed to be an interesting nonpharmacological alternative for treating chronic pain [70]. VR technology is rapidly advancing due to the introduction of headsets such as Oculus Rift (Meta) and HTC Vive. The recent advances in handheld controllers, cameras, and body trackers have enabled first-person body experience in VR. Embodiment (sometimes referred to as immersiveness) is the process by which a person identifies with and develops a sense of ownership over a body part that is not their own. In this experiment, virtual embodiment, the first-person perception of sensory feedback of the actions of an avatar, was used. Using VR, it is possible to display movements that are not possible for the patient in physical reality and modify the movement to appear more understated or exaggerated. When a person experiences embodiment, sensorimotor representation in the central nervous system of the painful limb is suggested to go through a reorganization. This modification in the representation and thus the perception of a painful limb allows people to perceive the limb as less painful [48,53–55]. Although the sense of embodiment was not measured in this feasibility study, the reduction in the pain intensity observed in this study before and after each intervention session could be due to the embodiment protocol used, as embodiment has the potential to shift the perception of pain to less intense and less threatening.

The reduction of pain in this experiment could also be partially due to the specific exercises used. Exercises used in this study’s intervention through KVET were designed based on the principles of GMI. GMI is a biopsychosocial approach to physical therapy aiming for a gradual reintroduction of motor actions and was developed as a systematic approach to ready
subjects with severe chronic pain to engage their sensorimotor pathways related to their painful bodily regions. In a meta-analysis of 6 randomized control trials, GMI was found to produce substantial pain reduction in chronic pain conditions of the upper limbs, more efficiently than usual physiotherapy [71]. GMI consists of 3 stages: (1) laterality testing, in which the participant identifies whether a limb is the right or left limb; (2) motor imagery, in which the participant imagines the movement of his or her affected limb while engaging motor pathways to “feel” as though they are actually moving; and (3) mirror therapy, or the mirror visual feedback (MVF) [72]. In this experiment, MVF was incorporated into the VR experience for treating chronic pain. Initially developed to treat phantom limb pain, MVF is a dynamic therapy that uses a mirror image of an unaffected contralateral limb to visually represent the function and motion of an affected or impaired limb. MVF has been shown effective in treating complex regional pain syndrome in randomized control trials, which found clinically significant reduction in pain above conventional physical therapy techniques [73]. The application of MVF through the virtual embodiment could have reinforced its effectiveness in this study. However, although pain intensity was reduced after each session, there was no change in pain intensity over the course of 4 weeks. Previously, GMI and similar interventions that aim for cortical remapping for neuroadaptive changes in chronic lower back pain have found short-term benefits on pain intensity and disability while long-term effects are yet to be confirmed [74]. The VR intervention in this experiment was designed based on GMI, suggesting that GMI through VR shows similar short-term benefits. Alternatively, the lack of reduction in the pain intensity over the course of 4 weeks could also be due to the number of sessions. Perhaps more durable improvements in pain intensity requires more sessions for attaining adaptive neuroplasticity in chronic pain. Future studies with more frequent sessions or longer interventions using VR could clarify the lack of long-term effects in this study.

In this study, the disability component of the ODI and the helplessness subscale of the PCS were reduced over the course of 4-week intervention, suggesting that the virtual embodiment protocol used was effective in improving psychological measures. Moreover, the observed effect size of reduction in disability was similar in magnitude to that reported for other nonsurgical treatments for chronic pain such as acupuncture, behavioral therapy, and exercise [75], suggesting that 8 sessions of KVET is an effective treatment for chronic low back pain disability. Similarly, the helplessness was also reduced over the course of 4-week immersiveness intervention, suggesting that 8 sessions of KVET are an effective treatment for improving the feeling of helplessness associated with chronic pain. Previously, digital rehabilitation therapy has been shown to reduce disability while increasing the range of motion in patients with chronic cervical neck pain [58]. Digital rehabilitation therapy was suggested to alter the activation of pain pathways in the central nervous system [76] with 2 approaches of distraction and immersiveness [48,51]. In this experiment, through the embodiment, the gradual exposure of patients could have helped the patients overcome learned behaviors and fear avoidance due to chronic pain. The immersiveness in a VR setting could be especially pertinent for individuals with chronic pain who experience reduced motivation for volitional movement. Previously, similar rehabilitation sessions of Fear Avoidance Beliefs Training have shown improvements in pain catastrophizing, pain-related mobility impairment, and pain intensity and disability [49]. Moreover, in this experiment, there was a negative correlation between the change over the course of 4 weeks in pain intensity and helplessness as well as pain intensity and disability, suggesting that the patients with the most severe pain may benefit the most from KVET therapy on psychological measures related to chronic pain. The immersive VR experience could have allowed the person to gain a sense of control and belief that they can influence the pain and improve movement. With improved movement, in turn, the brain could have dampened the threat response to movement. Thus, over time, adaptive neural pathways are established, allowing the patients to be desensitized to pain associated with movement.

Although this study provides evidence for safety and the benefits of a 4-week virtual embodiment training for chronic pain, there were also limits. Notably, the lack of a control group is an important limitation of this study. Further studies are to be carried with a control group who is following the standard of care for chronic pain. Alternatively, this study did not have a follow-up measure, hence the long-term efficiency and effects of the treatment are not known.

Finally, in this experiment, all 24 patients completed all planned sessions during the 4-week virtual embodiment training without adverse events, suggesting that when designed correctly, VR therapies that combine pain relief and movement may provide nonpharmacological, noninvasive, and nonaddiction modalities for treating chronic pain. This study’s results support the justification for a larger randomized controlled trial to assess the extent to which virtual embodiment training can exert an effect on symptoms associated with chronic pain. VR may be ubiquitous in homes within the near future, this could introduce the possibility of engaging in neurorehabilitation from the convenience of the home, with data from sessions accessible by prescribing clinicians. Further research is needed to explore the potential use and effects of VR for managing chronic pain.

Conclusion

This experiment showed that the gradual exposure of patients to functional rehabilitation through a rehabilitation medical device provided cognitive retraining, improving pain intensity after each session as well as pain-related disability and helplessness over the course of 4 weeks. When designed correctly, VR therapies that combine pain relief and movement may provide nonpharmacological, noninvasive, and nonaddiction modalities for treating chronic pain.
References


Contactless Monitoring System Versus Gold Standard for Respiratory Rate Monitoring in Emergency Department Patients: Pilot Comparison Study

Charlotte E Goldfine, MD; Md Farhan Tasnim Oshim, MSc; Brittany P Chapman, BSc; Deepak Ganesan, PhD; Tauhidur Rahman, PhD; Stephanie P Carreiro, MD, PhD

Division of Medical Toxicology, Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, United States
Manning College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, United States
Department of Emergency Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States
Halıcıoğlu Data Science Institute, University of California San Diego, San Diego, CA, United States

Corresponding Author:
Charlotte E Goldfine, MD
Division of Medical Toxicology
Department of Emergency Medicine
Brigham and Women's Hospital
75 Francis St
Boston, MA, 02115
United States
Phone: 1 617 732 5640
Email: cgoldfine@bwh.harvard.edu

Abstract

Background: Respiratory rate is a crucial indicator of disease severity yet is the most neglected vital sign. Subtle changes in respiratory rate may be the first sign of clinical deterioration in a variety of disease states. Current methods of respiratory rate monitoring are labor-intensive and sensitive to motion artifacts, which often leads to inaccurate readings or underreporting; therefore, new methods of respiratory monitoring are needed. The PulsON 440 (P440; TSDR Ultra Wideband Radios and Radars) radar module is a contactless sensor that uses an ultrawideband impulse radar to detect respiratory rate. It has previously demonstrated accuracy in a laboratory setting and may be a useful alternative for contactless respiratory monitoring in clinical settings; however, it has not yet been validated in a clinical setting.

Objective: The goal of this study was to (1) compare the P440 radar module to gold standard manual respiratory rate monitoring and standard of care telemetry respiratory monitoring through transthoracic impedance plethysmography and (2) compare the P440 radar to gold standard measurements of respiratory rate in subgroups based on sex and disease state.

Methods: This was a pilot study of adults aged 18 years or older being monitored in the emergency department. Participants were monitored with the P440 radar module for 2 hours and had gold standard (manual respiratory counting) and standard of care (telemetry) respiratory rates recorded at 15-minute intervals during that time. Respiratory rates between the P440, gold standard, and standard telemetry were compared using Bland-Altman plots and intraclass correlation coefficients.

Results: A total of 14 participants were enrolled in the study. The P440 and gold standard Bland-Altman analysis showed a bias of −0.76 (−11.16 to 9.65) and an intraclass correlation coefficient of 0.38 (95% CI 0.06-0.60). The P440 and gold standard had the best agreement at normal physiologic respiratory rates. There was no change in agreement between the P440 and the gold standard when grouped by admitting diagnosis or sex.

Conclusions: Although the P440 did not have statistically significant agreement with gold standard respiratory rate monitoring, it did show a trend of increased agreement in the normal physiologic range, overestimating at low respiratory rates, and underestimating at high respiratory rates. This trend is important for adjusting future models to be able to accurately detect respiratory rates. Once validated, the contactless respiratory monitor provides a unique solution for monitoring patients in a variety of settings.

(JMIR Form Res 2024;8:e44717) doi:10.2196/44717
KEYWORDS
cardiopulmonary monitoring; contactless monitor; radar; respiratory rate; vital signs

Introduction

Respiratory rate is a fundamental vital sign that serves as an indicator of physiologic function [1]. Changes in respiratory rate are often one of the first indicators of severe illness and clinical deterioration in a variety of disease states, such as sepsis, metabolic acidosis, respiratory distress, and drug toxicities [2,3]. Early identification of clinical decline by recognition of changes in respiratory rate is associated with improved outcomes, such as decreased intensive care unit admissions, decreased length of stay, and improved functional outcomes in patients [4]. However, respiratory rate is often considered a “neglected vital sign” due to inconsistent documentation [5]. There are 2 techniques currently used to measure respiratory rate. The gold standard for determining respiratory rate is manual counting. The current standard of care in hospital facilities and emergency departments (EDs) is transthoracic impedance plethysmography, which measures chest wall movement through cardiac telemetry electrodes. However, both techniques have several limitations. Manual counting is labor-intensive, needs to be done by visually examining the patient, and cannot practically be continuous for long periods of time [6,7]. Transthoracic impedance plethysmography is susceptible to inaccurate readings, as any aberrant movement of the patient makes the monitoring ineffective [8,9].

The development of technologies that accurately and effectively monitor respiratory rates is important to improve the care of patients. Additionally, with the widespread use of telemedicine, digital health interventions that can be used in outpatient settings or settings where a patient cannot be closely monitored are increasingly needed. One such technology is a contactless sensor system that uses an ultrawideband impulse radar-based contactless respiratory monitor PulseON 440 (P440) capable of detecting subtle movements in participants (eg, as the chest wall rises). The P440 works by sending an electromagnetic wave with a transmitter antenna, the reflections of which are caught by the receiver antenna. The device uses 2-way time of flight ranging to measure the distance between the radar and a target. In addition, this radar is a coherent radio transceiver that allows the energy in each transmitted pulse to be summed, improving the signal-to-noise ratio of received transmissions. This monitor has demonstrated high accuracy in laboratory testing, is less than 2 cm in size, and can perform scanning at rates up to 125 Hz at distances up to 30 meters [10]. The monitor operates at approximately 50 μW, which is considered a very low power transmission. For comparison, a standard incandescent lightbulb operates at 60 W. Its capability to detect small movements with extreme accuracy makes this device optimally suited to measuring respiratory rate.

As new technologies are being developed, it is important to compare them to current gold standards to assess whether these technologies are both accurate and practical for clinical use. This study aimed to pilot-test the P440 radar in a cohort of ED patients to (1) compare the performance of the P440 radar to the gold standard (manual counting) and to the standard of care (transthoracic impedance plethysmography) in a clinical setting and (2) compare performance in subgroups of interest (patients with cardiopulmonary diagnoses and by sex).

Methods

Recruitment

This pilot study was performed at a large, academic, tertiary-care, level-one trauma center in central Massachusetts that treats approximately 135,000 patients per year. A convenience sample was enrolled during the study period (June 2019-February 2020). Eligible participants were aged 18 years or older, were being monitored by cardiac telemetry for respiratory rate in the ED, and were able to provide informed consent. Individuals were excluded if they were pregnant or were prisoners, as these populations are routinely excluded in this phase of research. Participants were screened for eligibility through the ED electronic medical record and approached for consent after discussion with the treating ED clinical providers.

Study Design

Once the participant consented to participate, basic demographic and relevant clinical information were obtained from the electronic health record. A total of 3 P440 radar units were placed in a triangulated formation in the participant room for a 2-hour period (Figure 1). During the 2-hour period, the gold standard respiratory rate (manual counting) and the standard of care respiratory rate (transthoracic impedance plethysmography) were recorded at 15-minute intervals. Relevant clinical data (including medications given during the study period and significant events) were recorded.
Figure 1. Schematic of patient room set-up. Blue Ts represent the PulsON 440 radar units.

Hardware

The main study device was the P440, which used radar technology to calculate respiratory rate. Data were collected using application programming interfaces. The Raspberry Pi3 (Raspberry Pi Foundation) interfaced with the P440 (programmed in C language) and stored all collected data on a microSD card. Each radar unit consisted of the P440 module, an absorber, a Raspberry Pi unit, and a hard disk (Figure 2). All components were assembled in a 3D-printed box. Details of the radar have already been published elsewhere [10].

Figure 2. (A) PulsON 440 radar monitor with LF75 absorber behind the antenna. (B) Top view of the radar with absorber. (C) Radar box with Raspberry Pi and a hard disk drive. (D) Enclosed radar box.
**PulsON 440 Data Cleaning, Processing, and Fusion**

The radar data were collected from all 3 devices in raw radargram format. The files were marked with Uniplexed Information Computing System (UNIX) timestamps and were synchronized in order to have the same event observed by all 3 radars for the same time windows. Raw radargrams contained information about participant movement as well as reflections from all the static clutter present in the environment (e.g., walls, stretchers, and medical furniture). The first step in cleaning the raw data involved removing the static clutter [11]. A background subtraction technique was used to get rid of the clutter. The mean of the first 100 scans where there were no participants was subtracted from each window of 30 seconds. A bandpass filter was applied to remove direct current and high-frequency noise from the radargrams.

The system leveraged the different vantage points of the radars in 3 different positions in the room to observe the respiratory motion of the participant. As the orientation of the participant was not known and in a dynamic environment, we applied an independent component analysis (ICA) to find out which radar had the best respiratory signal [12].

Each radar signal and radargram were then filtered by an equiripple finite impulse response filter at cutoffs analogous to the normal breathing range from 6 to 30 breaths per minute. The location of the participant being monitored was extracted using a localization algorithm along with trilateration and Kalman filtering [13,14]. From each radar, the system focused on 50 range bins (45 cm) centered at the participants’ location bin. The focused range bin forms 50 individual time series of 30 seconds for a particular window. The 2D signal with 50 timeseries was then collapsed to a 1D signal using a sum-aligned function. The function found the timeseries that had the highest root-mean-square energy in that window and used it as a reference to align the remaining 49 timeseries with itself. After alignment, the function was summed to generate the global sum-aligned 1D signal, which preserved the breathing motion.

The ICA was applied to all 3 radars in 1D time series. The ICA output 3 independent components, which underwent a fast Fourier transform (FFT). The component with the strongest FFT peak was selected as the most dominant radar signal. We applied 2 different frequency detection methods to further identify the respiration rate from the selected radar—FFT and zero-crossing rate [10].

**Statistical Analysis**

Descriptive statistics were calculated for basic demographic and clinical information. We compared the respiratory rate measured by the P440 with the gold standard (manual) respiratory rate by the Bland-Altman plot method using GraphPad Prism (version 9; GraphPad Software). This approach used the means and differences between the pairs of readings to calculate the mean difference (bias) and the upper and lower limits of agreement [15,16]. We considered an acceptable upper and lower limit of agreement to be –2 to 2 breaths per minute. We subsequently calculated the intraclass correlation coefficient (ICC) for the 2 methods of measurement. We then repeated this procedure for the gold standard (manual) and standard of care (transthoracic impedance plethysmography) respiratory rates. The Bland-Altman plot method was also used to compare respiratory rate differences found between the P440 and gold standard in participants based on sex and the presence of a cardiopulmonary diagnosis.

**Ethical Considerations**

The study was approved by the University of Massachusetts Chan Medical School Institutional Review Board (protocol number H00016885). All participants provided informed consent. Study data were deidentified. Participants were not remunerated.

**Results**

**Participant Data**

A total of 14 participants were enrolled in the study. The demographics of the study participants are detailed in Table 1. The study population was comprised of 57% (8/14) male candidates with an average age of 58 years.
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (range)</td>
<td>58.64 (23-86)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Female</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1 (7)</td>
</tr>
<tr>
<td>White</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Declined to answer</td>
<td>4 (27)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Hispanic or Latino, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Home medications, n (%)</td>
<td></td>
</tr>
<tr>
<td>Opioid</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Benzodiazepine</td>
<td>3 (21)</td>
</tr>
<tr>
<td>Admitting diagnosis, n (%)</td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal bleed</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Diverticulitis</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Sepsis</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Hypoglycemia</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Pancreatitis</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Cardiac diagnoses&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Pulmonary diagnoses&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2 (14)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Cardiac diagnoses included chest pain, non-ST elevation myocardial infarction, and atrial fibrillation with rapid ventricular response.

<sup>b</sup>Pulmonary diagnoses include pneumonia and pulmonary edema.

Comparison of Respiratory Rate Measurements

The P440 and gold standard Bland-Altman analyses showed a bias of –0.76. The upper limit of agreement was 9.65, and the lower limit of agreement was –11.16 (Figure 3A). At lower respiratory rates, the bias was more positive, indicating a higher measured respiratory rate by the P440 than the gold standard. At higher respiratory rates, the bias was more negative, indicating a lower measured respiratory rate by the P440 than the gold standard. About 34% (30/88) of the differences in measurements were within the prespecified clinically significant limits of agreement of –2 to 2. The P440 and gold standard respiratory rates did not have a statistically significant correlation. The ICC was 0.38 (95% CI 0.06-0.60). The mean absolute error between the P440 and gold standard respiratory rates was an average of 3.89 (range 1.55-8.51).

The standard of care and the gold standard Bland-Altman analysis showed a bias of 0.69. The upper limit of agreement was 8.18, and the lower limit of agreement was –6.81 (Figure 3B). Around 59% (52/88) of the differences in measurements were within the prespecified clinically significant limits of agreement of –2 to 2. The gold standard and standard of care respiratory measurements did have a statistically significant correlation. The ICC was 0.82 (95% CI 0.72-0.88).
Figure 3. Bland-Altman plot for (A) PulsON 440 (P440) and gold standard and (B) gold standard and standard of care.

Subgroup Analyses
Participants grouped by cardiopulmonary chief concern plotted on a Bland-Altman analysis showed a bias of –2.39 (Figure 4). The upper limit of agreement was 7.89, and the lower limit of agreement was –12.67. Participants without a cardiopulmonary chief concern showed a Bland-Altman bias of 1.21 with an upper limit of agreement of 10.49 and a lower limit of agreement of –8.08. While there was no clear trend for the cardiopulmonary chief concerns, the noncardiopulmonary chief concern measurements followed a similar trend to the overall Bland-Altman, with the bias being more positive at lower respiratory rates and more negative at higher respiratory rates. There was not a statistically significant correlation for either the cardiopulmonary or noncardiopulmonary chief concerns. The ICC was 0.21 (–0.41 to 0.56) and 0.42 (–0.09 to 0.69) respectively.

When grouped by sex, the Bland-Altman analysis had a bias of –0.41 (–11.56 to 10.70) for male candidates and a bias of –1.52 (–10.05 to 6.881) for female candidates (Figure 5). The overall trend for male candidates was also positive at lower respiratory rates and negative at higher respiratory rate; however, there was no clear trend for the female candidates. There was no statistically significant correlation for either the male or female candidates. The ICC for male candidates was 0.45 (0.08-0.67) and for female candidates was –0.40 (-2.13-0.47).
Figure 4. Bland-Altman plot of (A) cardiopulmonary and (B) noncardiopulmonary chief complaint.
Discussion

Overview

Although overall there was no statistically significant correlation, the P440 agreed best within the physiologically normal respiratory range. There was less agreement at both increased and decreased respiratory rates as the P440 was underestimated at higher respiratory rates and overestimated at lower respiratory rates. One possible reason for not achieving statistical agreement is that fewer participants had episodes of bradypnea (slow respiratory rate) or tachypnea (high respiratory rate) during the study period and therefore the measurements were more prone to error. Additionally, the ED is a busy environment, and there were many events that could have affected the accurate measurement of respiratory rate, including patient movement, patient clothing or obstruction by blankets, people entering and exiting the room, and the need for bedside procedures. Understanding this trend is important to adjust future respiratory models to be more accurate in the abnormal ranges. However, further validation will be needed before it is able to be used in clinical settings.

There have been other recent studies using an ultrawideband to detect respiratory rates [17,18]. He et al [17] used a combination of a 3D depth camera and ultrawideband radar to use localization with the radar technology to detect respiratory rate in an experimental lab setting. However, the results were limited when movement (such as a person walking) interfered with the respiratory signal. Lauteslager et al [18] also used a ultrawideband radar to detect respiratory rates in a variety of controlled clinical settings. The radar was able to accurately detect respiratory rate; however, the study excluded time periods of irregular respiratory patterns, apnea, and high-motion artifacts. While lab and controlled patient scenarios are important for the initial testing of a monitor, it is also necessary
to understand how the device performs in real-world settings where there is unplanned motion and noise. Additionally, studying the device in a variety of clinical scenarios, such as patients with invasive mechanical ventilation, may help overcome some of the real-world barriers and understand the setting in which the device best performs. By piloting the P440 in the ED, we were able to gather a better understanding of how the device will perform in an uncontrolled setting.

There are several limitations to this study. First, this was a small study. This was designed to be a pilot study, and therefore a convenience sample was obtained during the enrollment period. Future studies will aim to recruit additional participants in order to validate the P440. Additionally, there are difficulties inherent in measuring respiratory rates that may have affected the accuracy of the P440. Gold standard respiratory rate monitoring requires direct observation that may not only affect a participant’s breathing pattern but also necessitate close proximity, which can cause interference with the monitor. Motion artifacts are also known to cause inaccuracies in respiratory monitoring, which is similarly seen in the standard of care telemetry monitoring that is currently used in practice.

Future iterations of the contactless monitoring system will focus on improving the respiratory rate algorithm. To achieve this, we plan to investigate the Eulerian phase magnification of the radargram signal in order to extract meaningful features [19]. The Eulerian phase magnification approach was first introduced for video magnification; however, its use for human motion estimation was limited by privacy concerns. Using the ultrawideband radar with this approach may provide a convenient, nonrestrictive, and unobtrusive means to detect motion, especially in noisy conditions like the ED. Our next step is to build a 1D radar signal magnification pipeline using the Eulerian phase-based magnification algorithm’s complex Gabor wavelet pyramid [20,21]. Different spatial wavelengths of the Gabor pyramid will give us different 1D signals from which we will get FFT peaks and zero-crossing rates as features. Using these informative features, we plan to fit a regression model to better estimate the respiratory rate. The motivation behind using a motion magnification-based algorithm is to leverage the amplification of subtle motions in a dynamic environment that might be difficult to observe by conventional temporal models. Additionally, by using machine learning models, we can get a better respiratory rate estimate by generalizing across different settings as well as diverse demographics.

Overall, the contactless radar respiratory monitor provides a unique solution to monitor patients in places that previously presented challenges, such as low-acuity outpatient settings, waiting rooms, and the home. The ability to monitor changes in respiratory rate accurately and continuously in these settings can help detect clinical deterioration without having direct contact with the person. This has become increasingly important, especially during the COVID-19 pandemic and with the increase in hospital crowding and long ED wait times.

Conclusions
This pilot study has provided important preliminary data that will be used to inform the development of future iterations of the P440 respiratory rate model. Once validated, the device can be miniaturized and used in a variety of settings to provide continuous respiratory monitoring that can be remotely accessed to detect clinical changes in a variety of disease states and improve the care of patients.

Acknowledgments
This work is in part supported by the National Science Foundation (grants SBE 1839999 and grant SCH 2124282) and start-up grant support from the Manning College of Information and Computer Sciences and the Institute for Applied Life Sciences at the University of Massachusetts Amherst. SPC is funded by the National Institutes of Health and the National Institute on Drug Abuse (K23DA045242). We would like to thank Dr Lindsay Walsh for assisting with data collection.

Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


Abbreviations

ED: emergency department
FFT: fast Fourier transform
ICA: independent component analysis
ICC: intraclass correlation coefficient
P440: PulsOn 440
UNIX: Uniplexed Information Computing System
An Intelligent Customer-Driven Digital Solution to Improve Perioperative Health Outcomes Among Children Undergoing Circumcision and Their Parents: Development and Evaluation

Zhi Yin Kwa¹², BSc; Jinqiu Li³⁴, BSc, MSc; Dale Lincoln Loh²⁵, BA, MB, ChB; Yang Yang Lee²⁵, MBBS; Guangyu Liu⁶, BSc; Lixia Zhu⁷, BSc, MSc, PhD; Minna Pikkarainen⁸⁹, PhD; Honggu He²⁴, MD, PhD; Vidyadhar Padmaka Mali²⁵, MBBS, MS

¹Department of Nursing, National University Hospital, Singapore, Singapore
²National University Health System, Singapore, Singapore
³Nursing Department, Zhuhai Campus, Zunyi Medical University, Zhuhai, China
⁴Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
⁵Department of Paediatric Surgery, National University Hospital, Singapore, Singapore
⁶Integrated Health Promotion, Ministry of Health Office for Healthcare Transformation, Singapore, Singapore
⁷Hôpital Chinois de Montréal, Centre Intégré Universitaire de Santé et de Services Sociaux du Centre-Sud-de-l’Île-de-Montréal, Montreal, QC, Canada
⁸Department of Health Technology and Rehabilitation and Department of Product Design, Oslo Metropolitan University, Oslo, Norway
⁹Martti Ahtisaari Institute, Oulu Business School, University of Oulu, Oulu, Finland

Corresponding Author:
Honggu He, MD, PhD
Alice Lee Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Level 2, Clinical Research Centre
MD 11, 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 65167448
Fax: 65 67767135
Email: nurhhg@nus.edu.sg

Abstract

Background: Circumcision as a common elective pediatric surgery worldwide is a stressful and anxiety-inducing experience for parents and children. Although current perioperative interventions proved effective, such as reducing preoperative anxiety, there are limited holistic solutions using mobile apps.

Objective: This paper aims to describe the development and primary evaluation of an intelligent customer-driven smartphone-based app program (ICory-Circumcision) to enhance health outcomes among children undergoing circumcision and their family caregivers.

Methods: Based on the review of the literature and previous studies, Bandura’s self-efficacy theory was adopted as the conceptual framework. A multidisciplinary team was built to identify the content and develop the apps. Semistructured interviews were conducted to evaluate the ICory-Circumcision.

Results: The ICory-Circumcision study was carried out from March 2019 to January 2020 and comprised 2 mobile apps, BuddyCare app and Triumf Health mobile game app. The former provides a day-by-day perioperative guide for parents whose children are undergoing circumcision, while the latter provides emotional support and distraction to children. In total, 6 participants were recruited to use the apps and interviewed to evaluate the program. In total, 4 main categories and 10 subcategories were generated from content analysis.

Conclusions: ICory-Circumcision seemed to lean toward being useful. Revisions to ICory-Circumcision are necessary to enhance its contents and features before advancing to the randomized controlled trial.

Trial Registration: ClinicalTrials.gov NCT04174404; https://clinicaltrials.gov/ct2/show/NCT04174404
Introduction

Background
Male circumcision is a surgery to remove the foreskin of the penis [1]. It is one of the most common day pediatric surgeries worldwide, with an estimated 30% incidence of circumcised males, of which two-thirds were Muslim [2,3]. Singapore’s male circumcision prevalence is about 15% [4]. Male circumcision is commonly performed in Singapore between ages 8 and 11 years often for religious reasons [3]. With Muslims comprising 14.7% of Singapore’s population [5], circumcision is likely common in Singapore.

While global and local trends can be ascribed to mainly religious or cultural reasons [6], it is also expected to rise due to growing evidence of health benefits such as up to 73% protection against acquiring HIV [7,8] and reduced risks of urinary tract infections [4]. Voluntary medical male circumcision may save US $16.5 billion by 2025 from averted HIV treatment and associated costs [8].

Rising preferences for elective male circumcision may also be explained by low complication rates of 0% to 30% in male circumcision [9] and benefits of elective day surgeries such as reduced hospital-acquired infection risks, financial burdens, and disruptions to daily commitments like school [10]. With shifts in male circumcision being done as elective surgeries, parents have to assume heavier parenting roles as their involvement in perioperative care increases [11]. These tasks include managing their child’s preoperative fasting and postoperative wound. Despite the advantages of elective male circumcision, surgeries, even minor ones, are still stressful and anxiety-inducing periods for both parents and children [12,13]. The unfamiliarity of settings and perioperative care and fear of their child’s death are reasons for parental preoperative anxiety [11,14]. In fact, parents of children who undergo day surgery have been found to experience higher parental preoperative anxiety than parents of hospitalized children [15]. This could be due to increased responsibilities and inadequate time to adjust to unfamiliar settings [11]. High parental preoperative anxiety often results in unfavorable somatic symptoms such as insomnia that can hinder parents’ everyday functions and impact work productivity [16]. Parental preoperative anxiety and lack of knowledge can incur unnecessary costs for families and hospitals through unnecessary visits to the emergency department after male circumcision [13,17]. Furthermore, parental preoperative anxiety affects children’s emotional responses and increases children’s preoperative anxiety, as children heavily depend on their parents, especially during foreign events like surgery [11,18,19]. Up to 84% of children undergoing male circumcision had experienced fear or worry, suggesting children’s preoperative anxiety is prevalent in pediatric male circumcision [20]. Children’s preoperative anxiety has been correlated with consequences such as increased postoperative pain, sleep-related problems, and hindered recovery [21-23]. Children’s preoperative anxiety also causes prolonged induction and further use of sedatives and requires additional nursing staff, incurring more costs for families and hospitals [13,24-26].

These combined findings suggested the need for a more comprehensive and effective solution to decrease children’s preoperative anxiety. This study aimed to develop an intelligent customer-driven solution for pediatric surgery care on the improvement of outcomes of parents and their primary school-aged children undergoing circumcision (ICory-Circumcision) and examine the feasibility of the program.

Review of Current Circumcision Clinical Practice in Singapore

Figure 1 shows the current pediatric circumcision routine care in the Singapore health system. At one of Singapore’s tertiary hospitals, about 12% of children who underwent male circumcision reverted to the emergency department before scheduled follow-up appointments [27]. However, only 2% of these children had postoperative problems that warranted medical intervention, while the remaining 10% did not require specialist care and, therefore, were avoidable [27]. As seen, parents’ lack of postoperative knowledge and communication with health care professionals (HCPs) led to what could have been avoidable costs. Additionally, parents in Singapore have expressed the desire for information provision through mobile apps [14]. Current practices of providing surgery-related information for male circumcision are through verbal or written mediums. Technological-based solutions have yet to be incorporated. Therefore, incorporating ICory-Circumcision into pediatric male circumcision settings in Singapore could potentially save resources for families and hospitals.
Review of Literature and Findings From Previous Studies

Parents and children undergoing elective surgeries experience stress and negative emotions [28]. As a result, studies have been conducted to explore their needs. Parents desire emotional support and perioperative information, involvement in their child’s perioperative care, and building good collaboration with HCPs [29-33]. Studies have shown that providing information helped decrease parental preoperative anxiety and encouraged parents’ participation in their children’s care [28,32,34]. Parents wished to know surgery indications, medications and fasting instructions, involvement in the operating theater, and pain and wound management [35-37]. Their strong desire for such information could be attributable to their major roles in assimilating information to their children [38]. Parents also hoped for such information to be individualized and disseminated to them via web-based mobile apps or literature [14,29,36]. They also preferred if postoperative information was given before surgery instead of just before discharge [39].

Children desire preparational information and tend to seek help from parents when they experience pain [40,41]. The majority of the children experience moderate to severe pain during the postoperative period despite their parents’ involvement in care [40]. This could suggest that pain was undertreated [33]. This further highlights the need to provide education on pain management to parents. Children desired parental presence and more distraction techniques to be used by their parents for pain management [41]. Parents also wished to monitor their children’s pain in addition to the strategies [31]. Finally, children hoped for more communication between parents and nurses to assist with postoperative pain [42].

Various technological-based interventions have been developed and aimed at parents of children undergoing elective surgeries. Videos aimed at educating parents about their children’s surgery have been used in several studies. However, the contents of the studies varied rather widely. Chow et al [43] conducted a systematic review and found that videos that included both preoperative and postoperative information were more effective. Two such studies focused on perioperative education; however, only one study showed a decrease in parental preoperative anxiety, while the other showed no significant changes [17,44]. The video contents of 2 studies were about the surgery day [45,46]. Chartrand et al [45] aimed to educate parents about the experience in the recovery room, and it improved parents’ knowledge but not anxiety. Berghmans et al [46] aimed at modeling a hospital tour for parents and children, but no significant changes in parental anxiety. Other studies focused on different surgery periods such as informed consent and postoperative pain management [47,48]. Two studies examined the effects of web-based preparation programs for parents and children undergoing elective surgeries, and both were effective in reducing parental preoperative anxiety [49,50]. Both interventions had elements of surgery-related educational modules for both parents and children. Children’s preoperative anxiety decreased in Fortier and Kain’s study [49] but did not in Wright et al’s study [50]. SMS text messages and mobile apps were also used in several recent studies in pediatric surgery settings [51-55]. Four studies used SMS text messages to convey perioperative education to parents of children undergoing elective surgeries [51-53,56]. These studies allowed real-time communication with HCPs via SMS text message or phone call. The programs were able to decrease parental preoperative anxiety, increase parental knowledge, reduce children’s preoperative anxiety, and improve parent satisfaction, which resulted in neither operation cancellations nor visits to the emergency department. While those 4 studies had no intraoperative texts, Kwan et al [57] examined the effectiveness of sending intraoperative texts, and it was effective in reducing parental anxiety. Ji et al [54], on the other hand, developed an app that uses drawings to explain procedures to parents, which resulted in reduced parental preoperative anxiety and improvement in parental satisfaction. Bailey et al [55] tested the effects of an educational video app on perioperative information and parents’ role in the operating theater.
Several studies have examined the effectiveness of mobile game apps on children’s preoperative anxiety [58-61]. These 4 studies used game apps that were available in app stores and were selected based on age appropriateness. There was a significant reduction in children’s preoperative anxiety after the children played the games in 3 studies. In addition, Cumino et al [58] also showed that a combination of strategies (parental leaflet+mobile game) was more effective in lowering the prevalence of anxiety in the operating room. Marechal et al [61] showed no significant difference in children and parental anxiety. A few studies also used mobile apps to prepare children for surgery [62-64]. All 3 studies aimed to simulate the operating room but through different presentations in the apps: medical clowning video, multimedia app presenting hospital procedures in stages and accompanying videos, and photographs and cartoons. All 3 studies led to a significant decrease in children’s preoperative anxiety. Fernandes et al [63] also showed decreased parental state anxiety.

Our review of the literature showed that perioperative needs of parents and children undergoing elective surgeries have been extensively researched, and as a result, many interventions have been developed to address their needs. However, there is a lack of technological-based interventions targeted at parent’s self-efficacy in children’s perioperative care. There is also a dearth of studies using mobile app–based education for parents, and none were conducted in Singapore.

### Methods

#### Content Development and Theoretical Framework

Taking all the gathered information into consideration, Bandura’s self-efficacy theory and interrelationships between self-efficacy, anxiety, knowledge, and satisfaction were adopted as the theoretical and conceptual framework to guide the development of ICory-Circumcision and methodology of this study (Figure 2).

**Figure 2.** Theoretical framework used for the study.

**Sources of self-efficacy**
- Enactive mastery experience
- Vicarious experience
- Verbal persuasion
- Physiological and affective states

**Intervention**

1. **BuddyCare** app for parents to use as a perioperative guide
2. **Triumf Health** game app which parents can give to their children

**Outcomes**
- Parents’ self-efficacy (primary)
- Parents’ preoperative anxiety
- Parents’ need for information
- Parents’ perioperative knowledge
- Parents’ satisfaction with perioperative care
- Children’s preoperative anxiety
- Children’s postoperative pain
- No shows or delayed shows, health care service use

Bandura posited that self-efficacy is derived from the integration of information from 4 sources of self-efficacy, namely, enactive mastery experience, vicarious experience, verbal persuasion, and emotional and physiological states [65]. Enactive mastery experience refers to parents’ prior experiences in taking care of children throughout the perioperative period, and it is the most influential source of self-efficacy. Vicarious experience was gained by observation and modeling, as it offers parents chances to judge their abilities against a reference point to master tasks [65]. Verbal persuasion refers to persuasive information parents receive from others to enhance parental self-efficacy. Emotional and physiological states influence self-efficacy as a person’s functions are affected [65].

Parental self-efficacy has been shown to negatively correlate with anxiety and child distress and positively correlate to child cooperation [66,67]. High parental preoperative anxiety has been positively correlated with children’s preoperative anxiety, while children’s preoperative anxiety has been positively correlated with higher postoperative pain [68-70]. Additionally, parental preoperative anxiety has been reported to increase the likelihood of surgical cancellations due to lower compliance with fasting instructions [70]. Based on Bandura’s theory, anxious parents could lower parental self-efficacy and subsequently affect children’s perioperative outcomes such as children’s preoperative anxiety and postoperative pain [71]. Parental preoperative anxiety has been shown to be positively correlated to the need for information, thus further reinforcing the need to develop interventions to provide the information parents require [72].

**ICory-Circumcision Components in Relation to Self-Efficacy Theory**

The Template for Intervention Description and Replication (TIDierR) checklist and guide was also recommended to be
used in the process of intervention development [73]. The components of ICory-Circumcision in relation to the self-efficacy theory are depicted in Figure 3.

Figure 3. ICory-Circumcision components in relation to self-efficacy theory.

Qualitative Evaluation of the ICory-Circumcision Program

A self-developed interview guide for field test of BuddyCare and TriumfHealth apps (Multimedia Appendix 1) was used to guide the semistructured interviews to explore the perceptions of the strengths, weaknesses, and the use of ICory-Circumcision from parents, children, and HCPs who used ICory-Circumcision. The qualitative data obtained from process evaluation were analyzed using inductive content analysis [74,75]. The analysis was done in 3 phases: preparation, organizing, and reporting [74], while steps were taken to achieve trustworthiness [76].

Ethical Considerations

Ethics approval (2019/00582) and amendment approval were obtained from the National Health Group Domain Specific Review Board before the commencement of the study. All research team investigators obtained the Collaborative Institutional Training Initiative certificate. Informed consent was obtained from the children’s parents, while assent was taken from the children. All potential participants were given information about the study using the participant information sheet to inform them about the study’s aim, potential benefits, risks, and responsibilities. Voluntary participation, the right to withdraw, and confidentiality were highlighted. Informed consent was not obtained from the HCPs who were interviewed as they were part of the study team. All data from questionnaires were entered electronically into the study hospital’s REDCap (Research Electronic Data Capture; National University Hospital) database, and the data were exported as nonidentifiable data into SPSS (IBM Corp) for data analysis. Only identified study team members with intranet access were able to enter, monitor, and export data. The audio recordings of the interview will be deleted from the audio recorder and stored in the principal investigator’s password-protected computer in the office of Alice Lee Centre for Nursing Studies. All physical records such as consent forms and questionnaires were stored in a locked cupboard at the Department of Pediatric Surgery in National University Hospital. The documents and electronic data will be destroyed after 6 years upon closure of the study by the Domain Specific Review Board. A brand-new SIM card was purchased for the study phone, and it will be disabled and destroyed at the end of the study as well. In addition, no identifiable information was entered in ICory-Circumcision’s apps; instead, pseudonyms and precreated emails were used. This ensured that no participant identifiers were captured by the apps’ companies to protect the participants’ privacy and data confidentiality. No compensation in terms of material or financial benefits was provided to the research participants who participated in this program.

Results

This study was carried out from March 2019 to January 2020 and comprised 2 mobile apps.

BuddyCare Mobile App for Parents

One of the eventual products was the BuddyCare mobile app that provides a comprehensive day-by-day perioperative guide for parents regarding their children’s surgery with an interface to communicate with HCPs. Parents were able to select the surgery date and time on the app, and then, the contents were arranged according to each participant’s timeline. The timeline of BuddyCare contents can be found in Multimedia Appendix 2. Two educational topics on the app were selected in accordance with the parental and children’s needs in the literature review and surgery pathway, one is circumcision-related information, including an overview of...
circumcision, tips on how to explain the surgery to their children, and what to expect about anesthesia; another one is caring for children, including pain management techniques (eg, emotional support, breathing techniques, positive reinforcement, and distraction), preoperative instructions (eg, fasting instructions), and wound management (eg, how to clean and when to bring their child to the emergency department; Figures 4 and 5). Positive quotes are refreshed periodically as emotional support to motivate the parents throughout the perioperative process (Multimedia Appendix 3). With the messaging function, participants are able to communicate with HCPs by sending SMS text messages through the messaging tab (Multimedia Appendix 4). The HCPs in the study team will be able to access the SMS text messages via a BuddyCare dashboard, and they can reply to the participant through this dashboard.

Figure 4. Screenshot of BuddyCare overview.
Triumf Health Mobile Game App for Children

Another product was the Triumf Health mobile game app that provides emotional support and distraction to children. The game allowed the children to customize their own characters and save Triumfland city from a diseased monster by finding one’s inner superpowers. The child was able to control their character to venture around Triumfland and gain points through completing quests in order to help the town doctor to eradicate the disease monster. One important aspect of the game was providing general surgery information to the child (Figure 6). The child could access the topics at any time on their own volition, and the information about each topic was displayed in levels to cater to the child’s reading and comprehension ability. To illustrate, once the child accessed the information in level 1, the information would be presented, and the app would prompt the child to ask if he understood the information. If the child says no, a short summary of the information from level 1 will be presented in short simple sentences. The app also rendered various psychological support to the children such as pain and mood. If the child responded with unfavorable answers such as severe pain or a negative emotion, the game provided appropriate words of encouragement to the child (Multimedia Appendix 5). The abovementioned features of the app made Triumf Health game user experience personalized and dynamic. Further gameplay, that is, accessing the educational module, entertainment games, and other elements of the intervention, was determined by the in-game choices made by the player. Furthermore, the provision of psychological support is dynamically dependent on the patient’s individual progress and in-game progress.
The Qualitative Evaluation on ICory-Circumcision

In total, 6 participants (2 boys who were going to take male circumcision, 2 of their parents, and 2 HCPs) were recruited to use the apps and were required to share their perceptions about the apps. An interview guide was developed and followed (Multimedia Appendix 1). In total, 4 main categories and 10 subcategories were generated from content analysis and presented in Textbox 1.
Textbox 1. Categories and subcategories.

**Strengths of ICory-Circumcision**
- BuddyCare content is useful
  - Comprehensive (n=3) and easy to understand (n=1)
  - Learning experience for parents (n=3)
  - Useful especially for parents with no experience (n=2)
- Mobile apps as useful platform
  - Convenient (n=2) and appropriate for the modern era (n=2)
  - BuddyCare supports routine care (n=4)
- Reasons for liking the Triumf Health game app
  - Follow-up on child’s postoperative status (n=1)
  - Enjoyed the game and its features (n=1)

**Factors for dissatisfaction in ICory-Circumcision**
- Reasons for disliking the Triumf Health game app
  - Boring (n=3) and frustrating (n=1)
  - Children preferred other means of distraction (n=4)
- Communication issues
  - Delayed and unsatisfied response in BuddyCare (n=1)
  - Inconvenience of BuddyCare dashboard (n=1) and difficulty in using (n=1)

**Outcomes of using ICory-Circumcision**
- Opinions of BuddyCare on perioperative outcomes
  - Reduction in parental and child anxiety (n=2)
  - Improved parental confidence in taking care of the child (n=1)
- Opinions of Triumf Health on perioperative outcomes
  - Minimal help in managing preoperative anxiety (n=3)
  - No help with coping with postoperative pain (n=1) versus little help (n=1)

**Suggestions for improvement**
- BuddyCare content suggestions
  - Less words (n=1)
  - Different languages for important information (n=1)
  - More visuals (n=2) versus sufficient visuals (n=1)
- BuddyCare technical aspects
  - Reduce reminders (n=1) versus adequate reminders (n=1)
  - Making a dashboard app (n=1)
- Fidelity of ICory-Circumcision
  - Training for health care professionals (n=2)
  - Intervention delivery suggestions (n=2)
Discussion

Main Findings
The principal aim of the program was to develop an intelligence solution to increase parental self-efficacy and decrease parental and children’s preoperative anxiety. Parents generally expressed positive reactions toward the BuddyCare app. They found BuddyCare to be comprehensive, convenient, and useful, and they would highly recommend it to other parents. Triumph Health app was useful in follow-up postoperative pain and emotional care for children. These findings align with the aims of ICory-Circumcision and the HCPs’ views. Participants also found ICory-Circumcision to be a good resource that complements routine care, which is similar to another study [49].

Feedback on Triumph Health and BuddyCare should be taken and revise ICory-Circumcision as an intervention. Based on the mixed reactions from the qualitative interview, it may suggest that ICory-Circumcision may not be individualized enough for participants. For example, more visuals such as videos could be added into BuddyCare, but they could be placed in a different tab, which allows parents the liberty to access that section or not. This is to cater to the different levels of comfort each parent has with seeing pictures of open wounds. For Triumph Health, the number of words could be reduced, and the mechanics of the game could be reviewed with the team in Finland to see if it could be better improved to suit the needs of the children in Singapore.

Lack of knowledge of pain management strategies and wound management techniques could affect the development of parental self-efficacy and increase negative emotions [29,72]. Studies showed that providing information about their children’s surgery to parents could reduce parental preoperative anxiety and showed an increase in parental self-efficacy [28,55,77]. Past experiences could have contributed to the high parental self-efficacy. Bandura [65] suggested that mastery experiences have the strongest influence on self-efficacy out of the 4 sources, and if caregivers had previous caregiving experience, they had high parental self-efficacy. On the contrary, parents would have higher anxiety when they have the first surgical experience due to medical reasons, which could impede the self-efficacy gained from physiological and affective states [65]. Therefore, providing adequate knowledge to parents is an efficient way to improve health-related outcomes.

Mobile apps are ubiquitous among parents and children, possibly due to the convenience brought by their easy accessibility [78-80]. The infiltration of mobile apps into pediatric settings is clear with the advent of mobile apps aimed at helping children with different health conditions [81,82]. Therefore, the number of mobile resources HCPs have access to has greatly expanded, improving efficiency and productivity [83]. Evidently, mobile apps have tremendous potential as a platform for information delivery.

As there are limited interventional studies presenting the development process, this study will contribute to the body of literature about intervention development [84]. This informs readers about the possible challenges that one can encounter should they decide to embark on similar intervention development [73]. This study also provided insights into the feasibility of ICory-Circumcision and the study’s methodology, which could improve the main trial’s processes and prove the effects of ICory-Circumcision. If the effects are then proved, it could potentially save nurses’ time, as nurses are heavily involved in providing education to parents and children about surgery [17,34]. Although our qualitative evaluation of ICory-Circumcision involved various users, including children, parents, and HCPs, the sample size was small due to the limited time for an honors student’s project and the COVID-19 pandemic occurrence in November 2019.

Conclusions
This paper detailed the development of a holistic technology–based intervention for parents and their children undergoing elective circumcision and examined its preliminary feasibility and evaluation. The qualitative evaluation identified strengths, weaknesses, and suggestions for improvement concerning ICory-Circumcision, suggesting its potential usefulness for parents and children in perioperative outcomes. Prior to proceeding with the randomized controlled trial, revisions to ICory-Circumcision to enhance its contents and features are recommended.

Acknowledgments
The authors wish to thank the clinicians, administrators, and those who have directly and indirectly contributed their expertise or public opinion to the development of this program. This study would not have been possible without the support of the National University of Singapore, Singapore General Hospital, National University Hospital, and Buddy Healthcare. This research work was funded by the Business Finland (grant 203/31/2018).

Conflicts of Interest
None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 17 KB - formative_v8i1e52337_app1.docx ]

https://formative.jmir.org/2024/1/e52337
Multimedia Appendix 2
Timeline of BuddyCare contents.
[DOCX File, 23 KB - formative_v8i1e52337_app2.docx]

Multimedia Appendix 3
Screenshot of BuddyCare positive quotes.
[PDF File (Adobe PDF File), 48 KB - formative_v8i1e52337_app3.pdf]

Multimedia Appendix 4
Screenshot of BuddyCare message function.
[PDF File (Adobe PDF File), 24 KB - formative_v8i1e52337_app4.pdf]

Multimedia Appendix 5
Screenshot of Triumph Health mood evaluation.
[PDF File (Adobe PDF File), 18 KB - formative_v8i1e52337_app5.pdf]

References


Abbreviations

HCP: health care professional

REDCap: Research Electronic Data Capture

TIDierR: Template for Intervention Description and Replication
Kwa et al

©Zhi Yin Kwa, Jinqiu Li, Dale Lincoln Loh, Yang Yang Lee, Guangyu Liu, Lixia Zhu, Minna Pikkarainen, Honggu He, Vidyadhar Padmakar Mali. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

Liang Zhang¹, MBBS; Yueqing Huang¹, PhD; Min Huang¹, PhD; Chun-Hua Zhao¹, MSc; Yan-Jun Zhang², MSc; Yi Wang³, MBBS

¹Department of General Practice, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, China
²School of Information Science and Engineering, Southeast University, Nanjing, China
³The First Clinical Medical College, Nanjing Medical University, Nanjing, China

Corresponding Author:
Yueqing Huang, PhD
Department of General Practice
The Affiliated Suzhou Hospital of Nanjing Medical University
16 Baitaxi Road
Gusu District
Suzhou, 215000
China
Phone: 86 13812757566
Email: huangyq_sz@163.com

Abstract

Background: The increasing prevalence of nonalcoholic fatty liver disease (NAFLD) in China presents a significant public health concern. Traditional ultrasound, commonly used for fatty liver screening, often lacks the ability to accurately quantify steatosis, leading to insufficient follow-up for patients with moderate-to-severe steatosis. Transient elastography (TE) provides a more quantitative diagnosis of steatosis and fibrosis, closely aligning with biopsy results. Moreover, machine learning (ML) technology holds promise for developing more precise diagnostic models for NAFLD using a variety of laboratory indicators.

Objective: This study aims to develop a novel ML-based diagnostic model leveraging TE results for staging hepatic steatosis. The objective was to streamline the model’s input features, creating a cost-effective and user-friendly tool to distinguish patients with NAFLD requiring follow-up. This innovative approach merges TE and ML to enhance diagnostic accuracy and efficiency in NAFLD assessment.

Methods: The study involved a comprehensive analysis of health examination records from Suzhou Municipal Hospital, spanning from March to May 2023. Patient data and questionnaire responses were meticulously inputted into Microsoft Excel 2019, followed by thorough data cleaning and model development using Python 3.7, with libraries scikit-learn and numpy to ensure data accuracy. A cohort comprising 978 residents with complete medical records and TE results was included for analysis. Various classification models, including logistic regression (LR), k-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), light gradient boosting machine (LightGBM), and extreme gradient boosting (XGBoost), were constructed and evaluated based on the area under the receiver operating characteristic curve (AUROC).

Results: Among the 916 patients included in the study, 273 were diagnosed with moderate-to-severe NAFLD. The concordance rate between traditional ultrasound and TE for detecting moderate-to-severe NAFLD was 84.6% (231/273). The AUROC values for the RF, LightGBM, XGBoost, SVM, KNN, and LR models were 0.91, 0.86, 0.83, 0.88, 0.77, and 0.81, respectively. Notably, the RF model exhibited the best performance. A simplified RF model was developed with an AUROC of 0.88, featuring 62% sensitivity and 90% specificity. This simplified model used 6 key features: waist circumference, BMI, fasting plasma glucose, uric acid, total bilirubin, and high-sensitivity C-reactive protein. This approach offers a cost-effective and user-friendly tool while streamlining feature acquisition for training purposes.

Conclusions: The study introduces a groundbreaking, cost-effective ML algorithm that leverages health examination data for identifying moderate-to-severe NAFLD. This model has the potential to significantly impact public health by enabling targeted...
investigations and interventions for NAFLD. By integrating TE and ML technologies, the study showcases innovative approaches to advancing NAFLD diagnostics.

(JMIR Form Res 2024;8:e53654) doi: 10.2196/53654

KEYWORDS

NAFLD; artificial intelligence; public health; transient elastography; diagnosis

Introduction

Nonalcoholic fatty liver disease (NAFLD) stands as the foremost chronic liver condition, impacting approximately 1.7 billion individuals worldwide. NAFLD manifests as a multifaceted chronic liver disorder marked by an excessive buildup of fat within liver tissue. As hepatic steatosis advances, it exacerbates the onset and progression of hepatitis and fibrosis, while heightening the likelihood of liver cancer. Furthermore, numerous studies provide compelling evidence linking NAFLD closely with metabolic disorders such as obesity, diabetes, and hypertension, significantly elevating the risk of cardiovascular disease [1]. A recent meta-analysis revealed that the prevalence of NAFLD in Asian countries mirrors that of Western nations. Notably, China exhibits the highest prevalence, incidence, and yearly mortality associated with NAFLD in Asia. If this trend persists, projections suggest that by 2030, the total NAFLD population in China will soar to 314.58 million. Consequently, China will emerge as the global leader in both patients with NAFLD and liver-related fatalities [2]. Given that the majority of NAFLD cases are asymptomatic, early diagnosis holds significant clinical importance. Furthermore, precise quantification of liver fat content and clarification of hepatic steatosis severity are crucial for determining appropriate clinical interventions, assessing disease progression, and evaluating treatment efficacy. Hepatic steatosis is typically categorized into minimal (<5%), mild (5%-33%), moderate (33%-66%), and severe (>66%) levels [3]. Patients exhibiting moderate-to-severe steatosis necessitate more intensive intervention and follow-up. Enhanced detection of individuals at high risk and early diagnosis can substantially aid in the diagnosis, treatment, and prevention of NAFLD.

Ultrasound is widely acknowledged as the preferred method for screening hepatic steatosis due to its cost-effectiveness, safety, convenience, and efficacy [4]. The characteristic ultrasound findings of fatty liver are either homogeneous or heterogeneous enhancement of liver echogenicity, along with liver enlargement and diminished visualization of intrahepatic ductal structures. However, the accuracy of ultrasound in diagnosing the disease is heavily reliant on the skill and expertise of the operator [5]. Studies have consistently found that ultrasound is highly operator dependent and lacks the capability to precisely determine the extent of hepatic steatosis or distinguish between steatosis and fibrosis, as both conditions lead to heightened liver echogenicity. Consequently, ultrasound is predominantly used for screening fatty liver disease (FLD) [6]. However, it is not recommended for tasks requiring accurate diagnosis and severity grading of early-stage fatty liver, liver transplantation assessments, or evaluation of short-term drug therapies.

At present, liver biopsy stands as the gold standard for diagnosing FLD and evaluating the severity of hepatic steatosis. However, its utility for dynamic monitoring of disease progression and efficacy assessment is limited by factors such as poor patient acceptance and high cost. Additionally, liver biopsy entails inherent risks of complications including invasiveness, bleeding, and infection. Moreover, it is prone to subjective evaluation bias and sampling errors, further impeding its effectiveness as a monitoring tool [7]. Currently, the noninvasive preliminary assessment of hepatic steatosis and the quantitative dynamic evaluation of hepatic fat content represent focal points in current research efforts. Techniques such as transient elastography (TE), computed tomography, and magnetic resonance imaging (MRI) have all been validated for the quantitative diagnosis of steatosis, with MRI demonstrating superior accuracy [6]. However, the high cost, poor patient acceptance, and lengthy examination times associated with biopsy, MRI, and computed tomography render them impractical for large-scale population screening. Consequently, obtaining sufficient sample data in outpatient services worldwide remains challenging. Controlled attenuation parameters in TE represent a quantitative diagnostic approach tailored for detecting steatosis graded as S1, S2, and S3, as well as fibrosis graded as F1, F2, F3, and F4. A meta-analysis has determined that controlled attenuation parameters exhibit good sensitivity and specificity for grading steatosis [7]. Additionally, a prospective study that used TE to assess disease progression in patients with NAFLD indicated that liver stiffness measurements (LSMs) can effectively monitor the degree of liver fibrosis in this patient population [8]. The study indicated that TE can serve as a comprehensive diagnostic tool for both hepatic steatosis and liver fibrosis. It also offers a rapid and noninvasive method to assess liver fibrosis in patients with diverse chronic liver diseases, encompassing chronic hepatitis C, chronic hepatitis B, and NAFLD. Moreover, TE shows promise in predicting complications associated with advanced compensated chronic liver disease [4,9]. However, the widespread adoption of TE in population-based health screenings faces obstacles, particularly in China, where TE is primarily used for evaluating patients with FLD in general hospital settings. There are persistent challenges concerning inadequate ultrasound equipment and insufficient specialized training for physicians in primary health services that require attention. Consequently, there is a pressing need for the development of more cost-effective and efficient methods to identify individuals in the population with FLD who warrant intervention and follow-up, particularly those reaching the threshold of needing medical attention (S≥S2).

Machine learning (ML) offers a promising avenue to tackle these challenges. ML, a branch of computer science, uses algorithms to discern patterns within extensive data sets and predict diverse outcomes [10]. Evolving from pattern recognition
and computational learning, ML uses computers to analyze interactions between variables, encompassing both nonlinear and complex relationships, while minimizing errors between predicted and actual outcomes. ML not only enhances predictive accuracy but also has the capacity to identify latent variables that might not be directly observable but can be inferred from other variables. Currently, various ML techniques, including logistic regression (LR), random forest (RF), artificial neural networks (ANNs), k-nearest neighbors (KNNs), support vector machine (SVM), and extreme gradient boosting (XGBoost), are being used in disease prediction with significantly higher accuracy compared with classical methods [11]. Previous studies have used clinical data to diagnose patients with NAFLD, often relying on traditional ultrasound results for FLD diagnosis [12,13]. Given the challenges posed by the unsatisfactory diagnostic accuracy of traditional ultrasound in FLD screening and its inability to provide early warnings for patients requiring follow-up and more stringent interventions, there is an opportunity to leverage data from TE in population-based health examinations. These data can be used to develop a new ML model with enhanced accuracy and the capability to classify patients based on severity thresholds.

Methods

Recruitment

All clinical data for our study were sourced from health examinations conducted at 7 health examination centers across Suzhou, encompassing 3 districts of Suzhou Municipal Hospital and its 4 affiliated community hospitals. The study included individuals who underwent health examinations from March to May 2023, with exclusion criteria applied to those lacking TE test results. Among the 1753 patients who underwent TE screening, 1344 were selected during their health examination. Ultimately, a total of 978 health examination records with complete medical files were included, accessible for querying in the case system.

Ethical Considerations

All participants who agreed to partake in the annual health examination were required to complete an informed consent form. Physical examination data were collected for the Suzhou Municipal Government and Suzhou Municipal Hospital. The authors take full accountability for all aspects of the work, ensuring that any questions regarding the accuracy or integrity of the study are thoroughly investigated and resolved. All procedures adhered to the ethical standards outlined in the Helsinki Declaration and received approval from the Ethical Committee of Suzhou Municipal Hospital. The study was approved by the Ethics Committee of Suzhou Municipal Hospital (ethical approval number K-2022-034-K01).

Machine and Operational Standard

The machine used in the health examination was the FibroTouch, specifically the Transient Elastography FibroTouch (FibroTouch-FT5000; Wuxi HISKY Medical Technologies). This device assesses the degree of hepatic fibrosis by measuring LSM through vibration-controlled instantaneous elastography. Hepatic steatosis is quantitatively evaluated by measuring the attenuation of ultrasound signals in the liver, known as the ultrasound attenuation parameter (UAP). To address detection errors in patients with obesity, the FibroTouch automatically adjusts the probe based on the thickness of subcutaneous fat following precise positioning and depth measurement. This adjustment ensures comparable diagnostic accuracy to FibroScan [14].

FibroTouch measurements were conducted by experienced and certified physicians, each having performed over 500 examinations. Following the manufacturer’s instructions, patients assumed a supine position with the right hand placed behind the head to facilitate the expansion of the intercostal space. An image-guided probe was carefully chosen to scan the region between the seventh and ninth intercostal spaces, avoiding cysts and blood vessels in the liver. The probe was maintained in a vertical position relative to the skin surface, with pressure applied within the appropriate range (Figure 1). Detection commenced once the M waveform intensity was uniformly distributed and the A waveform appeared linear. In this study, the representative measurement of FibroTouch was determined by calculating the median value of the 10 acceptable LSMS in kilopascals (kPa) and UAPs in decibels per meter (dB/m), along with their respective IQRs. LSM and UAP measurements were deemed reliable only if 10 successful measurements were obtained, with an IQR-to-median ratio of 30% and a success rate of at least 60% [15].
Figure 1. A simplified schematic diagram of TE testing showing an image-guided probe selected to detect the region through the intercostal space. ROI: region of interest; TE: transient elastography.

Statistical Analysis

When collecting data, we initially searched for the patient’s medical number to ensure that the accessed data did not contain any identifiable patient information. To enhance data quality and mitigate the impact of erroneous data on the model, we developed a comprehensive set of logic algorithms to systematically check for logical errors within the health examination data records. These checks included identifying unit errors, magnitude errors, format errors, and so forth. Any identified errors were then manually corrected following prompts from the algorithm. Additionally, to detect potential hidden errors arising from data entry issues, we generated scatterplots (Figure 2) to identify outliers, which were subsequently monitored and reviewed manually on a case-by-case basis. A professional medical doctor, with over 15 years of experience in the field, assessed the reasonableness of outliers. It is notable that some outliers exhibited values that were theoretically unlikely, such as low-density lipoprotein-cholesterol (LDL-C) levels exceeding 300 mmol/L. We conducted thorough verification of these unreasonable values and subsequently corrected or excluded data entries found to be erroneous due to researchers’ data entry mistakes or other factors.

Figure 2. Scatterplots for outlier detection. HDL-C: high-density lipoprotein-cholesterol; LDL-C: low-density lipoprotein-cholesterol; TC: total cholesterol.

In the lifestyle characteristics section, dietary information was omitted from the model. This decision was based on the understanding that many older individuals were unable to accurately report their daily food intake, while younger individuals often experienced irregular eating habits. Additionally, in terms of prevalence, conditions with low case numbers such as chronic obstructive pulmonary disease (n=5), myocardial infarction (n=28), and stroke (n=23) were excluded from the model. Osteoporosis was also excluded because bone mineral density assessments were not included in the health examination protocol, making it difficult to ascertain the presence of osteoporosis in the majority of patients undergoing...
health examinations. Furthermore, 56 patients were excluded due to data disorder errors. Specifically, variables such as red blood cell count, neutrophil count, white blood cell count, and lymphocyte count were mistakenly entered as percentages of red blood cells, neutrophils, white blood cells, and lymphocytes, respectively.

**Characteristic Processing**

**Exercise**

The PARs-3 (Physical Activity Rating Scale-3) is a commonly used exercise measurement scale in China [16]. In previous studies, the scale’s internal consistency and reliability were 0.86 and 0.82, respectively. The scale contains 3 dimensions, namely, time, intensity, and frequency of exercise. A score of 20 or higher on this scale is indicative of moderate physical activity, which equates to engaging in at least 150 minutes of moderate exercise per week. To simplify the analysis, moderate physical activity was categorized into 2 groups: individuals who participate in at least 150 minutes of moderate exercise per week versus those who engage in less than 150 minutes weekly.

**Smoking and Drinking**

The health examination records included information on the variety of wine, volume of drinking, and alcohol percentage consumed per day. We calculated alcohol intake using the following equation: alcohol intake (g) = volume of drinking (mL) × alcohol percentage (% vol/vol) × 0.8 (g/mL). Patients with alcohol intake above the recommended limits (male ≥30 g/day; female ≥20 g/day) were excluded. However, individuals with an alcohol intake of 0 were included in the analysis. We categorized alcohol intake into 2 groups: those with an alcohol intake of 0 and those with an intake greater than 0, treating them as 2-categorical variables. Additionally, we included hypertension, hyperuricemia, diabetes, and hyperlipemia as 2-categorical variables. All remaining data were normalized for analysis.

**Educational Attainment and Financial Situation**

Individuals who have never received formal education are classified as “uneducated.” Those with less than a high school education are categorized as “low,” whereas individuals with a high school diploma or specialized education are labeled as “mid.” Education at the college level or higher is defined as “high.” Regarding income, the medical report divides income into categories including below the social minimum wage, slightly below the average social wage, slightly above the average social wage, and significantly above the average wage. Individuals earning below the social minimum wage were categorized as “poor,” whereas those earning significantly above the average wage were classified as “rich.” Those with incomes falling between these extremes were categorized as “average.”

**Enrollment of Participants**

We screened individuals who ultimately met the diagnostic criteria for NAFLD based on the American Association for the Study of Liver Diseases (AASLD) Practice Guidance [17]. Cases with conditions known to substantially impact TE results, such as liver cancer and ascites, were excluded from the study. Similarly, individuals with conditions known to affect blood biochemistry analysis, such as infections and long-term glucocorticoid use, were excluded. Figure 3 presents a flowchart illustrating the enrollment process of patients.

---

**Metabolic Disease**

In contrast to previous studies where diastolic and systolic blood pressures were often analyzed as features, we chose not to include them. We believed that the blood pressure measurements of patients undergoing health examinations could be biased due to various factors such as changes in peak blood pressure, medication usage, and clinical hypertension. Similarly, random blood glucose levels are strongly influenced by diet. Instead, we opted to use fasting plasma glucose (FPG), postprandial plasma glucose (PPG), and hemoglobin A1c (HbA1c) as variables. Additionally, we included hypertension, hyperuricemia, diabetes, and hyperlipemia as 2-categorical variables. All remaining data were normalized for analysis.
Figure 3. Flowchart of patient enrollment and diagnostic standards of NAFLD from the AASLD Practice Guidance. AASLD: American Association for the Study of Liver Diseases; NAFLD: nonalcoholic fatty liver disease; TE: transient elastography.

All the collated data of the 916 people screened are summarized in Tables 1 and 2. Table 1 summarizes the descriptive characteristics of the study population, including gender, age, educational attainment, financial situation, residence, smoking and drinking state, exercise, and metabolic disease (hypertension, hyperuricemia, diabetes, and hyperlipemia). Table 2 presents a summary of blood biochemistry analysis of the study population.
Table 1. Summary of descriptive characteristics of the study population (N=978).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>503 (51.4)</td>
</tr>
<tr>
<td>Female</td>
<td>475 (48.6)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>20-60</td>
<td>498 (50.9)</td>
</tr>
<tr>
<td>≥60</td>
<td>480 (49.1)</td>
</tr>
<tr>
<td><strong>Finance</strong></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>268 (27.4)</td>
</tr>
<tr>
<td>Average</td>
<td>666 (68.1)</td>
</tr>
<tr>
<td>Rich</td>
<td>44 (4.5)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Uneducated</td>
<td>42 (4.3)</td>
</tr>
<tr>
<td>Low</td>
<td>262 (26.8)</td>
</tr>
<tr>
<td>Mid</td>
<td>478 (48.9)</td>
</tr>
<tr>
<td>High</td>
<td>196 (20.0)</td>
</tr>
<tr>
<td><strong>Residence</strong></td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>911 (93.1)</td>
</tr>
<tr>
<td>Village</td>
<td>67 (6.9)</td>
</tr>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>19 (1.9)</td>
</tr>
<tr>
<td>18.5-23.9</td>
<td>370 (37.8)</td>
</tr>
<tr>
<td>24-27.9</td>
<td>430 (44.0)</td>
</tr>
<tr>
<td>≥28</td>
<td>159 (16.3)</td>
</tr>
<tr>
<td><strong>Waist circumference (cm)</strong></td>
<td></td>
</tr>
<tr>
<td>Male (n=503)</td>
<td></td>
</tr>
<tr>
<td>≥85</td>
<td>338 (67.2)</td>
</tr>
<tr>
<td>&lt;85</td>
<td>165 (32.8)</td>
</tr>
<tr>
<td>Female (n=475)</td>
<td></td>
</tr>
<tr>
<td>≥80</td>
<td>299 (62.9)</td>
</tr>
<tr>
<td>&lt;80</td>
<td>176 (37.1)</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
</tr>
<tr>
<td>Ever</td>
<td>151 (15.4)</td>
</tr>
<tr>
<td>Never</td>
<td>827 (84.6)</td>
</tr>
<tr>
<td><strong>Alcohol (g/day)</strong></td>
<td></td>
</tr>
<tr>
<td>Male (n=503)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>343 (68.2)</td>
</tr>
<tr>
<td>0-30</td>
<td>120 (23.9)</td>
</tr>
<tr>
<td>≥30</td>
<td>40 (8.0)</td>
</tr>
<tr>
<td>Female (n=475)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>453 (95.4)</td>
</tr>
<tr>
<td>0-20</td>
<td>20 (4.2)</td>
</tr>
<tr>
<td>≥20</td>
<td>2 (0.4)</td>
</tr>
</tbody>
</table>
Table 2. Summary of blood biochemistry analysis of the study population (N=978).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting plasma glucose (mmol/L)</td>
<td>6.1 (1.6)</td>
</tr>
<tr>
<td>Postprandial plasma glucose (mmol/L)</td>
<td>8.9 (3.2)</td>
</tr>
<tr>
<td>Hemoglobin A1c (%)</td>
<td>5.9 (1.2)</td>
</tr>
<tr>
<td>Total bilirubin (μmol/L)</td>
<td>13.2 (7.4)</td>
</tr>
<tr>
<td>Uric acid (μmol/L)</td>
<td>330.8 (79.1)</td>
</tr>
<tr>
<td>Triglyceride (mmol/L)</td>
<td>1.8 (1.4)</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L)</td>
<td>4.8 (1.2)</td>
</tr>
<tr>
<td>High-density lipoprotein-cholesterol (mmol/L)</td>
<td>1.3 (0.5)</td>
</tr>
<tr>
<td>Low-density lipoprotein-cholesterol (mmol/L)</td>
<td>2.9 (0.9)</td>
</tr>
<tr>
<td>Alanine transaminase (U/L)</td>
<td>27.9 (16.0)</td>
</tr>
<tr>
<td>Aspartate transaminase (U/L)</td>
<td>26.4 (14.5)</td>
</tr>
<tr>
<td>γ-Glutamyl transpeptidase (U/L)</td>
<td>36.2 (23.7)</td>
</tr>
<tr>
<td>High-sensitivity C-reactive protein (mg/L)</td>
<td>2.2 (3.6)</td>
</tr>
<tr>
<td>Platelets (x10^9/L)</td>
<td>211.8 (54.8)</td>
</tr>
<tr>
<td>Hemoglobin (g/L)</td>
<td>136.9 (26.5)</td>
</tr>
<tr>
<td>Urea (mmol/L)</td>
<td>8.5 (31.4)</td>
</tr>
<tr>
<td>Creatinine (μmol/L)</td>
<td>70.5 (18.5)</td>
</tr>
</tbody>
</table>

**Model Building**

To predict NAFLD with different severities, we used several classification models, including LR, KNN, SVM, RF, light gradient boosting machine (LightGBM), and XGBoost.

RF is an ensemble classification algorithm developed by Leo Breiman and Adele Cutler in 1999. It operates by constructing a multitude of decision trees during the training phase and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. RF is widely used for classification, regression, and other tasks. Each decision tree in the RF is built independently by applying the general technique of bootstrap aggregating (bagging), where random samples are selected for the training set. RF determines the final result by aggregating the predictions of all individual trees through a simple majority vote. It has demonstrated high accuracy across various fields, including medical diagnosis. Additionally, RF is frequently used for feature selection in data science workflows. One of the reasons for its popularity in feature selection is due to the tree-based strategies used by RF. These strategies naturally rank features based on how effectively they improve the purity of the nodes in the decision trees. Features that result in the greatest decrease in impurity are typically encountered at the beginning of the trees, while those with the least decrease in impurity are found toward the end of the trees [11]. By selectively pruning trees below a certain node, a subset of the most important features can be derived.
LR is a type of discrete choice model that falls under multivariate analysis. It is extensively used in various fields such as sociology, biostatistics, clinical medicine, quantitative psychology, econometrics, and marketing. LR is often used for empirical analysis and is commonly used for comparison with ML studies [18]. This method offers several advantages, including high power and accuracy, making it a popular choice for modeling binary or categorical outcomes.

ANNs, a family of statistical learning algorithms, draw inspiration from biological neural networks. ANNs have demonstrated remarkable power in nonlinear modeling and have been proven for accurate predictions in many fields, including clinical decision support [19]. The operation of an ANN is akin to a biological neuron, where signals are received through dendrites. In ANNs, this process is replicated with an input layer that feeds into several hidden layers, ultimately leading to an output layer. Each layer consists of numerous perceptrons interconnected by adjustable weights. During training, the ANN iteratively adjusts these weights using a data set, aligning inputs with their desired outputs. This iterative learning process allows the ANN to refine its predictive capabilities over time.

The KNN classification algorithm is among the simplest methods in data mining classification techniques. KNN operates by searching the pattern space for k-training tuples that are nearest to the unknown tuple being classified. These tuples collectively form the KNN classifier for the unknown tuple. The concept of “nearest” is determined by a distance metric, such as the Euclidean distance, which measures the proximity between data points. One potential limitation of KNN classifiers is that they assign equal weight to all attributes based on distance, regardless of their relevance. Consequently, KNN classifiers may suffer from poor accuracy when confronted with noise or irrelevant attributes in the data.

XGBoost is a significantly enhanced implementation of the gradient-boosting supervised ML technique, known for its speed and performance. It shares similarities with RFs but uses a more regularized model formulation to control overfitting. XGBoost operates as a tree ensemble model, which involves the summation of predictions derived from a specific set of classification and regression trees. This regularization technique helps improve the overall performance of the model by mitigating overfitting issues. XGBoost is versatile and can be applied to both classification and regression tasks.

LightGBM is a gradient-boosting framework that uses decision trees as the base learner, similar to XGBoost. However, LightGBM is optimized for efficiency and performance, offering several advantages, including faster training speed and lower memory usage. Additionally, LightGBM supports single-computer multithreading, multicore parallel computing, and graphics processing unit training, and has the ability to handle large-scale data.

After cleaning the data, we constructed models (RF and LightGBM) to eliminate irrelevant features such as gender, urea, creatinine, and total cholesterol. The final model incorporated the following features: age, education, finance, alcohol intake, smoking, hypertension, hyperuricemia, diabetes, hyperlipemia, BMI, waist circumference, HbA1c, FPG, PPG, total bilirubin (TBil), uric acid (UA), triglyceride (TG), high-density lipoprotein-cholesterol (HDL-C), LDL-C, alanine transaminase (ALT), aspartate transaminase (AST), \( \gamma \)-glutamyl transpeptidase (\( \gamma \)-GT), high-sensitivity C-reactive protein (hs-CRP), platelets, and hemoglobin (HGB).

**Analysis Tools**

The basic patient information and paper questionnaire responses were manually entered by a researcher using Microsoft Excel 2019 as the information entry software. Data cleaning, model construction, and area under the curve chart output were performed using Python 3.7 (Python Foundation), with the packages scikit-learn and numpy. The editor used for this purpose was PyCharm (JetBrains). The flowchart was drawn using MyDraw (Nevron Software).

**Results**

**Overview of Data Comparison**

First, we compared the diagnostic rates of traditional ultrasound with TE and found that traditional ultrasound achieved a high diagnostic rate of 84.6% (231/273) in patients with TE-rated moderate-to-severe steatosis (S\( \geq S2 \)), which is consistent with previous reports on the accuracy of ultrasound diagnosis [20]. The comparison of hepatic steatosis stages produced by TE and traditional ultrasound results is shown in Table 3.
Table 3. Comparison of TE\textsuperscript{a} and traditional ultrasound results.

<table>
<thead>
<tr>
<th>TE</th>
<th>Ultrasound</th>
<th></th>
<th>Total, n (n=916)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diagnosed, n (n=502)</td>
<td>Undiagnosed, n (n=414)</td>
<td></td>
</tr>
<tr>
<td>S\textsuperscript{b}&lt;S1\textsuperscript{c}</td>
<td>81</td>
<td>305</td>
<td>386</td>
</tr>
<tr>
<td>S1\leq S\textsuperscript{d} &lt; S2 \textsuperscript{d}</td>
<td>190</td>
<td>67</td>
<td>257</td>
</tr>
<tr>
<td>S2\textsuperscript{d} \leq S3 \textsuperscript{e}</td>
<td>133</td>
<td>33</td>
<td>166</td>
</tr>
<tr>
<td>S\geq S3</td>
<td>98</td>
<td>9</td>
<td>107</td>
</tr>
</tbody>
</table>

\textsuperscript{a}TE: transient elastography.
\textsuperscript{b}S: hepatic steatosis stage.
\textsuperscript{c}S1: mild steatosis.
\textsuperscript{d}S2: moderate steatosis.
\textsuperscript{e}S3: severe steatosis.

Model Performance Comparison

Finally, the model incorporated the following features: age, education, finance, alcohol intake, smoking, hypertension, hyperuricemia, diabetes, hyperlipemia, BMI, waist circumference, HbA\textsubscript{1c}, FPG, PPG, TBil, UA, TG, HDL-C, LDL-C, ALT, AST, γ-GT, hs-CRP, platelets, and HGB. The features incorporated into the final model are presented in Table 4. We compared the study population with moderate-to-severe steatosis with those without it.
Table 4. Features that were incorporated into the final model and significantly contribute to moderate-to-severe steatosis.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>S≥S2</th>
<th>S&lt;S2</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (≥60 years), n (%)</td>
<td>165/273 (60.4)</td>
<td>291/643 (45.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Education (high), n (%)</td>
<td>25/273 (9.2)</td>
<td>163/643 (25.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Physical activity (moderate), n(%)</td>
<td>45/273 (16.5)</td>
<td>87/643 (13.5)</td>
<td>.25</td>
</tr>
<tr>
<td>Alcohol (male; 0), n (%)</td>
<td>117/137 (85.4)</td>
<td>224/306 (73.2)</td>
<td>.007</td>
</tr>
<tr>
<td>Smoking (male; ever), n (%)</td>
<td>42/137 (30.7)</td>
<td>71/306 (23.2)</td>
<td>.12</td>
</tr>
<tr>
<td>Finance (rich), n (%)</td>
<td>7/273 (2.6)</td>
<td>31/643 (4.8)</td>
<td>.12</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>138/273 (50.5)</td>
<td>259/643 (40.3)</td>
<td>.004</td>
</tr>
<tr>
<td>Hyperuricemia, n (%)</td>
<td>20/273 (7.3)</td>
<td>29/643 (4.5)</td>
<td>.08</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td>88/273 (32.2)</td>
<td>154/643 (24.0)</td>
<td>.009</td>
</tr>
<tr>
<td>Hyperlipemia, n (%)</td>
<td>53/273 (19.4)</td>
<td>83/643 (12.9)</td>
<td>.01</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>26.0 (2.8)</td>
<td>24.1 (3.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Waist circumference (cm), mean (SD)</td>
<td>90.5 (6.6)</td>
<td>82.6 (7.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fasting plasma glucose (mmol/L), mean (SD)</td>
<td>6.6 (1.8)</td>
<td>5.8 (1.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Postprandial plasma glucose (mmol/L), mean (SD)</td>
<td>10.1 (4.0)</td>
<td>8.4 (2.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hemoglobin A₁c (%), mean (SD)</td>
<td>6.2 (1.4)</td>
<td>5.9 (1.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Total bilirubin (µmol/L), mean (SD)</td>
<td>14.4 (10.0)</td>
<td>12.4 (5.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Uric acid (µmol/L), mean (SD)</td>
<td>351 (87.0)</td>
<td>317 (71.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Triglyceride (mmol/L), mean (SD)</td>
<td>2.0 (1.6)</td>
<td>1.7 (1.3)</td>
<td>.01</td>
</tr>
<tr>
<td>High-density lipoprotein-cholesterol (mmol/L), mean (SD)</td>
<td>1.2 (0.3)</td>
<td>1.3 (0.6)</td>
<td>.03</td>
</tr>
<tr>
<td>Low-density lipoprotein-cholesterol (mmol/L), mean (SD)</td>
<td>3.0 (0.9)</td>
<td>2.8 (0.9)</td>
<td>.03</td>
</tr>
<tr>
<td>Alanine transaminase (U/L), mean (SD)</td>
<td>29.8 (17.4)</td>
<td>26 (13.6)</td>
<td>.005</td>
</tr>
<tr>
<td>Aspartate transaminase (U/L), mean (SD)</td>
<td>27.7 (22.1)</td>
<td>25.4 (9.5)</td>
<td>.02</td>
</tr>
<tr>
<td>γ-Glutamyl transpeptidase (U/L), mean (SD)</td>
<td>39.1 (32.8)</td>
<td>33 (16.8)</td>
<td>.001</td>
</tr>
<tr>
<td>High-sensitivity C-reactive protein (mg/L), mean (SD)</td>
<td>2.7 (2.9)</td>
<td>1.8 (2.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Platelets (× 10⁹/L), mean (SD)</td>
<td>217 (56.5)</td>
<td>209 (53.9)</td>
<td>.03</td>
</tr>
<tr>
<td>Hemoglobin (g/L), mean (SD)</td>
<td>139 (28.6)</td>
<td>134 (26.2)</td>
<td>.02</td>
</tr>
</tbody>
</table>

*S*: hepatic steatosis stage.
*S2*: moderate steatosis.

Table 5 presents the performance of the classification models. The area under the receiver operating characteristic curve (AUROC) for RF, LightGBM, XGBoost, SVM, KNN, and LR was 0.91, 0.86, 0.83, 0.88, 0.77, and 0.81, respectively. Additionally, the accuracy for RF, LightGBM, XGBoost, SVM, KNN, and LR was 84%, 81%, 78%, 81%, 76%, and 77%, respectively. RF exhibited the best performance. Figure 4 displays the AUROC obtained on the test set of the moderate-to-severe fatty liver cohort using the final features.
Table 5. The AUROC<sup>a</sup>, accuracy, sensitivity, and specificity of the 6 classification models.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUROC</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.91</td>
<td>0.84</td>
<td>0.63</td>
<td>0.92</td>
</tr>
<tr>
<td>LightGBM&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.86</td>
<td>0.81</td>
<td>0.63</td>
<td>0.89</td>
</tr>
<tr>
<td>XGBoost&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.83</td>
<td>0.78</td>
<td>0.55</td>
<td>0.89</td>
</tr>
<tr>
<td>KNN&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.77</td>
<td>0.76</td>
<td>0.60</td>
<td>0.84</td>
</tr>
<tr>
<td>SVM&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.88</td>
<td>0.81</td>
<td>0.47</td>
<td>0.95</td>
</tr>
<tr>
<td>LR&lt;sup&gt;g&lt;/sup&gt;</td>
<td>0.81</td>
<td>0.77</td>
<td>0.52</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<sup>a</sup>AUROC: area under the receiver operating characteristic curve.
<sup>b</sup>RF: random forest.
<sup>c</sup>LightGBM: light gradient boosting machine.
<sup>d</sup>XGBoost: extreme gradient boosting.
<sup>e</sup>KNN: k-nearest neighbor.
<sup>f</sup>SVM: support vector machine.
<sup>g</sup>LR: logistic regression.

Figure 4. Receiver operating characteristics curve obtained on the test set of the moderate-to-severe fatty liver cohort using the final features. AUC: area under the curve; KNN: k-nearest neighbor; LGBM: light gradient boosting machine; LR: logistic regression; RF: random forest; SVM: support vector machine; XGBoost: extreme gradient boosting.

Model Simplification and Visualization

Furthermore, we attempted to build a more concise model. We repeated the process randomly 5 times, each time selecting the top 15 scored features to create a Venn diagram (Figure 5), resulting in a total of 11 filtered features. We ranked the importance of these 11 features and plotted them on a scree plot (Figure 6). The plot demonstrated a substantial change between FPG and ALT, leading us to choose the first 6 features as inputs, excluding PPG. We made this decision based on the fact that the PPG test takes 2 hours and is not typically performed in most population health examinations.
To our surprise, the 6-feature RF model maintained accuracy while simplifying the feature acquisition for training. Table 6 displays the performance of the 11-feature RF model and the 6-feature RF model. The AUROC of the RF (11 features) model is 0.90, with a sensitivity of 0.61 and specificity of 0.94, maintaining the same accuracy as the RF model before simplification. Meanwhile, the performance of the RF (6 features) model showed an acceptable decrease compared with the others, with an AUROC of 0.88, accuracy of 0.82, sensitivity of 0.62, and specificity of 0.90. Figure 7 provides a summary of the ROC curves for the 2 simplified RF models.
Table 6. Display of the performance of the 11-feature RF\textsuperscript{a} model and the 6-feature RF model.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUROC\textsuperscript{b}</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0.91</td>
<td>0.84</td>
<td>0.63</td>
<td>0.92</td>
</tr>
<tr>
<td>RF (11 features)</td>
<td>0.90</td>
<td>0.84</td>
<td>0.61</td>
<td>0.94</td>
</tr>
<tr>
<td>RF (6 features)</td>
<td>0.88</td>
<td>0.82</td>
<td>0.62</td>
<td>0.90</td>
</tr>
</tbody>
</table>

\textsuperscript{a}RF: random forest.

\textsuperscript{b}AUROC: area under the receiver operating characteristic curve.

Figure 7. Receiver-operating characteristic curve of simplified RF models. AUC: area under the curve; RF: random forest.

Figure 8 presents box plots depicting the distribution of the 6 features. We can clearly see that there is a significant difference (Table 4) in the features between the 2 groups, which again demonstrates the high importance of the 6 features. To present our models more intuitively, Figure 9 showcases an example tree from the dense RF tree used for classification in the analysis.

Figure 8. Box diagrams of 6 features in the simplified RF model. FPG: fasting plasma glucose; hs-CRP: high-sensitivity C-reactive protein; RF: random forest; Tbil: total bilirubin; UA: uric acid.
User-Friendliness and Cost-Effectiveness

In Suzhou, FPG, UA, and TBil can be tested through blood biochemistry analysis, hs-CRP can be tested through CRP analysis, and BMI and waist circumference are easy to obtain. Our decision tree clearly shows the eigenvalues and prediction accuracy of each node, which is convenient to use in clinical practice. Therefore, we aimed for a simplified decision tree model that can be widely applicable and generalized.

Waist circumference and BMI can be measured during routine physical examinations at a negligible cost. Blood biochemistry analysis costs approximately 168 yuan (US $23.6) per person, while CRP analysis costs around 35 yuan (US $5) per person. The total cost is approximately 203 yuan (US $28.5) per person. As blood biochemistry analysis and CRP analysis are routine checkup items in health examinations, our algorithm does not require additional testing items. Additionally, residents of Suzhou over the age of 40 years have free access to all of these tests once a year, making our model cost-effective.

TE screening, as a noninvasive quantitative assessment of liver fat deposition, poses challenges in determining its feasibility for widespread implementation in densely populated cities such as Suzhou (with >15 million residents), primarily due to the additional expense involved.

Discussion

Principal Findings

We used ML to differentiate the severity of NAFLD by incorporating body circumference, lifestyle, and blood indicators. The optimal results demonstrated that a 6-feature RF model could achieve an area under the curve of 88% (with 62% sensitivity and 90% specificity). The 6 features of the RF model were waist circumference, BMI, FPG, UA, TBil, and hs-CRP. The total cost of the indicator tests involved in the RF model was approximately 203 yuan (US $28.5) per person. By contrast, TE, as a noninvasive quantitative assessment of liver fat deposition, currently costs 260 yuan (US $36.5) per person. Compared with the cost of the indicator tests, TE incurs an additional expense of approximately 57 yuan (US $8) per person. Moreover, the Suzhou government offers free annual health examinations to residents over 40 years old within their jurisdiction. Consequently, people in Suzhou, with the assistance of our model, can obtain an almost free evaluation of NAFLD.

Therefore, we ultimately decided to use the RF model composed of these 6 features, as it not only serves as a routine component of health examinations but also has the advantages of being inexpensive and easily obtainable. In addition, we considered that some variables such as systolic blood pressure, diastolic blood pressure, and random blood glucose, which are included in similar studies, may be variable and unreliable. Therefore, we excluded these features to ensure that our model remains stringent and stable.

Metabolic associated FLD is a newer diagnosis; however, NAFLD was still used in this study. This decision was made because in the 43 patients we excluded who met the criteria for alcohol intake (30 g/day for men and 20 g/day for women), the diagnosis of FLD was significantly higher than in the rest of the cohort. Additionally, we considered that the grading data referenced by TE were generated from biopsies of patients with NAFLD, and the inclusion of patients with FLD who exceeded the alcohol intake limit would have led to less rigorous results. In the model, we included 4 metabolic diseases (hypertension, diabetes, hyperuricemia, and dyslipidemia) as 2-categorical variables. Although all 4 diseases showed a correlation with FLD, their importance was deemed lesser compared with features such as waist circumference and BMI.

Models presented in previous studies are typically stratified by age, with age demonstrating a high correlation in the final constructed model. It has been observed that from 2010 to 2018, the annual incidence of NAFLD was higher among those under 60 years of age (4.7%; 95% CI 4.0%–5.5%) than among those
over 60 years of age (2.4%; 95% CI 2.1%-2.8%). Additionally, the prevalence of NAFLD is parallel with the rising trend of obesity in China, increasing from approximately 2% in 2000 to 7% in 2014 [21]. Therefore, we believe that the high prevalence across all age groups is highly correlated with changes in lifestyle habits of the population. Consequently, BMI and waist circumference were heavily weighted in our model across all age groups. However, despite their potential significance, lifestyle variables such as smoking, alcohol intake, and physical activity showed low importance compared with the ones we ultimately used, and did not significantly improve the final accuracy. Additionally, a portion of the patients who participated in the TE examination had already been diagnosed with FLD and had undergone lifestyle adjustments.

In economically developed regions such as eastern China, the prevalence of FLD is higher. For example, the prevalence of NAFLD in Shanghai is 38.17% [22]. Ultrasound screening used in population-based health examinations does not allow for a clear diagnosis of FLD grading, and therefore, does not identify patients who need stricter lifestyle control and follow-up. Additionally, there is still a lack of awareness and perception of NAFLD as a chronic disease with serious consequences among the public. Surveys conducted in the 2000s reported that only 31% of the general population in China was aware of NAFLD [23]. If our results are appended to the health examination reports, it may catch people’s attention. Our next plan is to perform TE testing on patients with a model diagnosis of moderate-to-severe NAFLD, and we hope to screen out more patients who need follow-up through the combination of ML and TE examinations. At the same time, TE has the advantage of being less costly and more readily available compared with MRI and biopsy, allowing us to obtain more data to refine our model. We welcome the use of our model for validation. We hope that the use of ML to construct easy-to-use classification models for targeted population screening can be generalized.

Limitations
Our study has some limitations that should be addressed. First, while our model demonstrated a high specificity, the sensitivity was comparatively lower. This could be attributed to the complexity of our input data, indicating a potential need for higher-dimensional inputs. Second, although we used TE results to classify moderate-to-severe NAFLD along with other categories, it is important to acknowledge that TE itself may not be 100% accurate, necessitating liver biopsy as the gold standard. Incorrect classification could diminish the accuracy of our predictions. It is essential to test the model in real health examinations.

In addition, the clinical data in this study encompassed all age groups above 20 years old. Residents aged 40 years and above can avail themselves of free health examinations provided by the government, wherein related indicators can be included in the examinations to facilitate and reduce the cost of data acquisition. The population composition and dietary habits exhibit good representativeness in the East China region. The relevant research findings also show no obvious preference. However, residents aged 40 years and above generally have more chronic diseases and may be taking medications such as lipid-lowering drugs, which can influence the importance of lipid and other indicators in different age groups. These characteristics may play a significant role in the development of fatty liver, but they have not shown sufficient importance in the application of our model across a broader age range. The accuracy of using these indicators may vary across different age groups. Therefore, if the relevant conclusions of this study are widely promoted, they will require more representative data support to ensure applicability across diverse age demographics.

Finally, it is important to note that while TE offers improved precision and accuracy, studies suggest that obesity increases the risk of TE examination failure [24,25]. Additionally, research indicates that the presence of ascites can lead to failures in ultrasound examinations [25]. These potential failures underscore the need to consider alternative testing strategies when dealing with patients with obesity or ascites, ensuring comprehensive assessment and accurate diagnosis.

Conclusions
NAFLD has indeed emerged as a significant health burden in China. Unfortunately, many Chinese individuals pay little attention to the disease and are hesitant to undergo expensive tests such as MRI or TE. The proposed cost-effective algorithm using ML to identify moderate-to-severe NAFLD by screening health examination data is promising. This approach has the potential to address the limitations of ultrasound in staging hepatic steatosis and overcome the high cost and low accessibility of TE through the use of artificial intelligence.

Acknowledgments
The authors thank all the participants who contributed to this study.

Data Availability
The data sets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

References

https://formative.jmir.org/2024/1/e53654


Abbreviations

AASLD: American Association for the Study of Liver Diseases
ALT: alanine transaminase
ANN: artificial neural network
AST: aspartate transaminase
AUROC: area under the receiver operating characteristic curve
FLD: fatty liver disease
FPG: fasting plasma glucose
HbA1c: hemoglobin A1c
HDL-C: high-density lipoprotein-cholesterol
HGB: hemoglobin
hs-CRP: high-sensitivity C-reactive protein
KNN: k-nearest neighbor
LDL-C: low-density lipoprotein-cholesterol
LightGBM: light gradient boosting machine
LR: logistic regression
LSM: liver stiffness measurement
ML: machine learning
MRI: magnetic resonance imaging
NAFLD: nonalcoholic fatty liver disease
PARs-3: Physical Activity Rating Scale-3
PPG: postprandial plasma glucose
RF: random forest
SVM: support vertical machine
TBil: total bilirubin
TE: transient elastography
TG: triglyceride
UA: uric acid
UAP: ultrasound attenuation parameter
XGBoost: extreme gradient boosting
γ-GT: γ-glutamyl transpeptidase
reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Patient and Therapist Perceptions of a Publicly Funded Internet-Based Cognitive Behavioral Therapy (iCBT) Program for Ontario Adults During the COVID-19 Pandemic: Qualitative Study

Serena Thapar¹,², HBSc; Megan Nguyen¹, RN, PhD; Bilal Noreen Khan¹,³, MPH, MSc; Roz Fanaieyan¹, MPH; Vanessa Kishimoto¹, MPH; Rebecca Liu¹, PhD; Blanca Bolea-Alamañac¹,⁴, MD, PhD; Marisa Leon-Carlyle¹, MD, MSc; Anne O’Riordan¹,⁵, BSc; Maggie Keresteci¹,⁵, MA, CHE; Onil Bhattacharyya¹,⁶, MD, PhD

¹Institute for Health System Solutions and Virtual Care, Women’s College Hospital, Toronto, ON, Canada
²Department of Psychology, McGill University, Montreal, QC, Canada
³Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
⁴Department of Psychiatry, University of Toronto, Toronto, ON, Canada
⁵Patient Advisors Network, Toronto, ON, Canada
⁶Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Serena Thapar, HBSc
Institute for Health System Solutions and Virtual Care
Women’s College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 6474580101
Email: serena.thapar@mail.mcgill.ca

Abstract

Background: To address the anticipated rise in mental health symptoms experienced at the population level during the COVID-19 pandemic, the Ontario government provided 2 therapist-assisted internet-delivered cognitive behavioral therapy (iCBT) programs to adults free of charge at the point of service.

Objective: The study aims to explore the facilitators of and barriers to implementing iCBT at the population level in Ontario, Canada, from the perspective of patients and therapists to better understand how therapist-assisted iCBT programs can be effectively implemented at the population level and inform strategies for enhancing service delivery and integration into the health care system.

Methods: Using a convenience sampling methodology, semistructured interviews were conducted with 10 therapists who delivered iCBT and 20 patients who received iCBT through either of the publicly funded programs to explore their perspectives of the program. Interview data were analyzed using inductive thematic analysis to generate themes.

Results: Six salient themes were identified. Facilitators included the therapist-assisted nature of the program; the ease of registration and the lack of cost; and the feasibility of completing the psychoeducational modules given the online and self-paced nature of the program. Barriers included challenges with the online remote modality for developing the therapeutic alliance; the program’s generalized nature, which limited customization to individual needs; and a lack of formal integration between the iCBT program and the health care system.

Conclusions: Although the program was generally well-received by patients and therapists due to its accessibility and feasibility, the digital format of the program presented both benefits and unique challenges. Strategies for improving the quality of service delivery include opportunities for synchronous communication between therapists and patients, options for increased customization, and the formal integration of iCBT into a broader stepped-care model that centralizes patient referrals between care providers and promotes continuity of care.

(JMIR Form Res 2024;8:e50113) doi:10.2196/50113
KEYWORDS
depression; anxiety; cognitive behavioral therapy; digital health; internet-delivered cognitive behavioral therapy; iCBT; CBT; implementation; facilitators; barriers; interviews; qualitative

Introduction

The COVID-19 pandemic presented a significant challenge to the mental health of Canadians [1]. Studies have drawn attention to the negative impact of the pandemic on population-level mental health, highlighting increased levels of self-reported symptoms of depression and anxiety among Canadian adults [2-4]. Research conducted throughout the pandemic demonstrated barriers to accessing in-person mental health care in Canada, which occurred in the context of existing challenges, including lengthy wait times, limited availability of services in rural areas, high cost of services, a shortage of mental health professionals, stigma, and a lack of integration of mental health care services [5]. Due to the shortage of trained mental health professionals in Canada, the current workforce is unable to adequately meet the demands of one-on-one in-person psychotherapy [6]. Therefore, other solutions are required.

Digital mental health interventions, such as internet-delivered cognitive behavioral therapy (iCBT), are promising because they may address many of these barriers to care by providing increased reach at a lower cost than other modalities, subsequently improving access to evidence-based mental health treatment for depression and anxiety [7]. iCBT involves structured and predefined psychotherapeutic content organized in online modules. Patients are assigned homework to consolidate their learning, and brief therapist support is typically provided via secure in-app messaging and calls [8]. Randomized controlled trials have provided evidence in support of the effectiveness of iCBT across psychiatric disorders [9,10]. For example, a meta-analysis of transdiagnostic iCBT found medium to large controlled effect sizes for depression and anxiety outcomes (g=0.79 to 0.82) [11]. The treatment has also been successfully implemented in routine care settings in countries such as Australia, Sweden, Denmark, and Norway [12]. Despite its apparent efficacy and efforts to reduce barriers, iCBT programs face difficulties in maintaining user engagement, which is evident by limited uptake and high dropout rates [13]. Dropout in digital mental health therapies has been a long-standing issue, with dropout rates being typically lower in study trials compared with real-world implementations and routine care [14]. For example, some studies report that just more than half of patients complete a full course of iCBT [15]. A meta-analysis of 7313 participants across 40 studies found a dropout rate of 57%, with higher rates of dropout for iCBT programs for depression without support (74%) and lower rates for those with therapeutic support (28%) [16]. Real-world iCBT implementations in Australia and the United Kingdom (eg, This Way Up, MindSpot, and Improving Access to Psychological Therapies) show dropout rates ranging from 29% to 64% [17-19]. Future work should focus on identifying participants who will most likely benefit from iCBT and adhere to treatment protocols.

In addition, studies have shown that iCBT is a cost-effective way to improve accessibility to care, particularly for patients who face geographic and mobility limitations, such as those who live in remote or rural areas [20]. Moreover, Health Quality Ontario has suggested that the few publicly funded psychotherapy services in Ontario are not equitably distributed across the province, such as in rural areas in Northern Ontario where there are fewer psychiatrists [21], and cannot meet the needs of patients with mental health concerns [9]. Indeed, there is strong evidence that there is income-based inequity in access to mental health services, especially for psychologists who tend to be concentrated in private practice under the existing 2-tiered mental health care system in Canada [22,23]. Mental health services provided by general practitioners and psychiatrists can be billed through provincial and territorial health insurance plans; however, services offered by allied health professionals, such as psychologists, cannot. Estimates suggest that while two-thirds of Canadians can access mental health services through private employer benefits, the remaining one-third must access services through limited publicly funded services, pay out-of-pocket, or forgo seeking care [24]. Therefore, leveraging digital health technology can increase patient access to evidence-based psychotherapy, regardless of whether or not there is a pandemic.

To address the anticipated rise in mental health symptoms experienced at the population level, the Ontario government rapidly expanded digital mental health service offerings in May 2020 [25]. This included therapist-assisted iCBT, a form of guided iCBT where licensed mental health professionals provide regular support for patients by monitoring their symptoms, offering regular check-ins, and giving feedback on their homework assignments. This form of iCBT can be contrasted to coach-assisted iCBT, where nonregulated mental health workers are trained to provide support to patients throughout the program, or self-guided iCBT, where patients access a series of modules independently.

While publicly funded iCBT has been successfully implemented in Ontario during the pandemic by 2 service providers (ie, privately owned Canadian companies in the mental health and wellness sector), patient and therapist perceptions around its implementation have yet to be explored. Research on the experience of patients and therapists with publicly funded iCBT in other jurisdictions highlights accessibility, convenience, and the role of therapist support [26,27]. While these findings provide valuable insights, it is important to acknowledge potential differences that cultural factors, health care systems, and program variations may influence. Furthermore, few studies have examined the uptake and experience of publicly funded iCBT programs in Canada [28]. Therefore, the study objectives were to explore, from the perspectives of patients who accessed the iCBT program in Ontario and the therapists who provided the service, the facilitators and barriers related to iCBT delivery during the pandemic, and the proposed recommendations to address such barriers. The findings can inform future policy and funding decisions by providing insight into whether and
how the service should be expanded and better integrated into the mental health care system.

Methods

Study Setting

This study aimed to investigate the therapist-assisted iCBT program funded by the Ontario government during the COVID-19 pandemic to residents and offered in both English and French [25]. The 10- to 12-week program was promoted to individuals experiencing mild to moderate depressive or anxiety symptoms. The program could be accessed through self-referral or clinician referral to 1 of 2 service providers. Both service providers offered programs that adhered to the general principles of iCBT programs, which include a comprehensive intake assessment to identify primary mental health concerns, cognitive behavioral therapy techniques, and ongoing access to communication with a regulated mental health professional [29]. Both programs were based on the same principles but differed in their intake assessment process and method of communication between the therapist and patient. Program A (LifeWorks AbilitiCBT) offered a 5- to 7-minute online intake questionnaire, followed by a mandatory remote synchronous intake assessment with a therapist to identify the patient’s mental health needs and to develop a protocol tailored for the patient’s primary mental health concern [30]. Program B (Mind Beacon TAiCBT) used a 30-minute online intake questionnaire to achieve the same aims [31]. Both programs had a secure in-app messaging system for communication between patients and therapists, through which therapists could send messages to clients derived from a bank of predetermined messages. However, only program A offered optional synchronous phone or video calls that could be scheduled if needed throughout the program. Program A provided most patients with 10 content modules, irrespective of their specific mental health needs. One exception was for patients with posttraumatic stress disorder or trauma, who were provided with 2 additional modules for a total of 12 modules and an estimated completion time of 10 weeks. Program B provided patients with an average of 7 to 16 playlists (synonymous with modules), with additional playlists of content provided depending on the individual- or condition-specific needs, for an overall average completion time of 12 weeks. Despite the slight programmatic differences, it is reasonable to argue that the barriers to and facilitators of uptake and engagement in iCBT are unlikely to differ significantly between the 2 programs, both of which were implemented in the same context and aim to provide accessible and convenient iCBT interventions using online platforms and secure messaging system for patient-therapist communication. By conducting a combined analysis, a larger data set can be obtained, enabling a broader perspective on barriers and facilitators to iCBT implementation at the population level and the identification of overarching themes that go beyond program-specific differences. By focusing on common factors among programs, the study findings can provide valuable insights for the development and improvement of implementing iCBT programs across contexts.

Study Design

In this qualitative study, the research coordinator (MN) and 2 research assistants (BNK and ST) conducted semistructured, emergent interviews [32] with patients and therapists to gain a thorough understanding of their perspectives and experiences using or delivering iCBT. All interviewers had prior experience conducting semistructured interviews and received training prior to data collection. The study design and interview guides were developed based on input from multiple stakeholders, including patient advisors, psychiatrists, and experts in qualitative research. Furthermore, the interview guides were pilot-tested among the research team. The initial interview questions aimed to elicit contextual information, understand acceptance and satisfaction with the program, identify barriers and facilitators in using or delivering iCBT, and gather feedback on refining the service for future offerings. The interviews also focused on gaining insight into program completion and communication between therapists and patients. Follow-up questions were then tailored to participants’ responses to the initial questions in keeping with the emergent interview design. Finally, to ensure the relevance of the questions, separate interview guides were developed for patients and therapists (see Multimedia Appendix 1).

Recruitment and Data Collection

The 2 service providers assisted with recruitment by sending emails and in-app messages on behalf of the research team to the therapists and patients enrolled in the program. Both service providers sent recruitment emails to all therapists on an internal listserve and in-app message prompts through the platform to reach all users with access to the program on a weekly basis between September and November 2021. During the recruitment period, 39 patients (20 in program A and 19 in program B) and 29 therapists (17 in program A and 12 in program B) contacted the research team via email expressing interest in the study. Between October and December of 2021, 30 interviews were conducted with 20 patients (10 in program A, 9 in program B, and 1 in both program A and program B) and 10 therapists (5 in program A and 5 in program B). Interviews were conducted until data saturation was reached.

The interviews were conducted between the interviewer and interviewee who had no relationship established prior to study commencement via audio call using Microsoft Teams and lasted approximately 45 minutes. Inclusion criteria for patients included being 18 years or older, having mild to moderate depression- or anxiety-related symptoms at the time of registration with the iCBT program, and having accessed the publicly funded program during the COVID-19 pandemic (between May 2020 and December 2021). The inclusion criteria for therapists included being a health care professional who delivered the iCBT program from either provider during the COVID-19 pandemic.

Data Analysis

The interviews were audio recorded, transcribed verbatim, and analyzed using inductive thematic content analysis, allowing for the emergence of themes and patterns directly from the data [33]. Transcripts were not returned to participants for comment.
or correction, nor did participants provide feedback on the findings. MN, RF, ST, and VK all contributed to the data coding and analysis process, which involved iteratively reading and rereading the transcripts to identify meaningful patterns and themes. This data-driven approach ensured that the analysis was grounded in the participants’ experiences and perspectives. To ensure rigorous methodology, the research team engaged in multiple meetings to develop consensus and refine the emergent themes.

To analyze the data, the researchers first independently reviewed the transcripts and created initial codes. Together, they then confirmed the structure of the codes and looked for potential themes. Once themes were identified, they were named and defined, using exemplar quotes to illustrate key points. All research team members reviewed and provided input on the themes and their interpretations. The themes and codes were related back to the study objectives during paper preparation. To ensure rigor and trustworthiness, the researchers engaged in reflexivity and debriefing with peers throughout the analysis [34].

**Ethical Considerations**

The study received ethical approval from the Women’s College Hospital Ethics Assessment Process for Quality Improvement Projects (REB # 2021-0057-E). All participants provided written and verbal consent before participating in the semistructured interview, during which the research objectives were described. Participants were made aware that the interviewers were unbiased third-party evaluators of the program. All participants were compensated via a $25 CAD electronic gift card to their choice of one of several major retailers. The data has been de-identified.

**Results**

**Demographics**

Table 1 displays the demographic characteristics of the patient interviewees. Of the 20 patients, most (n=13, 65%) were between the ages of 20 to 50 years; 55% (n=11) of patients were female and 40% (n=8) of patients were male. Regarding ethnicity, 70% (n=14) of the patients identified as White, while 30% (n=6) of patients identified as belonging to a racialized minority group. Notably, 80% (n=16) of patients completed 10 or more modules or playlists, meaning that they completed most of the program content without dropping out prematurely. A total of 25% (n=5) of patients had an average comfort with technology, while 30% (n=6) and 45% (n=9) of patients reported advanced and expert comfort with technology, respectively. Furthermore, all patients were self-referred to the program. Many patients heard about the program through a self-directed web-based search (n=9, 45%) or via their social network (n=9, 45%), while only 10% (n=2) of patients heard about the program via advertisements.

Table 2 displays the demographic characteristics of the therapist interviewees. Of the 10 therapist interviewees, most were (n=6, 60%) between the ages of 20 to 35 years, with a higher proportion of female therapists (n=9, 90%). Most therapists were licensed social workers (n=8, 80%), while the remaining 20% (n=2) were registered psychotherapists. Half (n=5, 50%) of all therapists had been delivering iCBT for less than 6 months. Most therapists (n=6, 60%) had 5 years or less of experience in their current profession.

Facilitators of iCBT included the therapist-assisted nature of the program, the ease at which the service could be registered for and accessed on an ongoing basis due to a lack of cost, and the feasibility of completing the psychotherapeutic content given the online and self-paced nature of the program. However, the study identified 3 barriers to the program’s implementation: challenges with the online delivery of the program for the therapeutic alliance; the program’s generalized nature, which limited customization to individual needs; and a lack of formal integration between the iCBT program and the health care system.
<table>
<thead>
<tr>
<th>Demographic variables and categories</th>
<th>Number of interviewees (n=20), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>20-35</td>
<td>4 (20)</td>
</tr>
<tr>
<td>36-50</td>
<td>9 (45)</td>
</tr>
<tr>
<td>≥51</td>
<td>7 (35)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Female</td>
<td>11 (55)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Race or ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>14 (70)</td>
</tr>
<tr>
<td>Racialized minority</td>
<td>6 (30)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>2 (10)</td>
</tr>
<tr>
<td>College degree or diploma certificate</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Undergraduate degree or above</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Number of modules or playlists completed</strong></td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td>4 (20)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>16 (80)</td>
</tr>
<tr>
<td><strong>Comfort with technology</strong></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Advanced</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Expert</td>
<td>9 (45)</td>
</tr>
<tr>
<td><strong>How patients heard about the program</strong></td>
<td></td>
</tr>
<tr>
<td>Self-directed web-based search</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Social network or care provider</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Advertisement</td>
<td>2 (10)</td>
</tr>
</tbody>
</table>
Table 2. Therapist interviewee demographics.

<table>
<thead>
<tr>
<th>Demographic variables and categories</th>
<th>Number of interviewees (n=10), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>20-35</td>
<td>6 (60)</td>
</tr>
<tr>
<td>36-65</td>
<td>4 (40)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Female</td>
<td>9 (90)</td>
</tr>
<tr>
<td><strong>Professional designation</strong></td>
<td></td>
</tr>
<tr>
<td>Social worker</td>
<td>8 (80)</td>
</tr>
<tr>
<td>Registered psychotherapist</td>
<td>2 (20)</td>
</tr>
<tr>
<td><strong>Years in current profession</strong></td>
<td></td>
</tr>
<tr>
<td>5 years or less</td>
<td>6 (60)</td>
</tr>
<tr>
<td>&gt;6 years</td>
<td>4 (40)</td>
</tr>
<tr>
<td><strong>Length of time delivering iCBT</strong></td>
<td></td>
</tr>
<tr>
<td>6 months or less</td>
<td>5 (50)</td>
</tr>
<tr>
<td>More than 6 months</td>
<td>5 (50)</td>
</tr>
</tbody>
</table>

*a* iCBT: internet-delivered cognitive behavioral therapy.

**Facilitators of iCBT**

**The Therapist-Assisted Nature of the Program**

Patients valued the therapist-assisted nature of the iCBT program, with many reporting satisfaction with being able to connect with a mental health professional. Personal communication with and feedback from therapists were essential in supporting patients’ engagement with the program and creating accountability for completing the program content, as they regularly checked in on patient progress.

> I was really busy and I kind of left it unattended, and then [my therapist] just told me...don’t give up, find the time even though it might be busy, just take time aside and keep going, you’re almost done. I think if she wouldn’t be there or she wouldn’t send me that message, I would probably [have] just dropped it. [Patient 24, program A]

> I will get occasional emails, “Your therapist has checked in...it’s almost like a cue that, “Are you on track with your lessons?” Because I think if you leave people to their own devices, sometimes they’ll just fall off...the only accountability is to yourself. [Patient 7, program B]

In addition, the presence of a therapist allowed patients to feel connected to another human despite the online nature of the program. Patients felt supported knowing that an individual cared about their progress, read their messages, and reviewed their work.

> I think I also don’t want to just feel so clinically cold about it either...that I just do my thing, and hopefully something will come of it. It’s nice for me to know that there’s somebody checking in...you feel like somebody cares. I kind of like the messaging, I think that without it, it’s a little sterile. [Patient 7, program B]

> I felt like somebody was actually listening to me. I wasn’t just going through a computer program. [Patient 24, program A]

As several patients noted, the presence of a therapist was perceived as the most valuable component of the program.

> I felt the material that I was working on was helpful, but the therapist was most helpful for what was there...I don’t think it would have been of any value to me personally without a therapist involved...For me that was one of the biggest things, was just to have someone to lean on a little bit. [Patient 13, program B]

> That’s the best part...the conversation with the therapist. [Patient 2, program A]

However, as 1 therapist noted, not all patients take advantage of the ability to engage with a therapist.

Several patients appreciated the asynchronous communication modality, which allowed for a more accessible introduction to mental health treatment for those reluctant to engage in traditional therapy formats.

> It makes you not quite so vulnerable as one-on-one counselling either on the phone or face-to-face, for somebody who is new to it or may have been reluctant in the past, I think that was the initial appeal for me, was that there was that kind of buffer...having that initial message from your e-therapist...was a very gentle introduction to therapy. [Patient 11, program B]
Other patients similarly noted that they felt more comfortable openly expressing personal topics through messaging as opposed to face-to-face or over the phone (patient 7, program B).

The Accessibility of Registering for the Program

The iCBT program was found to be accessible for patients due to its free and immediate availability. Many patients interviewed were willing to enroll in the iCBT program because it was available almost immediately with no waiting period.

As soon as I finished my [intake assessment], I got the email almost immediately...within a week, I had a therapist send me an email to set up our initial interview. [Patient 28, program A]

The lack of cost enabled individuals experiencing financial difficulties to participate, as well as those who were on waitlists or could not afford mental health services. As 1 therapist noted, the low barrier to entry is especially important because “so many people are looking for support...on waitlists...having to pay out of pocket, [and] are feeling they can’t find anyone to talk to...This is a great base for them to get that step in the door to build that confidence to be able to get more support” (therapist 4, program A). For example, 1 patient noted that they were interested in pursuing traditional CBT but did not have the resources to do so. The program being offered free of charge reduced the barrier to engagement.

Doctors thought that CBT was the appropriate therapy for me, and so...I would go through an evaluation, and then it was a horrendous cost which I couldn’t afford...The cost is prohibitive. [Patient 26, program A]

For many patients, money is a factor in their ability to access mental health services. As 1 patient stated, “I would have had to look at my benefits to see if something else was covered, but when I read that [it was free], I knew that it was not going to be a financial burden on me.” (patient 13, program B). However, some therapists expressed concern that the free service made it noncommittal and may contribute to patient withdrawal or inactivity.

The Feasibility of Completing the Psychotherapeutic Content of the Program

Furthermore, the online nature of the program also allowed for easy access from home or mobile devices, making it more feasible.

It was the most practical thing ever. Everybody sits on their phone and scrolls through Instagram and Facebook and stuff, so having the module accessible on my phone was amazing because I was able to do that at any point in time throughout my day. [Patient 28, program A]

One patient noted that they had to travel outside the home to receive the services, they would not have been able to participate (patient 26, program A). Indeed, the accessibility of the program was appealing to many patients who were not generally available during business hours. For example, 1 patient noted that the online delivery “makes the service much more accessible because people who are busy or working...it makes it more feasible to access the program compared with having to go into a doctor's office” (patient 6, program A). This was especially important for patients who did not feel comfortable leaving their homes during the pandemic or lacked access to transportation. As 1 therapist noted, the program makes it easier for patients to access mental health support because it reduces the effort required to attend in-person therapy, a step which may compound the existing difficulties for patients experiencing challenges with their mental health.

Patients also appreciated the ability to work at their own pace and revisit their worksheets following the completion of the program. Patients similarly were satisfied with always having the program available.

I really enjoyed the immediate accessibility of it...just having the reassurance that it is immediately at hand is oftentimes, sometimes all you need. [Patient 11, program B]

Other patients noted the unique ability to spread out their therapy over time and work at their own pace.

I would just get up and walk away or maybe go for a walk and come back and you can’t do that in traditional therapies. [Patient 26, program A]

As 1 therapist noted, “I feel like one of the strengths is that it’s there when people need it, I think we are not telling people, ‘You have to be here on this particular day for your session’. We are saying, ‘It’s open. When you need it, it’s right there’” (therapist 19, program B). Furthermore, patients appreciated the absence of time pressure associated with traditional therapy sessions.

I can take time to read the modules. I would never have that luxury to take time and reflect on modules in [a] therapy session...This relaxation of time and availability of resources makes the mental health service delivery much more effective. [Patient 5, program A and program B]

Barriers to iCBT

Challenges With the Online Modality

Although therapists recognized the importance of developing a therapeutic alliance with their patients, some noted challenges in establishing rapport and communicating with patients online. Some therapists found it difficult to connect with patients due to the primarily asynchronous nature of communication and the limitations of the messaging platform. For example, 1 therapist noted that “it’s not the same as talking over the phone or in person...you have to wait for them to respond so that can be kind of difficult if they don’t respond...[Also]...it’s not a chat service so it can be hard to maintain a streamlined conversation that you would in person or on the phone because we are supposed to limit our interaction to one to two business days” (therapist 4, program A).

It can be difficult sometimes to build that connection and build that therapeutic relationship when you’re just a voice, or you’re just words on a screen talking to somebody, and you’re not in person getting to build more of that physical or that non-verbal connection
Therapists were required to refine their skills in connecting with patients through a digital modality.

Despite the reported challenges, some patients said that they were able to develop a strong therapeutic alliance with their therapist online. One patient provided evidence for the ability to develop the therapeutic alliance online, as they stated that “my therapist and I have such a great relationship” (patient 28, program A), while another shared that “I felt understood [and] heard...[the therapist] listened to me...” (patient 26, program A). However, a few patients would have preferred to speak to their therapist over the phone after completing each module. For example, 1 patient stated that they would have preferred to verbally speak to their therapist about their feelings after completing each module, as is their preferred mode of communication: “I would like to express it verbally rather than writing it.” (patient 2, program A)

**Limited Opportunities for Customization**

According to the feedback gathered, patients and providers of the iCBT program felt that it needed more opportunities for customization. They suggested that tailored protocols, feedback, and guidance would enhance the program. While some degree of customization was available through the chat functionality and the ability to add worksheets, tools, or readings, some therapists noted that more needs to be done to make the program more customized to patient needs.

"I think the chat functionality, and the ability to add in worksheets or tools or readings at a whim...gives a lot of flexibility, and I think CBT requires more flexibility than is given...we offer reflections, and we ask questions, and sometimes go a little deeper than just what the material presents." [Therapist 20, program B]

However, therapists also noted limited options for them to customize the program for their patients, such as the inability to change the patient’s protocol after treatment has started in program B.

"I don’t like the fact that when I have assigned a patient a protocol, let’s say, depression, and then we have worked together, and then, maybe, depression is not really the thing I should have assigned this person...but I cannot really change the heading of the treatment." [Therapist 19, program B]

Patients also expressed that the lack of tailoring made the program feel “superficial” and not customized to their needs (patient 1, program A). For some patients, the lack of tailoring made them feel “a little boxed into the format” and that the content was “not...necessarily customized for what I [they] needed” (patient 7, program B). Others suggested that given the broad nature of the content, it was not necessarily relatable.

Therapists noted that more needs to be done to make the program more customized. They suggested that tailored protocols, feedback, and guidance would enhance the program. While some degree of customization was available through the chat functionality and the ability to add worksheets, tools, or readings at a whim...gives a lot of flexibility, and I think CBT requires more flexibility than is given...we offer reflections, and we ask questions, and sometimes go a little deeper than just what the material presents. [Therapist 20, program B]

However, therapists also noted limited options for them to customize the program for their patients, such as the inability to change the patient’s protocol after treatment has started in program B.

"I don’t like the fact that when I have assigned a patient a protocol, let’s say, depression, and then we have worked together, and then, maybe, depression is not really the thing I should have assigned this person...but I cannot really change the heading of the treatment." [Therapist 19, program B]

They suggested the option for “patient-specific” content based on the patient’s needs (patient 28, program A). One therapist noted that patients have expressed challenges due to the lack of personalization, recognizing that the content has been created in bulk and is not personalized, termed “bulk-therapy” (therapist 10, program B). As 1 therapist stated, many of the messages they sent to their patients were obtained from a bank of prewritten messages, which they would send to their patients after changing the name in the message.

"You’re sending messages, but the messages are really short or they’re kind of formulaic...you also need to have the ability to connect with an individual." [Therapist 9, program A]

**Lack of Formal Integration With the Health Care System**

The iCBT program was noted to lack formal integration with the health care system by patients and therapists, who expressed a desire for the program to be integrated with other mental health programs and resources. The program currently does not allow therapists to make formal referrals for patients; instead, they can only suggest additional resources, which places the onus on patients to seek other services. Therapists emphasized the need for integration of services, especially given the siloed nature of different services.

"We don’t really refer. We make suggestions. We say oh, call this program, or I hear this hospital has an outpatient so-and-so, or you can try so-and-so...I know what referrals are. You fill it in, you phone and you fax it, you confirm, you do handover, and you send case notes. That’s a referral. That’s a solid connection with an appointment usually by the time you’re done for the patient. We don’t do any of that, and we say to the patient it’s all on you." [Therapist, program A]

The therapists also suggested that iCBT could be used as a triage process to ensure that patients are matched to the right level of care.

"If we’re connected to other mental health agencies, then we can all be able to communicate in a way that’s easy...If this isn’t what would be the best program...we can continue to help you...it’s easy and there’s a feeling that people are really being taken care of." [Therapist 9, program A]

They suggested that social workers were in an ideal position to refer patients to other services, but the existing infrastructure does not allow for this.

"I would encourage them to reach out to their family doctor [and] provide some resources...and then, I just hope that that’s what’s going to happen...It would be great to have a partnership where the person will have continuity of work, if we could refer them rather than just give them resources for them to self-refer." [Therapist 8, program A]

Patients and therapists both felt that patients might still require additional support after completing the iCBT program and desired the ability to stay connected with their therapist on an
ad hoc basis after completing the program. Some perceived the loss of contact with the therapist after completing the program to be “abrupt” and expressed the desire for continuity of care with the same provider in the short term or if they signed up for the program again.

**Discussion**

The study of patient and therapist perceptions of 2 free publicly funded iCBT programs implemented in Ontario during the COVID-19 pandemic revealed that while several aspects of the program were well-received and facilitated access to care, there were several barriers and missed opportunities. Patients and therapists reported that the following facilitators: (1) the therapist-assisted nature of the program, which enhanced participation, (2) the ease of registration and access, and (3) the online self-paced nature of the program, which increased the feasibility of completing the psychotherapeutic content. However, the study identified three barriers: (1) challenges with the online delivery of the program for developing a therapeutic alliance; (2) the program’s generalized nature, which limited customization to individual needs; and (3) a lack of formal integration between the iCBT program and the health care system.

The findings regarding the facilitators to implementation are largely consistent with what has been reported in the literature. Many patients who were highly engaged with the program felt supported by their therapist and had adequate contact with them, which aligns with previous studies [35]. Furthermore, meta-analyses have found that guided iCBT is more effective than unguided iCBT [36], suggesting that communication with a therapist is one of the active components contributing to the effectiveness of iCBT. Another strength of the program implementation was removing many of the financial, logistical, and emotional barriers for patients. For several clients, the asynchronous communication modality was appreciated as it reduced the barriers associated with stigma for those who were reluctant to engage in traditional therapy formats, many for the first time. These findings are consistent with the literature that posits a key advantage of iCBT is in reducing stigma and, thereby, improving service access [37,38]. Moreover, the absence of a waitlist in both iCBT programs ensured that clients could receive timely access to care, which is crucial given that waitlists are a large barrier to mental health care in Ontario. Across Canada, the average wait time for first-time community mental health services from 2019 to 2020 was nearly 1 month [39]. Last, many patients expressed that the self-paced format of the program allowed them to work through the program at a preferred pace that worked well for their schedule, allowing them to integrate the program into their lives feasibly. These findings are consistent with previous studies of the attitudes of therapists and patients regarding online treatment, for which they positively perceived the ability to perform assignments at the patient’s own time and pace [40,41].

When asked about barriers, some therapists acknowledged challenges in developing the therapeutic alliance with patients; findings that are consistent with other studies in Ontario during the pandemic, where mental health care providers expressed concerns that digital care impeded the therapeutic relationship [42]. However, these findings are contrary to previous studies that found a positive alliance that is similar to face-to-face psychotherapy [43] and with a similar relationship between therapeutic alliance and patient outcomes [44]. Indeed, previous qualitative studies have suggested that the treatment format may act as both a facilitator and barrier to engagement for different individuals [45-49]. While therapists noted that some degree of customization was possible, patients desired more tailored protocols, feedback, and guidance, and some patients expressed that the program felt superficial and not relatable. These findings are aligned with previous qualitative studies on iCBT, where patients desired more flexibility for the therapist to adjust the therapeutic content to their individual difficulties or life situations [50]. While greater customization and therapist involvement may increase effectiveness, there may be a tradeoff of increased costs and, thereby, reduced reach. The most significant finding was that there was a missed opportunity in formally integrating iCBT into the health care system during its implementation, hindering the continuity of care for patients. Furthermore, participants emphasized the need for therapists to provide direct referrals to patients to other services. Understanding for whom specific interventions are most beneficial remains a challenge, particularly for low-intensity interventions such as iCBT [36,51-53]. Last, patients desire continuous support post-iCBT, including necessary follow-up mechanisms and resources for relapse prevention. The frustration experienced by patients who lose access to their therapist has also been reported in qualitative experiences of patients engaging in iCBT [50]. Overall, integrating iCBT with existing services may ensure that adequate treatment options are provided to those for whom iCBT is inappropriate, as well as facilitate the provision of additional support when necessary.

While many iCBT programs have high rates of dropouts or noncompleters [54], the findings provide insight into what works for the minority of patients who are highly engaged with the program. Knowledge and understanding about the patient experience are limited for those with low engagement or who prematurely dropped out of the program. Additionally, despite efforts to reach nonusers, defined as individuals who discontinued treatment after completing the intake assessment but prior to any treatment or were deemed ineligible for the program, no one volunteered to participate in qualitative interviews. Furthermore, the results are not generalizable to all patients and therapists, given the limited variation of sample characteristics and that only those who were interested in sharing their experiences with iCBT chose to participate in the evaluation. Moreover, while the intent of the study was to assess implementation, the interview guide did not include questions to assess users’ perspectives on the perceived impact of the program on their mental health. However, a key strength of this study is that, unlike other real-world implementations of iCBT programs that have been evaluated by those involved in the program development and implementation, the present evaluation was conducted by a neutral third-party evaluation team contracted by the program funder.

The study’s findings have implications for policy, practice, and research. Participants proposed a stepped-care model, wherein
the iCBT program could serve as a triage mechanism to ensure that patients are matched with the appropriate level of care. Given the low barriers to access, the program can reach a large segment of the population and provide an entry point into other clinical services that may be more appropriate or act as an introduction to more intensive therapy for patients on waitlists for more specialized services. In clinical practice, iCBT may be used as an introduction to psychotherapy for those who are unsure about traditional CBT, to help prepare clients for more intensive psychotherapy, or when used as a companion to traditional CBT as a booster program following discharge. At the outset of the program implementation, the primary objective was to ensure timely and scalable implementation of iCBT with low barriers to care and broad reach. Understandably, initial decisions regarding the scope and design did not consider how the program could be better integrated into the mental health care system for continuity of care. Policy decisions have since been made based on preliminary findings from a pragmatic evaluation of this program to shift from a low-access model to a prescriptive service offering that requires a referral and is integrated into the Ontario Structured Psychotherapy program, coordinated by 10 partner hospitals across the province [55]. The findings also suggest that iCBT programs could benefit from more personalized approaches to treatment. For example, therapists may benefit from training to be more specific and tailored in their patient feedback rather than using generic responses that may be perceived as disingenuous. To ensure effective support and authenticity in iCBT therapy, further research is needed to determine how therapists can convey these qualities despite the limited communication opportunities in primarily asynchronous care, given the varied patient perceptions of support. In addition, as proposed by Cavanagh et al [56], it may be necessary to reconceptualize the therapeutic alliance in the context of iCBT, given the unique treatment modality, which may guide how therapists are trained to support patients within the context of the program. Furthermore, identifying individuals for whom less therapist interaction is necessary for therapeutic benefit can help allocate more resources for those who may need more frequent interaction and possible modifications to the program structure. However, the discrepancy between the quantitative studies with larger samples and patient experiences reported in qualitative studies suggests that while many patients who do not complete the program report missing face-to-face meetings and synchronous support [46], this may not make a clinical difference. Regardless, future large-scale investigations should examine how iCBT programs can be adjusted to reduce dropout rates for those less satisfied with the treatment format. To enable reaching out to individuals who have dropped out and are no longer actively using the platform, funders should require that vendors provide clients with the option to be contacted outside of the iCBT platform for research purposes by including an additional consent clause during the enrollment process and maintain additional contact details. This is necessary for future research that aims to examine the perspectives of clients who dropped out or disengaged from the program, as they are difficult to reach. Last, the study aims to support an emerging research culture evaluating mobile apps beyond traditional market research done by vendors, wherein therapeutic apps are held to the same standards as other therapeutic interventions, such as drugs, medical devices, or psychotherapy.

In conclusion, our study provides valuable insights into the benefits and limitations of iCBT programs. While the convenience and accessibility of iCBT programs have the potential to be transformative for the treatment of mental health disorders, there is still work to be done to address the concerns of patients and therapists regarding limitations to digital care, customization, and integration with the health care system. Future research should focus on developing more personalized approaches to iCBT treatment, as well as finding ways to better integrate iCBT programs into the existing mental health system.

Acknowledgments
The authors would like to express their gratitude to the research participants who took the time to share their perspectives and experiences, without whom this research would not be possible. The authors also greatly appreciate the funding provided by the Ontario Ministry of Health to carry out this research. The study funder was not involved in the design of the study and the collection, analysis, or interpretation of data.

Authors' Contributions
ST participated in the coordination of the study; collected, analyzed, and interpreted the data; and drafted the manuscript. OB, RL, MN, ST, and BNK led in the conception of the study design. BNK, OB, and MN revised the manuscript. BNK and MN oversaw the coordination of the study and contributed to data collection. MN, RF, and VK also analyzed the data. AO, MK, BBA, and MLC supported the conception of the study design. All authors read and approved the final manuscript submitted for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Interview guide.

[DOCX File, 19 KB - formative_v8i1e50113_app1.docx]
References


Abbreviations

iCBT: internet-delivered cognitive behavioral therapy
Assessment of a Daily Diary Study Including Biospecimen Collections in a Sample of Sexual and Gender Minority Young Adults: Feasibility and Acceptability Study

Stephanie H Cook1,2, DrPH; Erica P Wood1, MPH; Mariana Rodrigues1, MA; Janice Jachero Caldas1; Maxline Delorme1

1Department of Social and Behavioral Sciences, School of Global Public Health, New York University, New York, NY, United States
2Department of Biostatistics, School of Global Public Health, New York University, New York, NY, United States

Abstract

Background: Young sexual minority men (YSMM) engage in cardiometabolic risk behaviors (eg, substance use) at higher rates than their heterosexual counterparts. Theory and previous research suggest that these risk behaviors may stem, in part, from exposure to minority stress (ie, discrimination based on sexual identity and other identities such as race).

Objective: This pilot study examined the feasibility and acceptability of a virtual 2-day daily diary study that examined daily experiences with discrimination, cardiometabolic risk behaviors (ie, sleep, physical activity, and substance use behaviors), and patterns of physiological stress and inflammation among YSMM aged 18 to 35 years.

Methods: Participants (n=20) were recruited from the greater New York metropolitan area and engaged in a 2-day daily diary protocol wherein they provided web-based consent, took a web-based baseline survey, and then, starting the next day, provided 3 saliva samples a day for 2 consecutive days to measure salivary cortisol, engaged in 3 daily diaries per day, and provided 1 blood spot sample via the finger prick method to measure high-sensitivity C-reactive protein. At follow-up, participants were interviewed via videoconferencing to ascertain their experiences and feelings related to the study protocol. Qualitative analyses explored the feasibility and acceptability of the study protocol, and exploratory quantitative analyses explored the descriptive statistics and Pearson correlations among the main study variables of interest.

Results: The retention rate was high (19/20, 95%) in our study sample. Qualitative analyses demonstrated that participants were willing to engage in similar, longer-term studies (eg, studies that include both week and weekend days) in the future and suggested the feasibility and acceptability of our study protocol among YSMM. However, participants noted several areas for improvement (eg, redundancy of survey items and difficulty pricking one’s finger) that should be considered in future research. Preliminary quantitative analyses revealed a moderate negative correlation between everyday discrimination and mean cortisol levels (r = −0.51; P = .03). Furthermore, descriptive analyses suggest that daily cortisol curves differ across races or ethnicities among YSMM. White and other-identified YSMM experienced the highest cortisol awakening response (mean 0.39, SD 0.21 μg/dL for White participants; mean 0.34, SD 0.34 μg/dL for others) with the steepest decline around bedtime (mean 0.05, SD 0.04 μg/dL for White participants; mean 0.09, SD 0.13 μg/dL for others) followed by a lower cortisol awakening response (mean 0.31, SD 0.11 μg/dL for Hispanic participants; mean 0.23, SD 0.15 μg/dL for Black participants) and a slower decline around bedtime (mean 0.10, SD 0.09 μg/dL for Hispanic participants; mean 0.03, SD 0.02 μg/dL for Black participants) among Hispanic and Black YSMM.

Conclusions: Overall, the results suggest that similar study protocols are feasible and acceptable among YSMM. Future research should highlight the pathways through which cardiovascular disease risk may arise among YSMM using longer-term study designs and more diverse study samples.
Indeed, these authors found evidence to suggest that expectations of stress and discrimination lead to cardiovascular risks. Over time, this may result in physiological dysregulation (e.g., inflammation) and poor health coping behaviors. Dysregulation, in turn, can lead to physiological changes that contribute to increased risk of cardiovascular disease (CVD).

Experiences of sexual minority stress among racial or ethnic minorities are not equal and are framed by individual cultural realities and contexts. Intersectionality theory posits that the possession of multiple marginalized identities intertwines at the individual level and is reflective of structural-level power inequities and inequality, influencing health behavior and outcomes across the life course. In particular, the intersection of race- and sexual orientation–based discrimination may be detrimental to health. Indeed, many studies have documented the deleterious effects of discrimination among racial or ethnic minorities, including dysregulated cortisol rhythms, elevated C-reactive protein (CRP), and heightened ambulatory blood pressure. These important biological outcomes influence vascular inflammation, atherosclerosis, stroke, and other cardiovascular risk factors. In addition, there are health behaviors along the posited causal pathway between discrimination and CVD risk including substance use, diet, and physical activity, which are important mechanisms to study and understand. Nevertheless, the mechanisms linking discrimination and cardiovascular factors in racially diverse populations remain underexplored. Conducting this research is vitally important, considering the heightened rates of both stress from discrimination and subclinical CVD among these potentially vulnerable populations.

The COVID-19 pandemic has provided us with a unique opportunity to conduct our study on the web. Thus, the current investigation reports on the feasibility and acceptability of conducting a study to understand key pathways linking sexual minority and racial or ethnic discrimination to CVD risk, health behaviors (e.g., substance use and physical activity), and physiological dysregulation (stress reactivity) completely virtually.

The aims of this study were as follows: (1) to assess the feasibility and acceptability of a virtual daily diary study protocol; (2) to explore preliminary associations between discrimination, substance use, sleep, physical activity, physiological stress, and inflammation among young sexual minority men (YSMM) measured via surveys, ActiGraph technology, and biologics, respectively). The COVID-19 pandemic has provided us with a unique opportunity to conduct our study on the web. Thus, the current investigation reports on the feasibility and acceptability of conducting a study to understand key pathways linking sexual minority and racial or ethnic discrimination to CVD risk, health behaviors (e.g., substance use and physical activity), and physiological dysregulation (stress reactivity) completely virtually.

**Methods**

**Setting and Participants**

This single-site pilot study recruited YSMM between the ages of 18 and 35 years from the New York tristate area. Eligibility criteria included (1) identification as cisgender male; (2) being between the ages of 18 and 35 years, inclusive, at the time of screening; (3) identification as a sexual minority; (4) having no...
known heart conditions, diabetes, or high blood pressure; and (5) willingness to provide informed consent. Participants were identified via online and offline techniques, including posts to listservs and flyers posted on campuses around the New York metropolitan area (eg, New York University). The recruitment materials included information about the study, contact information for the principal investigator’s laboratory, and a link to a web-based screening survey via Qualtrics.

Ethical Considerations

Interested participants filled out a web-based screening survey, which ascertained relevant sociodemographic information, such as age, biological sex, gender identity, sexual orientation, race, and ethnicity. Eligible participants were contacted by trained research assistants (RAs) to schedule a baseline meeting via Zoom to confirm eligibility, provide more information about the study (eg, study protocol and duration of study), and provide electronic informed consent. The participants were reenumerated with up to US $95 for completion of the study protocol. The full details of the full study protocol are detailed below. This study was approved by the Institutional Review Board of New York University (FY2021-5355).

Procedure

Baseline

Participants who were screened eligible through the web-based screening survey were emailed by trained RAs to schedule a baseline meeting via Zoom. During this baseline visit, participants were given detailed information about the full study design, which included a baseline survey to be taken on the web via Qualtrics, a 2-day daily diary period that included biological specimen collection (see 2-Day Diary Period section), and a 30- to 45-minute follow-up and debrief interview via Zoom after completion of the baseline survey and the 2-day protocol. Participants who agreed to participate and who provided electronic informed consent were then sent a link to a Qualtrics survey to be taken within the proceeding 24-hour period as well as electronic copies of step-by-step instructions for saliva collection, HemaSpot collection, and ActiGraph use. Actigraphy was chosen for the purposes of this study because research demonstrates that it provides reliable estimates of sleep patterns in relation to polysomnography [27]. The baseline Zoom meeting took an average of 30-45 minutes, and the baseline survey took approximately 45 minutes to complete. Furthermore, participants who provided informed consent were mailed a study kit with materials needed to complete the 2-day diary period via the United Parcel Service (UPS). The study kit contained saliva collection aids for the 2-day period, a HemaSpot collection device, an ActiGraph watch, and printed copies of instructions for each section of the protocol.

Two-Day Diary Period

Within 48 hours of completing the baseline, RAs mailed study kits to participants’ homes via UPS overnight shipping. These study kits contained (1) a charged ActiGraph GT9X Link watch ready to start data collection immediately; (2) 2 Ziploc bags with 3 saliva collection tubes and 3 saliva collection aids per bag; (3) a HemaSpot blood spot collection device with an absorbent sheet, 2 gauze packs, 2 alcohol swabs, 2 safety lancets, and 2 bandages; (4) printed out instructions for the saliva collection procedure, HemaSpot blood spot collection procedure, and wearing the ActiGraph, as well as a sheet to record when the ActiGraph was placed on one’s wrist and when it was taken off, if applicable; (5) supplies for returning the study kit including a small plastic bag with a humidity indicator card and 2 desiccant packets for the HemaSpot device, a cold pack (which they were instructed to freeze immediately upon receiving the kit), a silver insulated bubble mailer, and a brown paper shipping envelope; and (6) a prepaid shipping label to return the ActiGraph, saliva samples, and HemaSpot kits back to the laboratory. Participants were instructed to start the 2-day protocol the day after receiving their study kit.

Over the 2-day study period, participants were instructed to wear the ActiGraph watch on their nondominant wrist consecutively, even when showering and sleeping, over the 2-day period. Furthermore, participants provided 3 saliva samples per day, a blood spot sample, and took 3 surveys each day via Qualtrics. The first 2 of these surveys, termed “momentary diaries” took 2-3 minutes to complete and asked participants to fill out relevant information related to saliva samples, mood, and experiences. The last survey of the day, termed the “nightly diary,” took about 10 to 15 minutes to complete and asked the same “momentary diary” questions as well as more detailed questions about their experiences throughout the day. Surveys were sent via email to participants at 6:00 AM, 6:30 AM, and 8 PM each day over a 2-day period, and they were instructed to take the surveys immediately after providing their saliva samples upon wakeup, 30 to 45 minutes after waking up, and around bedtime.

For the 2-day saliva collection period, participants were asked to provide 3 saliva samples per day via the passive drool method. The first saliva sample was taken immediately upon waking, the second saliva sample was taken 30-45 minutes after waking, and the third saliva sample was taken before bedtime. This procedure was repeated for 2 consecutive days. Participants were instructed to avoid brushing their teeth, eating, and drinking liquids high in sugar or acidity (eg, coffee) 30 minutes before taking their saliva samples. Participants were instructed to allow saliva to pool in their mouth and then use a saliva collection aid to collect 1 mL of saliva in the collection vial. The participants were also instructed to freeze their samples immediately after collection. For the HemaSpot blood spot collection procedure, participants were instructed to provide a blood spot sample on the afternoon of the second day of the 2-day study period. Participants were instructed to wash their hands before beginning the finger pricking procedure. Then, participants spread the absorbent sheet provided in the study kit on a table or countertop and placed gauze, alcohol swabs, safety lancets, and bandages on the sheet in close reach. When the participants were ready to start the procedure, they were instructed to open the pouch containing the HemaSpot device (which must be used within a few minutes of opening). They were then instructed to prepare a safety lancet for use by twisting the cap until it popped out. Participants then selected their ring or middle finger from their nondominant hand and selected a place to the side of the finger pad (avoiding their fingernail). Participants then wiped their finger with an...
alcohol swab, allowing it to air dry, and then placed their hand palm up on the absorbent sheet and pressed the safety lancet firmly against the site until it clicked. They were then instructed to squeeze gently to produce a drop of blood, wipe the first drop away with gauze, and then begin collecting with the second drop. Participants then positioned their finger over the HemaSpot device funnel and were instructed to collect 3-4 drops of blood (and to avoid oversaturation). If they were not able to obtain enough blood on the first attempt, they were told they were able to try again with the second safety lancet. They were then instructed to use the gauze pad to wipe away extra blood and to apply a bandage to the site immediately thereafter. Finally, the participants were told to close the HemaSpot device and wait 3-4 hours for the blood to completely soak into the paper and dry on a flat surface. After this waiting period, the participants were instructed to freeze their blood spot samples before mailing their study kit back. Figure 1 presents an overview of the study’s procedure.

Figure 1. Overview of the study procedures. RA: research assistant.

Financial Incentive Structure

In this study, a financial incentive structure was used to promote retention. Participants were compensated with US $10 for completion of the baseline survey. For wearing the ActiGraph and returning the ActiGraph, participants were compensated with US $15. To provide and return saliva samples, the participants were compensated with US $15. For providing and returning the blood spot sample, participants were compensated with US $15. Furthermore, the participants were provided with US $10 for completing the follow-up interview. Finally, if participants completed each component of the study, they received a US $30 bonus. Thus, the participants could earn up to US $95 for the completion of this study.

Qualitative Data

Procedure

Participants were instructed to mail their study kits to the laboratory the day after completing the 2-day study protocol by placing the frozen HemaSpot device in a small plastic bag with a humidity indicator card and the 2 desiccant packets. They were instructed to place the 6 saliva collection tubes and ActiGraph along with the sealed HemaSpot device in a silver-insulated bubble mailer with a frozen cold pack. Afterward, participants placed the silver-insulated bubble mailer into the brown paper silver envelope and were instructed to affix the prepaid return label to the envelope and drop it off at UPS to be sent back to the laboratory via overnight shipping.

After receiving the returned study kit back to the laboratory, RAs processed the saliva and blood spot samples by examining the features of the saliva (eg, denoting whether phlegm or blood was present) and blood spot samples (eg, if there was too much or too little blood). Furthermore, ActiGraph data were downloaded. Then, the RAs contacted the participants via email to schedule their 15 to 20 minute debrief interviews via Zoom. During this interview, the RAs described the purpose of the study and followed a semistructured interview guide to ascertain participants’ experiences with the study design, collecting saliva and blood spot samples, wearing an ActiGraph, and filling out diary surveys each day. After the completion of this interview, participants were compensated according to their compliance with the study protocol (see the Financial Incentive Structure section above) through an electronic Visa gift card via GiftBit.

The RAs took notes on participants’ responses to the follow-up interview questions (see Figure 1). The interview guide was structured to gather information pertaining to the feasibility and acceptability of the study protocol among the YSMM. In particular, we asked participants about their experiences with the RAs, saliva samples, ActiGraph watch, and HemaSpot collection device. We also asked participants to describe whether the study caused any stress or confusion, and if they had any areas of improvement to increase their willingness to engage in future research with similar designs.

Qualitative Analysis Plan to Assess Feasibility and Acceptability

A deductive directed content analysis approach was used to code the interview notes [28]. This approach uses prior theory and research to create an initial coding scheme consisting of overarching themes—in this case, our initial predetermined themes were feasibility and acceptability as well as potential areas of improvement. Three coders met to discuss the predetermined themes and then one coder applied the themes to the transcripts. The 3 coders met throughout the duration of the qualitative analysis to discuss themes and resolve any discrepancies in coding. After reading the transcripts, subthemes for each overarching theme were deduced from the interview notes to determine a finalized coding scheme organized by the themes of feasibility, acceptability, and areas for improvement. Debrief notes were analyzed using Atlas.ti qualitative software [29].
Quantitative Data to Explore Preliminary Associations

Baseline Survey
The baseline survey was conducted on the web using Qualtrics. Participants were assigned a unique participant ID and required to enter the ID to access each survey. The baseline survey asked questions pertaining to sociodemographics (eg, date of birth, race and ethnicity, education level, yearly income), family history of cardiovascular-related illness or disease, and general weekly exercise patterns (ie, the Global Physical Activity Questionnaire [30]). We then administered a series of scales related to experiences with stress (eg, Adverse Childhood Experiences Scale [31], Perceived Stress Scale [32]), discrimination (Everyday Discrimination Scale [EDS] [33], Major Experiences of Discrimination Scale [33]), substance use (Alcohol, Smoking, and Substance Use Involvement Screening [34]), and mental health (eg, Center for Epidemiologic Studies Depression Scale [35]). The baseline survey took an average of 45 minutes to an hour to complete.

Momentary Diaries
Momentary diaries were taken by participants twice per day over a 2-day period (once upon waking and once 30-45 min after waking) via Qualtrics. Participants were asked several questions related to their saliva samples (eg, what time they provided the sample and caffeine consumption) and questions pertaining to their current mood, stress, recent experiences of racial and sexual orientation discrimination, and substance use. Each momentary survey was sent to the participants via a secure Qualtrics link. The survey took an average of 2-3 minutes to complete.

Nightly Diaries
Nightly diaries were taken by participants once per day over a 2-day period (before bedtime) using Qualtrics. Nightly diaries contained the same questions as the momentary diary and included modified scales to assess experiences of mood (Positive and Negative Affect Schedule [36]), depressive symptoms (Center for Epidemiologic Studies Depression Scale), and discrimination (EDS). Nightly diaries were sent to the participants via Qualtrics via email.

Salivary Cortisol
Participants provided 3 saliva samples per day over the course of 2 consecutive days to assess diurnal cortisol [37]. They were instructed to provide samples upon awakening, 30 minutes after awakening, around lunchtime, and around bedtime. The participants were also instructed to take a brief diary survey after providing each saliva sample. After participants collected their first 3 saliva samples, they received a link to an additional “momentary” diary survey via text message or email, which took about 2 minutes to complete. This momentary survey assessed variables associated with cortisol (eg, caffeine consumption) as well as other psychosocial variables (eg, current mood). Finally, after each bedtime saliva sample, participants received a link via text message or email, depending on their preference, to a longer “nightly” diary survey. To increase adherence to sample collection and surveys, participants received text message reminders to provide their samples, which also contained the appropriate survey link.

High-Sensitivity CRP
The Center for Studies in Demography and Ecological Biodemography Lab analyzed serum high-sensitivity CRP (hsCRP) values using an enzyme-linked immunosorbent assay as described in detail elsewhere [38]. Microliter plates were coated with anti-CRP antibodies to measure CRP concentrations within serum samples and stored at −20 °C. This method has been validated for population health research as a robust method for detecting low concentrations of CRP [38-40].

hsCRP values were arranged into 3 categories according to clinical risk for CVD assigned by the American Heart Association and Centers for Disease Control and Prevention to create a nominal outcome variable for calculation: low (<1.00 mg/L), average (1.00-3.00 mg/L), and high (>3.00 mg/L) [41,42]. The categories described the results in terms of clinical significance, as CVD risk increases with higher CRP concentrations. Previous studies using CRP in heart disease have demonstrated this clinical significance using categorical CVD risk [43-45].

Other Covariates

Overview
BMI was calculated using a standard formula based on the participants’ height and weight (kg/m²), each of which was self-reported. Values were sorted into 2 categories: underweight or normal weight (<25.00) and overweight or obese (≥25.00). Furthermore, we also ascertained information pertaining to smoking status (never, former, and current), substance use (measured at baseline via Alcohol, Smoking, and Substance Use Involvement Screening and nightly diaries), and family history of CVD (yes or no).

Quantitative Analysis
First, sociodemographic and clinical characteristics were explored in our study sample (eg, means with SDs and frequencies). We then explored retention and feasibility through an examination of study dropout, loss to follow-up, and completion rates of each of the study components (ie, salivary samples, HemaSpot blood spot samples, momentary diaries, and nightly diaries). Finally, Pearson correlation coefficients were used to explore associations among our main independent variables (discrimination as measured by EDS), the dependent variable (hsCRP), and our mediators (sleep, physical activity, and physiological stress as measured by salivary cortisol). Time-varying variables (ie, sleep, physical activity, and cortisol) were averaged across the study period for each participant. Baseline discrimination (as measured by EDS) was used for the purpose of the correlations. Average daily cortisol curves were also examined and stratified by race and ethnicity using the profile plot command in Stata (version 18; StataCorp) [46].

Results

Sample Characteristics
Figure 2 displays an overview of the number of participants who were screened, ineligible, enrolled, and analyzed. Sociodemographic and clinical characteristics of the 19 YSMM who participated in the study are presented in Table 1. On
average, the participants were aged 24.37 (SD 5.46) years. With respect to racial or ethnic identity, 39% (7/19) identified as non-Hispanic white, whereas the remaining identified as Hispanic (3/19, 16%), Black (3/19, 16%), or another racial or ethnic identity. Most participants identified as gay or bisexual (8/19, 42%), and the remaining identified as having another sexual identity. Concerning health characteristics, most participants reported an average BMI (10/19, 53%) and no family history of CVD (16/19, 84%). The average hsCRP among participants was 0.20 (SD 0.17), and participants slept an average of 313.18 (SD 90.25) minutes per night and expended an average of 1.52 (SD 0.36) metabolic equivalent of tasks (METs) per day.

Table 2 presents the retention and feasibility measures. Overall, 20 YSMM enrolled in this study. Of these 20 YSMM, 1 participant was lost to follow-up after the informed consent meeting, and 2 participants were lost to follow-up after completion of the study protocol (ie, they did not participate in the follow-up interviews but completed all other parts of the study protocol). Of the 19 YSMM who engaged in the study protocol, 100% (n=19) sent back their HemaSpot samples (ie, blood spots) and all 6 of their saliva samples (3 saliva samples per day for 2 consecutive days). With respect to the momentary diaries, 89% (17/19) completed 4 out of the 4 momentary diaries (2 momentary diaries/d for 2 days). The remaining participants completed either 2 (5%; n=1) or 1 (5%; n=1) of the momentary diaries over the 2-day period. In contrast, 1 participant did not complete any nightly diaries over the 2-day period. However, of the remaining 18 participants, 94% (n=16) completed 2 out of the 2 nightly diaries over the study period, and the remaining 6% (n=2) completed 1.

Figure 2. Participants screened, excluded, enrolled, and analyzed.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD; range)</td>
<td>24.37 (5.46; 18-35)</td>
</tr>
<tr>
<td><strong>Race or ethnicity</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>7 (39)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3 (17)</td>
</tr>
<tr>
<td>Black, including Hispanic</td>
<td>3 (17)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (28)</td>
</tr>
<tr>
<td><strong>Sexual orientation</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Gay</td>
<td>8 (42)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>8 (42)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (16)</td>
</tr>
<tr>
<td><strong>Education</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>High-school diploma or GED</td>
<td>4 (21)</td>
</tr>
<tr>
<td>Some college</td>
<td>8 (42)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>7 (37)</td>
</tr>
<tr>
<td><strong>BMI category</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Normal (&lt;25)</td>
<td>10 (53)</td>
</tr>
<tr>
<td>Overweight or obesity (&gt;25)</td>
<td>9 (47)</td>
</tr>
<tr>
<td><strong>Family history of CVD</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>3 (16)</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>16 (84)</td>
</tr>
<tr>
<td>hsCRP&lt;sup&gt;d&lt;/sup&gt;, mean (SD; range)</td>
<td>0.20 (0.17; 0.03-0.63)</td>
</tr>
<tr>
<td>Number of min of sleep/night&lt;sup&gt;e&lt;/sup&gt;, mean (SD; range)</td>
<td>313.18 (90.25; 181-487)</td>
</tr>
<tr>
<td>METs&lt;sup&gt;f&lt;/sup&gt;/d, mean (SD; range)</td>
<td>1.52 (0.36; 1.09-2.73)</td>
</tr>
<tr>
<td><strong>Cortisol measures (µg/dL), mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Upon waking</td>
<td>0.33 (0.15)</td>
</tr>
<tr>
<td>30-45 min after waking (CAR&lt;sup&gt;g&lt;/sup&gt;)</td>
<td>0.33 (0.22)</td>
</tr>
<tr>
<td>Around bedtime</td>
<td>0.06 (0.08)</td>
</tr>
</tbody>
</table>

<sup>a</sup> One respondent had missing data regarding race.
<sup>b</sup> GED: General Educational Development.
<sup>c</sup> CVD: cardiovascular disease.
<sup>d</sup> hsCRP: high-sensitivity C-reactive protein.
<sup>e</sup> Two participants had missing actigraphy data on sleep.
<sup>f</sup> MET: metabolic equivalent of task.
<sup>g</sup> CAR: cortisol awakening response.
Table 2. Retention and feasibility of the study protocol (n=19a).

<table>
<thead>
<tr>
<th>Values</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HemaSpot samples returned, n (%)</td>
<td>19 (100)</td>
</tr>
<tr>
<td>Saliva samples returned, n (%)</td>
<td>114 (100)</td>
</tr>
<tr>
<td>Momentary diaries completed, mean (SD; range)</td>
<td>2.45 (1.13; 1-4)</td>
</tr>
<tr>
<td>Nightly diaries completed, mean (SD; range)</td>
<td>1.47 (0.51; 1-2)</td>
</tr>
<tr>
<td><strong>Momentary diary completion by day, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Average diaries completed; day 1 (n=19), mean (SD; range)</td>
<td>1.49 (0.51; 1-2)</td>
</tr>
<tr>
<td>Average diaries completed; day 2 (n=17), mean (SD; range)</td>
<td>1.50 (0.51; 1-2)</td>
</tr>
<tr>
<td><strong>Nightly diary completion by day, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Sample completed, day 1</td>
<td>18 (95)</td>
</tr>
<tr>
<td>Sample completed, day 2</td>
<td>16 (84)</td>
</tr>
</tbody>
</table>

*a20 participants enrolled in the study; however, only 19 (95%) completed the protocol.

Feasibility and Acceptability

Overall, 17 (89%) participants completed the follow-up qualitative interviews. Within each overarching prespecified theme of “feasibility,” “acceptability,” and “areas for improvement,” we found several subthemes. The overarching themes and subthemes, as well as their definitions and excerpts, are presented in Table 3. For feasibility, we found two subthemes: (1) informative onboarding experience and easy communication with RAs, and (2) easy protocol to follow. Most participants (14/17, 82%) reported that their onboarding experience (including informed consent and a detailed step-by-step protocol) was informative and that they were able to answer all of their questions. Several participants noted that the documents that were emailed (and hard copies mailed) to them during the onboarding call were clear and laid out the study protocol in a helpful and step-by-step manner. Furthermore, participants noted that the RAs were easy to communicate with and were available to answer any questions they may have had over the study protocol period, thus reducing any confusion or hesitancy when it came to completing the study’s steps. Furthermore, 76% (13/17) reported that the study protocol was easy to follow with little to no issues arising, especially with the instruction documents that were sent to the participant (eg, for the ActiGraph, HemaSpot, and saliva samples).

With respect to acceptability, we found three subthemes: (1) appropriate financial incentive structure, (2) desire to participate beyond financial incentives, (3) changed perceptions of mood or discrimination, and (4) willingness to participate in long-term studies with a similar protocol. First, 53% (9/17) reported that the financial incentive payment structure increased their willingness to complete the study and all steps in the protocol. These participants noted the importance of being compensated fairly for their time, especially because this protocol required the collection of biological samples (ie, saliva and blood). Furthermore, these participants also noted that they appreciated the breakdown of the payments—in other words, they appreciated being compensated for each component of the study separately rather than a lump contingent on completion of the entire protocol at the end of the study period. However, 41% (7/17) of the participants described internal and altruistic motivations for participation, which motivated them to complete the study regardless of the financial incentives attached to it. These participants described an interest in the scientific process, as well as wanting to contribute to their community’s health by getting involved in health-related research. Moreover, 47% (8/17) reported positive experiences directly because of participation, including a better sense of mood and becoming more aware of experiences of discrimination. These participants noted that reflecting on their experiences over the course of the day (eg, nightly diaries) required them to become more mindful of their feelings and moods throughout the day and to become more aware of their surroundings. Furthermore, these individuals were able to better capture experiences with discrimination and microaggressions and reflect on their subsequent moods. Finally, 100% (n=17) of the participants reported willingness to participate in long-term studies (eg, 30 days of data collection) using a similar protocol. In particular, participants felt that a long-term study would be more reflective of their day-to-day stressors and experiences. However, 2 of these participants noted that while they would be willing to participate in a longer-term study, they preferred that the study surveys (eg, baseline and nightly) be shorter in duration.
Table 3. Qualitative themes and subthemes (n=17²).

<table>
<thead>
<tr>
<th>Theme and description</th>
<th>Example quotes</th>
<th>Endorsed, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feasibility of intervention</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informative onboarding experience and easy communication process with RAs</td>
<td>Descriptions related to the ease of understanding the onboarding instructions and documents and being able to communicate in a timely manner with the RAs as well as get their answers answered promptly</td>
<td>14 (82)</td>
</tr>
<tr>
<td>• “Yes, [the onboarding experience] answered all the questions at that point and I had a few more questions [later] but I was able to get my questions answered.” (Hispanic, bisexual, 28 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• “Everything was fine, [it was] very pleasant to talk with research assistants. Instructions were clear granted how much information there was.” (White, gay, 35 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol easy to follow</td>
<td>Participant was able to comprehend and follow the study instructions with little to no problems</td>
<td>13 (76)</td>
</tr>
<tr>
<td>• “[The protocol was a] good experience, not very invasive. Labeled tubes and collection aid were helpful. Instructions were clear and helped.” (Asian, bisexual, 21 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• “Everything was very clear. Explanation of the blood spot sample was especially helpful.” (Asian, gay, 28 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Acceptability of intervention</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate financial incentive structure</td>
<td>Descriptions of being motivated to engage in the study due to the financial incentive structure and feeling that the financial incentive structure was appropriate for the time given</td>
<td>9 (53)</td>
</tr>
<tr>
<td>• “The incentive did help my willingness [to participate] and the breakup of the payments helped.” (Hispanic, bisexual, 25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• “[The financial incentive structure] made me want to ensure I completed all the steps, but it was not very hard so I would have done it either way. But the payments definitely helped.” (White, bisexual, 30 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desire to participate beyond financial incentives</td>
<td>Feelings and motivations of external reasons for participating in the study (eg, wanting to progress science)</td>
<td>7 (41)</td>
</tr>
<tr>
<td>• “From an academic perspective, I was interested in the study regardless. Realistically, I would be taking opportunity to be doing the study anyway.” (Black Hispanic, gay, 31 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• “[The financial incentive structure] did not really affect my participation, would have still done without compensation.” (Black, gay, 31 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changed perceptions of mood and discrimination</td>
<td>Increased awareness of one’s experiences with discrimination and mood throughout the day</td>
<td>8 (47)</td>
</tr>
<tr>
<td>• “It made me a bit more aware of my surroundings and did notice more [discriminatory] experiences once I was asked.” (Hispanic, gay, 18 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• “[I became] more sensitive to attuning to [discrimination]. Many of these things I encounter on a regular basis, but it became more salient when calling my attention to it.” (Black Hispanic, gay, 31 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingness to participate in longer-terms studies</td>
<td>Expressing interest and desire to participate in studies similar to this on a longer-term basis (eg, 30 d)</td>
<td>17 (100)</td>
</tr>
<tr>
<td>• “Yeah, it would be good to participate [in a longer term study] because I felt like study duration was very short. A longer study would be more insightful.” (Black Hispanic, gay, 24 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• “Absolutely I would participate in a longer-term study. I have participated in longitudinal studies before and recognize the challenge that goes into them.” (Black Hispanic, gay, 31 y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Areas for improvement</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Although participants overall had positive feedback with respect to the study design and their experiences with participation, there were several areas that they noted for improvement: (1) experiences with day-to-day stress unrelated to the study, (2) protocol-related criticisms related to the study procedures, and (3) stress directly resulting from the study protocol. First, 53% (9/17) of the participants noted that they dealt with various forms of daily stress unrelated to the study protocol (eg, relationship stress and school stress). Due to these various daily stressors, some participants noted that they had to take extra time to plan ahead and ensure that they were able to complete the steps of the study in the specified time frame. While not inhibiting their ability to complete the study, they noted that planning took extra time and mental energy. Second, 76% (13/17) of the participants specified criticism that was directly related to the study protocol. In particular, 9 of these participants noted frustration with the survey instruments. These participants noted that the surveys, particularly the nightly diary, were long and repetitive. Furthermore, some participants also noted the need to be more inclusive of other identities (eg, trans, nonbinary), although this was beyond the scope of this study. Moreover, 6 of these participants noted that it was difficult to collect their blood spot samples and had difficulty producing enough blood for the HemaSpot collection device, although they also noted that the inclusion of 2 lancets was helpful. Finally, a few (4/17, 24%) participants noted that they experienced stress stemming directly from engaging in the study.

In particular, participants noted stress when having to prick their finger for a blood spot sample (eg, needle anxiety) and having to produce saliva samples 3 times a day. Furthermore, one participant noted that they were not comfortable storing their saliva samples in freezers until they were ready to ship back to the laboratory.

**Preliminary Exploration of Hypothesized Pathway**

Correlations between main independent variable (discrimination), dependent variable (hsCRP), and the mediators (salivary cortisol, sleep, and physical activity) are presented in Table 4. Furthermore, Table 5 displays the average cortisol parameters by race, and Figure 3 displays the average cortisol curves over the 2-day period stratified by race or ethnicity. It is important to note that out of the 19 YSMM included in these analyses, 2 participants did not wear their ActiGraph watch over the 2-day period and thus were not included in the actigraphy measures. There was a moderate negative association between everyday discrimination and mean cortisol such that greater experiences of discrimination were associated with lower daily cortisol, on average. Moreover, we also observed a moderate, positive association between daily physical activity, on average (as measured by METs) and hsCRP such that greater levels of average physical activity were associated with higher hsCRP, on average. No other correlations were significant in our study sample. With respect to the average cortisol curves over the 2-day period (Figure 3), White YSMM experienced the strongest
cortisol awakening response (CAR; ie, the spike in cortisol that occurs in healthy individuals 30-45 min after waking), which steeply declined throughout the day, whereas Black YSMM’s cortisol started out lower and exhibited a weaker CAR and declined throughout the day.

Table 4. Correlation matrix (n=19a).

<table>
<thead>
<tr>
<th></th>
<th>Everyday discrimination</th>
<th>Mean cortisol (µg/dL)</th>
<th>Mean sleep (min)</th>
<th>Mean METs b</th>
<th>hsCRP c (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Everyday discrimination</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean cortisol (µg/dL)</td>
<td>—0.51 e</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean sleep (min)</td>
<td>—0.07</td>
<td>—0.14</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean METs d</td>
<td>—0.32</td>
<td>0.31</td>
<td>—0.16</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>hsCRP (mg/L)</td>
<td>0.04</td>
<td>0.29</td>
<td>—0.18</td>
<td>0.59 g</td>
<td>—</td>
</tr>
</tbody>
</table>

aTwo participants did not complete follow-up interviews.
bMET: metabolic equivalent of task.
chsCRP: high-sensitivity C-reactive protein.
dNot applicable.
eP=0.03.
fMET: metabolic equivalent of task.
gP=0.008.

Table 5. Cortisol by race (n=19a).

<table>
<thead>
<tr>
<th></th>
<th>White (n=3), mean (SD)</th>
<th>Hispanic (n=3), mean (SD)</th>
<th>Black (n=7), mean (SD)</th>
<th>Other (n=5), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waking cortisol (µg/dL), mean (SD)</td>
<td>0.35 (0.18)</td>
<td>0.28 (0.05)</td>
<td>0.24 (0.04)</td>
<td>0.36 (0.19)</td>
</tr>
<tr>
<td>Cortisol awakening response (µg/dL), mean (SD)</td>
<td>0.39 (0.21)</td>
<td>0.31 (0.11)</td>
<td>0.23 (0.15)</td>
<td>0.34 (0.34)</td>
</tr>
<tr>
<td>Bedtime cortisol (µg/dL), mean (SD)</td>
<td>0.05 (0.04)</td>
<td>0.10 (0.09)</td>
<td>0.03 (0.02)</td>
<td>0.09 (0.13)</td>
</tr>
</tbody>
</table>

aCorrelations for sleep include 17 out of 19 participants.
**Discussion**

**Principal Findings**

This pilot study found evidence for the feasibility and acceptability of a virtual daily diary protocol and explored preliminary associations between discrimination, cardiometabolic health behaviors, and inflammation among YSMM. Of the 19 participants who completed more than just the baseline survey, 100% (n=19) returned their HemaSpot and saliva samples, and 89% (n=17) completed all momentary measures.

Although some participants mentioned some areas for improvement, such as having to take extra time to plan ahead to ensure that they were able to complete the required steps and experiencing stress directly related to the study (eg, needle anxiety, having to produce saliva samples 3 times a day), 100% of the participants reported a willingness to participate in longer-term studies using a similar protocol. Such valuable findings greatly contribute to the literature by demonstrating the feasibility and acceptability of a fully virtual protocol with a component of collecting biological specimens. Our findings expand on current research that highlights the specific challenges of conducting a fully virtual protocol [47,48].

For instance, Feigelson et al [47] examined the feasibility of an at-home study involving cross-sectional survey data and stool collection and found that there is a need for significant personal contact and carefully timed follow-up to ensure participant willingness. In our study, only 2 individuals did not participate in follow-up interviews. However, they completed all the other parts of the study protocol. In addition, the subthemes found in the qualitative analysis regarded the informative nature of the onboarding experience, easy communication process with the RAs, and the fact that the protocol was easy to follow. Furthermore, some of the challenges encountered by the participants were concurrent with previous research that has described remote protocols. For instance, in-home sample collection and the inconvenience of fitting the protocol timeframe with participant schedules have been highlighted in previous studies [47,48]. In the current protocol, some of the participants mentioned that due to various forms of daily stress unrelated to the study, they had to dedicate time to ensure they would be able to complete the steps in the specified time frame. In addition, although 76% (13/17) of the participants reported that the study protocol was easy to follow, 24% (4/17) reported experiencing stress when collecting biospecimens. It was not clear if the experiences of stress outside the study protocol would still be a barrier if participants were to come to the laboratory to complete the study.

Furthermore, a notable finding of our protocol was that 47% (8/17) of participants reported better perceived mood and awareness of discriminatory experiences as a direct result of their participation in the study. The participants highlighted that the momentary diary prompts led them to reflect on their experiences throughout the day. As previously described in the study’s theoretical model, there is a pressing need for studies to take an intersectional approach and consider the unique experiences of stress among sexual minority individuals by racial/ethnic and sexual/gender status and by culture and context.
Although not a direct aim of our protocol, such results and reports greatly contribute to and expand on the current literature by highlighting the need for protocols and intervention development to take an intersectional approach, especially ones using ecologic momentary data collection. In addition, as mentioned above, 100% of the participants reported a willingness to participate in a longer-duration protocol with components similar to the current one. This was because participants believed that a longer-term study would better reflect their everyday stressors and experiences. Future research aimed at understanding key minority stressors and their effects on health behaviors should consider expanding the study duration to best capture individuals’ overall experiences. Similar study designs with longer durations have shown acceptance among sexual minority populations [49]. For example, one ecologic momentary assessment study examining tobacco, alcohol, and drug use in relation to daily discrimination experiences among 50 sexual and gender minorities found that participants completed an average of 68% of the 6 prompts sent to them daily over a period of 14 days [49]. Furthermore, in a daily diary study examining minority stress and daily mood among racially diverse sexual minority youth (n=94) over a 21-day period, participants responded to 83% of the daily diary prompts sent to them per day [50]. By capturing day-to-day experiences over a longer duration (eg, a 14- or 30-day period), researchers could not only improve generalizability and power but also capture variation in experiences of minority stress and cardiovascular risk behaviors across both week and weekend days.

Our preliminary descriptive findings also shed light on potential differences across racial or ethnic groups in the YSMM, which necessitates further exploration. For instance, we examined average daily cortisol curves across race or ethnicity and found evidence suggesting that Black YSMM may experience a more dysregulated cortisol curve compared with White YSMM. In particular, we found that Black YSMM experienced a less pronounced CAR and a slower decline throughout the day compared with White YSMM. CAR represents a brief period, typically 30 to 45 minutes after waking, where an individual experiences increased cortisol activity and is recognized as an important marker of the hypothalamic-pituitary-adrenal axis as it aids an individual in their transition from sleep to wake [51]. Typically, however, individuals experience a steady decline in cortisol activity throughout the day, with the lowest levels recorded around bedtime (ie, daily cortisol curves) [52]. Research has generally found that individuals with steeper cortisol curves (ie, cortisol that declines at a faster rate) are associated with better overall health outcomes compared with individuals with more blunted cortisol curves (ie, cortisol that declines at a slower rate) [52]. Although we were underpowered to detect significant differences within daily cortisol curves among our study sample (n=19), our descriptive findings shed light on potential differences across race or ethnicity among YSMM that warrant further exploration with larger, more diverse study samples.

Although our preliminary study highlights the feasibility and acceptability of a remote protocol examining momentary, daily, and biological measures, there are several limitations that should be considered. First, our sample consisted of YSMM who resided in the New York tristate area, and thus may not be generalizable to the wider population of YSMM in the United States. Second, our study sample was small, which also affected our ability to detect significant differences in our key measures of interest (eg, daily cortisol curves and CRP). To the best of our knowledge, this is the first study to integrate momentary assessment data and biologics using a fully remote protocol. Third, our study was, in part, disrupted by the COVID-19 lockdown, which may have impacted participants’ ability to fully engage with the protocol and may also have impacted participants’ daily cardiovascular health behavior patterns (eg, participants may not be walking or traveling as much as they used to prepandemic). Fourth, we may have missed important information within our surveys that could affect cardiovascular health, such as dietary patterns. Fifth, although the overarching literature supports the use of wearables, such as the ActiGraph, to measure sleep, it may not have wholly captured sleep patterns over the course of 2 days in comparison with gold standards, such as polysomnography [27,53]. Thus, it is important to carefully examine the updated literature when determining the best wearable device for capturing sleep patterns. Sixth, given that our pilot study only spanned a period of 2 days, our actigraphy measures (ie, sleep and physical activity) may not have reliably captured daily sleep and physical activity patterns among our sample of YSMM and we did not include data on napping in these analyses. Thus, future research should use longer collection periods to capture sleep and physical activity patterns more comprehensively among diverse samples of YSMM. For instance, research has found that a period of at least 3-5 days is necessary to capture accurate information pertaining to sleep [27]. Despite these limitations, the strengths of this study lie in its diverse sample of YSMM and its rigorous design that integrates surveys and biological metrics. By demonstrating the feasibility and acceptability of a fully remote protocol combining ecological momentary assessment and biological measures, our results pave the way for future work aimed at developing culturally tailored just-in-time interventions that might be delivered virtually. Such interventions should consider minority stressors and their influence on health behaviors and adverse outcomes throughout their lifespan. Doing so could lessen the mental and physical burden at the intersection of multiple minoritized identities as well as reduce inequities in the health care system.

Conclusions

Overall, the fully virtual protocol for assessing CVD among emerging adult sexual minorities was acceptable, indicating that it can be used in the future for similar work and with a larger time period and sample size. Furthermore, preliminary descriptive results suggest that there may be key racial or ethnic differences among YSMM in key cardiovascular health features (ie, daily cortisol curves) that warrant further research.
Acknowledgments
This work was supported by the National Heart, Lung, and Blood Institute (5R25HL105446-11; principal investigator, Mohamed Boutjdir).

Conflicts of Interest
None declared.

References


27. Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [Medline: 27164110]


29. ATLAS.ti homepage. ATLAS.ti. URL: https://atlasti.com [accessed 2024-02-03]


46. Statacorp. Stata statistical software: release 18. College Station, TX: StataCorp LLC; 2023. URL: https://www.stata.com/


**Abbreviations**

CAR: cortisol awakening response  
CRP: C-reactive protein  
CVD: cardiovascular disease  
EDS: Everyday Discrimination Scale  
hsCRP: high-sensitivity C-reactive protein  
MET: metabolic equivalent of task  
RA: research assistant  
UPS: United Parcel Service  
YSMM: young sexual minority men
Acceptance of Telemedicine by Specialists and General Practitioners in Cardiology Care: Cross-Sectional Survey Study

Felix Muehlensiepen1,2*, BA, MPH, Dr rer med; Marie Josephine Hoffmann3*, MD; Jonathan Nübel1, MD; Yury Ignatyev4, PhD; Martin Heinze4, Prof Dr; Christian Butter3, Prof Dr; Anja Haase-Fielitz1,3,5, PharmD

1Brandenburg Medical School (MHB) and Faculty of Health Sciences (FGW) Brandenburg, Neuruppin, Germany
2Autonomie, Gérontologie, E-santé, Imagerie et Société (AGEIS), Université Grenoble Alpes, Grenoble, France
3Department of Cardiology, Heart Center Brandenburg Bernau & Faculty of Health Sciences (FGW) Brandenburg, Brandenburg Medical School (MHB), Neuruppin, Germany
4Department for Psychiatry and Psychotherapy, University Hospital Immanuel Klinik Rüdersdorf, Brandenburg Medical School (MHB), Rüdersdorf, Germany
5Institute of Social Medicine and Health System Research, Otto von Guericke University, Magdeburg, Germany
*these authors contributed equally

Abstract

Background: In the coming years, telemedicine will play a key role in health care. Especially in rural areas with weak infrastructure, telemedicine could be crucial to providing adequate and personalized medical care.

Objective: We investigated the acceptance and preferences of telemedicine among cardiologists, internists, and general practitioners. In addition, we aimed to identify knowledge, explore factors that influence the decision to adopt or reject this technology, and create starting points for demand-oriented further research.

Methods: We conducted a web-based survey between May 2021 and February 2022. The 34-item questionnaire covered a wide range of questions regarding knowledge, acceptance, and use of telemedicine in cardiology care. Participants (cardiologists, internists, and general practitioners) were contacted through their professional email addresses, through a QR code published in a regional health journal, and through X (formerly known as Twitter). After exclusion of questionnaires with missed values, multidimensional scaling and k-means clustering were performed. Participants were divided into 3 clusters (C1, C2, and C3) based on their attitudes toward telecardiology. C1 uses telemedicine for personal health and clinical practice; C2 shows reluctance; C3 uses telemedicine mainly clinically.

Results: We contacted 929 physicians. Of those 12.1% (112/929) completed the questionnaires. Participants were 56% male (54/97), 29% female (28/97), and 2% (2/97) diverse (median age 50 years). About 16% (18/112) of the respondents currently use telemedicine daily, 14.3% (16/112) 3-4 times a week, and 43% (48/112) did not use telemedicine at all. Overall, 35.1% (34/97) rated their knowledge of telemedicine as very good or good. Most of the respondents replied that telemedicine could support cardiology care in monitoring of blood pressure and electrocardiograms (57/97, 58.8%, both), consultation (57/97, 58.8%), and extending follow-up time (59/97, 60.8%). Reported barriers to implementation were mostly administration (26/97, 26.8%), inadequate reimbursement (25/97, 25.8%), and the purchase of technology equipment (23/97, 23.7%). Attitudes toward telemedicine in clinical practice were closely related to the number of patients being treated per annual quarter: C3 (median 1350, IQR 1000-1500) versus C1 (median 750, IQR 300-1200) and C2 (median 500, IQR 105-825). The differences between clinical caseloads of C1-C3 members were significant: C1 versus C2 (P=.03), C1 versus C3 (P=.02), and C2 versus C3 (P<.001). Most participants (87/112, 77.7%) would like to expand telemedicine approaches in the future. In the field of cardiology, the participants reported a high suitability of telemedicine. The willingness to train in telemedicine is high to very high for > 50% of the participants.
Conclusions: Our results indicate generally moderate use but positive attitudes toward telemedicine among participating physicians with a higher clinical caseload. The lack of a structural framework seems to be a barrier to the effective implementation of telecardiology.

(KEYWORDS)

acceptance; adoption; cardiac; cardiology; cross sectional; health services research; heart; preference; survey; telecardiology; telehealth; telemedicine

Introduction

As the burden of patients with cardiovascular diseases (CVDs) is increasing [1], regions with an aging population, such as the German federal state of Brandenburg, are particularly affected. Demographic changes concern not only the patient population but also health care professionals (HCPs). The average age of physicians in the state of Brandenburg is 54.4 years. In the next 5-10 years, the number of physicians will decrease by one third, crucially impacting medical care in the sparsely populated states of Germany. There are already 683 inhabitants per contract physician in the state of Brandenburg [2]. At the same time, the digital transformation is radically changing health care delivery [3]. In rural areas, telemedicine could help to initiate treatment faster and might have a positive impact on quality of life [4]. In cardiology, telemedicine can be used in various ways, including remote patient monitoring, remote visits, and telecardiology consultations [5]. In patients with chronic heart failure, telemedical interventions were associated with optimized medical therapy, a significant reduction in hospital readmissions, and an improvement in quality of life [4,6,7]. As reported in other medical domains, telemedicine not only potentially extends the reach to underserved populations but also enhances opportunities to provide care within usual inpatient and outpatient settings [8]. Not to be underestimated for a future-proof and sustainable health care system, there are also first indications that telemedicine could make a significant contribution to reducing the carbon footprint of health care [9].

Yet, current data on the acceptance of telemedicine by HCPs in cardiology are lacking. Furthermore, the investigation of differences in telemedicine acceptance between urban and rural areas is a research priority. Thus, we aimed to assess the acceptance of telemedicine among cardiologists, internists, and general practitioners in Berlin and Brandenburg. In addition, we aimed to identify knowledge and explore factors that influence the decision to adopt or reject this technology. To create starting points for demand-oriented further research, we wanted to identify user types in the use of telemedicine.

Methods

Ethical Considerations

The study was approved by the local ethics committee of the Brandenburg Medical School (E-01-20210304). Data processing was based on the informed consent of the participants in the study. Participation in the study was not remunerated. Personal data were only collected from the participants to be able to process any requests in accordance with current law. These data were deleted after the end of the study. No other personal data were collected.

Questionnaire and Procedure

The authors developed a web-based survey that was pretested and validated among cardiologists (n=5) and general practitioners (n=5). The final 34-item questionnaire covered a wide range of questions regarding knowledge, acceptance, and use of telemedicine in cardiology care. Physicians were asked to participate in the survey if they met the following inclusion criteria: (1) working in inpatient or outpatient cardiology care, (2) working in the states of Brandenburg or Berlin, and (3) providing informed consent. Consequently, physicians who did not meet these criteria were excluded. Furthermore, questionnaires were excluded if less than half of the questions were answered. Participants were contacted through their professional email addresses, through QR code published in a regional health journal, and through X (formerly known as Twitter). Data were collected between May 28, 2021, and February 28, 2022, using the web-based survey application “LimeSurvey” [10], embedded in the domain [11].

Data Analysis and Statistics

The statistical analysis was conducted in several steps. First, questionnaire data were analyzed using descriptive statistics, including quantities, percentages, median scores, and ranges for ordinal variables. An exploratory quantitative analysis was performed to identify factors determining telecardiology use. To create a data matrix for the analysis, qualitatively recorded participants’ attitudes toward telecardiology as reported in the survey were recoded from left to right according to the direction of the hypothesis of positive attitudes, for example, the response options in relation to the a priori hypothesis that “Would you like to use telemedicine more often?” which ranged from the positive to the negative scale. The answers “yes, totally,” “yes,” “no,” and “no at all” were coded as “5,” “4,” “2,” and “1,” respectively. The undefined “I don’t know” response option was coded with an intermediate value of “3.” Recoding was undertaken for the characteristics indicated by the age groups, the clinical location, gender, medical specialty, and type of practice. The interval-scaled variable “number of patients” (quarterly) was recoded. To ensure sample homogeneity, participants who did not use telemedicine were not included in the analysis.

To classify survey participants based on their attitudes toward telecardiology, multidimensional scaling (MDS) and subsequent k-means clustering were used. MDS aims to represent input proximities (typically dissimilarities) between objects by means of fitted distances in a low-dimensional space [12]. It therefore
visualizes the level of dissimilarity among cases in a data set. Since we were interested in scaling the participants and not attitudes, the dissimilarity measure was applied across columns of the data matrix. Then, nonmetric MDS was applied to the dissimilarity matrix to obtain the coordinates of the sample in a representative low-dimensional space. The algorithm used for the MDS calculation was “Scaling by Majorizing a Convex Function” (SMACOF [13]). Starting with any initial configuration, the algorithm iteratively transforms proximities to estimated proximities (disparities) for calculating the configuration of the items in the context of the coordinates. This continues until the squared differences between the disparities and distances are minimized. These differences reflect a model’s (mis)fit, expressed by the Stress-1 index, which ideally has a value of 0; a value higher than 0.2 indicates a bad representation [14]. However, the larger the number of points (ie, participants in this study), the more difficult it is to map these into a low-dimensional space. This means that the Stress-1 index may become unacceptably high. For this reason, the Stress-1 index was criticized by Borg et al [15], who developed permutation tests for MDS solutions [16], whereas a significant test result ($P < .05$) allows to reject the $H_0$, hypothesizing that the stress and, subsequently, the configuration are obtained from a random permutation of dissimilarities. As with other dimension-reducing methods, however, the final decision for an MDS solution should not be made based on these indices but on their interpretability [17].

To create a classification of participants based on their attitudes toward telecardiology, k-means clustering was applied to the mapped sample [18]. Clustering consists of grouping objects that are, in some sense, similar to each other. The k-means is a nonhierarchical clustering method commonly used in data mining [19]. The algorithm starts with a collection of $s$ objects, where each object is a point in a q-dimensional space, and a given number of clusters, $K$, that is subjective and specified in advance by the user. However, to determine $K$, we also relied on the total Within Sum of Squares (WSS) index. The k-means groups the “$s$” objects into $K$s clusters to minimize the objective function given by the sum of distances between the points and the centers of their clusters. The k-means arrives at a solution in which objects within each cluster are as close to each other as possible and as far from objects in other clusters as possible. Finally, the chi-square test and the Kruskal-Wallis nonparametric test [20] (due to nonnormal data distribution), followed by the posthoc analysis using the Conover-Iman test with Holm’s correction [21], were applied to examine the difference in selected variables between each cluster. A value of $P < .05$ was considered statistically significant.

All data analyses and graphics were done within the R environment (R Core Team). To run MDS, SMACOF [13], ggpubr [22], magrittr [23], and dplyr [24] packages were used. The k-means clustering was done based on stats (version 3.6.2 [25]) and factoextra [26] packages. To run Conover Test with Holm’s correction, the package connover.test [27] was used.

**Results**

**Overview**
A total of 929 physicians were contacted, of whom 112 (12.1%) responded to the questionnaires. Of which, 15 (13.4%) participants were excluded from the analysis because less than half of the questions were answered.

**Sample Characteristics**
The data for this survey were obtained from 97 physicians. Most participants—48.5% (47/97) were cardiologists, 15.5% (15/97) were internists, 12.4% (12/97) were general practitioners, and 23.7% (23/97) were not yet specialists at the time of the survey or their status was not reported at the time of the survey (Table 1). Most respondents, 56.7% (55/97), were aged between 40 and 59 years and worked in the state of Brandenburg (69/97, 71%) in medium-sized cities (40/97, 41%) (Table 1). Additionally, more than half of the participants were men (54/97, 55.7%). The ratio of respondents from the participants practicing inpatient and outpatient care was 48% versus 42% (10% with no response), with most participants in outpatient care working in a single practice. Position types, hospital characteristics according to the number of beds, and patients treated per physician per quarter are shown in Table 1.
Table 1. Demographic data and characteristics of the participants.

<table>
<thead>
<tr>
<th></th>
<th>Cardiologists (n=47), n (%)</th>
<th>Other disciplines (n=50), n (%)</th>
<th>Total (n=97), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>0 (0)</td>
<td>5 (10)</td>
<td>5 (5)</td>
</tr>
<tr>
<td>30-39</td>
<td>6 (13)</td>
<td>10 (20)</td>
<td>16 (16)</td>
</tr>
<tr>
<td>40-49</td>
<td>15 (32)</td>
<td>8 (16)</td>
<td>23 (24)</td>
</tr>
<tr>
<td>50-59</td>
<td>21 (45)</td>
<td>11 (22)</td>
<td>32 (33)</td>
</tr>
<tr>
<td>60-69</td>
<td>5 (10)</td>
<td>5 (10)</td>
<td>10 (10)</td>
</tr>
<tr>
<td>70-79</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>&gt;80</td>
<td>2 (4)</td>
<td>2 (4)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Not stated</td>
<td>0 (0)</td>
<td>9 (18)</td>
<td>9 (10)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diverse</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Female</td>
<td>14 (30)</td>
<td>14 (28)</td>
<td>28 (29)</td>
</tr>
<tr>
<td>Male</td>
<td>29 (62)</td>
<td>25 (50)</td>
<td>54 (56)</td>
</tr>
<tr>
<td>Not stated</td>
<td>2 (4)</td>
<td>11 (22)</td>
<td>13 (13)</td>
</tr>
<tr>
<td><strong>Working area</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient sector</td>
<td>17 (36)</td>
<td>24 (48)</td>
<td>41 (42)</td>
</tr>
<tr>
<td>Hospital</td>
<td>30 (64)</td>
<td>17 (34)</td>
<td>47 (48)</td>
</tr>
<tr>
<td>Not stated</td>
<td>0 (0)</td>
<td>9 (18)</td>
<td>9 (10)</td>
</tr>
<tr>
<td><strong>Medical practice types</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solo practice</td>
<td>6 (35)</td>
<td>12 (50)</td>
<td>18 (44)</td>
</tr>
<tr>
<td>Group practice</td>
<td>5 (29)</td>
<td>6 (25)</td>
<td>11 (27)</td>
</tr>
<tr>
<td>Employed physician practices</td>
<td>2 (12)</td>
<td>5 (21)</td>
<td>7 (17)</td>
</tr>
<tr>
<td>Outpatient clinic</td>
<td>2 (12)</td>
<td>0 (0)</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (12)</td>
<td>1 (4)</td>
<td>3 (7)</td>
</tr>
<tr>
<td><strong>Position types</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resident</td>
<td>0 (0)</td>
<td>10 (58)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Medical specialist</td>
<td>5 (17)</td>
<td>1 (6)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Senior physician</td>
<td>15 (50)</td>
<td>3 (18)</td>
<td>18 (38)</td>
</tr>
<tr>
<td>Consultant</td>
<td>10 (33)</td>
<td>3 (18)</td>
<td>13 (28)</td>
</tr>
<tr>
<td><strong>Hospital characteristics (number of beds)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-50</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>51-100</td>
<td>1 (3)</td>
<td>1 (6)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>101-150</td>
<td>1 (3)</td>
<td>1 (6)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>151-200</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>201-250</td>
<td>2 (7)</td>
<td>1 (6)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>251-300</td>
<td>6 (20)</td>
<td>2 (12)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>301-350</td>
<td>1 (3)</td>
<td>1 (6)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>351-400</td>
<td>4 (14)</td>
<td>2 (12)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>401-450</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>451-500</td>
<td>4 (14)</td>
<td>2 (12)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>&gt;500</td>
<td>10 (33)</td>
<td>7 (40)</td>
<td>17 (37)</td>
</tr>
</tbody>
</table>

**Number of patients (per quarter)**
Telemedicine: Knowledge and Use

Overall, 64.9% (63/97) of respondents rated their knowledge of telemedicine as satisfactory, poor, or very poor, whereas 35.1% (34/97) rated their knowledge as very good or good (Table 2). The frequency of current telemedicine use is shown in Table 2, with 16% (18/112) of the respondents currently using telemedicine daily, 14.3% (16/112) using it 3-4 times a week, and 43% (48/112) reporting no use at all. However, 45.4% (44/97) answered that they would like to use telemedicine more often in the future. Overall, 52.6% (51/97) of the physicians surveyed indicated that there were barriers to the use of telemedicine. The top 3 barriers to the implementation of telemedicine, according to respondents, were administration (26/97, 26.8%), inadequate reimbursement (25/97, 25.8%), and the purchase of technology equipment (23/97, 23.7%).

<table>
<thead>
<tr>
<th>Location (number of inhabitants)</th>
<th>Cardiologists (n=47), n (%)</th>
<th>Other disciplines (n=50), n (%)</th>
<th>Total (n=97), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural community (&lt;5000)</td>
<td>0 (0)</td>
<td>8 (16)</td>
<td>8 (8)</td>
</tr>
<tr>
<td>Small city (5000-20,000)</td>
<td>7 (14)</td>
<td>6 (12)</td>
<td>13 (13)</td>
</tr>
<tr>
<td>Medium-sized city (20,000-100,000)</td>
<td>20 (43)</td>
<td>20 (40)</td>
<td>40 (41)</td>
</tr>
<tr>
<td>Big city (&gt;100,000)</td>
<td>20 (43)</td>
<td>7 (14)</td>
<td>27 (28)</td>
</tr>
<tr>
<td>Not stated</td>
<td>0 (0)</td>
<td>9 (18)</td>
<td>9 (10)</td>
</tr>
</tbody>
</table>
Table 2. Knowledge and use of telemedicine.

<table>
<thead>
<tr>
<th>Question and responses</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you rate your own knowledge of telemedicine?</td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>9 (9)</td>
</tr>
<tr>
<td>Good</td>
<td>25 (26)</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>41 (42)</td>
</tr>
<tr>
<td>Poor</td>
<td>18 (19)</td>
</tr>
<tr>
<td>Very poor</td>
<td>4 (4)</td>
</tr>
<tr>
<td>How often do you use telemedicine?</td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>16 (16)</td>
</tr>
<tr>
<td>3-4 times a week</td>
<td>14 (14)</td>
</tr>
<tr>
<td>3-4 times a month</td>
<td>14 (14)</td>
</tr>
<tr>
<td>3-4 times a quarter</td>
<td>13 (13)</td>
</tr>
<tr>
<td>Not at all</td>
<td>40 (43)</td>
</tr>
<tr>
<td>Would you like to use telemedicine more often?</td>
<td></td>
</tr>
<tr>
<td>Yes, totally</td>
<td>25 (26)</td>
</tr>
<tr>
<td>Yes</td>
<td>19 (20)</td>
</tr>
<tr>
<td>I don’t know</td>
<td>8 (8)</td>
</tr>
<tr>
<td>No</td>
<td>4 (4)</td>
</tr>
<tr>
<td>Not at all</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Not answered</td>
<td>40 (41)</td>
</tr>
<tr>
<td>Does anything prevent you from using telemedicine?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20 (21)</td>
</tr>
<tr>
<td>Rather yes</td>
<td>31 (32)</td>
</tr>
<tr>
<td>I don’t know</td>
<td>15 (15)</td>
</tr>
<tr>
<td>Rather no</td>
<td>21 (22)</td>
</tr>
<tr>
<td>No</td>
<td>10 (10)</td>
</tr>
<tr>
<td>What prevents you from using telemedicine? (Multiple selections were possible.)</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td>26 (27)</td>
</tr>
<tr>
<td>Present insufficient reimbursement</td>
<td>25 (26)</td>
</tr>
<tr>
<td>Purchase of technology equipment</td>
<td>23 (24)</td>
</tr>
<tr>
<td>No reimbursement</td>
<td>17 (18)</td>
</tr>
<tr>
<td>Data security</td>
<td>13 (14)</td>
</tr>
<tr>
<td>Poor internet connection</td>
<td>12 (13)</td>
</tr>
<tr>
<td>Lack of data for patients benefits</td>
<td>10 (11)</td>
</tr>
</tbody>
</table>

Implementation of Telecardiology

Most of the respondents replied that telemedicine could support cardiology care in the monitoring of blood pressure and electrocardiograms (57/97, 58.8%, both), consultation (57/97, 58.8%), and extending follow-up time (59/97, 60.8%) (Table 3). When asked which communication partners they should exchange through telemedicine, 80.4% (78/97) responded “physician-to-patient,” 72.2% (70/97) responded “physician-to-physician,” and 51.5% (50/97) responded “physician-to-assistant or other participants” (multiple replies were possible). According to the respondents, the diseases or conditions that are particularly suitable for telemedicine care include “cardiac arrhythmias” (78/97, 80.4%), “monitoring of various diseases and conditions” (75/97, 77.3%), and “therapy” (74/97, 76.3%) (Table 3).
Table 3. Implementation of telemedicine in cardiology care.

<table>
<thead>
<tr>
<th>Question and responses</th>
<th>Yes, n (%)</th>
<th>No, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Which partners should establish communication through telemedicine? (Multiple selections were possible.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician-patient</td>
<td>78 (80)</td>
<td>19 (20)</td>
</tr>
<tr>
<td>Physician-physician</td>
<td>70 (72)</td>
<td>27 (28)</td>
</tr>
<tr>
<td>Physician-assistant</td>
<td>34 (35)</td>
<td>63 (65)</td>
</tr>
<tr>
<td>Other participants and combinations</td>
<td>16 (17)</td>
<td>81 (83)</td>
</tr>
<tr>
<td>No communication</td>
<td>4 (5)</td>
<td>93 (95)</td>
</tr>
<tr>
<td><strong>At which stages can telemedicine support cardiological care? (Multiple selections were possible.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Widen the time of follow-ups</td>
<td>59 (61)</td>
<td>38 (39)</td>
</tr>
<tr>
<td>Consultation</td>
<td>57 (59)</td>
<td>40 (41)</td>
</tr>
<tr>
<td>Blood pressure monitoring</td>
<td>57 (59)</td>
<td>40 (41)</td>
</tr>
<tr>
<td>ECG\textsuperscript{a} monitoring</td>
<td>57 (59)</td>
<td>40 (41)</td>
</tr>
<tr>
<td>Acute situations (eg, sending ECG to the hospital)</td>
<td>53 (55)</td>
<td>44 (45)</td>
</tr>
<tr>
<td>Weight monitoring</td>
<td>49 (51)</td>
<td>48 (49)</td>
</tr>
<tr>
<td>Complications</td>
<td>30 (31)</td>
<td>67 (69)</td>
</tr>
<tr>
<td>At no stage</td>
<td>2 (2)</td>
<td>95 (98)</td>
</tr>
<tr>
<td><strong>Which cardiologic diseases could be monitored by telemedicine?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td>78 (80)</td>
<td>19 (20)</td>
</tr>
<tr>
<td>Monitoring of various diseases and conditions</td>
<td>75 (77)</td>
<td>22 (23)</td>
</tr>
<tr>
<td>Therapy</td>
<td>74 (76)</td>
<td>23 (24)</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>73 (75)</td>
<td>24 (25)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>56 (58)</td>
<td>41 (42)</td>
</tr>
<tr>
<td>No disease</td>
<td>2 (2)</td>
<td>95 (98)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}ECG: electrocardiogram.

**Telecardiology User Groups**

The exclusion of the cases with missed values yielded a data matrix of 53 cases. Figure 1 shows a 2D MDS solution for the distribution of survey participants. The Stress-1 index was borderline (stress 0.20). However, the permutation test indicated a well-fitting model ($P<.001$). Considering the fact that the next step of the analysis involved more detailed clustering using k-means, the 2D solution was sufficient and more complex solutions were not examined. Determining the cluster numbers with the WSS index resulted in a 3-cluster solution. The application of this solution for the segmentation of the study participants using k-means showed that all 3 clusters were localized and ordered according to the axes of the MDS diagram. Cluster 1 (C1, n=19) and cluster 2 (C2, n=21) were in the lower left and right quadrants, and cluster 3 (C3, n=13) was in the upper right quadrant of the diagram. Table 4 shows that physicians assigned to group C1 used telemedicine privately to improve their personal health, yet not only in their clinical practice. Physicians in group C2 showed reluctant attitudes toward telemedicine. Members of group C3 use telemedicine for clinical activities and not for their personal health. Attitudes toward telemedicine in clinical activities were closely related to the number of patients being treated. This could be confirmed by a higher patients’ number per annual quarter (median 1350, IQR 1000-1500) treated by C3 members compared with the corresponding numbers indicated by members of C1 (median 750, IQR 300-1200) and C2 (median 500, IQR 105-825). These differences were statistically significant with C1 versus C2 ($P=.03$), C1 versus C3 ($P=.02$), and C2 versus C3 ($P<.001$). The comparison of other characteristics did not have any statistical significance.
Figure 1. Segmentation of survey participants in a 2D group space. Numbers are identification numbers of participants. Cluster 1 (C1, n=19) and cluster 2 (C2, n=21) are in the lower left and right quadrants, and cluster 3 (C3, n=13) is in the upper right quadrant of the diagram. Telecardiology user groups were identified through cluster analysis of a web-based survey conducted between May 2021 and February 2022 to assess telemedicine knowledge, acceptance, and use. Analysis of 53 cases reveals 3 distinct clusters (C1, C2, and C3) based on use behavior. C1 uses telemedicine for personal health and clinical practice; C2 shows reluctance; and C3 uses telemedicine mainly clinically. Statistically significant differences were observed: C1 versus C2 \((P=.03)\), C1 versus C3 \((P=.02)\), and C2 versus C3 \((P<.001)\).
<table>
<thead>
<tr>
<th>Item</th>
<th>Cluster</th>
<th>K-W test&lt;sup&gt;a&lt;/sup&gt;</th>
<th>C-I test&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value</th>
<th>C-I test&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value</th>
<th>C-I test&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you rate your own knowledge of telemedicine?</td>
<td>C1 (n=19)</td>
<td>4 (3-4)</td>
<td>3.53 (0.96)</td>
<td>.12&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.11</td>
<td>.14</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2 (n=21)</td>
<td>3.48 (0.81)</td>
<td>4 (3-4)</td>
<td>3.40 (0.82)</td>
<td>4 (3-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>4.00 (0.82)</td>
<td>3 (3-4)</td>
<td>4.00 (0.82)</td>
<td>4 (3-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often do you use telemedicine?</td>
<td>C1 (n=19)</td>
<td>4 (3-4)</td>
<td>3.53 (1.02)</td>
<td>.02</td>
<td>.31</td>
<td>.02</td>
<td>.0096</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2 (n=21)</td>
<td>3.33 (1.15)</td>
<td>3 (2-4)</td>
<td>3.38 (0.96)</td>
<td>5 (4-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>3.43 (1.15)</td>
<td>3 (2-4)</td>
<td>3.46 (0.92)</td>
<td>5 (4-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Would you like to use telemedicine more often?</td>
<td>C1 (n=19)</td>
<td>5 (4-5)</td>
<td>4.47 (0.77)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.21</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2 (n=21)</td>
<td>4.52 (1.03)</td>
<td>4 (3-4)</td>
<td>4.62 (0.87)</td>
<td>5 (5-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>4.52 (1.03)</td>
<td>4 (3-4)</td>
<td>4.62 (0.87)</td>
<td>5 (5-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does anything prevent you from using telemedicine? (reverse coding)</td>
<td>C1 (n=19)</td>
<td>2 (2-3-5)</td>
<td>2.74 (1.41)</td>
<td>.22</td>
<td>.20</td>
<td>.17</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2 (n=21)</td>
<td>2.48 (1.17)</td>
<td>2 (2-3)</td>
<td>3.15 (1.28)</td>
<td>4 (2-4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>2.48 (1.17)</td>
<td>2 (2-3)</td>
<td>3.15 (1.28)</td>
<td>4 (2-4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How do you assess the need for relevant training on telemedicine am</td>
<td>C1 (n=19)</td>
<td>4 (4-5)</td>
<td>4.37 (0.60)</td>
<td>.001</td>
<td>.001</td>
<td>.09</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>mong colleagues?</td>
<td>C2 (n=21)</td>
<td>3.14 (0.85)</td>
<td>3 (3-4)</td>
<td>4.23 (0.60)</td>
<td>4 (4-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>3.14 (0.85)</td>
<td>3 (3-4)</td>
<td>4.23 (0.60)</td>
<td>4 (4-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How do you assess the willingness of colleagues to undergo further</td>
<td>C1 (n=19)</td>
<td>4 (2-5-4)</td>
<td>3.32 (1.06)</td>
<td>.21</td>
<td>.17</td>
<td>.20</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>training on telemedicine?</td>
<td>C2 (n=21)</td>
<td>2.86 (0.85)</td>
<td>3 (2-3)</td>
<td>3.38 (0.96)</td>
<td>4 (3-4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>2.86 (0.85)</td>
<td>3 (2-3)</td>
<td>3.38 (0.96)</td>
<td>4 (3-4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How high is your own willingness to participate in training courses</td>
<td>C1 (n=19)</td>
<td>5 (4-5)</td>
<td>4.26 (0.93)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.08</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>on telemedicine?</td>
<td>C2 (n=21)</td>
<td>3.05 (0.86)</td>
<td>3 (3-4)</td>
<td>4.15 (0.69)</td>
<td>4 (4-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>3.05 (0.86)</td>
<td>3 (3-4)</td>
<td>4.15 (0.69)</td>
<td>4 (4-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Would you be willing to financially invest in the application of</td>
<td>C1 (n=19)</td>
<td>4 (3-4)</td>
<td>3.79 (0.63)</td>
<td>.008</td>
<td>.007</td>
<td>.07</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>telemedicine in your everyday care routine?</td>
<td>C2 (n=21)</td>
<td>3.00 (0.89)</td>
<td>3 (2-4)</td>
<td>3.92 (0.86)</td>
<td>4 (3-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>3.00 (0.89)</td>
<td>3 (2-4)</td>
<td>3.92 (0.86)</td>
<td>4 (3-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you use telemedicine applications (privately) for your own health?</td>
<td>C1 (n=19)</td>
<td>4 (4-5)</td>
<td>4.32 (0.75)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2 (n=21)</td>
<td>2.10 (1.04)</td>
<td>2 (1-2)</td>
<td>1.69 (0.85)</td>
<td>2 (1-2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3 (n=13)</td>
<td>2.10 (1.04)</td>
<td>2 (1-2)</td>
<td>1.69 (0.85)</td>
<td>2 (1-2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>K-W test: Kruskal-Wallis-test; P values displayed.
<sup>b</sup>C-I test: Conover-Iman test with Holm correction; P values displayed.
<sup>c</sup>NS: Not significant.

**Discussion**

Cardiologists, internists, and general practitioners consider the overall use of telecardiology to be acceptable; two-thirds of respondents would like to use telemedicine in their daily practice. However, most physicians rate their knowledge as “satisfactory” or worse, and less than a third were using telemedicine at the time of the survey. Barriers to telemedicine adoption, such as “limited knowledge,” “administrative burden,” “purchase of technology,” and “inadequate reimbursement,” were clearly identified by both specialists and generalists. Direct communication with patients is preferred to information exchange with colleagues. In the exploratory analysis, we found 3 potential telecardiology user groups, which differ regarding the number of patients treated per quarter: The more patients treated, the higher the telemedicine acceptance rate. The results provide information on how telemedicine can support cardiology care from the physicians’ perspective.

In 2021, a number of changes were introduced for telemedicine in Germany, such as the mandatory electronic patient record, an increasing number of prescriptible digital health applications,
and video consultations for nonphysician therapists. The study TIM-HF II [7] has shown the world that telemedical interventional management can reduce mortality. Since January 2022, remote patient monitoring of patients with heart failure has been reimbursed by the statutory health insurance funds in Germany [28]. In addition to the reimbursement of costs, however, the expansion of telemedicine in the real world is the central topic of the implementation process that is now beginning. Yet, it is surprising that despite the successful study and the establishment of telemedicine infrastructure for cardiology care in the federal states of Brandenburg and Berlin, telemedicine acceptance among physicians in routine cardiology care is still heterogeneous. Considering further large-scale research activity on telehealth for prevention in hypertension care in this region [29], there seems to be a wide evidence-based practice gap for telemedicine in cardiology care [30]. Our results support this conclusion, as knowledge of telemedicine has been reported as low by the participants in this survey. Thus, we recommend high-quality training programs that reflect the multidimensionality of knowledge barriers by addressing the economic, organizational, and behavioral framework conditions of digital health implementation [31]. Furthermore, our results indicate that bureaucratic and infrastructural barriers hamper telemedicine implementation. These barriers were also identified in a similar study on telemedicine in rheumatology care, conducted before the COVID-19 outbreak [32]. This suggests the reported barriers to effective use of telemedicine have remained in Germany despite the pandemic and a massive digital health uptake globally.

COVID-19 has demonstrated the importance and acceptance of contactless approaches to medical care and, particularly, cardiology care [33-35]. Also, gold standard adherence measures in telemedical interventions need to be established so that study outcomes are more comparable [35]. As the survey was published in May 2021, it is not derivable from our data whether the willingness to use telecardiology has changed. Only a minority of the surveyed physicians currently use telemedicine, although two-thirds would like to implement telemedicine into their clinical routine.

It seems that the participants foreshadow and recognize a benefit that has already been described in the literature [7,36] but which does not yet seem measurable in its daily representatives due to poor global implementation of telemedicine. Due to that lack of everyday application, it seems understandable that most physicians regard their knowledge of telemedicine as poor.

As physicians reported barriers to the use of telemedicine, the structural framework for effective implementation of telecardiology is not yet in place. Significant administrative burdens and inadequate reimbursement structures prevented the physicians surveyed from using telemedicine. The greatest barrier seems to be physicians’ limited knowledge about “how to use telemedicine.” This underlines the need for clearly defined use cases for telemedicine in cardiology as well as the timely introduction of low-threshold training offers. Overall, this seems to reflect that the potential of telemedicine is not being fully reached. Further research should define use cases as well as specific interventions and evaluate the effects on patients’ outcomes and health and economic implications. Those seem particularly important because our data suggest that in the current health care system, only what is paid for is done. An increasingly aging society with an increasingly scarce resource of highly specialized doctors is catalyzing the need for enrollment in telecardiology under the aspects that PerplexityAI has already summarized: “telecardiology has the potential (…) to improve patient engagement and save time and money for patients and health care providers” [1]. Further research may therefore provide individualized patient- and clinician-adapted telemedicine options and triage mechanisms to select patients for either digital or analog consultations as appropriate. Based on our data, participants accept telemedicine and support its expansion if framework conditions such as reimbursement, the removal of existing usability barriers, and specialized training are optimized. The authors see this as a call to the health care system to create a framework for the use of telecardiology to optimize the use of an increasingly scarce resource with increasing demands and workloads. The provision of high-quality cardiology care using telemedicine will require urgent research, as well as the removal of existing barriers and training for specialists and generalists.

Due to the design of our questionnaire as a web-based survey, we assume a positive selection bias for physicians who are already interested in digitalization, telemedicine, or telecardiology. In addition to digital invitations to participate, participants were also recruited by an analog magazine report about the research project in the “KV-intern,” which was delivered to every physician registered with the Medical Association Brandenburg (Kassenärztliche Vereinigung Brandenburg). Either way, internet access and at least a certain level of digital expertise were required for participation. As the average age of the participants is comparable to the average age of physicians in Brandenburg [2], the group of participants surveyed nevertheless appears to be representative.

At this point, we have only explored the perspectives of physicians on telemedicine in cardiology. There is an urgent need to investigate the patients’ perspective on telemedicine implementation in cardiology care.

In summary, our results indicate low use but high acceptance among participating physicians. Although potential users report general willingness and the potential benefits of telemedicine, self-reported knowledge is limited. The lack of a structural framework seems to be a barrier to the effective implementation of telecardiology.

Acknowledgments

The authors would like to thank Prof Heiner Gembris for reviewing and advising on the web-based questionnaire. Funded by the Brandenburg Medical School (MHB) publication fund, supported by the Ministry of Science, Research and Cultural Affairs of the state of Brandenburg. Generative artificial intelligence was not used in any proportion of the manuscript.
Data Availability
The data sets generated during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
FM and AHF conceptualized the study. The methodology was developed by FM, AHF, JN, and MJH. Formal analysis and investigation were carried out by YI, FM, and AHF. The original draft of the manuscript was prepared by MJH, JN, FM, and AHF. YI, MH, and CB reviewed and edited the manuscript. AHF provided supervision for the entire project. All authors have read and approved the final manuscript. FM and JH contributed equally to this work.

Conflicts of Interest
None declared.

References
11. Telemedizinbrandenburg. URL: https://www.telemedizinbrandenburg.de/ [accessed 2024-02-01]


25. R Core Team and contributors worldwide. The R Stats Package. URL: https://rdrr.io/r/stats/stats-package.html [accessed 2024-02-06]


Abbreviations

CVD: cardiovascular disease
HCP: health care professional
MDS: multidimensional scaling
SMACOF: Scaling by Majorizing a Convex Function
WSS: Within Sum of Squares
©Felix Muehlensiepen, Marie Josephine Hoffmann, Jonathan Nübel, Yury Ignatyev, Martin Heinze, Christian Butter, Anja Haase-Fielitz. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Enhancing Health Care Accessibility and Equity Through a Geoprocessing Toolbox for Spatial Accessibility Analysis: Development and Case Study

Soheil Hashtarkhani¹, PhD; David L Schwartz², MD; Arash Shaban-Nejad¹, MPH, MSc, PhD

¹Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, The University of Tennessee Health Science Center, Memphis, TN, United States
²Department of Radiation Oncology, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

Corresponding Author:
Arash Shaban-Nejad, MPH, MSc, PhD
Center for Biomedical Informatics, Department of Pediatrics, College of Medicine
The University of Tennessee Health Science Center
50 N Dunlap Street, R492
Memphis, TN, 38103
United States
Phone: 1 9012875863
Email: ashabann@uthsc.edu

Abstract

Background: Access to health care services is a critical determinant of population health and well-being. Measuring spatial accessibility to health services is essential for understanding health care distribution and addressing potential inequities.

Objective: In this study, we developed a geoprocessing toolbox including Python script tools for the ArcGIS Pro environment to measure the spatial accessibility of health services using both classic and enhanced versions of the 2-step floating catchment area method.

Methods: Each of our tools incorporated both distance buffers and travel time catchments to calculate accessibility scores based on users’ choices. Additionally, we developed a separate tool to create travel time catchments that is compatible with both locally available network data sets and ArcGIS Online data sources. We conducted a case study focusing on the accessibility of hemodialysis services in the state of Tennessee using the 4 versions of the accessibility tools. Notably, the calculation of the target population considered age as a significant nonspatial factor influencing hemodialysis service accessibility. Weighted populations were calculated using end-stage renal disease incidence rates in different age groups.

Results: The implemented tools are made accessible through ArcGIS Online for free use by the research community. The case study revealed disparities in the accessibility of hemodialysis services, with urban areas demonstrating higher scores compared to rural and suburban regions.

Conclusions: These geoprocessing tools can serve as valuable decision-support resources for health care providers, organizations, and policy makers to improve equitable access to health care services. This comprehensive approach to measuring spatial accessibility can empower health care stakeholders to address health care distribution challenges effectively.

(JMIR Form Res 2024;8:e51727) doi:10.2196/51727

KEYWORDS
geographical information system; geoprocessing tool; health disparities; health equity; health services management; hemodialysis services; spatial accessibility

Introduction

The role of geography in understanding and addressing population health and health inequities is hardly deniable [1-3]. Access to health care is a critical indicator of health care system performance and directly impacts population health and disease burden [4,5]. Improving access to primary care, for instance, has been proven to lead to improved health outcomes and decreased potentially avoidable hospitalizations [6,7]. The concept of access plays a significant role in health services and policy research, including both spatial and nonspatial
dimensions. Nonspatial access refers to the factors unrelated to geography that influence access, such as affordability, timeliness, accommodation, acceptability, and awareness [8]. Spatial access, on the other hand, involves the geographic elements that influence the availability and accessibility of health care providers and services [9]. The calculation of spatial accessibility involves considering three key factors: (1) supply, (2) demand, and (3) mobility. Supply relates to the infrastructure’s locations (eg, health care providers); demand refers to the locations of individuals who are expected to use the infrastructure (eg, patients); and mobility considers the travel costs between demand and supply locations (eg, driving time) [10]. Identifying areas with limited spatial accessibility enables planners and policy makers to understand the distribution of health service locations and reveal and address spatial inequities [11].

Various methods have been used to evaluate spatial accessibility, including gravity models [12], regional availability models [13], and kernel density models [14]. Among the gravity models, the 2-step floating catchment area (2SFCA) method, initially introduced by Radke and Mu [15] and modified by Luo and Wang [16], has been widely used in the literature for measuring spatial accessibility. The 2SFCA approach measures spatial accessibility through a 2-step procedure based on the interaction between supply and demand within a certain catchment as a ratio of provider-to-population [16]. However, the 2SFCA technique has certain limitations. Locations outside the catchment area are entirely out of access, while population locations within the catchment area are assumed to have equal access to health care providers [17]. To overcome these limitations, Luo and Qi [18] proposed an enhanced version of the 2SFCA method known as the enhanced 2SFCA (E2SFCA) method. This enhanced approach differentiates accessibility within a catchment (usually 3 catchments) by incorporating multiple travel time zones and assigning weights based on a decay function within each catchment [18].

Although the advancements in geographic information system (GIS) technology have made the implementation of the 2SFCA-based models more feasible, researchers with limited GIS expertise still face challenges in gathering, preprocessing, analyzing required data sets, and implementing the model. However, GIS software like ArcGIS provides powerful tools, including model builders and Python programming tools, that enable developers to automate data processing by creating custom geoprocessing toolboxes. In this study, our objective is to develop and share Python script tools for implementing 2SFCA and E2SFCA methods in ArcGIS Pro (Esri). Each toolbox was developed in 2 ways: using a distance buffer and travel time (driving time or walking time) in catchment areas. Additionally, we will present a case study assessing the accessibility of hemodialysis services. In this study, we will measure the accessibility score using the 4 developed tools for the age-adjusted demand population in census tracts of the state of Tennessee to include an important nonspatial factor in the accessibility score.

### Methods

#### Theoretical Framework and Conceptual Explanations

A total of 4 geoprocessing tools were developed to work in ArcGIS Pro software based on 2SFCA [16] and E2SFCA [18] approaches with the following names:

1. 2SFCA01: 2SFCA with buffer distance catchments
2. E2SFCA01: E2SFCA with buffer distance catchments
3. 2SFCA02: 2SFCA with travel time catchments
4. E2SFCA02: E2SFCA with travel time catchments

After presenting the theoretical background of the 2SFCA and E2SFCA approaches, the development framework for each tool will be presented.

#### Models’ Theory

**Theoretical Background of 2SFCA**

The 2SFCA method assesses the relationship between resource availability and demand population distribution in 2 steps, resulting in an access score for each demand area.

**Step 1:** For each facility location (j), identify all demand locations (k) that fall within a specified catchment area ($d_j$) from the facility. The provider-to-demand ratio ($R_j$) within the catchment area is calculated using equation 1:

\[
R_j = \frac{\sum P_k}{\sum S_j}
\]

where $P_k$ represents the population at demand location $k$ within catchment area $j$ ($d_j \leq d_0$), $S_j$ is the capacity or number of providers at location $j$, and $d_{ij}$ is the distance (or travel time) between $k$ and $j$.

**Step 2:** For each demand location $i$, search for all facility locations (j) within the specified catchment area ($d_{ij}$) from location $i$, and calculate the summed provider-to-demand ratios ($R_i$) obtained in step 1 using equation 2:

\[
R_i = \sum R_j
\]

In equation 2, $R_i$ represents the accessibility at demand location $i$ based on the 2SFCA method. $R_j$ denotes the provider-to-demand ratio at facility location $j$ that falls within the catchment area of the demand location $i$ (ie, $d_{ij} \leq d_0$), and $d_{ij}$ is the distance (or travel time) between $i$ and $j$.

**Theoretical Background of E2SFCA**

The classic 2SFCA relies on a dichotomous distance decay function, assuming that individuals within catchment areas have equal access to services, while those outside catchment areas have no access at all. To overcome the distance decay limitation of the classic 2SFCA, we also used the E2SFCA procedure introduced by Luo and Qi [18] as follows:

In step 1, for each facility location (j), 3 distance or travel time catchment areas are created, including zone 1: 0-5 miles (0-8 km); zone 2: 5-10 miles (8-16 km); and zone 3: 10-15 miles (16-24 km) or minutes. Search all demand locations (k) that
were within the zones \((D_r)\) for location \(j\) and compute the weighted provider-to-demand ratio \((R_j)\) using equation 3:

\[
R_j = \frac{P_k}{S_j D_r} \sum_{j=1}^{D_r} W_r d_{kj}
\]

where \(P_k\) is part of the demand \(k\) falling within the catchment \(j\) \((d_{kj} D_r)\), \(S_j\) is the capacity or number of providers at facility \(j\), \(d_{kj}\) the distance (or travel time) between \(k\) and \(j\), and \(D_r\) is the \(r\)th catchment zone \((r \{1,2,3\})\) within the catchment. \(W_r\) is the distance weight for the \(r\)th zone calculated from the Gaussian function capturing the distance decay of access to the facility \(j\).

Step 2: For each demand location \(i\), search all facility locations \((j)\) within the distance (or travel time) threshold of the location \(i\), and summed up the provider-to-demand ratios \(R_j\) (calculated in step 1) as follows:

\[
\text{Accessibility at location } i = \sum_{j=1}^{D_i} R_j d_{ij}
\]

where \(\Delta\) represents the accessibility at the demand location \(i\), \(R_j\) is the provider-to-demand ratio at facility location \(j\) that falls within the catchment of demand location \(i\) \((d_{ij} D_r)\), and \(d_{ij}\) is the distance (or travel time) between \(i\) and \(j\). The same distance weights derived from the Gaussian function used in step 1 are applied to each zone to account for the distance decay.

**Tool Frameworks**

**Overview**

The accessibility tools in this study use 2 spatial data sets as input, provider and population data, and 1 output data set, as described in the following paragraphs. The framework for developing each tool is described in their respective subsections.

**Provider Data**

These are point data showing the location of health service providers (eg, hospitals) and contain an ID field and a capacity field. The capacity field is a numeric field including the number of providers (eg, number of physicians) or the number of resources (eg, number of hospital beds).

**Population Data**

These are polygon data that contain the geographical areas where the accessibility score is supposed to be calculated (eg, census tracts) and contain an ID and population field. The population field is a numerical field that represents the demand for service. It could be the total population of a census tract or the number of women, children, or older adults.

**Output Feature Class**

This polygon feature class is exactly similar to the population input feature class, with 1 added field named “final index.” This field demonstrates the accessibility score of service providers in the input regions.

**Framework for Developing 2SFCA01**

Figure 1A shows the simplified procedure used to develop the 2SFCA01 tool. Some preprocessing steps, including duplicating the input data file and creating temporary fields, have been done for each input. Then, in step 1, using the input parameters from the user, the output is created, named “Step 1 output.” This feature class is similar to input provider data, with a new field named provider to demand representing the ratio of providers to the demand population in the catchment of each provider facility. Step 2 is relatively straightforward and uses the same buffer size as step 1 to sum up the provider-to-demand values calculated in step 1 for each population area.

Figure 2A shows the screenshot of the 2SFCA01 tool. Users can easily select the input data and fields using combo boxes. As the value of the accessibility score is usually very small, the “per capita” parameter multiplies the score with a user-defined value.
Framework for Developing E2SFCA01

The procedure to create the E2SFCA01 tool is similar to the 2SFCA01 tool (Figure 1B). As in this model, 3 catchments are necessary for each step, and more parameters from users are required to be defined (circle shapes). The calculated values for each catchment size are combined using user-defined weights. In this tool, the output file not only includes the final index.
value but also includes the accessibility values for each of the 3 catchment sizes.

Figure 2B shows a screenshot of the toolbox. The default values for distance values are 5, 10, and 15 miles (8, 16, and 24 km) and weights: 1, 0.68, and 0.22, respectively. These default weight sets for distance decay were derived from the original study that developed the model [18], but users should choose the proper weights regarding the purpose and context of the study.

(a) Framework for Developing 2SFCA02

The accessibility tools developed with travel time catchments need 2 separate tools to conduct the analysis. The first user should create travel time catchment data for input data using a tool that we developed named “create travel time catchment areas” and then calculate the accessibility index using the 2SFCA02 tool. In order to run the “create travel time catchment areas” tool, users have 2 options. First, if they have a network data set for their study area, they can use it in this tool to create the necessary catchments. If not, they can select ArcGIS Online resources to capture the driving or walking time catchment areas. In this way, the user should have a Network Analysis license with enough credits to use this tool. As shown in Figure 3A, in addition to provider and population data, the ID fields should be defined by the user in this tool. It is necessary to include the output of this tool in the 2SFCA02 and E2SFCA02 tools as input. A value of 10 minutes of driving time is used as the default value, but users can change it to walking distance. Also, the time, date, and direction of the travel can be customized to consider the traffic, as the catchment sizes will be smaller during rush hour than at other times of the day. Two output files, 1 for the provider and 1 for the population data, will be exported. The main 2SFCA02 tool is shown in Figure 3B. This tool has 4 input files: 2 for the provider and population data, and 2 for the travel time catchments derived from the previous step. The ID fields that are specified in the “create travel time catchment area” tool should be specified. Other features of the toolbox and the process are similar to 2SFCA01.

Figure 3. Screenshots of the proposed tools: (A) Create Travel Time Catchments tool, (B) the 2-step floating catchment areas with travel time catchments (2SFCA02) tool, and (C) the enhanced 2-step floating catchment areas with travel time catchments (E2SFCA02) tool.

(b) Framework for Developing E2SFCA02

To use this tool, the user should use the create travel time catchment areas tool similar to adding 3 travel time values, for example, 10 minutes, 15 minutes, and 20 minutes of driving time. This will create 3 travel time rings around each provider facility and demand population point. Figure 3C demonstrates the E2SFCA02 toolbox to create accessibility.

Python scripts for all of the tools are available in Multimedia Appendix 1, and our implemented executable tools are available for download elsewhere [19].

Ethical Considerations

This study and the development of the presented tool did not involve data that requires ethical oversight. The case study in this manuscript used publicly available data focusing on dialysis center locations and population age distribution at the census tract level. This secondary analysis of aggregated, open-source data is exempt from institutional review board review, in accordance with the Federal Policy for the Protection of Human Subjects (45 CFR 46).
Results

Case Study: Access to Hemodialysis Services in the State of Tennessee

To demonstrate the practicality of the proposed tools in real-world scenarios, we aimed to assess the accessibility of hemodialysis services in different areas of the state of Tennessee.

Hemodialysis is a crucial treatment for individuals with end-stage kidney disease (ESKD), as it eliminates waste products and extra fluid from the blood when the kidneys can no longer do this on their own. Without this treatment, people with ESKD would quickly develop life-threatening complications. This is why the geographical accessibility of hemodialysis services is a critical issue. A study showed that patients living 60 minutes away from a hemodialysis center not only run an increased risk of mortality but also have a significantly lower quality of life compared with patients living 15 minutes or less away [20].

Input Data Sets

Provider Data

The address and location of hemodialysis centers and the number of machines in each center have been extracted from the Centers for Medicare and Medicaid Services [21]. The address data for the 192 hemodialysis centers have been geocoded into coordinates using the ArcGIS world geocoding service. The resulting shapefile includes ID and capacity (number of machines) fields imported to ArcGIS Pro software.

Population Data

The shapefile of census tracts in the state of Tennessee and their total population has been downloaded from the US Census Bureau website [22]. The state of Tennessee includes 1701 census tracts with an average population of 3981 (SD 1646) people. To have a better proxy of target demand, we adjusted the total population with the age distribution of each census tract. Researchers can adjust the target population based on demand and health care needs [23]. We derived the incidence rates of ESKD for various age groups from the 2020 report by the US Department of Health and Human Services [24]. According to the report, the incidence rate of ESKD among individuals aged between 0 and 12 years is 11 cases per million, whereas it is 2080 cases per million for those aged 75 years or older. Table 1 details the age-adjusted demand for ESKD health care services for each census tract, calculated using the following formula: Age-adjusted demand = \( N_{0-17} \times 1 + N_{18-44} \times 7 + N_{45-64} \times 51 + N_{65-74} \times 106 + N_{\geq 75} \times 189 \) where \( N_{a-b} \) represents the number of individuals in the census tract aged \( a \) to \( b \) years.

The shapefile of census tracts in Tennessee, including GeoID as an ID field and age-adjusted demand as a population field, was imported into ArcGIS Pro software. Figure 4 demonstrates the location of hemodialysis centers and the population density of census tracts in Tennessee.

Table 1. End-stage kidney disease (ESKD) incidence rate in different age groups of the US population and calculated weighted demand values.

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>ESKD incidence rate per million, n</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-12</td>
<td>11</td>
<td>1 (baseline)</td>
</tr>
<tr>
<td>18-44</td>
<td>77</td>
<td>7</td>
</tr>
<tr>
<td>45-64</td>
<td>561</td>
<td>51</td>
</tr>
<tr>
<td>65-74</td>
<td>1171</td>
<td>106</td>
</tr>
<tr>
<td>( \geq 75 )</td>
<td>2080</td>
<td>189</td>
</tr>
</tbody>
</table>
Mobility

To measure the accessibility indexes, we used a distance buffer size of 15 miles (24 km) for the 2SFCA01 tool and 5, 10, and 15 miles (8, 16, and 24 km) for E2SFCA01. Also, for the 2SFCA02 tool, we used 30 minutes’ drive-time catchments and 10, 20, and 30 minutes for the E2SFCA tool. The distance decay sets of 1, 0.68, and 0.22 have been used for weighting each catchment in enhanced versions. The resulting accessibility index for each tool was symbolized in a geographical map using natural break classification (Figure 5).

As depicted in Figure 5, each tool generates distinct accessibility scores in the state of Tennessee, although the overall trends remain largely consistent. Rural and suburban areas generally exhibit lower access scores compared to urban areas, where a concentration of hemodialysis centers is observed. The 2SFCA tool’s findings reveal that in areas with high access for every 100,000 people, there are between 12.9 and 27.7 dialysis machines available within a 15-mile (24-km) radius or from 15.7 to 21.2 machines accessible within a 30-minute travel time. Conversely, in regions with low access, the availability of these resources is nearly nonexistent. Interpreting the results from the E2SFCA tool is not straightforward due to its weighted measurement approach. Notably, it is evident that areas represented by white, indicating a lack of health care resources within the distance thresholds, should be prioritized for informed resource allocation efforts [25].

Figure 5. Access to hemodialysis centers in Tennessee using the following tools: (A) 2-step floating catchment areas with buffer distance catchments (2SFCA01), (B) enhanced 2-step floating catchment areas with buffer distance catchments (E2SFCA01), (C) 2-step floating catchment areas with travel time catchments (2SFCA02), and (D) enhanced 2-step floating catchment areas with travel time catchments (E2SFCA2); 1 mile=1.6 km.
Discussion

Overview

The primary objective of this study was to introduce GIS tools for measuring the accessibility of health care resources, specifically focusing on the widely used 2SFCA model and its enhanced versions. 2SFCA is the most popular model for measuring the accessibility of health care resources in the literature, and many extensions have been introduced to improve its functionality [26,27]. Our proposed tools aim to assist the health research community in identifying underserved areas in terms of health care accessibility. The development of these tools can significantly streamline the process of assessing and addressing spatial disparities in health care access.

The classic 2SFCA tools (2SFCA01 and 2SFCA02) offer the advantage of simplicity in interpretation. For policy makers, an access score of 20 per 100,000 with a 60-minute catchment size means that there are 20 health care providers accessible for every 100,000 individuals within a 60-minute drive time. This straightforward interpretation facilitates policy makers’ understanding of accessibility. On the other hand, the enhanced versions (E2SFCA01 and E2SFCA02) use weighted scores in each step, making the interpretation more complex. However, the use of a distance decay function in the enhanced versions helps overcome the limitations of the classic 2SFCA model and makes it more accurate for comparing the accessibility of different regions.

The use of travel time catchments in 2SFCA02 and E2SFCA02 tools has several advantages. First, travel time offers a more precise measure of accessibility as it takes into account factors such as traffic congestion, road type, and urbanization factors. It can also accommodate different modes of mobility, including walking time. However, it is essential to note that using travel time tools requires access to a Network Analysis license with sufficient ArcGIS Online credits. Each travel time calculation consumes approximately 0.5 ArcGIS Online credits. In the case study, we analyzed the E2SFCA02 tool with 2800 credits, considering 1701 census tracts and 192 hemodialysis centers in Tennessee.

We introduced the Create Travel Time Catchments tool as a standalone prerequisite for 2SFCA02 and E2SFCA02. This approach enables users to generate catchment areas for health care providers and the population, facilitating multiple runs of the access model without incurring additional time and cost for the initial step. To make travel time calculations more adaptable, we designed the tool with flexible options. Users with a network data set covering their study area can compute travel time catchments without consuming ArcGIS Online credits. For those lacking a local network data set but having sufficient ArcGIS Online credits, the tool can leverage web-based resources. However, in the absence of both a network data set and ArcGIS Online credits, the 2SFCA01 and E2SFCA01 tools are viable alternatives, as they use simple Euclidean distance buffers for analysis. Ideally, the E2SFCA02 tool, which incorporates both distance decay and travel time catchments, offers a more realistic measure that closely mirrors real-world health care accessibility dynamics.

We could not identify any peer-reviewed studies presenting a comprehensive spatial accessibility toolbox in ArcGIS. However, there have been a few attempts documented in the gray literature. One such effort was made by Langford et al [28], who shared a tool named USW-FCA2 on ResearchGate using an E2SFCA model. Their tool requires a network data set for the study area and a Network Analysis license in ArcMap. In comparison, our toolbox offers significantly more functionalities and options that cater to the specific needs of the target users. Additionally, some studies have used spatial accessibility tools on alternative platforms. Saxon et al [29] developed an open software environment based on the Python-based PySaL package for measuring spatial accessibility. They calculated travel costs by incorporating precomputed origin-destination distance matrices for all US census tracts and census blocks in the 20 major cities.

In the case study, we demonstrated the integration of a nonspatial factor, age, with spatial accessibility. Age is an important determinant of health care demand, and regions with older populations tend to have a higher demand for health services [10]. The procedure used in the case study to adjust the age of the demand population can be extended to consider other factors such as ethnic groups or disease distributions. The resulting geographical maps revealed disparities in access to hemodialysis services across the state of Tennessee. Urban areas, where hemodialysis centers are concentrated, generally exhibited higher accessibility scores. However, areas in the southern parts of the state displayed lower accessibility scores, indicating a need for attention and prioritization in resource allocation.

This study does have limitations to consider. The developed tools are designed specifically for ArcGIS Pro, which may limit their usability for researchers using other software platforms like QGIS. However, future studies could explore the adaptation of these tools to different GIS software to ensure broader accessibility and usability for researchers across various platforms.

Conclusion

In conclusion, the developed tools for measuring the accessibility of health care resources offer valuable benefits to researchers across various domains. For large-scale analyses, such as country-level assessments, the 2SFCA01 and E2SFCA01 tools provide fast analysis with basic software requirements, making them accessible and efficient options. Additionally, the 2SFCA02 and E2SFCA02 tools offer a more realistic measure of accessibility by incorporating travel time catchments that consider traffic and transportation modes. Among these tools, the E2SFCA02 tool stands out as a powerful option as it considers both distance decay and uses travel time catchments, providing a comprehensive approach to measuring health care accessibility. Overall, these tools empower policy makers and researchers to gain valuable insights into identifying underserved areas and formulating effective resource allocation strategies. By assessing spatial disparities in health care access, these tools contribute to improving equity and enhancing health care service delivery. In the future, we will use and evaluate the outputs of this study for various health resource allocation projects.
including primary care providers and cancer care services (eg, radiotherapy). Furthermore, we will explore the impact of nonspatial factors such as ethnicity, income levels, and different social determinants of health to better understand their contributions to health care accessibility.

Acknowledgments
This study did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability
The data sets generated during and/or analyzed during this study are available through Arcgis' website [19].

Authors' Contributions
SH authored the manuscript and conducted the data analysis. DLS conceptualized, reviewed, and edited the content. AS-N conceptualized and supervised the study, participated in writing, and managed fund acquisition.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Python scripts for the tools used in the study. [ZIP File (Zip Archive), 11 KB - formative_v8i1e51727_app1.zip]

References

15. Radke J, Mu L. Spatial decompositions, modeling and mapping service regions to predict access to social programs. Geogr Inf Sci 2000;6(2):105-112 [FREE Full text] [doi: 10.1080/10824000009480538]


22. United States Census Bureau. URL: https://www.census.gov/ [accessed 2024-01-23]


Abbreviations

2SFCA: 2-step floating catchment area
2SFCA01: 2-step floating catchment areas with buffer distance catchments
2SFCA02: 2-step floating catchment areas with travel time catchments
E2SFCA: enhanced 2-step floating catchment area
E2SFCA01: enhanced 2-step floating catchment areas with buffer distance catchments
E2SFCA02: enhanced 2-step floating catchment areas with travel time catchments
ESKD: end-stage kidney disease
GIS: geographic information system
Original Paper

Group, Blended and Individual, Unguided Online Delivery of Mindfulness-Based Cognitive Therapy for People With Cancer: Feasibility Uncontrolled Trial

Nasim Badaghi1; Mette van Krujsbergen1; Anne Speckens1; Joëlle Vilé1; Judith Prins2; Saskia Kelders3,4; Linda Kwakkenbos1,5,6

1Department of Psychiatry, Radboud University Medical Center, Nijmegen, Netherlands
2Department of Medical Psychology, Radboud University Medical Center, Nijmegen, Netherlands
3Department of Psychology, Health, and Technology, University of Twente, Enschede, Netherlands
4Optentia Research Unit, North West University, Vanderbijlpark, South Africa
5Department of IQ Healthcare, Radboud University Medical Center, Nijmegen, Netherlands
6Department of Clinical Psychology, Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

Corresponding Author:
Nasim Badaghi
Department of Psychiatry
Radboud University Medical Center
Geert Grooteplein Zuid 10
Nijmegen, 6525 GA
Netherlands
Phone: 31 0624835397
Email: nasim.badaghimoreno@radboudumc.nl

Abstract

Background: Online mindfulness based cognitive therapy (eMBCT) has been shown to reduce psychological distress in people with cancer. However, this population has reported lack of support and asynchronous communication as barriers to eMBCT, resulting in higher nonadherence rates than with face-to-face MBCT. Using a co-creation process, we developed 2 formats of eMBCT: group, blended (combination of therapist-guided group and individual online sessions) and individual, unguided (individual, unguided online sessions only). Group, blended eMBCT offers peer support and guidance, whereas individual, unguided eMBCT offers flexibility and the possibility of large-scale implementation.

Objective: The objective of this nonrandomized feasibility study was to assess aspects of feasibility of the group, blended and individual, unguided eMBCT interventions.

Methods: Participants were people with cancer who chose between group, blended and individual, unguided eMBCT. Both intervention conditions followed the same 8-week eMBCT program, including an introductory session and a silent day (10 sessions total). All sessions for individual, unguided eMBCT occurred via the platform Minddistrict, whereas group, blended eMBCT consisted of 3 online videoconference sessions guided by a mindfulness teacher and 5 sessions via Minddistrict. We assessed the feasibility of the intervention quantitatively and qualitatively by evaluating its acceptability among participants. Additionally, we assessed limited efficacy by looking at the number of questionnaires participants completed pre- and postintervention.

Results: We included 12 participants for each eMBCT condition. Participants in group, blended eMBCT completed, on average, 9.7 of 10 sessions, compared with an average 8.3 sessions for individual, unguided eMBCT (excluding dropouts). Of the 24 participants, 13 (54%) agreed to be interviewed (5 unguided and 8 blended). Participants in both conditions reported positive experiences, including the convenience of not having to travel and the flexibility to choose when and where to participate. However, among the barriers for participation, participants in the group, blended condition reported a preference for more group sessions, and participants in the individual, unguided condition reported a lack of guidance. Additionally, for the group, blended condition, the effect sizes were small for all outcome measures (Hedges g range=0.01-0.36), except for fatigue, which had a moderate effect size (Hedges g=0.57). For the individual, unguided condition, the effect sizes were small for all outcome measures (Hedges g range=0.24-0.46), except for mindfulness skills (Hedges g=0.52) and engagement with the intervention (Hedges g=1.53).
Conclusions: Participants in this study had a positive experience with group, blended and individual, unguided eMBCT. Based on the results from this study, we will adjust the intervention prior to conducting a full-scale randomized controlled trial to evaluate effectiveness; we will add 1 group session to the group, blended eMBCT using Zoom as the platform for the group sessions; and we will send reminders to participants to complete questionnaires.

Trial Registration: ClinicalTrials.gov NCT05336916; https://clinicaltrials.gov/ct2/show/NCT05336916

(JMIR Form Res 2024;8:e52338) doi: 10.2196/52338

KEYWORDS
cancer; eHealth; online interventions; mindfulness; psycho-oncology; qualitative research; oncology; CBT; blended; eMBCT; iCBT; cognitive therapy; unguided; psychotherapy; MBCT; co-creation; therapist; self-guided; peer-support; co-design; participatory

Introduction
The number of people with cancer is increasing at alarming rates. It has been estimated that, by 2040, the number of people with cancer will be almost double that of 2020 [1]. Additionally, approximately 1 in 3 individuals with cancer experiences severe psychological distress [2]. As a result, there is an increasing number of distressed people with cancer who could benefit from effective psycho-oncological interventions [3].

One kind of evidence-based psychological treatment for people with cancer is a mindfulness-based intervention (MBI). Mindfulness can be defined as moment-to-moment awareness, which is cultivated by purposely paying attention to the present experience without judgment [4]. Although mindfulness practices were originally developed centuries ago in the Buddhist traditions of Asia, it was not until the past couple of decades that they were implemented in health care. Different forms of MBIs (eg, mindfulness-based cognitive therapy [MBCT]) have been used across conditions [5], including cancer [6-8], and have been shown to have beneficial effects on psychological distress, quality of life, and well-being [5]. MBCT includes mindfulness components (eg, meditations, visualization exercises, movement exercises) and cognitive components from cognitive behavioral therapy (eg, identifying and reframing automatic thoughts, recognition that thoughts are not facts, habitual thoughts and behavioral patterns).

MBIs have been successfully adapted to online formats [9]. Although eHealth interventions are complex and relatively new [10], online MBIs offer multiple advantages over face-to-face interventions. For instance, online interventions are more easily accessible, more flexible in when and how participants can follow the program, and less costly [9]. A recent systematic review evaluated 9 randomized controlled trials (RCTs) and found that, although online MBIs generally have smaller effect sizes than face-to-face MBIs, they were still effective in reducing depression symptoms, anxiety, and stress, as well as improving mindfulness skills among people across different physical conditions [11]. In addition, a meta-analysis evaluated the effectiveness of different forms of online MBIs (delivered on a website or by an application) for people with cancer and found that online MBIs were effective in reducing distress, depression, and sleep disturbance and that they improved quality of life [12]. Plus, the authors of this meta-analysis concluded that online MBIs may provide unique advantage of increased accessibility and scalability. Online MBIs can offer a valuable alternative to face-to-face interventions, in particular for people with cancer who often already have to deal with frequent hospital visits, physical symptoms from the disease, and its treatment (such as intensive medical treatments and treatment-related fatigue and pain) [13].

Our research group previously conducted an RCT comparing the effectiveness and cost-effectiveness of online mindfulness-based cognitive therapy (eMBCT) and face-to-face MBCT with treatment as usual in reducing psychological distress in people with cancer (BeMind trial) [13]. The online condition in the BeMind trial consisted of an individual, 8-week, online mindfulness intervention supported by a qualified mindfulness teacher who provided feedback via email. The face-to-face condition was a prototypical 8-week group MBCT taught by a qualified mindfulness teacher.

Results from the BeMind trial showed that, in a heterogeneous sample of distressed people with cancer, both interventions were more effective at reducing psychological distress and were less costly than treatment as usual [13]. Nevertheless, nonadherence rates were higher in the individual eMBCT condition than in the group, face-to-face MBCT condition. Furthermore, qualitative analyses showed important barriers to participating in eMBCT, including insufficient peer support and asynchronous communication [14]. Additionally, mindfulness teachers had to invest more time for the individual, online condition than for the group condition, which may hamper large-scale implementation. Thus, the BeMind study showed that, although eMBCT is effective at reducing distress in people with cancer, there is room to improve the eMBCT intervention prior to implementation.

Considering the results from the BeMind trial and the social restrictions from the COVID-19 pandemic at the start of our project, we developed 2 new eHealth formats using a cocreation process. With experts in eHealth interventions, MBCT teachers, representatives from cancer patient organizations, and people with cancer, we explored how to give proper counseling, personalize the intervention, and make it more engaging. In addition, as adherence to online interventions without the guidance of a teacher (individual, unguided interventions) is often lower than intended, persuasive technology known to improve adherence, such as reminders and virtual coaches, was included [15]. By considering the different perspectives of the stakeholders, we aimed to develop a more appealing, persuasive, and participant-focused online intervention.

We developed the following 2 interventions using a cocreation process: group, blended and individual, unguided eMBCT.
Group, blended eMBCT consisted of 3 online, group sessions with a mindfulness teacher and 5 individual, teacher-assisted online sessions; this combination provided peer support and partly synchronous communication. Individual, unguided eMBCT consisted of 8 eMBCT sessions in which participants followed the intervention by themselves without teacher guidance; an unguided intervention could increase access and improve scalability at a lower cost for both participants and therapists. Both intervention conditions also included an introductory session and a silent day. We developed group, blended and individual, unguided interventions to optimize eMBCT delivery and efficacy by addressing the barriers we found in our previous study, by considering the target group needs and by constructing an online intervention that is engaging and attractive.

Although the 2 MBIs were carefully designed using a cocreation process with relevant stakeholders, aspects of their feasibility such as acceptability and preliminary efficacy needed to be established prior to conducting a full-scale RCT. In fact, pilot studies can support researchers with identifying possible challenges, weighing resources, and evaluating the feasibility of an intervention [16,17]. Moreover, pilot studies can assess preliminary efficacy of an intervention before moving on to a full-scale RCT, which involves more resources [16,17]. The objective of this pilot study was to assess the feasibility and preliminary efficacy of the group, blended and individual, unguided eMBCT interventions among people with cancer. The results from this pilot will help us improve the intervention conditions prior to testing their effectiveness in a full-scale, 3-arm RCT that will compare group, blended and individual, unguided eMBCT with care as usual [18].

**Methods**

**Study Design and Setting**

This was a mixed methods, nonrandomized feasibility study. Participants could choose to participate in either group, blended or individual, unguided eMBCT. All participants were invited to a semistructured interview postintervention. Participants were also asked to complete questionnaires before and after the intervention.

The study was conducted at the Radboudumc Center for Mindfulness in Nijmegen, The Netherlands. Although our study was not randomized, results are reported in accordance with the Consolidated Standards of Reporting Trials (CONSORT) extension for randomized pilot and feasibility trials [19], as many of the principles described apply.

**Participants**

Participants were eligible if they (1) were adults and had been diagnosed with cancer at any point in their life (irrespective of type or stage of cancer and time since diagnosis); (2) had internet access and were able to use a computer; and (3) had good command of the Dutch language. Participants were excluded if they (1) had participated in a mindfulness intervention before (>4 sessions); (2) had a severe psychiatric comorbidity that warranted acute treatment (ie, psychosis, mania, severe personality disorders, suicidal thoughts); (3) had dependence on drugs or alcohol; or (4) had severe cognitive impairments.

**Procedure**

Participants were recruited online via posts placed on websites of cancer-related organizations or patient group forums (online sites where people with cancer can hold conversations and find information), flyers and posters placed at Radboud University medical center, and social media platforms.

Interested participants contacted us via email, via phone, or by completing the contact form on the study website. After this, they received a phone call from one of the researchers to verify inclusion criteria and provide information about the study. Participants were allowed to choose their preferred intervention condition. For each eMBCT condition, a maximum of 12 participants were allowed, so once one eMBCT condition was full, participants were informed that they could only participate in the other one if they wished to be included in this study. Eligible participants were sent the written information about the study and the informed consent form by post and email. After participants signed and returned the informed consent form, they were enrolled in the study and asked to complete pre-intervention questionnaires via the secure Castor EDC system.

Within 1 week after the 8 weeks of the intervention, participants were invited to complete postintervention questionnaires and share their experiences in a semistructured interview. Participants in both conditions were allowed to have any form of medical, psychological (except for MBIs), or paramedical care they required during the study period.

See Figure 1 for the participant inclusion flow chart.
The content of both eMBCT interventions was based on the MBCT program developed by Segal et al [20]. For both conditions, the intervention consisted of 8 online sessions with mindfulness meditation exercises, psychoeducation, and reflections, plus a silent day. We included psychoeducation about mindfulness for cancer and grief and adapted the moving exercises for people with cancer. Participants were asked to do home practice for 30 minutes to 45 minutes a day. Although the content of the sessions did not differ between the 2 conditions, the delivery format did; in the group, blended eMBCT, sessions 1, 5, and 8 took place as online group sessions via the videoconferencing platform Zaurus. In both conditions, participants were allowed to join with a significant other. The specific content for each session has been published elsewhere [18].

**Group, Blended eMBCT**

Group, blended eMBCT consisted of 3 videoconference group sessions lasting 2.5 hours and guided by a mindfulness teacher (sessions 1, 5, and 8). The other sessions (sessions 2, 3, 4, 6, and 7 and the silent day) were followed individually via Minddistrict. Participants were provided with written online
feedback on the individual sessions from their mindfulness teacher within Minddistrict. The mindfulness teachers involved in the project were health care professionals experienced in psycho-oncology who met the qualification criteria of the Association of Mindfulness Teachers based in The Netherlands and Flanders, which are in line with the 2015 UK Network for Mindfulness-Based Teachers criteria. In addition, teachers had regular peer supervision sessions led by a senior mindfulness teacher (AS).

**Individual, Unguided eMBCT**

Participants in the unguided eMBCT condition were provided the entire training through Minddistrict. They received weekly access to one of the online mindfulness sessions, which involved the same themes, exercises, and homework as those in the group, blended eMBCT condition. However, there was no mindfulness teacher involved. Participants received automated feedback instead. Participants could contact the research team for technical support.

**Feasibility Outcomes**

Aspects of feasibility were assessed based on the areas of focus suggested by Bowen et al [21], including acceptability and limited efficacy. Acceptability focuses on how the participants react to the intervention (to what extent the intervention is suitable, satisfying, and attractive) [21]. Limited efficacy intends to test the intervention in a limited way [21]. For this pilot study, limited efficacy was evaluated with the pre-intervention and postintervention questionnaire scores (the same questionnaires will be used in the full-scale RCT).

Acceptability of the intervention was evaluated by how many participants chose each eMBCT condition, how many participants started the intervention, the participants’ clinical characteristics at baseline, the average number of sessions completed per condition—adherence—(through attendance lists for groups and login data in Minddistrict), and dropout rate (participants who discontinued the intervention).

In addition, acceptability was evaluated by conducting semistructured interviews postintervention. Participants from group, blended and individual, unguided eMBCT were asked program-specific questions to assess their experiences with the respective intervention condition; in addition, they were asked to express if they experienced barriers to or facilitators for group, blended and individual, unguided eMBCT (eg. Were there specific parts that were helpful/not helpful?). Questions were asked in an open way to permit participants to freely speak. Participants were interviewed via telephone within 3 months after the end of the intervention. Participant interviews were conducted by 2 researchers with previous experience in qualitative research who had not been involved in the delivery of the training (see Multimedia Appendix 1 for the complete interview guide).

Limited efficacy includes planned outcome measures for full-scale RCT. Consistent with the feasibility trial design [19,22], limited efficacy of the following measures of distress and secondary outcomes that will be used to evaluate the program in the full-scale trial are reported: Hospital Anxiety and Depression Scale [23], severity scale of the Fear of Cancer Recurrence Inventory-Short Form [24,25], fatigue severity subscale of the Checklist Individual Strength [26], rumination subscale of the Ruminating and Reflection Questionnaire [27], Five Facet Mindfulness Questionnaire-Short Form [28], Self-Compassion Scale-Short Form [29], Mental Health Continuum-Short Form [30], Twente Engagement with E-health Technologies Scale [31]. Detailed information about each questionnaire has been published previously [18].

**Sample Size**

There is no consensus about the optimal sample size for pilot and feasibility studies [32]. Guidance varies between 12 and 30 participants or more per trial arm [33,34]. To reach the study objectives, including sufficient feedback on the acceptability of both conditions, we aimed to include 12 participants in the individual, unguided condition and 12 participants in the group, blended condition.

**Data Analyses**

Descriptive statistics were used to characterize participant demographics, recruitment numbers, and login data. Baseline and postintervention mean scores, standard deviations, and effect sizes from all the questionnaires were calculated for both conditions using SPSS (version 27; IBM Corp). The number of sessions completed was calculated by summing the number of sessions that each participant completed. In this count, we included the introductory session and the silent day, which made a total of 10 sessions. The proportion of sessions completed in Minddistrict was represented as percentages based on each participant’s usage login data. We transformed the percentages into units to aggregate the number of sessions completed by each participant.

Interviews were transcribed verbatim and analyzed by means of an iterative process of thematic analysis in which coding categories were derived directly from the text data (inductive coding) [35]. We followed a form of data-driven thematic analysis and followed the different phases of thematic analysis as suggested by Braun and Clark [35]. First, we familiarized ourselves with the data. Second, 2 researchers independently did a first round of coding 2 interviews (1 from a participant in the group, blended intervention and 1 from a participant in the individual, unguided intervention), during which initial codes were generated. Third, codes between the 2 researchers were compared and reviewed. Fourth, the 2 researchers created a common coding map for analysis and coded the rest of the interviews. All remaining interviews were coded by one researcher and reviewed by a second. Discrepancies were discussed and resolved by consensus. Fifth, the codes from step 4 were reviewed and categorized into broader themes with a larger group. The group consisted of 3 senior researchers (AS, JP, and LK) who have extensive experience in the fields of mindfulness, cancer, and research methodology, a mindfulness teacher with more than 20 years of experience teaching different mindfulness courses (including mindfulness for people with cancer), and a PhD candidate with a background in mindfulness and clinical experience (NB). Definitions and labels for each theme and subthemes were generated. Finally, we selected vivid and compelling examples from the interviews that clearly portrayed the themes and subthemes identified.
**Ethical Considerations**

The Buddy feasibility trial was approved by the ethical review board, CMO Arnhem-Nijmegen (number: NL73117.091.20), prior to data collection. The study was conducted according to the principles of the Declaration of Helsinki (6th edition, 2008) and in accordance with the Medical Research Involving Human Subjects Act (WMO). All participants who joined the study signed an informed consent form prior to enrollment. Participation was free of charge and voluntary. Participants were informed that they could withdraw at any time without consequences, that their anonymity was ensured, and that there was no monetary compensation for participation.

**Results**

**Acceptability**

Enrollment took place between August 2020 and January 2021. A total of 29 participants were assessed for eligibility, of which 2 were excluded; 1 participant had previously followed an MBI, and the other participant was never diagnosed with cancer. Of the 27 eligible participants, 1 participant did not want to participate, 1 participant preferred to join a group, face-to-face MBI program, and we could not contact the other person. In total, 24 participants started the program. The first 12 eligible participants who contacted us preferred to join the group, blended condition. Therefore, subsequent eligible applicants were only offered individual, unguided eMBCT. See Figure 1 for the complete flow of participant selection procedures.

All participants were Dutch and had at least a secondary education; their mean age was 51 (SD 11.5) years, and most (21/24, 87%) were female. Participants had the following types of cancer: breast (9/21, 42%), ovarian (5/21, 24%), colon (3/21, 14%), leukemia (2/21, 10%), lymphoma (1/21, 5%), and bowel (1/21, 5%). Most participants (15/20, 75%) received treatment with curative intent. In addition, overall, it appeared that participants in the blended group had more working hours than participants in the unguided group (21 hours vs 13 hours) and were more often treated with curative intent (10/11, 90% vs 5/9, 56%), and a larger proportion was diagnosed with ovarian cancer (5/11, 46% vs 0/10, 0%). Demographic and disease characteristics of the participants for each condition are shown in Table 1.

For the group, blended condition, the mean number of sessions completed was 9.7 of 10; all participants completed at least 9 of the 10 sessions (minimum 9 sessions and maximum 10 sessions); adherence was high; and there were no dropouts.

In total, 5 of the 12 (42%) participants in the unguided, online condition dropped out. These participants completed fewer than 4 sessions: 3 participants completed 3 sessions, and 2 participants completed 1 session only. We could not contact 2 participants after they dropped out, and the other 3 reported that the program was too hard to follow because of personal circumstances. Participants in the unguided, individual condition who dropped out and completed baseline assessments (n=4) had metastatic cancer and were receiving palliative anticancer treatment. The mean number of sessions completed for the individual, unguided condition, excluding dropouts, was 8.3 of 10 (minimum 1.5 sessions and maximum 10 sessions). The mean number of sessions completed for the individual, unguided condition, including dropouts, was 6.1 of 10.
Table 1. Demographic and clinical characteristics of the included participants at baseline.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Blended (n=12)</th>
<th>Unguided (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>51 (2.3)</td>
<td>51 (4.4)a</td>
</tr>
<tr>
<td>Female sex, n (%)</td>
<td>11 (92)</td>
<td>10 (83)</td>
</tr>
<tr>
<td>Married or living as married, n (%)</td>
<td>9 (82)a</td>
<td>9 (90)</td>
</tr>
<tr>
<td>Dutch nationality, n (%)</td>
<td>12 (100)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>Work per week (hours), mean (SD)</td>
<td>21 (11)b</td>
<td>13 (14)c</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>6 (54)a</td>
<td>4 (40)d</td>
</tr>
<tr>
<td>Tertiary</td>
<td>5 (45)a</td>
<td>6 (60)d</td>
</tr>
<tr>
<td>Anticancer treatment, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curative</td>
<td>10 (90)a</td>
<td>5 (56)d</td>
</tr>
<tr>
<td>Palliative</td>
<td>1 (10)a</td>
<td>4 (44)d</td>
</tr>
<tr>
<td>Duration since first cancer diagnosis (months), mean (SD)</td>
<td>14 (14)a</td>
<td>18 (24)b</td>
</tr>
<tr>
<td>Type of cancer, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>3 (27)a</td>
<td>6 (60)b</td>
</tr>
<tr>
<td>Ovarian</td>
<td>5 (46)a</td>
<td>0a</td>
</tr>
<tr>
<td>Colon</td>
<td>1 (9)a</td>
<td>2 (20)b</td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td>0a</td>
<td>2 (20)b</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>1 (9)a</td>
<td>0b</td>
</tr>
<tr>
<td>Bowel</td>
<td>1 (9)a</td>
<td>0b</td>
</tr>
<tr>
<td>Previous experience with meditation (yes), n (%)</td>
<td>6 (55)a</td>
<td>4 (40)b</td>
</tr>
</tbody>
</table>

Results From the Interviews

Participation and Themes

All 24 participants were invited for semistructured interviews, and 13 (54%) agreed to participate (5 from the individual, unguided condition and 8 from the group, blended eMBCT condition). Interviews lasted between 35 minutes and 90 minutes. Responses from the semistructured interviews showed that there were overlapping as well as different barriers and facilitators for group, blended and individual, unguided eMBCT. The barriers and facilitators were organized into the following 4 emergent themes (each one with multiple subthemes): program content; program format; group, blended condition; and individual, unguided condition.

Program Content

Factors categorized as program content were the specific online program components (eg, exercises, videos, diaries) in Minddistrict that were used in both intervention conditions. Program content facilitators that participants reported included the possibility to choose different exercises that were suitable to personal needs. For instance, one participant reported:

*What I also like is that you could choose which exercises yourself (...) that’s just really nice, that bit of freedom you had.* [group, blended condition; completed 10 sessions]

In addition, participants reported that the identification with other peer participants in the videos and the normalization of their experiences through video stories were useful and made them feel less alone:

*I really liked those videos, to experience what other people thought about it—for them experienced it (...) I often recognized myself in it, so that was helpful.* [individual, unguided condition; completed 7 sessions]

Reported program content barriers included too many exercises with no clear explanation about their rationale, too many forms to fill out, and reflection on emotions in diaries that were
challenging and confrontational; for instance, 1 participant reported the following:

*It was sometimes quite intense to fill in your diary every day, and every week there was also another diary that you had to keep.* [group, blended condition; completed 10 sessions]

**Program Format**

Factors categorized as program format were the arrangement of the online program that facilitated participants' participation (eg, structure, time, place) to both intervention conditions. Not having to travel and being able to follow the program at one’s own pace were the most reported program format facilitators. One participant reported:

*I liked doing it at home so I could do it in my own time and place, and also with no travelling times.* [individual, unguided condition; completed 10 sessions]

Additionally, the presentation of visual information that complemented the written and spoken exercises and having the choice of a physical booklet as an additional source and future reference were positively valued, as participants indicated:

*I always really liked the information pieces with those drawings, because then I could make it visual instead of everything being spoken, then I really had an image and I really liked that.* [individual, unguided condition; completed 7 sessions]

*...a book you can easily pick up in addition to your exercise. I personally prefer it.* [individual, unguided condition; completed 10 sessions]

However, some other participants said that they found the program structure unclear and that it was difficult to navigate. In addition, they reported that it required too much time investment and that it was difficult to follow it at home with constant interruptions of family members. For instance, one participant from the individual, unguided condition said:

*I just don’t know when to schedule it. Then, I had just found a moment, and another child came downstairs and asked loudly: ‘Can I have an apple?’ Yes. Or then, the partner comes and gets some tea, and he would say… ‘Oh, sorry, I see you are doing mindfulness.’* [individual, unguided condition; completed 7 sessions]

**Group, Blended Condition and Individual, Unguided Condition**

There were also barriers and facilitators that were specific to the intervention conditions. Participants in the group, blended condition liked the group sessions because they had connection with others, peer support, the possibility to ask questions, and synchronicity in communication. A participant in the group, blended condition who completed 10 sessions reported that “it’s also nice to hear that other people are struggling with the same things and yes, you know, you're suddenly not crazy anymore.”

Another participant reported that:

*I was really looking forward to it when it was finally time for another group session of, oh yes, nice, just talking to people.* [group, blended condition; completed 10 sessions]

Moreover, one participant emphasized that:

*You know, if you have any questions, at least you can ask questions. Then, you will get an answer right away.* [group, blended condition; completed 9 sessions]

Barriers that were specific to the group, blended sessions included having to be sitting for a long time during the group sessions, the intensity and length of the sessions, the infrequent number of group sessions, and the fact that they were online rather than in person. A participant commented that “and with the 3 times you only have together, that was quite short” [group, blended condition; completed 9 sessions]. Another participant, also from the group, blended condition, reported that “it’s quite a long time to sit behind a screen like that. And I wasn’t that far into my recovery yet, so I especially thought the first session was really exhausting” [completed 10 sessions]. An outstanding result was that almost all participants from the group, blended condition indicated the need to have more group sessions. They mentioned that the group sessions were not enough to get to know each other properly. One participant reported:

*I was really looking forward to when it was finally time again for a group session (...) I think for me, if it had been a group session 8 times, then that would also just be very nice.* [group, blended condition; completed 10 sessions]

Participants in the individual, unguided condition reported lack of peer support and lack of feedback from a therapist as barriers for participation. One participant from the individual, unguided condition reported that “the fact that there is no contact with a person or with a group and that there is also no concrete agreement that we will meet each other—even if it is only online...that made it very difficult for me to keep it up” [completed 6 sessions].

For all the barriers and facilitators for both conditions across themes and subthemes, see Table 2.
Table 2. Barriers and facilitators experienced by participants during group, blended and individual, unguided online mindfulness-based cognitive therapy (eMBCT).

<table>
<thead>
<tr>
<th>Themes and subthemes</th>
<th>Facilitators</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(1) Program: content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exercises</strong></td>
<td>• Multiple options: different exercises and different voices for the meditations • Pleasant and clear voices that become familiar with time • Good and short exercises</td>
<td>• Too many exercises. • No instructions explaining the goal of the meditation exercises</td>
</tr>
<tr>
<td><strong>Silent day</strong></td>
<td>• Pleasant silent day</td>
<td>• Not feasible to do it independently at home without distractions • Difficult to separate from daily disturbances and quotidian environment • Too long silent day • Lack of peer support and guidance from a therapist</td>
</tr>
<tr>
<td><strong>Diaries</strong></td>
<td>• Promotion of personal reflection • Support personal processes</td>
<td>• Too many different forms to fill out every week • Confronting to fill out a diary every day</td>
</tr>
<tr>
<td><strong>Automatic feedback</strong></td>
<td>• Recognition with peer participants and normalization</td>
<td>• Too impersonal • Participants forced to choose an answer before being able to proceed</td>
</tr>
<tr>
<td><strong>Videos</strong></td>
<td>• Encouraging to see that the program is helpful for other people with cancer • Explanations that clarify what is meant by the elements of the program • Relate to other people with cancer experiences</td>
<td>• Set too high standards • Videos not inclusive enough</td>
</tr>
<tr>
<td><strong>Reminders</strong></td>
<td>• Helpful reminders</td>
<td>• Too many reminders</td>
</tr>
<tr>
<td><strong>(2) Program: format</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Initial contact research team</strong></td>
<td>• Very helpful to have a personal introduction into the online program, makes it accessible</td>
<td>• Business-like communication style • Too impersonal</td>
</tr>
<tr>
<td><strong>Help desk</strong></td>
<td>• Supportive if you ran into problems • Helpful to have the option of personal contact</td>
<td>• Sometimes, it took too long to respond to participants.</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td>• Logical structure: the sessions build on each other consistently. • Lot of suggestions • Clear structure of the platform</td>
<td>• Unclear where to write notes or not write them at all • Unclear structure, repetition; what do I need to do?</td>
</tr>
<tr>
<td><strong>Navigating through program</strong></td>
<td>• Easy to move forward in the program • Possible to look back at own notes • Did not get stuck • Being able to fill things out yourself</td>
<td>• Navigating the program was difficult. • Not clear how to save entered information • Unclear where to put notes in both daily and weekly forms • Not being able to go back to the exercises to do them again or to the diaries to add information later on • Getting stuck, not being able to move forward</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>• Very relaxed, own time, own planning • No travelling time • Possibility to combine eMBCT with cancer treatment, rehabilitation, household chores • Option to adapt the time invested in the program to the energy levels</td>
<td>• Time-consuming program, took too much time</td>
</tr>
<tr>
<td>Themes and subthemes</td>
<td>Facilitators</td>
<td>Barriers</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------</td>
<td>----------</td>
</tr>
</tbody>
</table>
| **Place**            | • Pleasant to do it at home  
                      • No traveling, does not cost energy | • Interruptions for family members  
                      • Difficult to find a room in the house where you will not be disturbed |
| **Infographics and avatar** | • Possibility to choose the coach and answers | • The avatar was not of any added value for some participants.  
                      • Getting stuck if an avatar was not selected |
| **Physical booklet** | • Having the choice of a physical booklet  
                      • Being able to look back in the physical booklet to previous sessions, besides the online program  
                      • Having an additional source for reading  
                      • Future reference | • Too many different things: online program, booklet; unsure where to go  
                      • For some people, the app was very clear, and they used the app only. There was no added value from the booklet. |

(3) **Group, blended condition**

**Group sessions**
- Connection with others  
  - Peer support  
  - Possibility to ask questions  
  - Synchronicity in communication  
  - Recognition that others struggle with the same things
- Being stressed about not being able to log in in time  
  - Not being able to see people properly in the screen (Zaurus), no speaker perspective  
  - Prefer to meet people in person rather than on a screen  
  - Very tiring to sit behind a screen for a long time  
  - Intense, long, and tiring group sessions  
  - Too infrequent  
  - Confrontation with other participants’ cancer

**Feedback from mindfulness teacher**
- Good quality, elaborated, and personalized feedback  
  - Trustworthy, accessible, and supportive
- Even though people got written feedback, this was less stimulating to some.  
  - Asynchronous written feedback and not clear timing of receiving it

(4) **Individual, unguided condition**

**Lack of peer support**
- Need for self-discipline  
  - Difficult to maintain engagement without appointments  
  - Lack of support from a community

**Lack of feedback therapist**
- Feels unsafe to share personal information with unknown recipient

---

AThe themes for the program content and format and their respective subthemes applied to both intervention conditions.  

BNo response.

**Planned Trial Outcomes**

Overall, 21 of the 24 (88%) participants completed baseline questionnaires, and 19 of the 24 (79%) participants completed posttreatment questionnaires. More specifically, most participants (11/12, 92%) in the group, blended condition completed both baseline and posttreatment questionnaires, while in the individual, unguided condition, only 7 (7/12, 58%) completed both baseline and posttreatment questionnaires. Although this study had a small sample size and tests of significance were not included, effect sizes were calculated to evaluate changes between pre- and postassessments. For the group, blended condition, the effect sizes for change before and after treatment were small for all outcome measures (Hedges g range=0.01-0.36), except for fatigue, which had a moderate effect size (Hedges g=0.57). For the individual, unguided condition, the effect sizes for change before and after treatment were small for all outcome measures (Hedges g range=0.24-0.46), except for mindfulness skills (Hedges g=0.52) and engagement with the intervention (Hedges g=1.53). Table 3 shows the baseline and postintervention scores for the planned trial outcome measures for both groups. No adverse events were reported.
Participants in the blended condition valued flexibility of the sessions; they felt connected with others, experienced good peer support, and appreciated the synchronicity in communication. People in the group, blended condition even indicated that there were too few group sessions compared with 5 participants who discontinued the intervention in the individual, unguided eMBCT. Acceptability and adherence seemed to be higher in the group, blended condition than in the individual, unguided condition. It is important to critically evaluate and weigh the feedback messages were not even read all the time [38].

In fact, it was reported that human feedback was the most important, it should be critically evaluated and weighed. Participants in this study reported that they preferred the group, blended condition, they still accepted the individual, unguided condition. It is important to carefully think about what is effective. Therefore, it is crucial to carefully think about how to use different kinds of input in a cocreation process when developing online interventions.

In terms of preferences and dropouts, the first 12 participants who enrolled in this study preferred the group, blended condition, and there were no dropouts in this condition, compared with 5 participants who discontinued the intervention in the individual, unguided eMBCT.

**Discussion**

**Principal Findings**
This nonrandomized study evaluated aspects of feasibility of group, blended and individual, unguided eMBCT for people with cancer. Overall, participants were positive about their experiences in both conditions. This supports the progression to a full-scale RCT in which the effectiveness of group, blended and individual, unguided eMBCT will be assessed.

We found that participants in both intervention conditions valued practicing at their own time, at any place. This flexibility of eMBCTs among people with cancer has been previously reported as a facilitator [14,36,37]. It is evident that many people with cancer need flexible psycho-oncological interventions.

Participants in the group, blended condition particularly valued the group component of the sessions; they felt connected with others, experienced good peer support, and appreciated the synchronicity in communication. People in the group, blended condition even indicated that there were too few group sessions and that they would have liked more. Based on these results, a fourth group session will be added to the group, blended condition in the full-scale RCT. It should be noted that, although participants’ feedback has been mentioned as relevant in the cocreation process [10] and we obviously considered it important, it should be critically evaluated and weighed. Participants in this study reported that they preferred the group sessions, but if we were not able to offer the group, blended condition, they still accepted the individual, unguided condition.

**Table 3. Baseline and postintervention scores after the 8-week intervention for the planned Buddy trial outcome measures.**

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Score in the blended group (n=12), mean (SD)</th>
<th>Effect size, Hedges g (95% CI)</th>
<th>Score in the unguided group (n=12), mean (SD)</th>
<th>Effect size, Hedges g (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychological distress (HADS&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>13.6 (7.1)&lt;sup&gt;b&lt;/sup&gt; 12.1 (6.8)</td>
<td>0.21 (–0.61 to 1.03)</td>
<td>18.4 (8.3)&lt;sup&gt;c&lt;/sup&gt; 17.4 (5.8)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.13 (–0.84 to 1.10)</td>
</tr>
<tr>
<td>Fear of cancer recurrence (FCRI-SF&lt;sup&gt;e&lt;/sup&gt;)</td>
<td>78.4 (18.6)&lt;sup&gt;b&lt;/sup&gt; 78.5 (14.2)</td>
<td>0.01 (–0.82 to 0.81)</td>
<td>94.7 (18.5)&lt;sup&gt;c&lt;/sup&gt; 93.1 (15.1)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.09 (–0.88 to 1.05)</td>
</tr>
<tr>
<td>Fatigue (CIS&lt;sup&gt;f&lt;/sup&gt;)</td>
<td>33.7 (5.5)&lt;sup&gt;b&lt;/sup&gt; 36.7 (4.6)</td>
<td>0.57 (–1.41 to 0.26)</td>
<td>33.2 (7.3)&lt;sup&gt;c&lt;/sup&gt; 32.6 (4.4)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.09 (–0.88 to 1.06)</td>
</tr>
<tr>
<td>Rumination (RRQ&lt;sup&gt;g&lt;/sup&gt;)</td>
<td>37.4 (7.4)&lt;sup&gt;b&lt;/sup&gt; 35.2 (5.9)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.32 (–0.52 to 1.16)</td>
<td>40.1 (4.9)&lt;sup&gt;c&lt;/sup&gt; 39.4 (4.9)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.14 (–0.83 to 1.10)</td>
</tr>
<tr>
<td>Mindfulness skills (FFMQ-SF&lt;sup&gt;h&lt;/sup&gt;)</td>
<td>75.5 (7.8)&lt;sup&gt;b&lt;/sup&gt; 77.2 (5.9)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.24 (–1.08 to 0.60)</td>
<td>77.6 (3.3)&lt;sup&gt;c&lt;/sup&gt; 75.6 (4.1)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.52 (–0.46 to 1.50)</td>
</tr>
<tr>
<td>Self-compassion (SCS-SF&lt;sup&gt;i&lt;/sup&gt;)</td>
<td>47.8 (7.6)&lt;sup&gt;b&lt;/sup&gt; 48.9 (7.2)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.14 (–0.98 to 0.69)</td>
<td>53.2 (7.2)&lt;sup&gt;c&lt;/sup&gt; 54.7 (3.0)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.24 (–1.21 to 0.73)</td>
</tr>
<tr>
<td>Positive mental health (MHC-SF&lt;sup&gt;j&lt;/sup&gt;)</td>
<td>37.5 (15.9)&lt;sup&gt;b&lt;/sup&gt; 43.3 (15.0)</td>
<td>0.36 (–1.19 to 0.46)</td>
<td>39.3 (12.7)&lt;sup&gt;c&lt;/sup&gt; 33.7 (9.9)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.46 (–0.52 to 1.43)</td>
</tr>
<tr>
<td>Engagement with intervention (TWEETS&lt;sup&gt;k&lt;/sup&gt;)</td>
<td>24.9 (3.0)&lt;sup&gt;b&lt;/sup&gt; 23.4 (5.7)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.32 (–0.52 to 1.16)</td>
<td>25.5 (2.4)&lt;sup&gt;c&lt;/sup&gt; 16.4 (8.4)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.53 (0.44 to 2.63)</td>
</tr>
</tbody>
</table>

<sup>a</sup>HADS: Hospital Anxiety and Depression Scale.
<sup>b</sup>n=11.
<sup>c</sup>n=10.
<sup>d</sup>n=7.
<sup>e</sup>FCRI-SF: Fear of Cancer Recurrence Inventory-Short Form.
<sup>f</sup>CIS: Checklist Individual Strength.
<sup>g</sup>RRQ: Rumination-Reflection Questionnaire.
<sup>h</sup>FFMQ-SF: Five Facet Mindfulness Questionnaire: Short Form.
<sup>i</sup>SCS-SF: Self-Compassion Scale-Short Form.
<sup>j</sup>MHC-SF: Mental Health Continuum-Short Form.
<sup>k</sup>TWEETS: Twente Engagement with E-health Technologies Scale.
note that all participants who discontinued the intervention were in palliative treatment. Although we could not determine the exact reason, these preliminary insights suggest a proclivity for group, blended eMBCT and questions the acceptability of the individual, unguided eMBCT for people receiving palliative treatment. In addition, it should be considered that, although the individual, unguided condition may be easier to implement and cheaper, people with cancer may still prefer a group, blended intervention format.

Another finding is that the completion rates for the postintervention questionnaires and interviews were low, in particular for the individual, unguided group. Participants were not reminded to complete questionnaires, and no incentive to participate in the interviews was provided. In addition, people in the individual, unguided condition had no contact with a mindfulness teacher or peers throughout the intervention. It might be that these participants did not feel as engaged in the study as the people in the group, blended condition. It has been shown that the use of electronic reminders and real-time monitoring among people with cancer can contribute to a very high completion rate [39]. In the full-scale RCT, we therefore plan to include prompts, such as emails, calls, and WhatsApp messages, so participants are reminded to complete questionnaires.

Strengths and Limitations

In this pilot study, we included representatives from the target group as well as experts in the field of cancer, mindfulness, and eHealth to develop an effective intervention. This study highlights the importance of assessing relevant stakeholders’ opinions before developing an intervention and prior to going through the efforts of conducting a full-scale RCT. Based on the results of the cocreation process, we developed an app that is visually attractive; user friendly; low cost; and flexible in how, when, and where to participate. We developed an intervention that is in line with the participants’ needs and wishes and that considered expert opinions. It has increasingly been mentioned in the emerging field of eHealth interventions [10] that it is crucial to carefully consider and understand the target group when developing an effective online intervention. In this pilot study, we not only carefully addressed the target group’s desires and needs before the intervention but also evaluated their experiences after, to develop an optimal intervention that is acceptable to the end user.

This study also has limitations that should also be considered when interpreting its results. First, because of the nature of pilot studies, this study had a small sample size; in addition, it was nonrandomized, limiting our ability to assess limited efficacy. Second, our sample was rather homogeneous (eg, all Dutch, mostly highly educated women), and these participants were self-selected. The participants who had a choice between both conditions all chose the group, blended format. Consequently, findings cannot be generalized to all people with cancer. It may require more research to be able to apply online MBIs across people with cancer with different characteristics (eg, type of cancer, age, sex, language). Moreover, some participants who were invited for the interviews did not reply, and some declined participation, which further limits the generalizability of our results and calls for further research: More attention needs to be paid to people who are not reached or who do not choose to participate. In addition, in our study, we only assessed barriers and facilitators for the interventions among those who had already agreed to participate. Gaining more in-depth knowledge about those who declined participation in the program could have provided additional information about the acceptance of the program.

Research Implications

MBCTs have proven to be effective for people with cancer [7,11], and here, we showed that participants felt positively about the 2 formats of eMBCT. Although all interviewed participants considered the intervention conditions acceptable, there were differences in their experiences both between and within intervention conditions. Participants experienced the same components of an online intervention in different ways, which is in line with the findings of similar studies [14,37]. For instance, a study about an online MBI among people with cancer found that some participants found the meditations too long, whereas others liked how they enabled them to have time for themselves [37]. In our previous study, we also found that many aspects of the eMBCT (such as the treatment setting and format) were mentioned both as a facilitator and a barrier [13]. In this pilot study, we did not assess specific participants’ characteristics that might explain these differences. Exploring which type of program delivery works for whom can help to establish the best fit for individual patients, balancing effectiveness and the resources required. The subsequent full-scale RCT with a larger and more varied sample will enable us to conduct mediation and moderation analyses to help clarify some of these uncertainties.

It should be noted that participants valued the possibility of following the program at their own time and place. Being able to decide when and where to participate in online interventions among people with cancer has been reported as a positive characteristic among other pilot studies too [14,37,38]. In addition, to our knowledge, there are no studies comparing preferences of people with cancer between online, group, blended eMBCT and individual, unguided eMBCT. This highlights the importance of research on effectiveness among online MBIs for people with cancer.

Conclusions

The main goal of this study was to assess aspects of feasibility of group, blended and individual, unguided eMBCT for people with cancer. This study showed that both intervention conditions were positively received and could potentially be effective. The results of this investigation inform adjustments to the intervention and study process prior to conducting a full-scale RCT to evaluate its effectiveness [18].
Acknowledgments
This project received funding from the Dutch Cancer Society (KWF) under the project number 12125.

Data Availability
The data sets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
NB analyzed the data, interpreted the results, and wrote the first draft of the manuscript. MvK set up the cocreation process and pilot study, developed and adjusted the intervention, and collected data. AS contributed to the design of the study and the application for funding. She was involved in the cocreation process, supervision of the mindfulness teachers supporting the program, and analysis and reporting of the paper. JV analyzed the quantitative part of this study. JP contributed to the study plan and research design and was involved in coding the interview themes and subthemes. SK contributed to the study plan and research design and coconducted the cocreation sessions. LK analyzed the qualitative part of this study, supported the analysis for the quantitative part, and edited all versions of the manuscript. All authors critically reviewed and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Description of the intervention components and interview guide.

References


Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
eMBCT: online mindfulness-based cognitive therapy
MBCT: mindfulness-based cognitive therapy
MBI: mindfulness-based intervention
RCT: randomized controlled trial
Comprehensive Assessment and Early Prediction of Gross Motor Performance in Toddlers With Graph Convolutional Networks–Based Deep Learning: Development and Validation Study

Sulim Chun¹*, BA; Sooyoung Jang¹*, MD; Jin Yong Kim¹, MS; Chanyoung Ko¹, MD; JooHyun Lee¹, BSN; JaeSeong Hong¹, BBA; Yu Rang Park¹, PhD

Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

*these authors contributed equally

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
6th floor
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 2493
Email: yurangpark@yuhs.ac

Abstract

Background: Accurate and timely assessment of children’s developmental status is crucial for early diagnosis and intervention. More accurate and automated developmental assessments are essential due to the lack of trained health care providers and imprecise parental reporting. In various areas of development, gross motor development in toddlers is known to be predictive of subsequent childhood developments.

Objective: The purpose of this study was to develop a model to assess gross motor behavior and integrate the results to determine the overall gross motor status of toddlers. This study also aimed to identify behaviors that are important in the assessment of overall gross motor skills and detect critical moments and important body parts for the assessment of each behavior.

Methods: We used behavioral videos of toddlers aged 18-35 months. To assess gross motor development, we selected 4 behaviors (climb up the stairs, go down the stairs, throw the ball, and stand on 1 foot) that have been validated with the Korean Developmental Screening Test for Infants and Children. In the child behavior videos, we estimated each child’s position as a bounding box and extracted human keypoints within the box. In the first stage, the videos with the extracted human keypoints of each behavior were evaluated separately using a graph convolutional networks (GCN)–based algorithm. The probability values obtained for each label in the first-stage model were used as input for the second-stage model, the extreme gradient boosting (XGBoost) algorithm, to predict the overall gross motor status. For interpretability, we used gradient-weighted class activation mapping (Grad-CAM) to identify important moments and relevant body parts during the movements. The Shapley additive explanations method was used for the assessment of variable importance, to determine the movements that contributed the most to the overall developmental assessment.

Results: Behavioral videos of 4 gross motor skills were collected from 147 children, resulting in a total of 2395 videos. The stage-1 GCN model to evaluate each behavior had an area under the receiver operating characteristic curve (AUROC) of 0.79 to 0.90. Keypoint-mapping Grad-CAM visualization identified important moments in each behavior and differences in important body parts. The stage-2 XGBoost model to assess the overall gross motor status had an AUROC of 0.90. Among the 4 behaviors, “go down the stairs” contributed the most to the overall developmental assessment.

Conclusions: Using movement videos of toddlers aged 18-35 months, we developed objective and automated models to evaluate each behavior and assess each child’s overall gross motor performance. We identified the important behaviors for assessing gross motor performance and developed methods to recognize important moments and body parts while evaluating gross motor performance.
Introduction

For the continuous and proper development of children, an accurate and timely assessment of their developmental levels is essential [1]. Early diagnosis during the toddler stage allows for early intervention, which can significantly impact children’s later life outcomes [2,3]. Previous research has shown that early intervention in vulnerable populations, such as those with low birth weight and prematurity, leads to significant improvements in later childhood developments compared to those who do not receive early intervention and that these differences persist into adolescence [4,5]. Numerous studies have shown that the influence of early intervention extends beyond adolescence to adulthood, with significant socioeconomic benefits [6-8]. A recent study about the development of children exposed to lead showed that early intervention before the age of 3 years benefited their future academic performance [9].

As a result, many countries recommend the need for regular developmental screening of infants and young children, and South Korea has implemented the National Health Screening Program for Infants and Children for children under 6 years of age since 2007 [10-14]. The National Health Screening Program for Infants and Children in South Korea developed the Korean Developmental Screening Test for Infants and Children (K-DST) in 2014, which assesses gross motor, fine motor, cognitive, language, social, and self-help skills in children aged 4-71 months [15].

In various areas of development, gross motor development begins earlier than other areas of development, such as fine motor and language development, and therefore, it is possible to assess the risk of developmental delay at a younger age by monitoring gross motor development. Studies have shown that gross motor development at an early age is predictive of subsequent developments [16] and is also associated with future academic achievement [17,18].

However, a global shortage of pediatric health care providers hinders the proper developmental assessment of children. In a report published in 2016, one-third of pediatricians in the United States did not use standardized screening tools in their pediatric practice because of issues such as limited clinic hours and a shortage of medical staff to perform developmental screenings [19]. As an alternative to pediatric health care professionals, many countries, including South Korea, rely on parental reports to determine developmental milestones. However, parental reports are based on subjective opinions, and parents may respond positively even when they have observed their child’s activities only once, leading to false positives [20]. Based on these factors, there is a need for an objective, labor-free, and automated tool to assess the development of children.

Recently, there have been several studies using deep learning to assess gross motor development in children. A study reported that a deep learning model can predict cerebral palsy progression from videos of spontaneous movements taken in infancy [21]. However, this study did not use a previously validated metric such as the K-DST, which may limit the explainability and generalizability of the model. Liu et al [22] evaluated the gross motor skills of children with autism with an average age of 5 years, and Suzuki et al [23,24] assessed gross motor skills on a video-by-video basis using a deep learning model with behavioral videos of 4- to 5-year-old children. However, since these studies were conducted on children aged ≥4 years, there is a limitation in that they could not validate the model effectiveness in the <3 years age group, where early intervention is expected to be more effective.

Considering these factors, we developed an automated and accurate pediatric developmental assessment model using videos of toddlers aged 18-35 months performing gross motor movements that have been validated with the K-DST. Our 2-step model assesses each behavior and evaluates each child’s overall gross motor performance based on the performance level of each behavior. In addition, we identified behaviors that contribute to the overall gross motor skills assessment and detected critical moments and important body parts for the assessment of each behavior.

Methods

Study Design and Participants

In this study, we used behavioral videos of toddlers aged 18-35 months, when most of them could walk, perform a wide range of gross motor actions, and minimally understand the examiner’s instructions to perform the task [16,25]. We selected 4 behaviors frequently used by the K-DST to assess gross motor development in this age group: climb up the stairs, go down the stairs, throw the ball, and stand on 1 foot [15]. These 4 movements were chosen as core tasks based on existing child development guidelines and in consultation with 3 pediatricians and 15 child development experts, considering the physical and cognitive abilities of this age group [26,27]. The participant performed multiple trials for each behavior. For each of these trials, the raters watched the video and rated the performance as “bad,” “good,” or “perfect.”

We also categorized participants into “relatively slow” and “relatively fast” groups based on their overall performance: if their performance was rated as “bad” on 2 or more behaviors, we categorized them as “relatively slow”; the remaining cases were categorized as “relatively fast.”
**Ethical Considerations**

The data set used in this study is from our previous study and it was constructed while adhering to the ethical principles of the Declaration of Helsinki [28]. The construction of the data set was approved by the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine (4-2021-0845), and the requirement for informed consent was waived due to the retrospective nature of the study. Participants of the data set were recruited from daycare centers, kindergartens, primary pediatric hospitals, and internet communities. Written informed consent for data collection and subsequent analysis was obtained from all caregivers of the participants. Participants received ₩50,000 (approximately US $38) and were provided with an intelligence scale test valued at around ₩300,000 (US $232) as compensation. To ensure the confidentiality and privacy of the participants, each study participant was deidentified via an alphanumeric code.

**Experimental Setting**

The videos were recorded in the presence of caregivers, examiners, and children. Depending on the behavior, a staircase or a ball was used as the apparatus. The video recordings for each child were conducted for approximately 1 hour. A camera was positioned to capture the entire body of each child. Using a frontal angle camera, the child’s behavior was recorded as an RGB (red-green-blue) video. All videos were collected using a Sony DSC-RX100 with 1920x1080 resolution and at 30 frames per second. The collected videos were rated by human raters based on the K-DST criteria, and these values were used as true labels in the stage-1 model.

**Data Preprocessing**

To assess the behavior of the children in the RGB videos, we estimated the position of each child as a bounding box and then extracted 17 human keypoints within the box [29]. To detect the participants, we estimated the bounding boxes using Faster-RCNN [30] with the ResNet 50 backbone in the RGB videos. HRNet was then used to detect human keypoints in the detected bounding boxes [31]. Skeleton data were generated at a rate of 30 frames per second.

**Model Construction**

We divided the data into training, validation, and test sets in a 6:2:2 ratio for each behavior, ensuring that data from the same individual were not allocated across multiple sets. To predict the overall gross motor performance of the children, we designed a 2-stage model. The overview of our model is shown in Figure 1. The first stage is the action evaluation stage, in which each behavior is evaluated separately using a graph convolutional networks (GCN)–based deep learning algorithm. To improve the performance of the stage-1 model, we performed transfer learning with pretrained weights. These pretrained weights are released by PYSKL and are trained with the channel-wise topology refinement graph convolution networks (CTR-GCN) model on the NTU RGB+D dataset by detecting 17 skeleton nodes with HRNet [32-34]. The CTR-GCN model is a stacked structure of 10 basic blocks, 8 of which were frozen during the training on our data. Augmentation using random flipping and scaling was applied to our training data. The training task was repeated 5 times for the same data, and 80 frames were randomly selected each time. The model training strategy of this study and the architecture of the CTR-GCN is shown in Figure 2. A total of 4 CTR-GCN models were trained to generate the predicted probabilities for the 4 gross motor skills, 1 for each behavior [35]. Although these 4 models can assess the performance of each behavior, it was necessary to integrate all 4 models to have a comprehensive assessment of the child’s gross motor development. Accordingly, to assess overall gross motor performance, the stage-2 model aggregated the outcome probability values of each label per behavior. The extreme gradient boosting (XGBoost) algorithm was used for the stage-2 model [36]. The validation process was performed using a 10-fold cross-validation strategy. The parameters used to train our models are shown in Multimedia Appendix 1.
Figure 1. An overview of the suggested 2-stage model for predicting and evaluating comprehensive gross motor performance of children. Faster-RCNN and HRNet were used to extract the skeletal joints from the 4 behavioral videos. The evaluation of each behavior in the stage-1 was performed by graph convolutional networks model separately, and Grad-CAM was used for analyzing the influence of each joint and time segment of the video. In stage-2, the XGBoost algorithm was used for overall performance evaluation, and the SHAP method was used to recognize the contribution of each behavior to the evaluation. B: behavior; C: class; CTR-GCN: channel-wise topology refinement graph convolution networks; Grad-CAM: gradient weighted class activation mapping; GMS: gross motor skills; SHAP: Shapley additive explanations; XGBoost: extreme gradient boosting.

GMS data set

B1: Climb up the stairs
B2: Go down the stairs
B3: Throw the ball
B4: Stand on one foot

Preprocessing

Object detection (Faster-RCNN)
Skeletal detection (HRNet)

Stage 1

GMS level evaluation (CTR-GCN)
Channel-wise topology modeling
Channel-wise aggregation

Stage 2

Overall performance evaluation (XGBoost)
C1: Relatively slow = 4.6%
C2: Relatively fast = 95.4%

Explainability (SHAP value)

B1: +8.3
B2: +3.2
B3: +5.0
B4: -2.1
Figure 2. A detailed architecture of the suggested model and CTR-GCN. We applied augmentation methods and frame sampling strategies. The pretrained weights for the CTR-GCN model were applied, and 8 basic blocks of the model were frozen during training. The Grad-CAM was generated from the gradients and feature map from the last block. The Grad-CAM was then interpolated to align with the input frames. A basic block of CTR-GCN consists of 3 CTR-GCs, which use temporal pooling to aggregate temporal features of skeleton graph sequences and pairwise subtraction and concatenation for correlation modeling between skeletal joints. C: channel dimension of the data; CTR-GCN: channel-wise topology refinement graph convolution networks; GMS: gross motor skills; Grad-CAM: gradient-weighted class activation mapping; MLP: multilayer perceptron; N: number of skeletal joints; T: temporal dimension of the data; Tanh: hyperbolic tangent function; XGBoost: extreme gradient boosting.

Evaluation of Model Performance and Verification of Explainability

The stage-2 model–assessed performance was compared with human panel–assessed performance on a fixed-test data set. A panel consisting of 1 pediatrician and 2 nonexperts assessed the participants’ overall gross motor status. Sensitivity and specificity for each panel were calculated.

For the interpretability of the stage-1 action evaluation model, gradient-weighted class activation mapping (Grad-CAM) was used to identify critical time points and body parts in behavioral videos [37]. To create a Grad-CAM heatmap, we obtained weights for each label through gradient calculation and extracted feature maps from the final graph convolutional layer of the CTR-GCN model. The heatmap was then generated by linearly combining the derived weights and feature maps and applying the ReLU function [38]. We then visualized the heatmap along with the original input, which is a sequence of positions of skeletal joints. To understand the influence of each body part on model decision, we determined the top-1 activated joints that had the highest Grad-CAM values per frame and grouped their frequency by body part [39]. The 17 joints were grouped as belonging to the head, left arm, right arm, left leg, and right leg [39].

In the second stage of overall performance prediction, we used the Shapley additive explanations (SHAP) method to identify the actions that contribute more to the total developmental assessment [40]. The mean absolute SHAP value was obtained to estimate the contribution of each feature to the model output.

Statistical Analysis

The performance of the models was evaluated using the area under the receiver operating characteristic curve (AUROC) score, which was calculated as the average of all folds and presented with SD. The optimal cutoff value for the overall gross motor skill assessment was determined based on receiver operating characteristic analysis using the Youden index. The receiver operating characteristic curve was also plotted with the average of the folds within the threshold intervals and the area between the SDs. All statistical analyses were performed in Python (version 3.6.8; Python Software Foundation) using sci-kit-learn (0.24.2 version).

Results

Characteristics of Cohort Participants

Behavioral videos of the 4 gross motor skills were collected from 141 children, of which 71 (50.4%) were boys, and 70 (49.6%) were girls. The average age of the children in this study was 29.6 (SD 4.3) months. The characteristics of the behavioral data for each gross motor skills are listed in Table 1. A total of 2502 behavioral videos were collected, with 698 (23.9%) rated as “bad,” 581 (23.2%) rated as “good,” and 1321 (52.8%) rated as “perfect.”
Table 1. Characteristics of cohort participants. The 141 participants consisted of 71 (50.4%) boys and 70 (49.6%) girls, and the average age was 29.6 (SD 4.3) months. A total of 2502 behavioral videos were collected, with 698 (23.9%) rated as “bad,” 581 (23.2%) rated as “good,” and 1321 (52.8%) rated as “perfect.” The distribution of the demographics of the population and the number of videos by label for each behavior are represented as the total number and its percentage.

<table>
<thead>
<tr>
<th>Parameter and variable</th>
<th>Climb up the stairs (N=141)</th>
<th>Go down the stairs (N=141)</th>
<th>Throw the ball (n=140)</th>
<th>Stand on 1 foot (N=141)</th>
<th>Total (N=141)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (months), mean (SD)</td>
<td>29.5 (4.3)</td>
<td>29.5 (4.3)</td>
<td>29.5 (4.3)</td>
<td>29.5 (4.3)</td>
<td>28.6 (4.3)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>70 (49.6)</td>
<td>70 (49.3)</td>
<td>69 (49.3)</td>
<td>70 (49.6)</td>
<td>70 (49.6)</td>
</tr>
<tr>
<td>Boys</td>
<td>71 (50.4)</td>
<td>71 (50.7)</td>
<td>71 (50.4)</td>
<td>71 (50.4)</td>
<td>71 (50.4)</td>
</tr>
<tr>
<td>Number of videos by label, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad</td>
<td>137 (21.9)</td>
<td>144 (23.1)</td>
<td>95 (15.0)</td>
<td>222 (35.9)</td>
<td>598 (23.9)</td>
</tr>
<tr>
<td>Good</td>
<td>106 (16.9)</td>
<td>107 (17.2)</td>
<td>156 (24.7)</td>
<td>212 (34.2)</td>
<td>581 (23.2)</td>
</tr>
<tr>
<td>Perfect</td>
<td>384 (61.2)</td>
<td>372 (59.7)</td>
<td>380 (60.2)</td>
<td>185 (29.9)</td>
<td>1321 (52.8)</td>
</tr>
</tbody>
</table>

Performance of the Evaluation of Each of the 4 Gross Motor Skills

Table 2 shows the results of the first-stage model. The AUROC values with each behavioral evaluation were from 0.79 to 0.90. We found that the model for the “climb up the stairs” behavior performed the best, with an AUROC score of 0.90, followed by “go down the stairs” with an AUROC score of 0.86; subsequently, the models for “throw the ball” and “stand on 1 foot” performed similarly, with AUROC scores of 0.79 and 0.80, respectively (Figure 3).

Table 2. Results of the evaluation of the 4 gross motor skills.

<table>
<thead>
<tr>
<th>Performance metric</th>
<th>Gross motor skill, mean (SD)</th>
<th>Climb up the stairs</th>
<th>Go down the stairs</th>
<th>Throw the ball</th>
<th>Stand on 1 foot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.78 (0.02)</td>
<td>0.76 (0.03)</td>
<td>0.68 (0.02)</td>
<td>0.60 (0.02)</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.71 (0.03)</td>
<td>0.67 (0.03)</td>
<td>0.61 (0.02)</td>
<td>0.63 (0.02)</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>0.86 (0.02)</td>
<td>0.85 (0.02)</td>
<td>0.78 (0.01)</td>
<td>0.76 (0.02)</td>
<td></td>
</tr>
<tr>
<td>F1-score</td>
<td>0.72 (0.04)</td>
<td>0.67 (0.04)</td>
<td>0.62 (0.03)</td>
<td>0.60 (0.03)</td>
<td></td>
</tr>
<tr>
<td>AUROCa</td>
<td>0.90 (0.01)</td>
<td>0.86 (0.02)</td>
<td>0.79 (0.02)</td>
<td>0.80 (0.02)</td>
<td></td>
</tr>
</tbody>
</table>

aAUROC: area under the receiver operating characteristic curve.
Figure 3. Performance scores and confusion matrices of the stage-1 gross motor skill evaluation model for 4 behaviors. The bar chart of scores is shown with error bars indicating the range between minimum and maximum scores observed in the cross-validation. The AUC scores range from 0.79 for “throw the ball” to 0.90 for “climb up the stairs.” The confusion matrices show the model’s ability to distinguish between “bad” and “good” labels. AUC: area under the curve.

Grad-CAM on the Visualization of Human Keypoint

Our keypoint-mapping Grad-CAM visualization showed the differences in the activated joints for each behavior and label (Figure 4). By observing the highlighted areas in the heatmap, we could identify the contribution of the joints to the evaluation of each behavior. The horizontal axis, labeled as “time,” indicates the moments of the behavioral video that contributed to the classification across the selected 80 frames of the videos. The vertical axis, labeled as “skeletal joint,” shows the critical joints related with behavior evaluation. For the “climb up the stairs” behavior, the Grad-CAM results of the behavior evaluated as “bad” showed that the Grad-CAM scores of the arms and head increased as the child falls and grabs the stairs with their hands. On the other hand, for behavior evaluated as “perfect,” the child’s legs scored consistently high as they walked up.

It was also observed that the Grad-CAM score was higher when a given task was being performed. In the Grad-CAM results for the child who was rated “bad” for “climb up the stairs,” we could observe that the moment when the child wandered and looked back to the assistant has a lower Grad-CAM value than the moment when they climbed the stairs. To compare the importance of each body part, we determined the top-1 activated joint, which is the joints with the highest Grad-CAM value per frame, and grouped the frequencies by body part. (Multimedia Appendix 2) [39].
Figure 4. Grad-CAM heatmap with frame-by-frame mapped keypoints for each behavior. The change of Grad-CAM values over time for 17 human keypoints was displayed as a heatmap. For each behavior, the Grad-CAM heatmap for a given participant was compared between a “perfect” and “bad” performance. The actions of the participant over time were visualized as human keypoints and shown above the heatmap. The age and gender of each child were displayed together. Grad-CAM: gradient-weighted class activation mapping.

Overall Performance Status Prediction

The results of the stage-2 overall performance prediction model and the human panels on a fixed-test data set are shown in Figure 5. The model had an AUC score of 0.90, and the specificity and sensitivity of the optimal cutoff points were 0.83 and 0.82, respectively. For the human panels, sensitivities of 0.90 and 0.91 and specificities of 0.59 and 0.81 were recorded by nonexperts and an expert, respectively. Comparing each of these showed that the model performed better than the nonexpert panel and was similar to the expert panel. Table 3 shows the overall results of the model.

According to the grouped SHAP value obtained from variables for each action, the action “go down the stairs” contributed the most to the prediction, with a SHAP value of 1.28 (Figure 5). The next highest values were “climb up the stairs,” “throw the ball,” and “stand on 1 foot,” with values of 0.73, 0.65, and 0.36, respectively.
Figure 5. ROC curves, confusion matrix, and grouped SHAP values of the stage-2 overall performance status prediction model. The stage-2 overall performance status prediction model had an AUC score of 0.90, and the specificity and sensitivity of the optimal cutoff points were 0.83 and 0.82, respectively. For the expert panel, the sensitivity and specificity were 0.91 and 0.81, respectively. For the nonexpert panels, the mean sensitivity and mean specificity were 0.90 and 0.59, respectively. In the confusion matrix, we displayed the relative ratio of the predicted values to each actual value. To identify the highly contributed behaviors in the stage-2 overall performance status prediction model, we obtained the SHAP value of each label in 4 behaviors and summed the SHAP values for each behavior. AUC: area under the curve; ROC: receiver operating characteristic; SHAP: Shapley additive explanations.

Table 3. Results of the stage-2 overall performance status prediction model.

<table>
<thead>
<tr>
<th>Performance metric</th>
<th>Performance, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.82 (0.04)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.71 (0.09)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.88 (0.04)</td>
</tr>
<tr>
<td>$F_1$-score</td>
<td>0.74 (0.06)</td>
</tr>
<tr>
<td>AUROC$^a$</td>
<td>0.90 (0.02)</td>
</tr>
</tbody>
</table>

$^a$AUROC: area under the receiver operating characteristic curve.

Discussion

Principal Findings

In this study, we evaluated each gross motor behavior and assessed each child’s overall gross motor performance status using movement videos of toddlers aged 18-35 months. To the best of our knowledge, this study is the first to predict the overall gross motor behavior status using pediatric gross motor movement videos at ages younger than 3 years.

Several previous studies have attempted to predict pediatric development using digital phenotype data, such as detecting developmental disabilities using drag-and-drop data in games [41], identifying visual impairments using gaze patterns and facial feature data in response to visual stimuli on a smartphone, and measuring fine motor skills in children using sensor-augmented toys [42]. Suzuki et al [23,24] conducted studies that collected the behavioral videos of 4- to 5-year-old children and extracted skeletal data through OpenPose to evaluate behavioral performance on a per-video basis using a convolutional neural network and autoencoder model. Liu et al [22] proposed a method to evaluate the initial gross motor skills of children with autism with an average age of 5 years using velocities, trajectories, and angles of upper and lower limb joints based on skeleton data extracted through OpenPose. However, unlike these previous studies of gross motor skill assessments, our study focused on gross motor function in toddlers younger than the age of 3 years, which may allow us to quickly identify developmental delays in children younger than the age of 3 years for early intervention. Additionally, this study not only assessed each behavior but also built a model to evaluate the overall performance of each individual by aggregating the assessments of each behavior.

In this work, we performed action recognition using CTR-GCN on skeleton data extracted through human pose estimation with Faster-RCNN and HRNet [30,31,35]. Recently, many studies have been published on action recognition, which is broadly categorized into RGB-based methods and skeleton-based methods [43]. In this study, instead of RGB-based methods, which directly use RGB video, we used a skeleton-based method using Faster-RCNN and HRNet to estimate the location of human presence as a bounding box and extract human keypoints [30,31]. These skeleton-based methods are not only computationally efficient but also have the advantage of focusing on the child’s behavior and deidentifying the study participants by removing background information [43,44].

For human pose estimation, we used HRNet and Faster-RCNN compared to the studies by Suzuki et al [23,24] and Liu et al [22], which used OpenPose [30,31,45]. In human pose estimation, there are 2 types of methods: the bottom-up method (eg, OpenPose), where each body part is detected first and subsequently the body parts are combined, and the top-down method (eg, HRNet + Faster-RCNN), where the person is detected and then each body part is searched within the detected bounding box [29-31,45]. The HRNet method is known to be more accurate than OpenPose, and the top-down method is expected to be more accurate in detecting body parts, especially...
when there are multiple people in the video [31,32]. Since children are often filmed with their caregivers in the developmental test, the HRNet was more suitable for our study.

The types of behaviors assessed in this study have been used in the K-DST for the corresponding age group, and previous research has shown that these types of gross motor behaviors are good predictors of childhood developmental disorders, such as intellectual disability, autism spectrum disorder, and cerebral palsy [15]. Furthermore, the model we developed in this study provides more objective assessments of gross motor skills than the K-DST, which relies on parental reports and enables the assessment of gross motor skills to be automated without requiring trained pediatric health care providers.

Of the 4 behaviors evaluated, “go up the stairs” was the most accurately classified; however, in the actual model, “go down the stairs” had a higher contribution in SHAP values (Figures 2 and 4). When viewing videos of actual children’s behaviors, we found that while performing the “go down the stairs” behavior, the examiner placed the child on the stairs, and the child subsequently performed the action of going down the stairs to return to the caregiver at the bottom of the stairs without the examiner’s intervention. Other behaviors required frequent intervention by the investigator to encourage the child to perform the behavior successfully, because the children sometimes did not understand the investigator’s instructions (eg, holding up 1 leg for more than 1 second) or had a variety of alternative actions at the onset of the behavior (eg, returning to the caregiver instead of climbing the stairs).

We also aimed to validate the explainability of the model by calculating the Grad-CAM values of each joint for each behavior, frame by frame (Figure 4). This allowed us to identify specific joints that had high importance values at critical points in the child’s behavior. For example, in a video of a child performing the “stand on 1 foot” behavior, when we analyzed the Grad-CAM of each joint on a frame-by-frame basis, we could observe that the importance of the leg joints increased as the child stood on 1 leg. The importance of each joint across the videos was determined by counting the number of times each joint was the most important in a particular frame (Multimedia Appendix 2) [39]. This allowed us to identify the vital body parts for evaluating each behavior. In the case of “climb up the stairs,” for example, it was found that the values in the arm area increased when the child was performing the behavior poorly. This finding can be attributed to the child’s tendency to resort to crawling instead of standing when the child had difficulty climbing, thereby increasing the values in the arm. The analysis of Grad-CAM values per joint in the children’s behavioral videos allowed us to identify which joints were important for certain behaviors and which body parts were more deficient in each child during specific behaviors.

One limitation of this study was that we could not validate the model’s performance in different patient populations. The study used data from participants aged 18-35 months, as this is the developmental stage when children can perform a wide range of gross motor movements, such as walking and running, and can understand simple verbal instructions from the examiner. Therefore, further research is needed to determine which gross motor activities in different age groups can be used to assess gross motor development in children. In addition, because this study was limited to Korean children, we suggest that its applicability should be studied in various settings, including other ethnicities and cultural settings.

Additionally, this study did not collect long-term follow-up prognostic data on the participants, such as the subsequent occurrence of developmental delays. If prospective data had been collected on the occurrence of future developmental disabilities (eg, cerebral palsy and autism spectrum disorders), more thorough studies could have been conducted using our model. Therefore, it is necessary to consider the long-term prognosis follow-up of participants in future studies.

Conclusions
We developed a model to assess 4 behaviors using behavioral video in children aged 18-35 months and to assess each child’s overall gross motor performance. This is the first study to assess the overall gross motor behavioral status of children younger than 3 years of age using gross motor video for automated and objective prediction of child development. We also identified important behaviors during the model's assessment of overall gross motor performance. Furthermore, we developed a method to identify important moments and key body parts during behavioral assessment using Grad-CAM. We anticipate that a more accurate and automated assessment of gross motor development will be possible with this model if more data are available in a variety of settings.

Acknowledgments
This research was supported by a grant from the R&D project funded by the National Center for Mental Health (grant MHER22A01). Funding was used toward labor costs for data collection and analysis. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability
We used publicly available data from our previous research [28]. These data are available on GitHub [46].

Authors' Contributions
SC, SJ, YRP, JYK, and CK contributed to conceptualization. SC, SJ, YRP, JYK, CK, and JSH contributed to methodology. SC, SJ, JHL, and YRP contributed to literature research. SC and JY contributed to data curation. SC, JYK, and JSH contributed to...
artificial intelligence modeling and validation. SC and SJ contributed to statistical analysis. SC, SJ, and YRP contributed to manuscript writing. YRP contributed to supervision. All authors have read and agreed to the published version of the manuscript. All authors had full access to all the data in the study and accepted the responsibility for the decision to submit it for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Parameters used to train the model.
[PDF File (Adobe PDF File), 73 KB - formative_v8i1e51996_app1.pdf]

Multimedia Appendix 2
Distribution of top-1 activated joints per frame grouped by body parts.
[PDF File (Adobe PDF File), 211 KB - formative_v8i1e51996_app2.pdf]

References


Abbreviations

AUROC: area under the receiver operating characteristic curve

CTR-GCN: channel-wise topology refinement graph convolution networks

GCN: graph convolutional networks

Grad-CAM: gradient-weighted class activation mapping

K-DST: Korean Developmental Screening Test for Infants and Children

RGB: red-green-blue

SHAP: Shapley additive explanations

XGBoost: extreme gradient boosting

©Sulim Chun, Sooyoung Jang, Jin Yong Kim, Chanyoung Ko, JooHyun Lee, JaeSeong Hong, Yu Rang Park. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Perception of Apps for Mental Health Assessment With Recommendations for Future Design: United Kingdom Semistructured Interview Study

Erin L Funnell1,2*, BSc; Benedetta Spadaro1*, MPhil; Nayra A Martin-Key1, PhD; Jiri Benacek1, PGDip; Sabine Bahn1,2, MD, PhD

1Cambridge Centre for Neuropsychiatric Research, Department of Chemical Engineering, University of Cambridge, Cambridge, United Kingdom
2Psyomics Ltd, Cambridge, United Kingdom
* These authors contributed equally

Abstract

Background: Mental health care provision in the United Kingdom is overwhelmed by a high demand for services. There are high rates of under-, over-, and misdiagnosis of common mental health disorders in primary care and delays in accessing secondary care. This negatively affects patient functioning and outcomes. Digital tools may offer a time-efficient avenue for the remote assessment and triage of mental health disorders that can be integrated directly into existing care pathways to support clinicians. However, despite the potential of digital tools in the field of mental health, there remain gaps in our understanding of how the intended user base, people with lived experiences of mental health concerns, perceive these technologies.

Objective: This study explores the perspectives and attitudes of individuals with lived experiences of mental health concerns on mental health apps that are designed to support self-assessment and triage.

Methods: A semistructured interview approach was used to explore the perspectives of the interviewees using 5 open-ended questions. Interviews were transcribed verbatim from audio data recordings. The average interview lasted 46 minutes (rounded to the nearest min; SD 12.93 min). A thematic analysis was conducted.

Results: Overall, 16 individuals were interviewed in this study. The average age was 42.25 (SD 15.18) years, half of the interviewees identified as women (8/16, 50%), and all were White (16/16, 100%). The thematic analysis revealed six major themes: (1) availability and accessibility, (2) quality, (3) attitudes, (4) safety, (5) impact, and (6) functionality.

Conclusions: Engaging in clear communication regarding data security and privacy policies, adopting a consent-driven approach to data sharing, and identifying gaps in the app marketplace to foster the inclusion of a range of mental health conditions and avoid oversaturation of apps for common mental health disorders (eg, depression and anxiety) were identified as priorities from interviewees’ comments. Furthermore, reputation was identified as a driver of uptake and engagement, with endorsement from a respected source (ie, health care provider, academic institution) or direct recommendation from a trusted health care professional associated with increased interest and trust. Furthermore, there was an interest in the role that co-designed digital self-assessments could play in existing care pathways, particularly in terms of facilitating informed discussions with health care professionals during appointments and by signposting individuals to the most appropriate services. In addition, interviewees discussed the potential of mental health apps to provide waiting list support to individuals awaiting treatment by providing personalized psychoeducation, self-help tips, and sources of help. However, concerns regarding the quality of care being affected because of digital delivery have been reported; therefore, frequent monitoring of patient acceptability and care outcomes is warranted. In
addition, communicating the rationale and benefits of digitizing services will likely be important for securing interest and uptake from health care service users.

(JMIR Form Res 2024;8:e48881) doi:10.2196/48881

KEYWORDS
app design; digital health; eHealth; interviews; mental health; mHealth; mobile phone

Introduction

The demand for mental health care is steadily rising [1], with mental health disorders being among the leading causes of global disease [2] and disability burden [3]. Despite the documented consequences of delays in providing support [4,5], there is inadequate service coverage and poor quality of care within most health care systems globally [6], with documented disparities between the quality of physical and mental health care [7].

In this regard, mobile and internet technologies, such as web and smartphone apps, have been identified as potential tools to improve access to care by mitigating some of the economic, geographic, and human resource constraints posed by in-person care [8]. Evidence gathered from numerous studies supports the view that the internet and mobile apps have become a source of health information, support, screening, and disease management, as well as being the entry point for referral processes with various degrees of success in implementation.

In the United Kingdom, there is an urgent need for innovative mental health technologies, as despite a lifetime prevalence of 1 in 4 adults experiencing a mental health disorder [9], many individuals who seek help face considerable barriers to effective support. First, an estimated 1 in 35 people are waiting for specialist mental health care in the United Kingdom [10], with the reported wait for specialist care often exceeding a month [4]. Patients who spend more time on waiting lists report worsening mental health symptoms [4,5]. In addition, there are reports of impairments in social or relationship domains, with some individuals reporting that their mental health deteriorated to the extent that they needed emergency care [4].

Second, there is well-documented evidence of mental health misdiagnosis. It is estimated that 90% of mental health concerns are managed entirely in primary care settings [11]. Despite this, there is frequent misidentification of common mental health disorders, such as major depressive disorder [12,13], bipolar disorder [14], and anxiety disorders [15]. There is also evidence that some general practitioners (GPs) feel their surgeries are underprepared to provide adequate mental health care [16]. A plethora of interconnected reasons may explain the misdiagnosis of mental health conditions, such as the overlapping symptom profiles of different psychiatric disorders [17], short consultations [18], and a lack of appropriate training [19]. These assessment difficulties are compounded by person-level barriers, such as difficulties in accurately communicating symptoms to health care professionals [20]. Additional barriers to diagnosis include geographic variability in resources and the availability of trained mental health care professionals.

In contrast, smartphone ownership is becoming widespread, all-time high, and the increasing number of mobile device users has created an unprecedented opportunity to develop evidence-based mobile apps for remote delivery of mental health care. There is evidence of interest from individuals with mental health concerns in digital tools designed for mental health assessments, particularly when they are integrated into care pathways, with results delivered directly to a health care professional before their appointment [5]. Individuals appear to feel more comfortable disclosing sensitive information digitally [21] than health care professionals. In addition, digital self-assessment and triage support tools may be useful in augmenting existing services by ensuring timely response and intervention to urgent cases [22,23] and signposting appropriate services outside formal health care.

However, despite the interest in and promise of mental health apps, prolonged engagement and use remains an issue [24]. This poses the question of how to design tools that are usable and useful for individuals experiencing mental health concerns. Iterative user-centric research and design are key to addressing this challenge. Indeed, work has been done to explore the opinions of apps designed for mental health concerns and disorders [25-29]. However, given the interest in digital tools intended for mental health self-assessment and triage, more work is required to understand the perspectives and acceptability of such tools designed and implemented to augment existing care pathways.

Therefore, in this study, we conducted semistructured interviews with individuals who represented potential app users: those with lived experiences of mental health concerns. The key objective of this study is to advance the understanding of potential users’ views and perspectives on mental health apps, with a specific focus on apps designed for self-assessment and triage. To this end, the semistructured interview included eight questions broadly focused on the following topics: (1) previous use and perception of mental health apps; (2) perspectives of mental health assessment apps; (3) perspectives of opportunities for integration of mental health apps, including mental health assessment apps, into traditional care; (4) perspectives of the safety and privacy aspects of mental health apps; and (5) desired app features for an ideal mental health app.

Methods

Overview

The methods and results presented in this study are reported in-line with COREQ (Consolidated Criteria for Reporting Qualitative Studies [30]; Multimedia Appendix 1). The research team comprised 3 research assistants (ELF, BS, and JB), a research associate (NAM-K), and a practicing psychiatrist and

https://formative.jmir.org/2024/1/e48881
professor of neurotechnology (SB). All authors had previous experience in qualitative data analysis. JB is identified as male, and all other authors are identified as female. This study was a follow-up to a previous web-based survey study [5].

**Participants and Recruitment**

Participants from a previous web-based survey study [5] who expressed interest in participating in follow-up studies were contacted via email between April and June 2022 and invited to participate in this interview study. The email addresses collected in the prior survey study [5] used to contact potential participants for this study were stored in a locked (password-protected) Excel file, only accessible to members of the research team who were named after ethical approval. The inclusion criteria were as follows: (1) aged ≥18 years, (2) living in the United Kingdom, and (3) having to visit a health care professional after 2016 to discuss their mental health symptoms. The first 2 rounds of recruitment were blinded to the demographic characteristics. In the second and third rounds of recruitment, only participants who had identified as men or nonbinary in the survey study were contacted to increase the representation of these groups in this study sample.

**Study Procedures and Materials**

Participants were first provided with written details of the aims, methods, and requirements of this interview study via email. They were then asked to express their interests and provide consent to participate via email. Upon the expression of interest and receipt of consent, the date and time for the interview agreed with the researchers (ELF and BS). The interviews were conducted by an interviewer (ELF or BS) in the presence of an observer (ELF or BS).

All study materials, including the participant information sheet, consent form, interview guide, and debriefing, were developed in consultation with the senior author (SB) of a practicing psychiatrist. In addition, some questions in the interview guide were adapted from previous relevant literature investigating attitudes toward digital interventions for mental health [31,32]. All study materials were then further amended and finalized in consultation with members of the Cambridge University Hospitals Patient and Public Involvement panel, who have lived experiences of mental health concerns. This strategy was used to ensure the suitability and relevance of the study materials to the target population.

A semistructured interview guide was used to facilitate the conversation, which included open-ended questions to encourage participants to discuss their perceptions of digital tools (eg, applications) for mental health. The guide included eight questions (with further prompts) focusing on (1) previous use of mental health apps, (2) perceptions of mental health apps, (3) perspectives of apps for mental health assessment, (3) perspectives of receiving an indication of a mental health diagnosis from an app, (4) perspectives of information collected by a mental health assessment app being sent to a health care professional, (5) perspectives on whether apps can improve access to mental health services, (6) views on safety and privacy related to mental health apps, and (7) desired features for future mental health apps. In addition, the participants were asked (8) if they had any additional thoughts relevant to mental health and digital technologies (eg, apps). A copy of the semistructured interview script is available in Multimedia Appendix 2. The interviews were adaptive, such that only relevant questions were asked based on previous responses. The interviewer could reformulate or clarify questions during the interviews to gain a deeper understanding of the participants’ thoughts and opinions, or delve into relevant details that were mentioned in relation to the questions asked. The audio of the interview was recorded for subsequent transcription. After completion of the interview, all participants were provided with debrief via email and offered a £15 (~US $18) Highstreet voucher for their time.

**Data Collection**

Interviews were conducted between May 5, 2022, and June 22, 2022. Overall, 14 (88%) of the interviews were conducted using Zoom (Zoom Video Communications Inc) videoconferencing, 1 (6%) using Microsoft Teams videoconferencing, and 1 (6%) via telephone call. Participants were informed that they did not need to have their cameras on during the interview. Audio data were recorded for all the interviews. In total, 12 hours and 25 minutes (rounded to the nearest min) of interview audio were recorded, with the average length of the interview being 46 minutes (rounded to the nearest min; SD 12.93 min). The interviews were transcribed verbatim from audio recordings using cloud-based AI-powered software Otter [33]. The researchers (ELF and BS) then reviewed the transcripts by listening to the audio recordings and amending the transcripts where necessary. Any unclear audio segments were labeled “unintelligible.” Interview transcripts were numbered and not connected to any identifiable information (ie, email address for recruitment and participation reimbursement). Any names or identifying information (eg, city of residence) disclosed during the interviews were removed from the transcripts. The transcripts were downloaded as PDFs and analyzed as described below.

**Data Analysis**

The data were analyzed using a bottom-up (data-driven) thematic analysis approach based on the Braun and Clarke framework [34]. A total of 2 authors (ELF and JB) analyzed all the transcripts under blinded conditions using the following process: the first interview transcription was analyzed, and initial codes were identified. The second interview was analyzed by checking for the presence of codes identified from the analysis of the first interview and adding any new codes identified. This process was continued for each interview transcript, each time adding or refining existing codes. Upon completion of the analysis, the authors were unblinded and compared their lists of identified codes. Any inconsistencies were discussed with a third author (BS or NAM-K) until consensus was reached. Once the codes were agreed upon by all authors, they were organized into themes under blinded conditions by 2 authors (ELF and JB). The resulting themes were discussed with a third author (BS and NAM-K) until a consensus was reached. Data analysis (ie, code creation and assignment, theme creation, and assignment) was performed using Google Sheets spreadsheets. A copy of the codebook organized into their respective themes is available in Multimedia Appendix 3.
Ethical Considerations

This study was approved by the University of Cambridge Human Biology Research Ethics Committee (approval number: PRE.2021.053). The participants provided informed consent electronically via email to participate in the study. In one case, a participant provided consent at the start of the interview because they had experienced difficulties in sending a complete consent form via email. On the day of the interview, the interviewer (ELF or BS) verified that participants understood the information that had been provided to them and gained verbal consent that they were happy to continue with the interview.

Results

Sociodemographic Characteristics

A total of 16 individuals were included in this study. The average age was 42.25 (SD 15.80), with 50% (n=8) of the participants identifying as women and all White (16/16, 100%). English was the native language of 87% (14/16) of the interviewees. More than 69% (n=11) had at least one undergraduate degree. A total of 43% (7/16) were single and 43% (7/16) were cohabiting. Regarding accommodation characteristics, living alone or with a partner was the most common arrangement, with 38% (n=6) living alone, 25% (n=4) living with a partner, and 19% (n=3) living with a partner or children. A total of 37% (6/16) were employed, and 62% (10/16) had a household income of less than £35,001 (approximately US $43,608) before tax.

Mental Health Characteristics

The majority (15/16, 94%) of respondents had discussed their mental health with a GP in the last 5 years, with 88% (n=14) having also seen a therapist or counselor. More than half (9/16, 56%) of the participants had also seen a psychiatrist. Mental health care visits were typically provided free of charge via the National Health Service (15/16, 94%). A total of 81% (13/16) of the interviewees were diagnosed with a mental health disorder (see Table 1 for a breakdown of the frequency of diagnoses). The most common diagnosis in the sample was major depressive disorder (13/16, 81%), followed by generalized anxiety disorder (5/16, 31%), and posttraumatic stress disorder (5/16, 31%).

Table 1. Frequency of mental health diagnoses in the sample (N=16).

<table>
<thead>
<tr>
<th>Mental health condition</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major depressive disorder</td>
<td>13 (81.25)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>2 (12.50)</td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>5 (31.25)</td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>3 (18.75)</td>
</tr>
<tr>
<td>Panic disorder</td>
<td>1 (6.25)</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>1 (6.25)</td>
</tr>
<tr>
<td>Sleep disorder</td>
<td>2 (12.50)</td>
</tr>
<tr>
<td>Posttraumatic stress disorder</td>
<td>5 (31.25)</td>
</tr>
<tr>
<td>Personality disorder</td>
<td>2 (12.50)</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>2 (12.50)</td>
</tr>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>3 (18.75)</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>1 (6.25)</td>
</tr>
<tr>
<td>Learning disability</td>
<td>1 (6.25)</td>
</tr>
</tbody>
</table>

Thematic Analysis

Thematic analysis revealed 6 major themes comprising 25 minor themes (Figure 1).
**Availability and Accessibility**

Availability and accessibility emerged as key themes in this study, with issues surrounding (1) app stores, (2) exclusion, (3) environment, (4) convenience, and (5) employee benefits emerging as important minor themes.

Regarding the first minor theme, app stores, interviewees reported that there is too much variability in what app store searches return, which could make app stores difficult to navigate and, in turn, be overwhelming for users:

*It’s overwhelming and it can be just really difficult to find what you’re looking for as well. You know, if you just put in a search term, and then you just get kind of a load of random stuff. That’s not necessarily categorized very well, I don’t think.* [Transcript 5]

In addition, concerns were raised about groups of individuals with specific conditions being excluded from the adoption of a digital approach to mental health, with gaps in the market for apps for certain mental health conditions, such as neurodevelopmental disorders:

* [...] specialized things like ASD specifically, I’ve not really seen anything in terms of like apps to help with anything like that sort of helping with meltdowns and things like that.* [Transcript 4]

Interviewees also noted a lack of diversification in-app offerings, expressing concerns that available apps are basic and essentially all offer the same thing:

*I think the ones that the NHS or the GP or the CPN have recommended they seem to be quite basic. I suppose cause the NHS makes them, they make it for everybody, so you can’t really fine tune it, I don’t think.* [Transcript 15]

*But in terms of my perception, I think, I don’t think they have a lot of options out there. I mean, there’s a broad range. I mean, you’ve got apps, that you know, under different names, but they all, they all work the same.* [Transcript 13]

In terms of the second minor theme, exclusion, interviewees raised concerns related to severe or complex mental health symptoms impeding users’ ability to actively engage with a mental health app:

*I don’t think I’d have the mental abilities to read a question, understand it, put in an answer, look at my...*
answers, look for any pattern there might be or anything like that. I think I wouldn’t be able to do it. [Transcript 6]

In addition, with regard to exclusion, not being able to afford a potential one-off or subscription payment was raised as a concern:

I know some of them have like paid options, which sometimes can be a bit of a barrier. I’m in a position now where I can like pay for subscriptions, but in the past like when I was a university student I think that was one thing that sometimes put me off was having to pay for them. I used to use one calm. I found it was actually quite expensive and there’s like other apps that are completely free, that do the same thing. [Transcript 7]

Moreover, exclusion due to digital delivery, particularly in the elderly, those with limited digital literacy, or individuals with physical health conditions such as visual impairments or dyslexia, was raised as an important barrier:

[...] it would exclude some people who haven’t got the internet. And it would exclude a lot of older people who might have it and haven’t got a clue how to do it. So, so that would be one of the disadvantages of excluding some people [...] [Transcript 6]

The environment was also identified as a minor theme within the availability and accessibility theme, with interviewees raising the issue of not having the right environment to complete a digital mental health assessment:

[...] when you’re in a house when everybody else is stuck at home as well you don’t get peace and quiet. You actually need to concentrate on it. [...] When you’re trying to sit down at home in a busy house doing it, you’re never guaranteed to have the time you need to actually finish the section in one go, having to restart things three sometimes four times just to actually work through it. [Transcript 8]

In contrast, interviewees also reported that digital technologies are convenient and improve accessibility to mental health services:

[...] a lot of people can’t, can’t get to places. I don’t just mean COVID I mean, they haven’t got the money or it takes longer or whatever.” “But then the same things that are advantages are disadvantages. Because like people who are physically disabled, it must, it’ll be a lot easier for them to use to do things online. [Transcript 6]

as well as fitting well into modern daily routines:

I find that the good thing of them, is that you can use them, do them anytime. Can do it at your own pace and at home so you don’t have to depend on anyone to do it or, or have a scheduled time to. [Transcript 14]

Finally, regarding the role of the workplace, it was noted that engagement with mental health apps could improve if provided as an employee benefit:

Yeah, I mean, I got it free from my work. They had an agreement with them, so I feel I’ll try it anyway. [Transcript 11]

**Quality**

The quality theme was identified in this study, encompassing the minor themes of (1) subjectivity, (2) scientific rigor, (3) background and development, and (4) perceived value.

Interviewees mentioned the subjective nature of mental health symptoms. Because of their reported subjective nature, symptoms were deemed difficult to consolidate with the formal diagnostic labels and descriptions of mental health disorders, causing distrust regarding the validity of diagnoses:

I must say it’s all very subjective, isn’t it? Your own feelings are all what you feel and then you think just that is really what they mean? Am I feeling what they, what they mean? [Transcript 3]

Related to this aspect of subjectivity, interviewees mentioned concerns related to self-reporting of mental health symptoms in self-assessments. Some interviewees expressed that the subjective and hard-to-define nature of mental health symptoms can create a risk that users of mental health apps may unintentionally report their symptoms inaccurately:

So[...] you know the thing is people are expressing their symptoms, but they might be inaccurate symptoms. It might only be that person’s interpretation of what they’ve got. [Transcript 3]

Others have stated that app users may intentionally incorrectly report or exaggerate their symptoms to support self-diagnosis of a specific mental health disorder:

I guess there’s always the risk of people not being honest. And maybe people exaggerating symptoms or something like that. I know there’s a lot of talk at the minute, isn’t it about social media and people like self-diagnosing from social media, and I know it’s only a small percentage of people, but I guess that would be a concern. [Transcript 7]

The subjective nature of mental health symptoms and concerns related to self-reporting are both associated with potential difficulties in using digital tools for accurate mental health assessment, with interviewees reporting concerns that mental health symptoms may be difficult to capture in assessments that rely on predefined multiple-choice questions:

[...]it’s difficult if you’re, what you’re experiencing doesn’t fit in with the responses that have been configured. [Transcript 1]

I’d be concerned about using the app on its own to do a mental health assessment just because I think it can lack nuance sometimes. [Transcript 16]

One potential method, offered by an interviewee, to address the challenges of subjectivity outlined here was to allow for clinician-facilitated symptom reporting, wherein the app user and clinician could collaboratively complete the mental health assessment. This collaboration would allow the user to provide
details of their symptoms or experiences, which the clinician can use as a basis to answer the questions within the assessment:

[... ] the GP, for example, could go up to [...] his or her clients and say, ‘Oh, what do you want me to put to this question’ and that, they’ll just give you, the client will just give you a tidbit and then they’ll just put that down. I think that would help. [Transcript 13]

In terms of scientific rigor, interviewees expressed concerns about a mental health assessment app’s diagnostic accuracy, stating that apps may be able to suggest a condition that is a probable explanation for a user’s mental health symptoms but that it would likely not be completely accurate:

I don’t think an app will ever truly get a diagnosis correct. It’ll probably get a rough estimate. [Transcript 13]

Interviewees also raised concerns that due to the overlapping symptom profiles of many mental health disorders, there is a potential risk of misclassification of symptoms:

I think not too sure about the apps for online diagnoses in general as well. I suppose with a lot of the conditions, the elements of different mental health issues, it might not be that you’ve got say bipolar [...] you might have a different mental health condition and you seem to fit the criteria for bipolar. [Transcript 15]

With these concerns about diagnostic accuracy, interviewees reported a preference for apps that provide mental health screening to indicate probable present conditions, which can then be confirmed by a mental health care professional, over formal mental health diagnosis:

I don’t think a doctor ought to diagnose someone on the basis of an app. I think they both have to go together. [Transcript 6]

In addition, a preference for an indication of where their symptoms lay on a severity scale was also mentioned by several respondents as part of the reported desire for screening, as opposed to a binary “present” or “not present” assessment outcome more akin to a diagnosis:

I think I’d be more comfortable with saying like, you’re showing depressive symptoms of moderate severity. Here’s a list of the conditions that typically show these symptoms. [Transcript 16]

Interviewees also expressed the importance of evidence-based assessment in encouraging trust in both the assessment itself and the results:

So yeah, I think it’s helpful if it would, if it was like a regulated test and things like that. [Transcript 4]

The importance of an evidence-based assessment was also highlighted by interviewees if the results of the assessment were intended to be shared with the user’s health care provider (eg, GP, psychiatrist):

That’s [sending results directly to a health care professional after a digital MH assessment] a really good idea. If the app had been developed by professionals and it had been approved as being scientific, meeting all the criteria. [Transcript 15]

One source of apprehension disclosed by interviewees regarding engaging with mental health apps is the potential for nefarious digital tools (eg, apps, websites). This included tools that were either making recommendations not based on clinical evidence or tools that were perceived to have been designed with the sole intention of making a profit for the developer, rather than to help app users:

There’s got to be something like that out there, you know, like dodgy websites, which are not quite breaking the law, but they’re not following the proper recommended guidance. Some of them can look really professional but the real interest is money really. [Transcript 15]

I don’t like the idea of people profiting off of people’s data about mental health. [Transcript 2]

A concern shared by interviewees was the lack of regulation and quality assurance:

I don’t know who would monitor or approve or disapprove of these apps, because there’s so many thousands, tens of thousands of apps for everything and there’s no sort of overseeing body which could say “you can’t say that.” [Transcript 15]

Some interviewees expressed concerns about the appropriateness of the self-help tips offered in mental health apps. This concern focused on self-help advice offered in mental health apps often being designed for general well-being, and thus, typically not being beneficial for individuals experiencing symptoms of a mental health disorder:

[Mental health apps] try to provide solutions, and they say that “oh, just do a breathing exercise.” And you know, I used to be a yoga instructor. So I know a lot about breathing and breathing exercises, and they’re really good, but if you’re if you’re struggling, if you know if you’re having a panic attack, breathing is not going to help. [Transcript 2]

Concerning the minor theme of background and development, there was interest in ensuring that key stakeholders were included in the development and design of mental health assessment apps. In terms of including clinicians in the design, it was reported as important to verify that mental health assessment apps are designed to best support existing care services:

I guess I’d [be] keen to, yeah, to want mental health professionals to have sort of contributed to the app to make sure it is going to reduce waiting lists, rather than just do nothing. [Transcript 16]

Beyond designing an app that can offer benefits within the clinical pathway, there was a mention of the importance of engaging with the intended population during development to establish which features and designs are required for accessibility:
So I guess in the design phases, checking with those individuals with those conditions and sort of saying, is this accessible to you? What would make it accessible to you? [Transcript 16]

Perceived value is a potential facilitator of app uptake. That is, the benefits conferred by demonstrating the efficacy of the mental health app, especially when the app offers therapeutic or self-help functionalities:

I think it always helps to know that it’s been effective for people like [...] especially for doing like therapy skills and therapy techniques. [Transcript 7]

However, the caliber of the evidence required to demonstrate the app’s efficacy was not disclosed, so it is unclear whether positive reviews or word-of-mouth recommendations are sufficient or if a more formal evidence base is necessary to promote engagement with mental health apps.

The perceived value of the app as determined by the user was deemed to impact app use and linked to drop out. Interviewees stated that they had forgotten to use or lost motivation to use apps if they did not find them valuable:

But yeah, I didn’t really find much value in it and I, I forgot to use it [...]. So I just kind of in the end, I just didn’t, didn’t bother. [Transcript 11]

With interviewees additionally stating that if the novelty of the app wears off over time there is a risk of dropout:

I only really used that for a few days. It’s one of those things, you know, download something and it looks really cool and then the novelty wears on really fast. [Transcript 5]

In contrast, interviewees discussed how apps perceived as fun encouraged continued use by incentivizing users to return through gamification:

there’s sort of incentive to play every day and you can buy him [habit tracker app mascot] outfits and stuff like that, and it’s just a bit more fun. [Transcript 4]

Attitudes

Attitudes toward using digital technologies for mental health were revealed to be an important theme, with issues surrounding (1) openness to digitalization, (2) interest in using a digital mental health assessment, (3) apprehension, (4) the value of reputability, and (5) stigma emerging as key minor themes.

In this study, attitudes regarding being open to digitalization were overwhelmingly positive:

Yeah, I’m more than happy with that don’t have an issue at all. I think we use apps all the time, don’t we for pretty much everything. [Transcript 11]

In contrast, the importance of being interested in using a mental health app in the first place was highlighted:

[...] don’t want to [use a mental health app], I’m not interested. I’ve never thought about any of them, and not being interested in using a mental health app when symptoms are not present. [Transcript 6]

I’m not exactly sure why but have a huge desire not to monitor anything when, when I’m feeling fine. [Transcript 6]

Apprehension was revealed to be an important factor in determining mental health app use, particularly regarding the lack of face-to-face contact:

I find the lack of human contact, I find it, I find it the downside of it. [Transcript 14]

and distrust for apps that appear in advertisements:

I’m suspicious of some of the things that I see pushed in ads and things like that. [Transcript 4]

In contrast, regarding the minor theme of the value of reputability, interviewees stated that app endorsement from a source that the individual perceives as reputable, such as a trusted institution (ie, a university, health care service, such as the National Health Service, a mental health charity) or a direct recommendation (ie, from a health care provider, or a trusted friend or family member) could drive their intention to use the app:

I suppose anything reputable. [...] any institution that I feel comfortable with, be it a service or a, or research department, that would influence me? Yeah. [Transcript 6]

If it was given under an accredited body like NHS Wales, and Mind, then that will put a slightly different complexion on it for me. [Transcript 3]

But also friends and family. If they’ve used it, that’s important. [Transcript 11]

Finally, stigma was seen as an important driver of app use, with interviewees expressing that a mental health app could reduce feelings of judgment and isolation:

I don’t know if that maybe helps with the stigma a little bit. If you know that all of your friends have got that app downloaded, then you know that you’re not the only one that’s really struggling. [Transcript 7]

Safety

A major theme identified in this study was safety, comprising four minor themes: (1) data security; (2) disclaimers; (3) concerns about the potentially triggering nature of digitally delivered mental health assessment results; and (4) data sharing, consent, and safeguarding.

Interviewees expressed the importance of data security of mental health apps:

Well, all I can say is obviously privacy, confidentiality is extremely important. [Transcript 6]

I can see that the GDPR and privacy concerns would be high. [Transcript 1]

The level of trust interviewees expressed regarding apps’ data security policies varied, with some stating that mental health apps are unable to provide basic data security:

At the moment, don’t quite think we’re here. Not quite there. Purely and simply, from my perspective, in and around the areas of control and security. That’s my
Interviewees verbalized concerns over the potential triggering nature of receiving mental health assessment results digitally, using the words “frightened,” “worried,” and “overwhelmed.” There were 2 dimensions related to this issue, namely, concerns about the assessment’s diagnostic accuracy:

[...]

and a potential lack of timely access to clinicians with whom one could discuss and validate the results:

And I can’t see why anybody else would not find it frightening because if somebody was going to use this method, then let’s just you know, you can’t get a doctor’s appointment, a physical appointment about three weeks down here. So I would have to sit on that and then get a doctor’s appointment. So I, I would find it, and also seeing it in black and white what was wrong with me. No, I think that would send me into a spiral. [Transcript 3]

In contrast, another interviewee expressed that receiving the results of a mental health assessment app may be reassuring rather than triggering based on their own previous experience using digital mental health self-assessment tools:

Yes, it [receiving mental health results from a digital tool] wasn’t intimidating if you like. When I got the high score, I thought “Yeah, this is good,” I didn’t see it as a negative thing at all. [Transcript 15]

Interviewees praised the benefits of having an app that is integrated into the existing health care pathway where the results of an assessment would be directly sent to a health care professional:

I’d prefer that [mental health assessment results being sent directly to a clinician] rather than letters really. It’s much quicker, instant information. [Transcript 11]

However, related to the minor theme of data sharing and consent, interviewees indicated that the integration of app data into the existing care pathway must be authorized by the app user through a consent-driven approach to result sharing:

[...] you shouldn’t be forced into presenting this information [results from a digital mental health assessment to a health care professional]. [Transcript 13]

with a preference for multiple opportunities to provide consent (eg, at the point of download of the app and at the point of sending the data to a clinician) to share their mental health data with a relevant health care professional:

I would just like to consent at the moment of sending. So I think there’ll [have to] be two ways of consent. There is a general consent when I joined the app, right? But there is also a consent before it’s actually, actually shared. Yeah, if there are these two layers of consent. [Transcript 13]

A preference for risk safeguarding was expressed with the suggestion that mental health apps should be built with...
functionality to override the consent requirement if the user discloses a substantial level of risk:

I also think it should be a bit of a disclaimer like if you disclose a level of risk, we will send it on automatically to a mental health care professional just so people who are disclosing levels, like levels of risk aren’t [..] just left to it. [Transcript 16]

as well as including crisis signposting in app:

...when you are logging your mood, if you log you know, really bad, it might prompt you to say, you know, I noticed that you’re not feeling so good, would you like to call this number or would you like to look at this website. [Transcript 4]

In addition, there was a suggestion to pair requesting consent to share mental health data with health care clinicians to facilitate informed decision-making, wherein the user has the agency to decide which care pathway they prefer to receive help from:

[...] you could have [an] option saying “okay, we’ve identified this, where would you, we recommend that you should go to a GP, you know, but we also have these, these options. Where do you consent for us to send your information? [Transcript 11]

Impact

The impact of a mental health assessment app on mental health care provision was deemed an important theme in this study, with interviewees discussing issues about (1) the effects on the existing care pathway and (2) support and signposting outside formal health care. Regarding the former, interviewees positively viewed the impact that mental health assessment apps could have on existing care pathways:

So to actually have that [results of a mental health assessment completed in an app] sort of sent to someone [a clinician] and have that resource available, I think that’d be really, really valuable, actually. [Transcript 5]

I have [had] GPs in the past that haven’t really had much training or like knowledge in mental health. So I think maybe if I could then go to my GP and say, “Well, I’ve done this and it’s said that it indicates this diagnosis” I think it could maybe look like make that easier. [Transcript 7]

However, the importance of the role of the health care professional in reviewing any output from a mental health assessment app:

I think it’s really important to pass on this information for the counselor of GP or whatever to look over and then they’ll decide whether that information is accurate [Transcript 13]

It means the human being can just kind of clarify answers and make sure that whatever decision the app has come to is actually accurate [unintelligible] just checking up with the actual person. [Transcript 16]

It should be emphasized that any app integrated into existing care pathways should support clinicians and not replace them:

It’s about accelerating my access to a person or a group of people that can help me and not replacing that with a digital application because I don’t believe in that. [Transcript 2]

Interviewees expressed that completing a digital mental health assessment could facilitate discussions with their health care providers:

One of my big worries is booking doctor’s appointments and struggling to word what I am trying to put across. I think a report sent across with an opinion or something, yeah, to my GP so that I do not have to do it myself and it sounds more reliable. I suppose, would be really, really helpful actually personally. [Transcript 4]

and could help triage patients to appropriate health care professionals or services:

I think on the whole, the app could make an accurate assessment of yeah, which pathway was best. [Transcript 16]

[...] for lower risk perhaps you can still defer, defer them to somewhere else to see if you can offload the pressure load of the GPs and leave it for for higher needs patients, I guess. [Transcript 14]

as well as offering tracking of mental health symptoms during treatment, which can be shared with clinicians to inform:

I can show that [data from mental health tracking app] to her [community psychiatric nurse], and she can take a screenshot for my records, and then we can just discuss if there’s any external triggers or you know how I feel about noticing the slow change in my mood by going back to the app seeing the graphs and the change. [Transcript 15]

In contrast, interviewees expressed concerns regarding the potential of a mental health assessment app to further overwhelm existing services, creating a bottleneck where clinical needs are revealed, which cannot be met by existing mental health services:

[...] is this going to open the floodgates and then there aren’t the services to meet all the huge need that you uncover. [Transcript 6]

There are additional concerns that the quality of mental health care may worsen because of digitization, with face-to-face care being replaced with digital care:

Like if it’s [an app] an option that is there as well as, but I would, I would be, my worry will be if that one, one day that[...] would be the substitute to be seen a person [Transcript 14]

Even when it is potentially inappropriate for higher-severity or complex mental health symptoms:

I guess my like, that my concern is that sometimes would it just become like, they just start signposting
With this in mind, interviewees stated that there was a need for clinicians to refer patients to mental health apps to justify digitization:

*I think I would ask why the doctor is presenting me with an app. When it’s something we couldn’t do with ourselves.* [Transcript 3]

In terms of the impact a mental health assessment app could have on support and signposting outside of formal health care, interviewees discussed the opportunities that mental health assessment apps can offer in terms of encouraging help-seeking:

*I think that would be the sort of ideal outcome that I would look for if I was completing some sort of assessment. Obviously, you’re completing it because you think that there might be an issue. And then, at the end of the day, that’s the kind of first port of call that you know, you would go and see a health professional.* [Transcript 5]

*I think it can be difficult sometimes to actually go and ask for it. Even if you know you’re really struggling, it can be hard to actually go and have a conversation. So I think you could maybe help people who are quite anxious about going to see their GP if they know that they’ll get some information first.* [Transcript 7]

In addition, interviewees expressed the potential for support to be provided to users while they were on a waiting list for further assessment or appropriate treatment:

* [...] if you’re on a waiting list being told that there are these apps that might be able to help you while you’re on that waiting list to manage, you know, managing emotions or whatever it is, then that would be good.* [Transcript 1]

In addition to improving access to faster help by signposting support outside of traditional care services:

*it might make it [wait times] shorter because if they are then being directed to other quicker options. That might be better for the traditional services.* [Transcript 11]

as well as the app having the potential to provide users with sources of help and self-help tips, rather than solely focusing on an indication of a diagnosis:

* [...] as long as it had additional information and useful links and things like that, that go alongside it, and indicators of where they can go to maybe get help, or you know, to help themselves while they’re waiting to get an appointment, for example, then that would be that would be more useful than just an indication of a diagnosis and nothing else.* [Transcript 1]

Potentially, like, self-help type resources as well. So like information, and I guess there are sometimes like courses you can do on like anxiety management, or whatever.* [Transcript 16]

### Functionality

This major theme included the following minor themes: (1) usability, (2) features, (3) customization, (4) personalization, and (5) interoperability.

In terms of usability, interviewees commented on the importance of app design, with a focus on a preference for a professional interface that may foster a degree of trust in the reputation of the mental health app:

*I guess, just the, you know, the app, if the app looked professional, and it looked well done, and you know, it doesn’t look like someone had coded it together in their bedroom or something like that.* [Transcript 5]

In addition to the importance of app design, interviewees expressed that ease of use facilitates app uptake and continued app use:

*I think there should be an app that’s quite basic and as well as functional but easy to understand.* [Transcript 13]

with perceived difficulties in using the app constituting a potential barrier to disengagement with the mental health app:

*I think being user friendly is to the real... is the real key and anything that was too wordy to read on a screen I was just ‘Oh well I can’t deal with that.* [Transcript 5]

As part of usability, interviewees commented on the importance of in-app guidance to support users in their interactions with the app:

*I think it would have been better if it was more guided. You know, so perhaps, set a set schedule of tasks to do or something. I am sure there was a lot of information in it, but it just was the format that[...] I didn’t like really.* [Transcript 11]

Therefore, this could be a feature included in the app to promote ease of use, particularly in self-help apps that often reportedly rely on self-motivation:

*I think a lot of them are more like self-help geared so I suppose if you struggle to do sort of stuff on your own, then you might struggle to use the apps, because they are very self-guided.* [Transcript 7]

Finally, in terms of usability, during the first use of the app after downloading, there was a preference for the ability to explore the app before any sort of commitment was asked (eg, creating an account or having to enter payment information). Part of this preference was related to being able to ensure that the features included in the app were relevant to the needs of the user and suited their preferences:

*I open one and it’s immediately like, sign in, well I might just delete that one because I don’t want to sign in and I’ll go for the next one. Whereas if it’s open up, and I could take a look around the features and then decide, oh, actually, there’s something that I’d like to use, then I’d be happy enough to sign up. Yeah.* [Transcript 5]
A common feature mentioned by interviewees was tracking functionality, including monitoring of the user’s mood and mental health symptoms:

[...] it would be best if you could sort of track your mood over weeks or months or days. Maybe have graphs or a little bit of information available on [...] what’s average for you. [Transcript 4]

In addition, interviewees expressed that an app with functionality to monitor both physical and mental health data would be useful for identifying triggers for changes in mood or mental health symptoms:

It’s all in one which is quite useful because then you can see like... hmm... I think it helps identify triggers sometimes like, yeah, if [I] had a really bad night’s sleep, then that might explain why the next day, I wasn’t feeling very good. It’s quite helpful having it all in one place. [Transcript 7]

The final type of tracking that interviewees regarded as helpful was habit tracking, that is, a functionality where a user can set tasks to complete throughout the day with the help of reminders:

[...] you can set tasks every morning like brush your teeth, get out of bed, go outside, then like by ticking all of those off your little bird buddy goes off on an adventure. [Transcript 4]

Interviewees expressed that in-app reminders are a helpful feature and, in some cases, are the most useful features:

I think the most useful things for me are setting reminders to take my medication on, on an app. [Transcript 4]

Interviewees also highlighted the importance of social connectedness and the value of a mental health app that offers peer support:

And so from my perspective, my personal opinion, is that an app would be a value if it helps me connect with other people. [Transcript 2]

In terms of features specific to mental health assessment, interviewees emphasized the importance of comprehensiveness:

definitely more comprehensive assessments that kind of go through different types of diagnosis and criteria, like sounds like would be a lot more very valuable than just putting everyone in like either depression or anxiety buckets [Transcript 10]

And having multiple answering modalities beyond only offering predefined answer options:

It depends on how am I answering these questions? I mean, is it a free text? Is it voice? Is it video? Is it typing? Is it selecting? Because if it’s yeah, if the questions are open and I can be very subjective about it, then I would like that, probably. If it’s just selecting, like PHQ and GAD, that’s kind of, I don’t find value in those. [Transcript 2]

With free-text options being important to add nuance:

if you had a multiple choice for you to be for there to be an option at the bottom of it to say, you know, why did you answer this way or is there any additional information that you want to give because often there is [Transcript 10]

However, despite the positive remarks provided about the possible features offered within mental health apps, interviewees reported that some features included within apps previously used by interviewees were counterproductive. Interviewees commented that some app features interfered with the usability of the app:

[The app] helps you plan your day and to get less distractions and feel more organized throughout the day. It was beeping and it was saying ‘Have you completed the task you set this morning? How do you feel about the task? Did it go well?’ It was all these bullying questions. [...] It’s too intrusive. [...] I just found it really annoying. Sometimes I didn’t want to answer any questions. But if you didn’t, you’d get a prompt an hour later or something. [Transcript 15]

Therefore, customization of app features to users’ requirements and preferences (eg, changing reminder frequency, personalizing a crisis plan, changing the design) was also highlighted as an important aspect of functionality to avoid the features becoming irritating or counterproductive, leading to potential discontinuation of use:

I control the level of notifications. You know, some people might like lots, some might not want any. You know, times of notifications, hints, tips, that sort of thing. [Transcript 11]

I think having the ability to set notifications at a specific time or just once a day, is really helpful to remind me to actually log what I’ve done and things like that. [Transcript 4]

In addition, customization can offer users the freedom to choose their preferred method of engagement with a mental health app. Interviewees expressed interest in various modalities of information delivery within apps, beyond written text. Specifically, there was an interest in apps offering videos that provide information about mental health disorders:

I think having a whole library of videos [for mental health information] would help. [Transcript 13]

In addition, interviewees expressed that they would find personalization of a mental health app output (eg, self-help recommendations, psychoeducation) valuable:

Yeah, I think as I mentioned, sort of programmed approach or maybe personalized in some way, depending on maybe an initial assessment so it’s you know more individualized. [Transcript 11]

Interviewees reported an interest in apps that are interoperable, providing the ability to link mental health apps with other health data collection devices, specifically physical monitoring devices such as Fitbit:

[...] it could link with, you know, Google Fit, you know, because obviously exercise and you know,
mental wellbeing do go together. So, it’d be good to just link in with that. [Transcript 11]

**Discussion**

**Principal Findings**

This study aimed to explore the perspectives of individuals with lived experiences of mental health concerns on mental health apps designed for self-assessment and triage. A semistructured interview approach was used, and the findings revealed 6 key themes: availability, accessibility, quality, attitude, safety, impact, and functionality. These themes provide insight into how potential users experiencing mental health symptoms may perceive and use mental health self-assessments and triage apps. These findings can help app developers design and improve these technologies to better support the mental health needs of current and potential future users.

**Availability and Accessibility**

Concerns pertaining to the theme of “availability and accessibility” were raised by interviewees in this study. They noted that mental health symptoms can make it difficult for individuals to engage with digital mental health assessments and triage tools. Indeed, mental health symptoms and cognitive deficits associated with conditions such as depression and psychosis may make it challenging for people to use such technologies [36]. This highlights the importance of designing accessible and user-friendly digital mental health tools. To achieve this, app developers should consider involving real-life patient populations in the design and evaluation of new technologies [37] to ensure that the tools are appropriate and effective for intended users.

In addition, the risk of digital exclusion was raised by the interviewees, particularly in relation to older individuals who may not be digitally literate. Previous research has shown that older adults are less likely to use technology and the internet compared with younger individuals [38]. To address this, training in digital tools for mental health may prove effective in supporting older adults to engage with these technologies and avoiding exclusion from accessing important support and resources [39]. Interviewees also mentioned that it was important to ensure that digital tools were not the only “front door” to care services, guaranteeing that those who may be excluded from engaging with such tools are still able to access help through existing pathways.

Moreover, interviewees raised concerns about the lack of diversification within the app landscape, in terms of which mental health disorder apps were available. Specifically, many interviewees referred to the perceived lack of available apps designed for neurodevelopmental disorders. To address this, app developers should avoid contributing to the perceived oversaturation of apps for general common mental health conditions (ie, depression and anxiety disorders) and instead identify opportunities to design apps for a more diverse range of conditions, especially as this study demonstrates a clear interest in accessing such apps if they are available. In addition, investigating gaps in the app market and investing in co-designing approaches with users could help improve uptake among typically excluded populations [40,41].

Finally, interviewees suggested that offering mental health apps as an employee benefit can improve accessibility. Previous research has shown that digital interventions for employees experiencing psychological distress can be highly scalable and cost-effective [42-44] conferring positive impacts on well-being and productivity outcomes such as sleep, stress, and presenteeism [44].

**Quality**

Interviewees identified the “quality” of available digital mental health assessments as an important consideration. They mentioned that the subjective nature of mental health symptoms could make it difficult to accurately assess and diagnose mental health conditions. However, it is worth noting that this has also been mentioned in traditional face-to-face mental health assessments. The risks of inaccurate self-reporting due to recall and perceptual biases are widely recognized [45]. For example, previous work has demonstrated only a moderate correlation between self-reported length of sleep and actigraphy measurements, with individuals overestimating sleep duration [46]. In addition, prior research on bipolar disorder shows that individuals have trouble remembering the details of earlier manic or hypomanic episodes [47,48]. This subjectivity and difficulty in accurate reporting can lead to the misclassification of symptoms and, in turn, misdiagnosis.

Furthermore, interviewees commented on their concerns that the available digital tools designed for mental health self-assessment could provide inaccurate diagnostic results. Indeed, many of the currently available apps designed for mental health do not provide evidence of their efficacy, accuracy, or effectiveness [49,50]. There is limited high-quality evidence on the diagnostic accuracy of available digital mental health assessments, which vary in their performance when compared with a gold standard clinical interview, with some showing poor discriminatory and differential diagnostic performance and others demonstrating excellent accuracy [51]. To address this, interviewees emphasized the importance of ensuring that digital mental health assessments are clinically validated and evidence-based, particularly if the results are intended to be shared with health care professionals. A further way to address this concern is for app developers to create apps that offer screening and an indication of where a user’s symptoms fall on a severity scale rather than a diagnosis, particularly given that interviewees in this study expressed a preference for screening overdiagnosis. To increase engagement and build trust in the app and its assessment results, there should be functionality to easily share results with a health care professional for evaluation and confirmation.

Another concern raised by the interviewees was the potential for nefarious digital tools and inappropriate information or self-help advice targeting vulnerable individuals. Previous work has demonstrated that these concerns may be well founded, as some available mental health apps inappropriately promote the medicalization of normal mental states [52]. These concerns also reflect the current lack of regulation and quality assurance in the digital mental health field [53].
Attitudes

The theme of “attitudes” toward digital mental health technologies was also identified in the study as an important factor influencing the use of mental health assessment and triage apps. Similar to previous research, openness to digital technologies was identified as a key driver of engagement in digital mental health technologies [54]. In addition, and perhaps unsurprisingly, this study demonstrated the importance of interest in using these technologies, with a lack of interest constituting a significant barrier to the initial uptake. In fact, some interviewees described lack of interest as the fundamental reason for not currently engaging or not intending to engage with mental health apps. Previous research has shown that providing relevant and customizable content can increase interest, and in turn, the uptake of digital mental health technologies [55]. Therefore, engaging with stakeholders in the design phase of such technologies is critical to ensure that the content is relevant to the intended user population, driving interest, and engagement. However, it is important to note that in some groups, their lack of overall interest in such digital tools will always remain a barrier, and, as previously stated, nondigital access to services must be maintained so as not to inadvertently exclude individuals.

In addition, the value of reputation was identified in this study, supporting previous studies [55]. Interviewees emphasized the importance of the app being endorsed by a reputable source, such as a research institution, health care provider, trusted friend, or family member. This can help increase trust in the app and reduce stigma. Interviewees also expressed distrust of mental health apps that are advertised in paid advertisements, as advertising may indicate that the app was designed primarily for commercial gain rather than therapeutic benefits. Therefore, the initial engagement and use of an app can potentially be increased by improving the prospective user’s perception of the app’s reputation through endorsements from trusted individuals or organizations and by ensuring that any paid advertisements are not perceived negatively.

Furthermore, beyond just passive reputation, this study additionally determined that an active recommendation of a specific app by one’s health care provider is an excellent strategy to encourage uptake, as this fosters a sense that the app will be both highly relevant and effective in managing the patient’s conditions and needs. This reflects previous qualitative findings regarding the potential importance of a health care professional’s recommendation to encourage interest in and uptake of mobile apps, specifically for managing depressive symptoms in primary care [56]. Given this evidence demonstrating the potential influence of health care professionals on the uptake of digital mental health tools, ranging from assessment and triage to management (ie, symptom tracking and self-care as an adjunct to formal pharmacological or psychological treatment), primary care clinicians should be aware of their capacity to signpost patients and identify appropriate opportunities to do so. There are resources available to support clinicians in identifying high-quality apps to recommend to patients, such as the Orcha app library [57] and the American Psychiatric Association’s app evaluation model screener [58], which could be valuable assets in encouraging digital tool uptake in patients.

Safety

With respect to the “safety” theme, some interviewees raised concerns regarding trust in digital mental health technologies, particularly in terms of data sharing and anonymity. Some individuals may be wary of sharing their sensitive, personal information or seeking support through digital mental health technologies because of doubts regarding the confidentiality or anonymity provided by these apps. Unfortunately, these concerns may be well founded, as despite mental health apps collecting some of the most sensitive personal information, their data security provisions often do not differ from those of nonmental health apps [59]. Interviewees expressed that to address these concerns, digital mental health apps must provide clear and transparent information on how they handle user data. Lamentably, despite interviewees conveying a desire for this information, many mental health apps do not offer a privacy policy to users [60]. Of those that do provide a privacy policy, many demonstrate low readability scores [61], potentially fostering a sense of mistrust in how collected data are being analyzed and used. Conversely, some interviewees expressed a more nonchalant attitude regarding data security.

Interestingly, this study identified how one’s interaction with nonmental health apps can influence one’s perception of the data security of mental health apps, with interviewees claiming that they have minimal concerns about sharing sensitive data as nonmental health apps, specifically social media and banking apps, already collect data of a perceived similar sensitivity. Despite this apparent lack of concern on the part of app users, app developers are still responsible for upholding the required levels of data security.

In addition, interviewees communicated the importance of providing users with the option to opt out of data sharing. Interviewees recommended that, ideally, this opting-in would include asking for explicit consent at multiple time points during the use of the app, before any data were inputted, and then again before sharing any calculated results with a care professional. Currently, many publicly available apps treat continued use as a proxy for the user’s consent [61] to collect and analyze mental health data rather than asking explicitly for consent periodically, despite potential changes to the app’s data collection and analytic strategy, data sharing, or privacy policy. In addition, even if mental health apps were to adopt a consent-driven approach (ie, where consent is obtained at multiple time points), problems related to consent would persist because of the low readability scores of app privacy policies. As discussed above, users may not truly understand what they are consenting to [61].

However, interviewees expressed that, in some cases, it may be necessary to override the user’s consent to share results with a health care professional if they disclose a substantial level of self-harm or suicide risk.

Another requirement raised by the interviewees was the need for a diagnostic disclaimer. This means that nondiagnostic mental health apps should clearly state that their assessments are intended for screening purposes only and should not be used as a substitute for a professional diagnosis. Many available mental health apps lack such disclaimers [62]. Providing such a disclaimer clearly in an app store description can help prevent
users from overreliance on the results of these assessments and may encourage them to seek confirmatory assessments or support from qualified professionals.

Impact

The theme of “impact” was also identified in the study, with interviewees commenting on the effects that digital mental health assessments can have on current care pathways. They suggested that digital mental health assessments can facilitate informed discussions with health care professionals, echoing a sentiment identified previously in the user feedback of a novel digital mental health assessment [63].

However, some interviewees also expressed concerns that mental health apps could further overwhelm existing mental health services, particularly if they led to an influx of new users seeking support. One method proposed by interviewees to proactively avoid overwhelming services was to co-design apps intended to be implemented in existing clinical pathways with relevant care professionals. This would ensure that they are a valuable tool and not a hindrance to the established delivery of care [64]. Although digital mental health technologies will not immediately solve issues of long waiting lists and a lack of trained mental health care professionals, they can allow cost-effective and time-efficient collection of patient and symptom data. In addition, interviewees expressed concerns that the digitization of existing mental health services may result in poorer care, particularly for those with more severe or complex mental health symptoms. This view supports the notion that digital tools should be used for augmentation, rather than replacement, of existing services to support clinicians in the delivery of care.

A concept that was discussed positively by interviewees in this study was the potential of apps to support the waitlist management of mental health services. Mental health triage apps have the potential to direct patients to the most appropriate formal care pathway or other services (ie, local charity, self-help resources, psychoeducation) based on their symptom profile, potentially alleviating the burden on the health care system by providing individuals with mild or subthreshold symptoms with self-help tips and psychoeducation, reserving GPs, or specialized services for more severe or complex cases. Moreover, interviewees overwhelmingly expressed an interest in mental health apps that can support users while they are on the waiting list before receiving formal treatment by offering not only an assessment and triage, but also self-help tips, psychoeducation, and sources of help. Psychoeducation has been demonstrated to increase mental health literacy, decrease feelings of stigma, and increase intention to seek help [65]. In addition, psychoeducation has been shown to improve outcomes in bipolar disorder, as measured by hospital admissions [66]. Considering the wait time between assessment and psychological treatment in the United Kingdom and the associated potential for the deterioration of symptoms and well-being [4,5], any opportunity to arm individuals with resources to support their own mental health should be explored. However, beyond the potential to improve patient experience while on the waitlist, more work should be done to investigate the therapeutic benefits of providing self-help information to patients awaiting formal treatment.

Functionality

Finally, in terms of “functionality,” similarly to previous work, interviewees mentioned that ease of use and in-app guidance would increase app use [55], especially for apps that require a high level of motivation, such as mental health apps. In terms of assessment specific features, interviewees mentioned the importance of comprehensive assessments to ensure that the complete picture of mental health symptoms and contributing factors is captured. In addition, interviewees highlighted the importance of including varied answering modalities to capture the nuances of experience associated with specific symptoms. This reflects the sentiment from user feedback of a digital mental health assessment, which found that questions with predefined answer options cannot always correctly capture symptoms with requests for the ability to add free-text data [63]. This study demonstrated that users are keen to engage with apps that offer a wide range of complementary functions beyond self-assessment to support mental well-being. The desirable app functions mentioned by the interviewees included mental health symptoms and habit tracking. Mental health symptoms and mood tracking are popular features of mental health apps [29] and can be valuable to users who wish to increase their awareness of their mental health concerns [67]. Beyond only mental health symptom tracking, interviewees expressed a preference for apps with the ability to track physical health alongside mental health in a single app, thereby providing more comprehensive insights (ie, menstrual cycle, exercise). Further demonstrating the importance of ease of use from the perspective of interviewees in this study, apps that offered interoperability and facilitated the ability to link mental health apps to other health data collection devices (eg, Fitbit) were preferred. This data sharing between devices could combine physical health data (ie, heart rate, sleep amount, and quality) with manually entered mental health symptom data to help individuals identify mood triggers and patterns as well as self-management of their mental health symptoms.

In addition, the opportunity to gain peer support through mental health apps was considered important by interviewees as it has the potential to address the missing “human” aspect of digital mental health technologies. In addition, some interviewees mentioned that in the past, using social media for peer support enabled them to learn more about their illness from other individuals who had similar symptoms, which is why they thought it would be beneficial to have a mental health app. Peer support has been shown to improve feelings of hope, empowerment, and social functioning [68], with the consolidation of early stage evidence investigating digital peer support demonstrating acceptability and positive effects on functioning and outcomes [69].

Although reminders were viewed as an important and useful feature of mental health apps, interviewees stated that too many reminders and notifications were also considered intrusive and could lead to app discontinuation. However, facilitating the customization of these features could promote engagement, further corroborating previous findings [63]. In addition, offering
different modalities (eg, text, audio, and video) is seen as an important factor in engagement and app usage. Therefore, app developers should ensure that such features are easily customizable to optimize usefulness from the user's perspective (ie, by ensuring that reminders to use the app are delivered when the user can do so) and are not overwhelming.

As discussed above, there was interest from interviewees in resources to support self-management of mental health symptoms while on a waiting list (ie, self-help tips and psychoeducation). However, it is important to note that personalization was expressed as important to interviewees, with a preference for personalized rather than generic self-help tips and psychoeducational information. Therefore, although there may be temptation from app developers to create apps using generic self-help, this may impede interest and eventual engagement with their app. Therefore, offering self-help tailored to specific conditions or populations is preferable. Doing so may also go some way toward alleviating some of the concerns expressed by interviewees in the quality theme concerning the appropriateness of self-help offered in apps, providing app users an opportunity to more confidently choose apps based on what is most relevant to them, as well as ensuring that the information is suitable for their mental health symptoms beyond general mental well-being.

Creating an account by entering personal or payment information could be a barrier to app use identified in this study due to a sense of suspicion in users about data mining, reflecting previous work [29]. Interviewees expressed an interest in being able to access the app initially without creating an account, wherein they could ascertain the relevance of the app to their individual needs and preferences before choosing to provide personal details and payment information where applicable.

Implications for Practice

Apps adopted for mental health assessment and monitoring present an interesting use case of mobile health technology, and our understanding of their acceptability and perception of benefits and barriers to adoption continues to evolve. Findings from these interviews generated valuable insights into putative users' perspectives, concerns, and preferences regarding the use and design of apps for mental health assessment. Taken together, these results demonstrate that there is real interest in tools designed for mental health self-assessment and triage, particularly when:

1. The self-assessment and triage tool is offered with additional complementary app features that are personalized or can be customized to the individual user.
2. Self-assessment is comprehensive and includes answering modalities beyond selecting predefined answer options to capture the nuances of symptoms and experiences.
3. The tool is easily accessible, as it does not require the entry of sensitive or financial information for use or is offered via a workplace well-being scheme.
4. The tool is recommended by or associated with a reputable source, either passively (ie, developed with expert academics) or actively (ie, a direct recommendation from a friend or trusted health care professional).

However, despite this interest, individuals have real concerns that may impede their uptake and prolong engagement. This study also elucidates some simple ways in which these concerns can be addressed:

1. It appears that by engaging in clear, transparent, and accessible communication about the app’s evidence base and privacy policy in the app description, some of the concerns related to “quality” and “safety” can be addressed before the user even downloads the app.
2. Apps that collect sensitive mental health data and have the functionality to share such data with an individual’s care provider should ask for consent multiple times to establish permission to collect, analyze, and then share.
3. Engaging in co-design activities with both users and clinicians can ensure that the app is (1) widely accessible to many users regardless of the severity or complexity of their mental health symptoms, their level of digital literacy, or the specific condition, and (2) can be effectively integrated into care pathways to support care professionals in the delivery of care. Ensuring access to publicly available digital mental health assessment and triage tools can be achieved through a diverse app marketplace offering that includes typically overlooked conditions such as neurodevelopmental disorders.

Technology integration into mental health care presents opportunities and challenges rooted in the intersection of technical and human factors. Building trust with key stakeholders, such as clinicians, patients, and commissioners, on whom the success of digital integration into existing services will depend, is crucial [70]. Careful design, evaluation of tool performance, and establishing relevance to the population of interest [71] are essential to harness a technology’s potential for improving mental health care delivery and minimizing the possible integration challenges arising from the interaction between technical and human factors. In this study, interviewees seemed to positively view the role that digital tools designed for mental health self-assessment and triage could play in established care pathways if responsibly co-designed with health care professionals for successful integration to not overwhelm services. Indeed, technology can sometimes create a barrier to forming a strong therapeutic alliance between patients and health care professionals. In this regard, efforts should be made to ensure that technology augments rather than replaces human interactions. A crucial aspect of this will be interdisciplinary collaboration and co-design of tools and integration strategies. In addition, although the potential for service digitization is certainly attractive, it should be approached cautiously. Interviewees in this study reported concerns regarding lower quality of care resulting from the integration of digital delivery, and continuous monitoring and appraisal of care quality are vital following the deployment of digital tools. In addition, engaging in conversations to explain the rationale and benefits of using digital tools in the delivery of digital care will likely build trust and drive uptake in its use.

Strengths and Limitations

This study has several strengths. For instance, the use of a qualitative approach allows a nuanced and in-depth exploration
of a range of potential users’ perspectives on mental health apps specifically designed for self-assessment and triage. Qualitative research is a powerful tool for generating insights and informing practical solutions and real-world actions. By emphasizing participant perspectives, qualitative research can capture thoughts, feelings, and context, pinpointing subtleties that quantitative methods may miss. In addition, the use of PPI in this study allowed for optimization of the study design and materials, including the suitability and relevance of the interview questions to the population of interest.

Despite these strengths, the results of this study should be viewed through the lens of the questions asked in the semistructured interview as they may have shaped the findings. For example, as we specifically asked interviewees about their views regarding the privacy and security of digital tools, it may have overrepresented the importance of this theme when some individuals would not have mentioned it if not prompted.

Furthermore, insights from this study were drawn from a small cohort of participants and therefore cannot be generalized more broadly, particularly as the entirety of the sample was White. In addition, these exploratory interviews provide insights from a population recruited through web-based social media, mainly through Facebook advertisements. Thus, the resulting cohort may be biased toward individuals who are familiar with digital technologies; thus, the findings may indicate a more positive perspective than that observed in a formal health care setting. However, the cohort in this study offered a wide range of perspectives, including individuals without previous experience interacting with mental health apps and services.

Conclusions
The adoption of mental health assessment and triage apps presents a significant use case for mobile health technologies. Insights from this study indicate user preferences for mental health apps with personalized features, easy accessibility, and that are recommended by or associated with institutions or individuals perceived as reputable. Concerns about quality, safety, and data privacy can be addressed through clear communication, consent-driven data collection and sharing processes, and co-design with users and clinicians. The positive perception of digital tools in established care pathways highlights potential opportunities for commissioning mental health care services and waiting-list management. However, further research is needed to assess the suitability of digital assessment and triage tools for different psychiatric populations, and to determine their impact on clinical outcomes. In addition, steps must be taken to ensure that concerns regarding the potentially detrimental impact of digitization on care quality are addressed when referring or signposting to digital mental health tools.

Acknowledgments
This study was funded by Stanley Medical Research Institute (grant 07R-1888). The authors are most grateful to all participants for their time and for openly sharing their views and opinions with them. The authors also thank Dr Amanda Stranks and Bethany Vincent at Cambridge University Hospitals Patient and Public Involvement (CUH PPI) for connecting us with the members of the CUH PPI panel. The authors thank the CUH PPI panel members for their time and valuable input in designing the study material.

Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
ELF, BS, NAM-K, and SB conceived of the study’s focus and materials. The recruitment was conducted by BS with the aid of ELF. The interviews were conducted by ELF and BS. Thematic analysis was performed by ELF and JB as the first reviewer and BS and NAM-K as the second reviewer. ELF, BS, and NAM-K prepared the manuscript, with revisions from JB and SB. All authors have contributed to the manuscript and approved the submitted version.

Conflicts of Interest
Funding for this study was provided by the Stanley Medical Research Institute (grant 07R-1888). SB is a cofounder and chief medical officer of Psyomics, Ltd. SB holds shares at Psynova Neurotech Ltd. and Psyomics Ltd. ELF is a paid consultant for Psyomics, Ltd.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Studies) checklist. [DOCX File, 22 KB - formative_v8i1e48881_app1.docx]

Multimedia Appendix 2
Semistructured interview guide. [DOCX File, 23 KB - formative_v8i1e48881_app2.docx]
References


4. Two-fifths of patients waiting for mental health treatment forced to resort to emergency or crisis services. Royal College of Psychiatrists. URL: https://www.rcpsych.ac.uk/news-and-features/latest-news/detail/2020/10/06/two-fifths-of-patients-waiting-for-mental-health-treatment-forced-to-resort-to-emergency-or-crisis-services [accessed 2023-02-25]

5. Spadaro B, Martin-Key NA, Funnell E, Benáček J, Bahn S. Opportunities for the implementation of a digital mental health assessment tool in the United Kingdom: exploratory survey study. JMIR Form Res 2023 Aug 07;e43271 [FREE Full text] [doi: 10.2196/43271] [Medline: 37549003]


53. Wykes T, Schueller S. Why reviewing apps is not enough: transparency for trust (T4T) principles of responsible health app developers and academics in the co-design of a digital information tool. PLoS One 2021 Jul 16;16(7):e0253448 [FREE Full text] [Medline: 36339864]


Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Studies
GP: general practitioner

© Erin L Funnell, Benedetta Spadaro, Nayra A Martin-Key, Jiri Benacek, Sabine Bahn. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Co-Production of a Flexibly Delivered Relapse Prevention Tool to Support the Self-Management of Long-Term Mental Health Conditions: Co-Design and User Testing Study

Alyssa Milton\textsuperscript{1,2,3}, PhD; Ingrid Ozols\textsuperscript{4,5}, MMHS, MSuicidology; Tayla Cassidy\textsuperscript{6,7}; Dana Jordan\textsuperscript{1}, MPH; Ellie Brown\textsuperscript{8,9}, Grad Cert, D CounsPsy; Urska Arnautovska\textsuperscript{10}, PhD; Jim Cook\textsuperscript{11}, MSc; Darren Phung\textsuperscript{11}, BDesComp; Brynnnor Lloyd-Evans\textsuperscript{12}, MSc, PhD; Sonia Johnson\textsuperscript{12}, MBChB, MSc, MD; Ian Hickie\textsuperscript{2}, AM, MD; Nick Glozier\textsuperscript{1,3}, MA, MSc, MBBS, PhD

\textsuperscript{1}Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia
\textsuperscript{2}Brain and Mind Centre, University of Sydney, Camperdown, Australia
\textsuperscript{3}ARC Centre of Excellence for Children and Families over the Life Course, Sydney, Australia
\textsuperscript{4}mentalhealth@work (mh@work), Melbourne, Australia
\textsuperscript{5}Faculty of Medicine Nursing and Health Sciences, Monash University, Melbourne, Australia
\textsuperscript{6}One Door Mental Health, Sydney, Australia
\textsuperscript{7}School of Social Work and Arts, Charles Sturt University, Canberra, Australia
\textsuperscript{8}Orygen, Parkville, Australia
\textsuperscript{9}Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia
\textsuperscript{10}Faculty of Medicine, The University of Queensland, Woolloongabba, Australia
\textsuperscript{11}TechLab ICT, University of Sydney, Sydney, Australia
\textsuperscript{12}Division of Psychiatry, University College London, London, United Kingdom

Corresponding Author:
Alyssa Milton, PhD
Faculty of Medicine and Health
The University of Sydney
Level 5 Professor Marie Bashir Centre, 67-73 Missenden Rd
Camperdown, 2050
Australia
Phone: 61 02 9515 1461
Email: alyssa.milton@sydney.edu.au

Abstract

Background: Supported self-management interventions, which assist individuals in actively understanding and managing their own health conditions, have a robust evidence base for chronic physical illnesses, such as diabetes, but have been underused for long-term mental health conditions.

Objective: This study aims to co-design and user test a mental health supported self-management intervention, My Personal Recovery Plan (MyPREP), that could be flexibly delivered via digital and traditional paper-based mediums.

Methods: This study adopted a participatory design, user testing, and rapid prototyping methodologies, guided by 2 frameworks: the 2021 Medical Research Council framework for complex interventions and an Australian co-production framework. Participants were aged ≥18 years, self-identified as having a lived experience of using mental health services or working in a peer support role, and possessed English proficiency. The co-design and user testing processes involved a first round with 6 participants, focusing on adapting a self-management resource used in a large-scale randomized controlled trial in the United Kingdom, followed by a second round with 4 new participants for user testing the co-designed digital version. A final round for gathering qualitative feedback from 6 peer support workers was conducted. Data analysis involved transcription, coding, and thematic interpretation as well as the calculation of usability scores using the System Usability Scale.

Results: The key themes identified during the co-design and user testing sessions were related to (1) the need for self-management tools to be flexible and well-integrated into mental health services, (2) the importance of language and how language preferences vary among individuals, (3) the need for self-management interventions to have the option of being supported when delivered in
services, and (4) the potential of digitization to allow for a greater customization of self-management tools and the development of features based on individuals’ unique preferences and needs. The MyPREP paper version received a total usability score of 71, indicating C+ or good usability, whereas the digital version received a total usability score of 85.63, indicating A or excellent usability.

Conclusions: There are international calls for mental health services to promote a culture of self-management, with supported self-management interventions being routinely offered. The resulting co-designed prototype of the Australian version of the self-management intervention MyPREP provides an avenue for supporting self-management in practice in a flexible manner. Involving end users, such as consumers and peer workers, from the beginning is vital to address their need for personalized and customized interventions and their choice in how interventions are delivered. Further implementation-effectiveness piloting of MyPREP in real-world mental health service settings is a critical next step.

KEYWORDS
self-management; serious mental illness; self-care; digital health tools; blended interventions; peer support; mobile phone

Introduction

Background

Serious mental health conditions, including schizophrenia, bipolar disorder, and major unipolar depression, are associated with longer term use of mental health services [1]. Despite this substantial need, service provision is limited by funding and workforce constraints, and additional ways of supporting individuals with serious mental ill-health are required. Self-management programs have been developed to assist individuals with serious mental health conditions to actively understand and manage their own health [2]. Core components of self-management include psychoeducation, relapse prevention; the identification and avoidance of stressors; the development of effective coping strategies; and, often, a recovery element [3,4]. There is now substantial meta-analytic evidence that the provision of supported self-management programs (ie programs with guidance from a health professional or another helper) alongside standard care improves outcomes for people experiencing serious mental health conditions, including significant reductions in symptom severity, shorter length of admission, improved functioning, and better quality of life [4].

As a result of this robust evidence base [4], global health policies increasingly emphasize the importance of self-management interventions to support individuals with severe mental illness in managing their own health [5,6], and they are now included as a best practice recommendation in clinical guidelines (eg, the United Kingdom’s 2014 National Institute for Health and Care Excellence (NICE) Guidelines for Schizophrenia [7]). Despite this recommendation, self-management interventions are not always offered as a standard in services for people experiencing serious mental ill-health [4,8]. This is in contrast to services for people experiencing chronic physical illnesses, such as diabetes [9], where self-management programs are a core component of routine practice. Overall, self-management programs have the potential to benefit individuals experiencing serious mental ill-health and reduce strain on health services and have an economic rationale for reducing treatment costs [10], but implementation challenges remain.

Self-management programs typically involve the use of a collaborative learning process that encourages people experiencing serious mental ill-health to become experts in their own recovery [2,11]. Evidence suggests that supported self-management programs, with guidance from a health professional or another helper, are preferable to independent self-management for people with serious mental health conditions [2,12]. Self-management programs in which support is provided by a peer support worker (PSW) who has experienced a mental illness themselves have demonstrated promising evidence of effectiveness [12-15].

An example of a program that showed effectiveness was the Crisis Resolution Team Optimisation and Relapse Prevention (CORE) peer-supported self-management program [2,12,16], which was implemented in a large-scale (N=441), UK-based, randomized controlled superiority trial. This trial found that peer support using a structured workbook aimed at helping consumers develop self-management strategies to support their recovery beyond the immediate crisis led to a significantly lower rate of readmission to acute care within 1 year compared with self-management alone (29% for peer-facilitated self-management vs 38% for self-management control) [12]. Replication of the CORE study findings in routine settings across the UK National Health Service services and internationally has the potential to substantially reduce the burden on the acute care system. However, it is recognized that different countries may have specific nuances based on their cultures and systems [17,18]. As such, optimizing the delivery of the intervention in these different settings would benefit from consumer-driven consultation and co-production to translate the consumers’ needs into intervention components and refine the intervention accordingly before any research and implementation occur.

The peer-supported self-management intervention used in the CORE trial was systematically adapted in a stepwise co-production and piloting process in partnership with PSWs [2] from an existing recovery resource developed by Julie Repper, Miles Rinaldi, and their colleagues in South West London [19]. This existing paper-based resource “Taking Back Control” was itself co-produced with people with lived experience expertise; has a strong recovery focus; and incorporates self-management tools, including relapse.
prevention planning, goal setting, wellness planning, and a component for recovery from a mental health crisis [2].

Since the original resource was developed in 2007, the proliferation and advancement of technology, particularly smartphones, have provided an opportunity for digital interventions to become more accessible and acceptable to people experiencing serious mental ill-health [20]. People experiencing psychosis have adopted digital technology comparably to the general population [21], and mental health interventions delivered via smartphones are acceptable and feasible for people with psychosis and have the potential to support recovery [22-24]. However, access to technology has been found to vary among people experiencing schizophrenia by age, and a proportion do not use technology to manage their condition [25]. Therefore, it is imperative to allow people to choose between paper-based and digital self-management resource mediums depending on their preferences and circumstances. Further, as the delivery of self-management interventions does not have a typical medium (digital, paper-based, verbal, or flexible delivery), delivery mode (face to face, digital, telephone, or hybrid), or support (self-directed, clinician, peer supported, or blended) [4], offering a suit of flexibly delivered personalized programs may help maximize the reach, acceptability, and appropriateness of self-management interventions when delivered in real-world services.

This study reports the translation of the CORE study self-management resources from the United Kingdom context to the Australian context, including the adaptation of paper-based resources used in the original CORE trial and the development of a digitally based resource guided by the original paper-based tools to provide flexible delivery options for consumers. Following the strong tradition of co-production [2,19], participatory design methodologies were selected for this study to adapt both the paper-based resources and develop the digital self-management resource. Integrating user feedback into the design of digital mental health interventions is the gold standard [26-30] and improves engagement with digital tools for serious mental ill-health [31]. Best practice recommendations emphasize that researchers should publish descriptions of development work, including a description of how design features are influenced by user feedback [32].

Goals of This Study

The overarching aims of this research study were to (1) translate the CORE paper-based and peer-facilitated self-management resources to the Australian context for successful implementation and (2) co-design and user test a digital prototype of the CORE peer-facilitated self-management resources.

Methods

Study Design

This study adopted a participatory design, user testing, and rapid prototyping methodologies. Two frameworks (Figure 1) provided an evidence-based structure for the co-design process: (1) the 2021 Medical Research Council framework for developing complex interventions [18], including developing, testing the feasibility, evaluating, and implementing the intervention (note that this study focuses on the development and feasibility stages of this framework) and (2) the co-production framework [33], which seeks consumer leadership and input from the outset, including co-planning, co-design, co-delivery, and co-evaluation. In this study, co-planning involved collaborators and researchers with lived experience informing the protocol and study design, ethics application, and recruitment; co-design involved the recruited people with lived experience defining, conceptualizing, evaluating, and designing the prototype; co-delivery involved lived experience researchers facilitating the aforementioned co-design research; and co-evaluation involved the knowledge transition team having representation from people and researchers with lived experience for analyzing and writing up the results.
Participants
The inclusion criteria for study participation required participants to be aged ≥18 years; self-identify as having a lived experience of using a mental health service, experiencing a mental health condition, or working in a mental health peer support role; have English proficiency; and be able to complete the informed consent processes.

Co-Design and User Testing Stage 1
The first round of co-design and user testing of the original co-designed UK version of the CORE paper-based self-management resource (which had five sections: (1) moving on again after a crisis, which focused on resuming routines and community support; (2) keeping well, which focused on activity scheduling and health-promoting behaviors; (3) managing ups and downs, which focused on relapse prevention; (4) goals and dreams, which provided goal-planning tools; and (5) making a personal recovery plan, which provided psychoeducation in a recovery-focused manner [2,19]) was conducted with 6 participants through audio-recorded, one-on-one, and 90-minute user experience sessions, which were held face to face or via a digital video chat depending on participant needs. Interviews were conducted by either a lived experience facilitator or a mental health clinician. During the sessions, facilitators engaged the participants in 3 phases of participatory design processes, namely discovery, evaluation, and prototyping. In the discovery phase, the facilitators used open and prompted discussions to explore participants’ practices, goals, values, and needs in relation to the self-management of serious mental health conditions. In the evaluation phase, participants tested and evaluated the self-management resource, focusing on its strengths and weaknesses. In this phase, the think-aloud methodology [34] was used to gauge the usability and desirability of the design of the paper-based self-management workbook. After this, the facilitator focused on early prototyping with participants, discussing what a digital version might look like to inform the development of a potential digital prototype. Before the end of the first round of co-design and user testing, participants were asked about their basic demographics and their views on the usability of the CORE paper-based resource through an adapted version of the System Usability Scale (SUS) [35].

Knowledge Translation Stage 1
A knowledge translation team (which included representative stakeholders with lived experience, peer support, clinical, research, and technology backgrounds) was formed and regularly met with the lead researcher (first author) via digital meeting platforms. The knowledge translation team updated the UK version of the CORE paper-based self-management resource to an Australian version based on the stage 1 feedback. Further, the knowledge translation team engaged in an interactive process of synthesizing, exchanging, and applying knowledge [36]. The ultimate goal was to translate user testing feedback into practice.
organizational management, technology development, and policy reform [36]. During stage 2, the tool was renamed My Personal Recovery Plan (MyPREP, which will be used henceforth) by the knowledge translation team.

User Testing Stage 2

A subsequent round of user testing of the high-fidelity digital prototype took place with 4 new participants using the think-aloud methodology to gauge the usability and desirability of the design of the Australian paper-based self-management workbook. After the tool was adapted based on the identified problems and suggestions, a group of 6 PSWs who would be piloting the tool in their community mental health service were provided with access to the digital and paper-based tools for a final round of qualitative feedback.

Data Analysis

The audio-recorded sessions were transcribed and anonymized. The qualitative data were subsequently interpreted using a previously established knowledge translation process for participatory design studies [37]. Specifically, the knowledge translation team developed a coding framework outlining all key concepts. Data were coded in the NVivo software (version 12; QSR International) using this framework. Data interpretation followed established thematic techniques [38], which involved an iterative and reflexive process of reading, coding, exploring the pattern and content of coded data; reflection; and discussion. Similarities and differences in opinions were examined, and differences were dealt with through discussion to reach a consensus. The knowledge translation team also identified themes and key learnings to inform the customization and configuration of the paper-based MyPREP program and the digital high-fidelity prototype. Acceptability scores were calculated using the standard SUS process [35]. Frequency and descriptive analyses of the quantitative data generated in user acceptance–testing sessions were conducted in SPSS (IBM Corp).

Ethical Considerations

Ethical approval for this study was obtained from the University of Sydney’s Human Research Ethics Committee (HREC reference number 2019/571). Informed written consent was obtained in advance from all the participants in the qualitative sessions of this study. We assert that all the procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Results

Participant Demographics

All 10 participants in the first 2 rounds of user testing identified as having lived experiences of a mental health condition and mental health service use, with 3 (30%) of the 10 participants identifying as male, 3 (30%) residing in regional areas of Australia, and 6 (60%) having a job as a PSW. The final prototype was then appraised by 6 (male: n=3, 50%; female: n=3, 50%) additional PSWs, all working in urban and suburban areas of Sydney, New South Wales, Australia.

Main Themes

The key themes identified throughout the co-design and user testing sessions were related to (1) the need for self-management tools to be flexible and well-integrated into the mental health services used by participants, (2) the importance of language and how language preferences vary among individuals, (3) the need for self-management interventions to have the option of being supported when delivered in services, and (4) the potential of digitization to allow for a greater customization of self-management tools and the development of features based on individuals’ unique preferences and needs. Further, summaries of the full recommendations of adaptations to the MyPREP paper-based and digital prototypes are presented in Multimedia Appendix 1.

The Need for Integrated and Flexible Self-Management Tools

Most participants confirmed that there was a lack of routine provision of self-management tools supporting recovery in community mental services, and these tools should be offered as a standard as early as possible:

Honesty, I think it might have been really beneficial to start [using self-management tools] while I was inside [the hospital]. You know that that way I’m not sort of just being discharged into sudden loss of support. I’m being discharged with a plan I have got a set of actions and a structure to go back to. [P3, user testing round 1]

These participants expressed that there was a strong need for the flexible delivery of self-management tools in mental health services. That is, services should not provide just one type of medium to deliver interventions, as the needs of individuals are not uniform:

Individuals may experience barriers [to using self-management tools] due to their lack of use in technology. Or finding it challenging to adapt to technology this resulting in becoming fearful of using technology. [P4, user testing round 2]

From what I’ve observed, especially if it’s like 60 and above, they love paper versions, even though like, I mean, some are tech savvy. Yeah, they do feel that comfort in what they know as well. But I guess it’s just about educating them a little bit more about the technology and allowing them to adapt as well. So having both options is always, always better than not having an option. [P4, user testing round 2]

Self-management tools also need to be well integrated with other plans and documentation offered in services, and digital platforms could enable this integration:

We get given a lot of stuff when we’re coming out of the crisis, particularly for going to hospital discharge. You’ve got a discharge plan, you’ve got all your appointments, you got your medication sheets. You may even have a safety plan depending on which...
program you’ve gone through. So under a suicide attempt or you, we thought there might be a safety plan that’s developed as well. So I think we just need to consider how this works in conjunction with all the other hits and pieces that might be provided. [P4, user testing round 1]

Language

A major theme identified in the co-design and user testing sessions was language. Some words need to be changed to reflect the common language in Australia compared with the United Kingdom. People preferred less formal and less clinical language. Participants also liked that the language used generally did not make assumptions about the person’s situation:

I like that it’s like everyday language ‘the things I need to sort out.’ Yeah, I like that. [P1, user testing round 1]

[When speaking about the ‘people and places I can turn to’ page] I really like it. And what I like about it is it’s not assuming that those people are family members. You you’re not making people feel worse by, you know, having to put down friends or colleagues or neighbors or whatever; because family is not in your life. So ‘what support would I like?’ Yeah, that’s great. Nice basic phrasing there. [P4, user testing round 1]

The participants also discussed the importance of not having too much text on each page, as it can become overwhelming and distracting for users, especially as impaired cognition and concentration can be symptoms of mental health conditions. Decreasing the number of words on the screen should be a priority for increasing usability. This could be done by incorporating images, especially infographics, and additional features, such as read-out-loud audio and avatars:

Language, is really, really important. The less that’s on a screen or on a paper or on a document is better. Yeah, because you can’t take it in. It’s too much noise. It’s too busy. And if my concentration is already impaired, it’s not going to help me stay engaged. I’m just, I just won’t get engaged. [P2, user testing round 2]

One of the participants said the following about the name of the tool:

I don’t like the words. Actually, at first I don’t like the word recovery. I don’t like the personal, and I certainly don’t like plan... [the workbook needs] something more casual...it’s the balance between not being too positive that can be invalidating too. [P1, user testing round 1]

Further, participants emphasized that the language used in self-management tools needs to be empowering, promote self-agency, and be sensitive to the mental health situations of people, as mental health can fluctuate. Although it was generally agreed the language was used in MyPREP met this need, there were few words identified by some participants as having the potential to be disliked by some users of the tool, such as “recovery,” “triggers,” “moving on after a crisis,” “plans,” and “goals”:

And now going back to the word recovery, if we’re going to be changing it in certain instances, personal recovery. And it’s interesting because now we are having this challenge around the word recovery. Yeah, some of us are now saying, hang on, many of us are not recovering from anything. If we want to look after health and wellbeing proactively, yeah. ...where are you on your journey? Exploring your health journey? Or your life journey, whatever it is that you want to use. [P2, user testing round 2]

Views on language were mixed, with some reporting that they liked the language that others did not: “oh I like the words ‘moving on’” (P1, user testing round 1). Moreover, some participants reported that some terms should remain in the tool. For example, some felt that the term recovery should remain, as it promoted autonomy:

You know, I think recovery is like very self-empowering. And that puts a lot of independence on people. ...And so just thinking about perhaps placing the autonomy a little bit more on the individual as well. So that is self-directed, that is more self-directed yeah. [P1, user testing round 1]

Working on the tool with support (particularly from a peer worker) would enable these conversations about language to occur.

Supported Facilitation

Self-management tools need to be customized and personalized to the individual, and many participants felt that this needs to be carried out through conversations with facilitators and supporters, such as PSWs:

Yeah, I think, um, this [using MyPREP] would be excellent to do again, through a conversation. Yes like where, if the person accessing the service would be talking with a peer worker and then, so, yeah, ‘I believe this is what recovery is,’ ‘what you think recovery is about,’ ‘what sort of thing.’ [P2, user testing round 1]

These conversations “made it [the paper-based self-management workbook] so much more human” (P1, user testing round 1), especially because facilitators such as peer workers can tailor conversations based on individual needs while using the tool as a guide:

Yeah, I would tailor it. So it’s. Yeah, that is more for a conversation. Yeah, where the peer worker could be like, ‘So what are some signs that I could look for in you that might make me think you’re not going so well?’ [P2, user testing round 1]

Further, the accompanying psychoeducation component that helped contextualize the workbook was located at the end of the booklet in the original UK resource. Participants highlighted that this needed to be integrated with the MyPREP module activities (note that this was actioned for both the digital and paper-based Australian versions of MyPREP). Moreover, it was
emphasized that the content of this psychoeducation component could be discussed with someone such as a peer worker, clinician, or supportive other:

I think that psychoeducation needs to be at the front [of the MyPREP tool]. Yeah, the way I would work with it is in that context, I’d be looking through it and go, oh okay, I’ll go to the next page and then I’ve probably filled in a lot of that stuff before I’ve got to the end. It is where all the guidance around this. ...It might not be read, so just all of this content could be put it in within the chapters. Yeah, that or a statement at the front of the elements that you feel from the information on how to help you fill it in was confined to the back of the book. A plain guide just say, Look, don’t just launch into it. But there’s some further guidance and support there on the back of the book that you can work with a friend or someone through it all with your worker. You know, because that is a lot of text there, and we are assuming the people have the ability to read and comprehend all that information. [P4, user testing round 1]

Additional Desirable Features and Functions When Digitized

Additional features and functions beyond those of the paper-based MyPREP workbook were highlighted by participants during the co-design process. Exemplar additional features and functions are displayed in Figure 2 and described in detail below (refer to Multimedia Appendix 2 for enhanced visibility of features on each page).

Figure 2. Exemplar additional features and functions. MyPREP: My Personal Recovery Plan.

On the basis of the feedback gathered during the user testing sessions, customization by the users themselves was enabled in the digital version so that they could adapt the content to their preferences. End users could make changes by configuring the color pallet and the voice and image of their avatar and were able to alter and flexibly select the MyPREP modules. The ability of end users to customize MyPREP was seen as very important, as it enabled a feeling of ownership over the tool. For example, the color pallet could be changed to reflect not only the preference but also the mood of the participant, which was seen as important in mental health:

I think it’s like there’s dark colors and light colors pastel. Yeah, I like that. I think that’s really nice because it’s like I said, you’re going towards personal, like being personal. And this gives choice of being personal as well. It’s cool. [P4, user testing round 2]

When you are unwell, you only see fog, color can break through. [P1, user testing round 2]

The customization of the image and voice of the avatars in the program was seen as trauma informed (as it had the potential to decrease inadvertent traumatization if someone had a previous traumatic experience with a person of a particular sex) and gender inclusive (as voice and image could be changed) and helped tailor the avatar according to the individuals’ backgrounds, such as their age. This was seen as an important feature, and additional languages could be easily integrated into the MyPREP package in the future:

You just have to think of the LGBTQ plus community as well... I think that’s because if you do like a said male or female, it’s just like ‘why only two?’ Yeah, you know, but then if it’s like I said a plant [an alternative avatar on MyPREP], it’s very cute and it has a smiley face. And I really like that because I think especially in the new generation, if you think
about it, a lot of like animations are popping up as well. So, if you want to adapt to the younger generation, they might really enjoy this thoroughly. [P4, user testing round 2]

A new feature, which was not possible in the paper version but enabled through digitization, was a “need help now” button, which was suggested by participants and provides the contact details of crisis support and additional mental health support services. This feature was developed after participants in the user testing sessions recognized and acknowledged that creating a MyPREP plan could potentially be confronting, and the individual may want additional support from others to process the information or additional mental health crisis support:

If I’m not in a good space and I’m trying to read it, I want to know I’ve got a safety button. I can go to hey, look, I’m feeling like shit, OK, and I might be reading this and you know what? I actually want to talk to someone, or I want to write to someone or chat to someone. Yeah, and I get blocked. Yes. I think, you need to have a safety, safety spot, do you need help now? Something. The hope is you read it, but it’s also a risk thing that, yeah, I think you need this. [P2, user testing round 2]

Two other new features were “my emergency contacts” and “supporter circle.” The emergency contacts feature allows users to enter the contact details of people and services they can contact during a mental health crisis. In addition, the supporter circle function was designed to allow users to share their recovery plan entries with people in their support circle (such as carers, health professionals, and peer workers) so that they could view their MyPREP entries. One of the participants expressed that it was important that the user be prompted to input contacts for both features when signing in for the first time to provide a safety net if the user became distressed in the process of completing the plan and suggested displaying these important contacts on the dashboard at all times to provide a sense of support to users:

OK, I would want them to have their emergency contacts sorted and their support circle sorted because they can’t move on after a crisis. If they don’t, if they’re unaware of their support circle, then we can’t help them to do that either if we don’t know. Especially and then managing up and ups and downs like they’re going to need their support circle and their emergency contacts. [P3, user testing round 2]

All participants acknowledged the usefulness of the supporter circle feature in providing access to the user’s professional support network, including psychologists and support workers, if the user wanted their MyPREP entries to be shared:

People that are supporting you that you want to know what’s going on? Mm-Hmm. Yeah. So, I mean, you can have them invited by, or you can share a copy, so I mean, that would be really useful for clinicians and obviously peer workers and mental health workers... [P3, user testing round 2]

It was emphasized that some may not use this feature because of privacy concerns, particularly with family or carers: “some people really don’t like family seeing this stuff” (P3, user testing round 2). Importantly, permission for viewing access could also be withdrawn by the end user at any stage:

The major function of the digital self-management tool is to assist users in systematically mapping and recording personalized self-management strategies in response to prompts to assist in their recovery from a mental health crisis. A feature that facilitated this was the avatar for read-out-loud functionality, which was viewed as especially important for those who held a preference for this or who had lower literacy skills. Further, each activity could be completed in a variety of ways based on the individual’s needs. Specifically, end users could engage in MyPREP activities by uploading voice recordings, images, emojis, and text. This was seen as a highly inclusive feature:

Oh, we got little icons. I would probably put a smiley face. Awesome, and then you can upload stuff. That’s cool, that’s like really personalized. Awesome! ...I think people get really anxious when they feel like they’re being recorded or they don’t like hearing their voice playing back, I know I don’t. So I would always go for the text. ...I think some people might like it. Some people might not want to type. [P3, user testing round 2]

These features were also seen as important motivating factors that promote better end user engagement with MyPREP:

I don’t know if people are going to struggle with doing these entries, because if they might think, OK, what is the purpose of it? But if it’s more fun and personal, it’s more fun. So, it’s like okay, I’m going to do this one. I might even put a photo of me in it and get that done. But if it’s always just bland, I don’t know. People might be like, well, especially, you know, you want that motivation. But I think this is really cool. [P4, user testing round 2]

The MyPREP module activity entries made by the end user were also editable, and there was a function in which entries could be saved so that the end user could keep a record of the changes:

This is really awesome. And it will keep the track as well. You can kind of say you can kind of go back, it can be like okay I was feeling not great on this day, what happened that day as well? ...Yes. So, you can kind of see that progress, especially for consumers who do enjoy that progress. [P4, user testing round 2]

The SUS

At the end of each user testing session, participants completed the SUS. The mean and range are listed in Table 1. The MyPREP paper version received a total SUS score of 71, indicating C+ or “good” usability. The digital version received a total SUS score of 85.63, indicating A or “excellent” usability.
indicated that the end users involved in this study rated the tool highly. The resulting tool was subsequently rated using the SUS, which is a widely used method to evaluate usability. The SUS scores for the paper-based and digital versions of MyPREP (My Personal Recovery Plan) are presented in Table 1.

### Table 1. System Usability Scale (SUS) usability scores for the paper-based and digital versions of My Personal Recovery Plan (MyPREP).

<table>
<thead>
<tr>
<th>SUS items</th>
<th>MyPREP paper-based version, mean (SD; range)</th>
<th>MyPREP digital version, mean (SD; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think that I would use the Personal Recovery Plan frequently.</td>
<td>2.4 (1.3; 0-3)</td>
<td>3.25 (0.5; 3-4)</td>
</tr>
<tr>
<td>2. I found the Personal Recovery Plan unnecessarily complex(^d).</td>
<td>2.4 (1.3; 0-3)</td>
<td>3.75 (0.5; 3-4)</td>
</tr>
<tr>
<td>3. I think the Personal Recovery Plan would be easy to use.</td>
<td>3.0 (2.1; 0-5)</td>
<td>3.5 (0.6; 3-4)</td>
</tr>
<tr>
<td>4. I think I would need the support of a support person (e.g., a peer support worker) to use the Personal Recovery Plan(^d).</td>
<td>2.2 (1.5; 0-4)</td>
<td>2.75 (1.0; 2-4)</td>
</tr>
<tr>
<td>5. I felt the various sections in the Personal Recovery Plan were well integrated.</td>
<td>3.6 (0.5; 3-4)</td>
<td>3.5 (0.6; 3-4)</td>
</tr>
<tr>
<td>6. I thought there was too much inconsistency in the Personal Recovery Plan(^d).</td>
<td>3.2 (1.3; 1-4)</td>
<td>3.5 (1.0; 2-4)</td>
</tr>
<tr>
<td>7. I would imagine that most people would be able to use the Personal Recovery Plan easily.</td>
<td>2.2 (1.3; 0-3)</td>
<td>3.0 (0.8; 2-4)</td>
</tr>
<tr>
<td>8. I found the Personal Recovery Plan very cumbersome(^d).</td>
<td>2.8 (1.1; 1-4)</td>
<td>3.5 (1.0; 2-4)</td>
</tr>
<tr>
<td>9. I would need to learn a lot of things before I could start using the Personal Recovery Plan(^d).</td>
<td>3.2 (1.3; 1-4)</td>
<td>3.25 (1.0; 2-4)</td>
</tr>
<tr>
<td>10. I would feel very confident using the Personal Recovery Plan myself.</td>
<td>3.6 (0.8; 2-4)</td>
<td>3.75 (0.5; 3-4)</td>
</tr>
</tbody>
</table>

\(^a\)The total SUS scores for the paper-based and digital versions were 71.5 and 84.4, respectively.

\(^b\)Of the 6 participants in the first round of user testing, 5 (83%) completed the SUS for the paper-based version.

\(^c\)Of the 4 participants in the second round of user testing, all 4 (100%) completed the SUS for the digital version.

\(^d\)Reverse scored item.

In a final meeting before endorsing MyPREP for piloting, the researchers met with 6 PSWs, and the tool was presented and discussed; these peer workers were then given access to the MyPREP intervention to gather their feedback. At this meeting and at follow-up, only positive additional feedback was relayed, and the peer workers emphasized that they were excited to use the tool in their service:

> I have reviewed the Workbook and honestly couldn’t find much wrong with it at all!! It’s great and I am very excited to put it into practice. My team had no feedback to give on the digital version other then it looks great and can’t wait to see it used with clients in practice!! [Peer worker 1]

**Discussion**

### Principal Findings

In this study, we present the user testing of an Australian version of the paper-based MyPREP and the co-design and user testing of a digital version of MyPREP. The key themes identified throughout the co-design and user testing sessions were related to (1) the need for self-management tools to be flexible and well-integrated into mental health services, (2) the importance of language and how language preferences vary among individuals, (3) the need for self-management interventions to have the option of being supported when delivered in services, and (4) the potential of digitization to allow for a greater customization of self-management tools and the development of features based on individuals’ unique preferences and needs. The resulting tool was subsequently rated using the SUS, which indicated that the end users involved in this study rated the paper-based tool as “good” and the digital version as “excellent.”

Taken together, these results suggest that the digital prototype has valuable potential for use in mental health services and support the progression to piloting MyPREP in mental health services to inform a future large-scale randomized controlled trial (RCT). Early studies providing detailed descriptions of intervention development, such as this study, are now called for as best practice [32]. Specifically, it is important to include details such as important decision-making steps across all stages of development, explanations of how the information gathered from intended end users was carried out, and how intended end user input influenced how design features were incorporated into the design.

### The Need for Implementation Research

The participants in our research also highlighted the need for the integration and coordination of self-management interventions and tools within services, as there can be a multitude of staff, documents, and plans (such as suicide prevention plans, discharge plans, wellness plans, and medication adherence plans) involved at various points throughout a consumer’s or a service user’s journey in a mental health service. Further, although the digital version of MyPREP was seen as being able to be used independently by end users at their home or within health settings, participants emphasized the need for such tools to be delivered with support from others, particularly PSWs. Although this may be explained by the fact that many of the participants were peer workers or consumer advocates themselves, this finding mirrors the current movement within mental health services toward the holistic inclusion of lived experience expertise [39]. Further, integrated staffing
models may also be associated with better recovery, including better social functioning [40]. Taken together, implementation-focused research is needed to determine how MyPREP is actually delivered and how it can be optimized once it is introduced in real-world settings.

A recent systematic review of supported self-management interventions for people with serious mental health conditions found that there is a current lack of studies focused on implementation and that even fewer studies are based on implementation science theories [8], which are summarized in the taxonomy of implementation outcomes [41]. Although RCTs remain the gold standard approach to informing clinical decision-making and drawing causal inferences [42,43], there remain unacceptable research-to-practice gaps and a disconnect in generalizability and performance between an extremely controlled clinical trial environment and highly complex real-world mental health environments [44]. Implementation science, which is the formal study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice, can bridge the clinical evidence-to-practice gaps by improving the quality and effectiveness of health services [44]. The hybrid implementation-effectiveness design is an innovative and rapid solution to provide high-quality evidence on effectiveness and implementation simultaneously [44,45], making it the ideal next stage of the research pipeline for the MyPREP project to determine effectiveness and implementation. This step is particularly important in our case, as the basic content of MyPREP has already been tested successfully in the CORE trial in the United Kingdom [16].

Accessibility of MyPREP

The digitization of MyPREP may have allowed it to become more accessible to end users when compared with the paper-based version, and this may partly explain the increase in usability scores, as measured using the SUS. For example, the use of the read-out-loud feature via avatars was viewed very positively by participants, as was the provision of options for writing, recording audio, and uploading images when completing MyPREP module activities. A recent systematic review suggested that developers of self-management interventions should adapt interventions to ensure greater inclusivity for participants with less formal education, as it was found that educational level was associated with engagement [8]. This is particularly critical for individuals with severe mental health conditions, given the increased rate of lower education in this cohort [46]. Indeed, the participants in our study repeatedly emphasized that plain and informal language was preferred. Further, we used Design for Dignity Principles, Web Content Accessibility Guidelines, and International Organization for Standardization standards for process improvement, safety, and quality (eg, 9241-11 ISO standard) to promote digital accessibility and ensure that the features are usable and acceptable for people with accessibility issues.

Customization of MyPREP

Customization embedded within MyPREP, such as the ability to change the color palette and choose an avatar, was a clear theme throughout the co-design process and was rated very positively in the user testing sessions of the digital tool. Customization was important, as MyPREP is a self-management tool; the participants emphasized that these self-directed changes to MyPREP fostered a sense of autonomy, control, choice, and ownership, and this should be embedded as a standard in all features. Reviews in this area recommend that the content of self-management interventions should be tailored to the service users and have the flexibility to be personalized and customized, especially as interventions were found to not always fit end users’ needs [8].

In the future, there are plans for MyPREP to be customized further. A major example is enabling control over language. Specifically, feedback concerning the language used for the names of modules (eg, “my recovery plan,” “moving on after a crisis,” and “goals and dreams”) was liked by some participants in the user testing sessions, but not by others. This was despite MyPREP being co-produced and going through iterative co-design cycles with people with lived experience of serious mental ill-health [2,19]. The solution to this lies in customization, through which end users can adapt the names of the modules to suit their own set of beliefs and even remove modules that are not relevant to them from their MyPREP dashboard.

Limitations

This is a preliminary iterative co-design and user testing study and should be viewed as such. Similar to most user testing studies, our sample of service users was small (eg, the study by van der Krieke et al [47]). Further, we used not only advertisements across networks and services but also snowball sampling, which is a type of convenience sampling. A major disadvantage of such convenience sampling is that it risks a nonrepresentative study sample. In our case, the study sample was quite diverse in terms of age and sex. However, a large proportion of participants were consumer advocates and peer workers with considerable mental health knowledge and expertise. Further, those recruited for this study might have had a particular interest in working with digital health tools, which may have introduced avidity bias and may explain the very high SUS acceptability scores for the digital version of MyPREP. The next stage is piloting MyPREP in services to increase the representativeness of our sample and make iterative adaptations to MyPREP based on user feedback. However, overall, as the original MyPREP was trialed in the United Kingdom with at least 275 crisis care service users in a large-scale RCT (with 441 service users enrolled in the trial), this may suggest that the representativeness of our sample at this point does not pose a concern for the next stage of piloting of the Australian version in real-world mental health services.

Another limitation that is common in user testing (eg, the study by van der Krieke et al [47]) is that the presence of the facilitators over a digital meeting platform during the testing sessions may have affected the views of the participants, as they might have felt reluctant to be critical. We do not expect this to be a major limitation; however, throughout user testing, the facilitators continuously emphasized that this was an opportunity to improve the paper-based and digital prototypes and
encouraged discussion around problems and areas for improvement.

Conclusions
The co-production of the MyPREP self-management intervention and associated research in Australia are currently in their early stages. However, this co-design and user testing phase is a crucial step in adapting MyPREP to the Australian mental health setting and digital context. However, the current findings may remain relevant to implementation in any setting. Overall, the co-production process is vital, as service-wide implementations that fail to consider end user needs and organizational structures often encounter problems. Indeed, Killikelly and colleagues [31] suggest that both co-design and support from mental health staff or researchers when using the tool are 2 features that are associated with successful implementation and improve engagement with digital tools for people experiencing serious mental health conditions. To avoid implementation issues, it is essential to involve consumers and peer workers, who may support the delivery of MyPREP from the outset. In this study, feedback from these end users highlighted a strong desire for the personalized delivery of self-management interventions that offer choices and options, considering individual end users’ different needs and circumstances. MyPREP has worked toward addressing this need by offering digital and paper-based mediums, providing options for how delivery is supported (ie individuals can choose their supporters), increasing accessibility (eg, avatars and voice-recording options), and allowing customization (eg, customization of the color palette and choice of avatars) based on users’ preferences and needs. To strengthen MyPREP's implementation in real-world Australian mental health service settings, implementation-effectiveness piloting and robust trialing are required to test and refine the tool.

Lived Experience Commentary
In our view, co-production is an umbrella term for co-family, co-creation, co-planning, co-implementation, and co-evaluation. This is a concept and philosophy in which collaboration with lived experiences is paramount. It is a way to cocreate new interventions, improve systems, and solve problems and is now being adopted in many public health policy arenas, including research.

This has moved far beyond “consumer participation,” which was enshrined in 1992 in the Australian mental health policy (1992 National Mental Health Policy endorsed by Australian Health Ministers). Back then, this was ad hoc and tokenistic. Now, as “co” is becoming more widespread, there are more genuine attempts to learn to integrate co-production as a way of doing research across the board.

In research, with co-production, we are witnessing a move from consumers and carers (or those with lived and living experience) voices from being “subjects of” to being equal collaborators, who work with researchers to influence change that benefits the service and system user. This research project has done just that. By having lived experience experts lead, participate in, and contribute to the design and facilitation of the research, we actively used lived experience knowledge and expertise in recovery planning through various stages. The recruitment of participants, data collection, workshops, one-on-one interview formats, and questions were developed and led by lived experience.

In this project, a trusting respectful alliance between research and lived experience evolved organically and naturally, providing a solid foundation to “do” this project with passion and enthusiasm, creating a safe and supportive atmosphere emulated in dealings with others. The interviewees felt more than comfortable to offer their time generously, sharing their insights into and thoughts on what would work well and what should be improved to increase engagement with and the use of this recovery-based intervention.

During different phases of this project, reciprocal positives emerged for the research team. Empowering and enabling each member to consider different perspectives and interpretations. Embracing new ways of thinking created a richer understanding of issues not considered previously. This was articulated in the ways in which the research was focused and conducted. On reflection, lessons on how this research approach will inform similar ongoing collaborations should be considered, where research projects embrace and welcome different skill sets, experiences, and knowledge.

Acknowledgments
The copyright for the Personal Recovery Plan workbook is held by South West London and St George’s National Health Service (NHS) Foundation Trust. Written permission to adapt and use the plan and distribute the adapted plan was obtained from the copyright holders by the first author at the University of Sydney.

The authors are grateful for the time and help of peer support workers (PSWs) and people with lived experience of mental health service use and mental health conditions who participated in the study. They are grateful to One Door Mental Health and Flourish Australia for supporting the recruitment for this study. They would also like to thank the Australian PSWs at One Door Mental Health, who contributed to the intervention development in Australia and will deliver the upcoming pilot. They also remain grateful to all the original PSWs, Crisis Resolution Team Optimisation and Relapse Prevention (CORE) service user and carer working groups, and the University College London (UCL) research team in the United Kingdom who originated this work (note that further information about the original study intervention, the PSWs’ training program, and the intervention’s operating procedures is available from the UCL CORE study website). The CORE study was led by professor Sonia Johnson from the Division of Psychiatry, UCL. The authors would also like to thank Aaron Schokman, Viji Venkataraman, Lei Gu, and Yang Wu for their contributions to this project.

https://formative.jmir.org/2024/1/e49110
This study was funded by an Early and Mid-Career Researcher Grant from the Brain and Mind Centre and the Moriya Elizabeth Vine Fund for Research into Schizophrenia. The preparation of this manuscript was supported partially by the Australian Government through the Australian Research Council’s Centre of Excellence for Children and Families Over the Life Course (project ID CE200100025). The original CORE study was funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research program (reference number RP-PG-0109-10078). The views expressed in this paper are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health.

Data Availability
The qualitative data sets generated and analyzed during this study are not publicly available, as they are lengthy interview transcripts from which quotations used in this publication were selected, to ensure that individual services or participants cannot be identified. The data sets are available from the corresponding author upon reasonable request.

Authors' Contributions
AM and NG conceived this phase of the research and its design, with support from BL-E and SJ, based on the original UK Crisis Resolution Team Optimisation and Relapse Prevention (CORE) study. AM, IOAM and DP collected the data. AM, DJ, DP, and IOAM synthesized and analyzed the data. AM drafted the manuscript with support from IOAM, TC, EB, and UA. DP, JC, and AM oversaw the development of the digital resource for user testing. All the authors read and approved the final manuscript.

Conflicts of Interest
None declared.

References


https://formative.jmir.org/2024/1/e49110 JMIR Form Res 2024 | vol. 8 | e49110 | p.1127 (page number not for citation purposes)


Abbreviations

**CORE**: Crisis Resolution Team Optimisation and Relapse Prevention

**HREC**: Human Research Ethics Committee

**MyPREP**: My Personal Recovery Plan

**NICE**: National Institute for Health and Care Excellence

**PSW**: Peer Support Worker

**RCT**: randomized controlled trial

**SUS**: System Usability Scale
© Alyssa Milton, Ingrid Ozols A M, Tayla Cassidy, Dana Jordan, Ellie Brown, Urska Arnautovska, Jim Cook, Darren Phung, Brynmor Lloyd-Evans, Sonia Johnson, Ian Hickie, Nick Glozier. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Help-Seeking, Support, and Engagement in Gestational Diabetes Mellitus Online Communities on Facebook: Content Analysis

Sheila Pham1, MPH, MA; Kate Churruca1, PhD; Louise A Ellis1, PhD; Jeffrey Braithwaite1, PhD
Australian Institute of Health Innovation, Macquarie University, North Ryde, Sydney, Australia

Corresponding Author:
Sheila Pham, MPH, MA
Australian Institute of Health Innovation
Macquarie University
75 Talavera Road
North Ryde, Sydney, 2113
Australia
Phone: 61 423078104
Email: sheila.pham@mq.edu.au

Abstract

Background: The prevalence of gestational diabetes mellitus (GDM) has drastically risen in recent years. For some, self-management includes the use of GDM online communities on Facebook. Such communities can fill gaps in information and support that participants are not able to access elsewhere to address unmet needs. Given the popularity of sharing information about pregnancy on Facebook and the documented benefits of diabetes online communities, the same may be true of GDM online communities.

Objective: This study aimed to categorize and quantify what is being discussed in GDM Facebook groups, including informational and emotional help-seeking behavior, and how this support and engagement may be demonstrated by peers through comments and reactions.

Methods: We sourced the data from the 2 largest Facebook groups focused on GDM in Australia. A summative content analysis was conducted on original posts across the 2 groups and coded for topics as well as help-seeking types. The coding scheme was based on the previous work of Liang and Scammon. Visible indicators of engagement, including the number of comments and “reactions,” were tabled and manually evaluated.

Results: There were 388 original posts, and the analysis produced 6 topics: GDM self-management (199/388, 51.3%), GDM clinical management (120/388, 30.9%), preparing for birth (40/388, 10.3%), mental distress (35/388, 9%), birth announcement (29/388, 7.5%), and GDM journey reflections (21/388, 5.4%). Secondary coding of help-seeking type revealed more than half of the posts were informational help-seeking (224/388, 57.7%), while a small proportion were both informational and emotional help-seeking (44/388, 11.3%), and some (12/388, 3.1%) were emotional help-seeking only. Self-disclosure was identified as a fourth category, comprising almost a quarter of all posts (90/388, 23.2%). A total of 6022 comments were posted in response to the original posts, and there were 4452 reactions across all posts. Emotional help-seeking attracted the most comments per thread (mean 21.5, SD 19.8), followed by informational and emotional help-seeking (mean 20.2, SD 14.7), informational help-seeking (mean 15.6, SD 14.6), and self-disclosure (mean 14.3, SD 21.8). Across all help-seeking categories, few reactions occurred compared to comments; in contrast, self-disclosure attracted a large number of reactions (mean 9.4, SD 45.3).

Conclusions: This is one of the first studies to examine peer support in a GDM online community on Facebook. Our findings suggest that active participants’ needs around information and support in relation to GDM are being somewhat met by peer-led online communities. Given the practical limitations of formal health care, including the provision of ongoing social support, it is important to recognize how GDM online communities can complement formal health care and help address unmet needs.

(JMIR Form Res 2024;8:e49494) doi:10.2196/49494

KEYWORDS

clinical management; communication; content analysis; engagement; Facebook; gestational diabetes; health communication; help-seeking behavior; mental distress; online communities; peer-support; self-disclosure; self-management; support

https://formative.jmir.org/2024/1/e49494

JMIR Form Res 2024 | vol. 8 | e49494 | p.1130
(page number not for citation purposes)
Introduction

Accessing health information on the web nowadays includes social media such as Facebook and, increasingly, its “group” function. Globally, the number of people engaging with Facebook groups equates to around 1.8 billion people per month [1]. The group function of Facebook is described as “a place to connect, learn, and share with people who have similar interests” [2]. Among the many “similar interests” people have are health concerns as well as life experiences such as pregnancy.

Research on pregnancy and the internet suggests that Facebook is used by some for supportive and informational purposes [3]. Given this, it is not surprising that many pregnancy Facebook groups now exist, as well as those focused on complications of pregnancy such as gestational diabetes mellitus (GDM). GDM is defined as any degree of hyperglycemia recognized for the first time during pregnancy [4]. As a condition, it affects a significant and growing proportion of pregnant women around the world each year [5]. Although GDM prevalence has drastically risen, there has been limited examination of the attendant growth of GDM online communities, including Facebook groups [6].

People may join online health communities because their family and community support networks do not include relatable others undergoing similar experiences [7], and thus they do not receive the benefit of “peer-to-peer health care” [8]. Research on diabetes online communities has found they fill gaps in information and support that participants are not able to access elsewhere [9]. Online health communities can provide both informational and emotional support, which helps people actively cope with health-related problems [10]. A study about breast cancer for women suggested that patients specifically seek out discussion groups on the web due to “unmet needs” [11], while another study suggested online health communities are where a range of desires and needs can be met by peers [12].

A scoping review of 47 studies focused on the use of diabetes online communities found a variety of psychosocial benefits, and although reports of negative consequences were low, it was also noted that diabetes online communities may not be beneficial for all [13].

In online health communities, users often demonstrate support-seeking behavior through explicitly stated requests, with posting itself a signal that the person is a potential support provider to others [10]. It is also common for users to “share” or self-disclose as a coping mechanism [14]. On Facebook, in addition to initiating posts, active engagement occurs through comments and reactions such as “likes.” The meaning of these reactions is not necessarily explicit beyond face value but can be broadly interpreted as support [15], though arguably a comment is a stronger indicator of support given the greater time and effort required to produce it.

Given the popularity of sharing information about pregnancy on Facebook and the documented benefits of diabetes online communities, the same may also be true of GDM online communities. Furthermore, health information-seeking online can also improve the patient-physician relationship if the patient discusses the information with the physician and they have a positive prior relationship [16]. This may be worthwhile considering the context of GDM, which generally requires additional health care compared to pregnancies without complications.

The aims of this study about Facebook posts and interactions within GDM Facebook groups were to examine: (1) the issues being discussed, (2) evidence for informational and emotional help-seeking behavior, and (3) how this support and engagement is demonstrated through comments and reactions.

Methods

Study Design and Data Collection

This study sourced data from 2 peer-led closed Facebook groups focused on GDM, founded, and run independently by private individuals. All original posts (ie, the first post in a thread) during a 1-week period were included, as well as replies published during the collection week. A limited period was chosen because the large volume of posts was considered sufficiently robust for the purposes of this study. These particular Facebook groups were chosen as they were the 2 largest groups focused on GDM in Australia; at the time of data collection, the combined membership of the 2 groups was over 6500 members. For this study, a “snapshot” approach was taken, with the data set copied verbatim by the first author, then fully deidentified and recollated for analysis to protect the privacy of participants. The data were then analyzed using content analysis and descriptive statistics.

Analysis

First, summative content analysis [17] was used to identify key topic areas from all “original posts” in a thread (ie, the first posts). This inductive approach to analyzing qualitative data started with reading through the data and identifying and quantifying certain words and content to understand contextual use before applying latent content analysis. To this end, the first author independently read and reread each post and identified keywords (eg, blood sugar and insulin) as the basis of topics. Multiple topics were allowed within a single post (ie, categories were not mutually exclusive). All authors then compared and confirmed the identified categories and interpretations.

Second, the original post in every thread was coded in terms of help-seeking type (Table 1). Here, a deductive coding scheme was used, following Liang and Scammon [10]. Visible indicators of engagement, including the number of comments, “likes,” and “reactions,” were tabled and manually evaluated by the first author (SP). A total of 2 secondary authors (KC and LAE) verified a sample (n=10) of the first author’s coding to ensure consistency. The depth of analysis was further consolidated by the research team, which compared and discussed codes to provide additional perspectives.
Finally, as the data from the 2 Facebook groups were combined to enlarge the sample size and potentially increase heterogeneity, it was important to determine whether there were statistically significant differences in proportions between the help-seeking categories of the 2 groups. A Fisher exact test was deemed an appropriate test given the likelihood of small category sample sizes.

**Ethical Considerations**

Research based on Facebook posts raises important ethical questions, given the implications for privacy. Before the commencement of data collection, approval was sought and gained from Macquarie University’s Human Research Ethics Committee (5201827734364). When the first author (SP) requested permission from the administrators of both groups to join in order to conduct research, she disclosed her positionality as someone who had experienced GDM. As stipulated by the terms of the ethics approval, no identifying data would be published, including verbatim quotes.

**Results**

**Topic Areas**

A total of 388 original posts were extracted across the 2 groups, with 63 posts from one group and 325 posts from the other. From the content analysis, 6 topic areas were identified (Table 2). These were not mutually exclusive, as some longer posts were coded for more than 1 topic.

A number of residual topics were excluded from the main analysis given the relatively small number of posts: “other pregnancy experiences” (n=18), “humor and memes” (n=9), “postpartum concerns” (n=7), “food and diet” (n=11), and “group management” (n=3).

**Help-Seeking and Engagement**

Secondary coding of help-seeking type is captured in Table 3. The process identified mutually exclusive categories, where more than half of the posts were classifiable as informational help-seeking (224/388, 57.7%). A small proportion were classifiable as both informational and emotional help-seeking (44/388, 11.3%), while a minority (12/388, 3.1%) were emotional help-seeking only.
Through the process of secondary coding we identified a distinct fourth category: self-disclosure. The intention behind such posts was not overt in terms of help-seeking (eg. “Just wanted to tell you all I gave birth to a healthy baby boy last week”). Almost a quarter of posts (90/388, 23.3%) were classifiable as self-disclosure.

A small number of posts (n=18) did not fit into any of the above 4 categories, such as posts sharing a recipe without comment or other practical matters such as offering to pass on medical supplies.

Visible indicators of engagement, namely the number of comments and reactions (including “likes”), were also tabled across all threads. A total of 6022 comments were posted in response to the original posts. The length of threads ranged from 1 to 179, with the median number of comments being 11. There were 4452 reactions across all posts. Emotional help-seeking posts were less prevalent but attracted the most comments per thread (mean 21.5, SD 19.8), followed by informational and emotional help-seeking (mean 20.2, SD 14.7), informational help-seeking (mean 15.6, SD 14.6), and self-disclosure (mean 14.3, SD 21.8).

Overall, across the 3 help-seeking categories, relatively few reactions occurred compared to comments, regardless of whether it was informational (mean 2.4, SD 20.1), informational and emotional help-seeking (mean 1.4, SD 3.9), or emotional help-seeking (mean 2.7, SD 4.4). In comparison, self-disclosure attracted a very large number of reactions (mean 39.4, SD 45.3).

**Significant Differences Between the Groups**

A Fisher exact test was applied to determine if there were any statistically significant differences in the proportions of help-seeking categories between the 2 Facebook groups. There were no significant differences found between the groups except for “emotional help-seeking;” with 6 posts identified for both groups, which represented a statistically significant (P=.006) difference in the proportions of 9.5% (group 1) and 1.8% (group 2).

Upon closer examination of how emotional help-seeking posts were responded to in each group, there were other notable differences that further qualified this significant difference. In the smaller group (group 1), emotional help-seeking posts attracted far more comments in response (mean 26.8, SD 27) compared to the larger group (group 2), which had fewer comments in response to emotional help-seeking posts (mean 16.2, SD 7.9). Conversely, there were fewer reactions in group 1 (mean 1.8, SD 2.6) compared to group 2 (mean 3.5, SD 5.8).

### Discussion

**Overview**

GDM self-management was the prevalent topic in over half the posts (199/388, 51.3%), which likely reflects a key motivation for both joining a GDM online community as well as an important reason for sustaining membership and engaging. The second most prevalent topic, GDM clinical management (120/388, 30.9%), is suggestive of the inadequacy of care provided in formal health care settings, including information provision, hence the need for additional discussion on the web with peers. This accords with how online health communities have been described as “communities of practice” due to the way learning occurs through a combination of experiential knowledge and other expert sources [18]. The topic “preparing for birth” alludes to a desire for information from peers (and expert sources), whereas “GDM journey reflections” points to expressly stated individual learning coupled with a desire to share and pass on knowledge to peers.

Informational was by far the most popular type of help-seeking, and this categorization largely overlapped with the most popular topics, demonstrating how critical information is in a peer-support context outside of formal health care. There were fewer posts where emotional help-seeking was the sole focus or in combination with informational help-seeking, but it is not surprising that these attracted the most comments per thread as empathic peers made a concerted effort to engage and offer reassuring words and engagement.

The statistically significant difference between the 2 Facebook groups in terms of the proportion of emotional help seekers warrants discussion. A possible explanation is the difference in size of the Facebook groups. Smaller-sized groups, in general, are friendlier and promote more contributions from members, with greater opportunities to speak [19], and emotional help-seeking posting also encourages supportive peers to show reciprocity by being more supportive. Examining the data from the 2 groups, there is a clear difference in terms of the volume of comments. This suggests greater intimacy and engagement in the smaller group, with comments being a better indicator of support than reactions, which are more impersonal and require less time and effort.

When we look at the posts categorized under “self-disclosure,” there are fewer comments but a much larger number of reactions. In such cases, engaging seems to be primarily enacted through a “reaction,” as peers do not necessarily see a need to comment. The general popularity of self-disclosure in GDM online...
communities is evident from this data and supports a more general observation that self-disclosing in an unprompted manner is intrinsic to the parlance of social media; that this is true of GDM online communities as many others. It has been found that self-disclosure in itself can be valuable, and with supportive conversation partners, there are positive psychological benefits, including reducing stress and improving self-affirmation [20]. In addition, previous research has found that even when someone only reads a poster’s self-disclosure without interaction, they can still develop a sense of personal connection [14]. It is useful to consider that self-disclosure can be further classified as per Malloch and Zhang [14], with factual self-disclosure revealing factual information, cognitive disclosure revealing thoughts and reasoning, and emotional disclosure revealing the poster’s feelings. This deeper categorization of self-disclosure was not undertaken in this study as it was beyond its scope, but merits further exploration in future research.

A final but nonetheless important consideration is the role of “lurkers,” who comprise the majority of participants in online communities; it is difficult to measure the true impact of interactions on all users of the groups because lurkers are not obvious [14]. For the most part, the visible support and engagement through comments and reactions is what users of a group are able to see. Given the large number of lurkers, however, we can only assume that the true engagement and utility of posts and comments are greater than what has been captured here. Furthermore, comment threads in groups can spark private messages between users, with discussions not visible to anyone else.

Strengths and Limitations
There were a number of strengths in this study. Given the dearth of literature about GDM online communities, this research illuminates an emergent phenomenon and activity experienced by many thousands around the world. The findings suggest important avenues for further inquiry in relation to GDM, in both online and offline settings.

A key limitation is that the data were only analyzed based on visible comments and reactions. Another limitation is that emoticons and photos were not systematically coded, even though they were part of the data set, as the semiotics of both were beyond the scope of this analysis. Finally, only 1 week of data were analyzed, and the results may vary depending on the collection period.

Conclusion
This study affirms the value of peer support that can be found in an online community. The large volume of posts and comments as well as high levels of positive engagement suggest that active participants’ needs around information and support in relation to GDM are being somewhat met by a peer-led online community. Given the practical limitations of formal health care, including the provision of ongoing social support, it is important to recognize how GDM online communities can complement health care and help address unmet needs. Furthermore, examining what information is being sought and shared by participants in GDM online communities is suggestive of gaps in information delivered through formal health care.

Acknowledgments
SP led and executed the study with design support, input, and advice from KC, LAE, and JB. KC and LAE provided statistical and methodological expertise alongside the other authors, and JB provided strategic advice. All authors reviewed and provided editorial suggestions on SP’s draft and agreed with the final submitted version.

Conflicts of Interest
None declared.

References


Abbreviations

GDM: gestational diabetes mellitus.
How to Identify e-Cigarette Brands Available in the United States During 2020-2022: Development and Usability Study

Shaoying Ma1*, PhD; Aadeeba Kaareen1*, BSocSci; Hojin Park1, PhD; Yanyun He1, PhD; Shuning Jiang2, BS; Zefeng Qiu3, MS; Zidian Xie3, PhD; Dongmei Li3, PhD; Jian Chen3, PhD; Richard J O’Connor4, PhD; Geoffrey T Fong5,6,7, PhD; Ce Shang1,8, PhD

1Center for Tobacco Research, The Ohio State University Wexner Medical Center, Columbus, OH, United States
2Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, United States
3Department of Clinical and Translational Research, University of Rochester Medical Center, Rochester, NY, United States
4Department of Health Behavior, Roswell Park Comprehensive Cancer Center, Buffalo, NY, United States
5Department of Psychology, University of Waterloo, Waterloo, ON, Canada
6School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada
7Ontario Institute for Cancer Research, Toronto, ON, Canada
8Department of Internal Medicine, Division of Medical Oncology, The Ohio State University Wexner Medical Center, Columbus, OH, United States
*these authors contributed equally

Corresponding Author:
Shaoying Ma, PhD
Center for Tobacco Research
The Ohio State University Wexner Medical Center
3650 Olentangy River Road, 1st Floor, Suite 110
Columbus, OH, 43214
United States
Phone: 1 6148976063
Email: shaoying.ma@osumc.edu

Abstract

Background: Prior studies have demonstrated that the e-cigarette market contains a large number of brands. Identifying these existing e-cigarette brands is a key element of market surveillance, which will further assist in policy making and compliance checks.

Objective: To facilitate the surveillance of the diverse product landscape in the e-cigarette market, we constructed a semantic database of e-cigarette brands that have appeared in the US market as of 2020-2022.

Methods: In order to build the brand database, we searched and compiled e-cigarette brands from a comprehensive list of retail channels and sources, including (1) e-liquid and disposable brands sold in web-based stores, (2) e-cigarette brands sold in brick-and-mortar stores and collected by the Nielsen Retail Scanner Data, (3) e-cigarette brands compiled by Wikipedia, (4) self-reported e-cigarette brands from the 2020 International Tobacco Control Four-Country Smoking and Vaping (ITC 4CV) US survey, and (5) e-cigarette brands on Twitter. We also estimated the top 5 e-cigarette brands by sales volume in brick-and-mortar stores, by the frequency and variety of offerings in web-based shops, and by the frequency of self-reported brands from the 2020 ITC 4CV US survey.

Results: As of 2020-2022, a total of 912 e-cigarette brands have been sold by various retail channels. During 2020-2022, the top 5 brands are JUUL, vuse, njoy, blu, and logic in brick-and-mortar stores; blu, king, monster, twist, and air factory for e-liquids in web-based stores; hyde, pod mesh, suorin, vaporlax, and xtra for disposables sold in web-based stores; and smok, aspire, vaporesso, innokin, and eleaf based on self-reported survey data.

Conclusions: As the US Food and Drug Administration enforces the premarket tobacco market authorization, many e-cigarette brands may become illegal in the US market. In this context, how e-cigarette brands evolve and consolidate in different retail channels will be critical for understanding the regulatory impacts on product availability. Our semantic database of e-cigarette brands can serve as a useful tool to monitor product and marketplace development, conduct compliance checks, assess manufacturers’ marketing behaviors, and identify regulatory impacts.
Introduction

Electronic cigarettes or e-cigarettes, were introduced in the US market in 2007 [1]. As of 2022, a total of 2.55 million US high school (14.1%) and middle school (3.3%) students reported using e-cigarettes in the past 30 days [2]. Unlike cigarettes, the e-cigarette market is highly diverse, with a wide range of product configurations (eg, rechargeables vs disposables) and features (eg, flavors and nicotine concentrations). E-Cigarette market has also been characterized as highly dynamic, with leading brands constantly changing [3-7]. As of 2018, the e-cigarette market was valued at US $3.5 billion, with a large number of brands (>460) and flavors (>7700) [8,9].

The number of e-cigarette brands is an important indicator of market concentration and competitiveness, which measures the ability of a single firm, or a small group of firms, to exert monopoly power [10,11]. For example, the cigarette market is characterized as a highly concentrated oligopolistic cigarette market, with a few large cigarette companies accounting for the majority of market shares [10,12]. In contrast, before JUUL uptake, the US e-cigarette market was very competitive and comprised hundreds, if not thousands, of small manufacturers selling products to consumers and retailers [13]. However, in recent years, cigarette companies have increasingly owned or invested in e-cigarette products, taking over a significant share of the e-cigarette market [14]. As a result, the e-cigarette market may have become less competitive over time and the e-cigarette and e-cigarette markets are growing integrated [15].

Constructing a database of e-cigarette brands is crucial to monitor market concentration, competitiveness, and development, especially for the US e-cigarette marketplace. However, such a database has not been established yet. As prior studies demonstrated, e-cigarette retails are dispersed in various retail channels, including brick-and-mortar stores, web-based stores, and vape shops or specialty stores [16,17]. However, only products and brands sold in brick-and-mortar stores are well-tracked by the Nielsen Retail Scanner Data. Little is known about e-cigarette brands and products sold in vape shops and web-based stores. Moreover, the few studies that examine products sold in these alternative channels suggest that products found in vape shops and web-based stores are significantly different from those in brick-and-mortar stores in terms of prices and product characteristics [18,19]. Therefore, a comprehensive database of brands is needed to identify brand differences by retail channels and assess accurately the e-cigarette market concentration and competitiveness.

A database of e-cigarette brands can also be used to identify regulatory impact by tracking the life cycle of a product and conducting compliance checks by identifying illegally sold products. In response to the substantial rise in e-cigarette use among US youth, the regulatory environment has evolved with many policy changes or proposals in recent years [20,21]. The available brands and types of e-cigarette products in the US market have been affected by the federal-level regulatory actions taken by the Food and Drug Administration (FDA) as well as state- and local-level policies that aim to curb e-cigarette use [3,21,22]. It is therefore highly warranted to have a live database of brands sold in the marketplace, which can demonstrate whether products with prohibited attributes are dropped from the market and how brands are evolving to comply with regulations.

Another function of a brand database is to assist in compliance checks. As the US FDA enforces the premarket tobacco market authorization (PMTA), many e-cigarette brands may become illegal in the US market [23]. A database of e-cigarette brands will allow the FDA and state and local agencies to identify if a certain brand is illegal or has been approved by the FDA. The brand database can also inform how e-cigarette brands evolve and consolidate with the enforcement of PMTA and answer the question of whether the PMTA makes the e-cigarette market less competitive in terms of limiting the availability of products [24,25].

The US marketplace of e-cigarette products is dynamic and rapidly changing. Therefore, a database of e-cigarette brands is also critical to tracking social media brand mentions and identifying emerging brands [26-29]. There have been attempts to identify mentions of e-cigarettes’ brand names and flavors on social media, such as Twitter (Twitter, Inc) [5,30], Reddit (ReddIt Inc) [30], YouTube (Google) [29], and Instagram (Meta Platforms) [31], and to identify and classify emerging flavors in the web-based market [32], using machine learning algorithms. However, the lack of a comprehensive semantic database of e-cigarette brands that capture different purchasing channels has hindered the efforts of using these language-based algorithms to efficiently identify brands.

Given the importance of an e-cigarette brand database, this study aims to construct such a database and further assess popular and common brands reported from different sources. The database also provides a snapshot of the products that have appeared in the market as of 2022, allowing for future studies that analyze marketplace development and policy impacts.

Methods

Data Sources

We searched comprehensively and identified 6 sources that we used to create the semantic database of e-cigarette brands. These include (1) existing brand websites surveillance [8] that reported on a pre-determined list of e-cigarette brands; (2) e-cigarette brands sold in the brick-and-mortar stores reported by the Nielsen Retail Scanner Data, accessed through the Kilts Center for Marketing at the University of Chicago Booth School of Business [33]; (3) Wikipedia’s “list of electronic cigarette and
e-cigarette liquid brands,” which was established by a community of volunteers who add brand names from data sources including peer-reviewed journal articles, news articles, reports from antitobacco organizations, and other sources, and we accessed it on November 29, 2022 [34]; (4) a list of e-cigarette brands mentioned on Twitter from May 2021 to December 2021 [5]; (5) self-reported brands from the 2020 International Tobacco Control Policy Evaluation Project’s Four-Country Smoking and Vaping (ITC 4CV) US survey; and (6) e-cigarette brands collected using web scraping of 5 popular web-based stores. The scraping data captured a wide range of e-liquid and disposable e-cigarette products and brands sold on the web, which were traditionally not well captured by other sources. The scraping data contained information of over 16,000 unique e-liquid or disposable products collected during 2021-2022. Additional details are available in Multimedia Appendix 1.

### Brand Identification

For web-based store brands, the same brand could be written in slightly different ways across stores, and we observed variations caused by (1) spaces or hyphens (such as “sad boy” vs “sad boy”); (2) pluralization (such as “bad drip” vs “bad drips”); (3) suffixes (such as “barista brew” vs “barista brew co.”); and “mr. good” vs “mr. good vape”); (4) abbreviations or aliases (such as “naked 100” vs “NKD 100”); and (5) misspelling (such as “coastal clouds” vs “costal clouds”). Using algorithms, we identified 237 unique e-liquid brand 97 unique disposable e-cigarettes in 2021-2022.

We used STATA/SE (version 17.1; StataCorp) software to convert all brand names to lowercase and remove duplicates. For brands with more than 1 product line, a research specialist from our team looked up relevant information on the web and identified them so that we extracted each brand name with multiple product lines and collections and further cleaned up the brand database. A total of 4 members from our research team reviewed the brand list to ensure accuracy. We then used this list of e-cigarette brands (based on Nielsen, Wikipedia, and data collected from web-based stores), to identify and match with the self-reported e-cigarette brands in ITC 4CV Wave 3 (2020) survey data.

Among 1696 observations from the self-reported brand variables in the ITC survey, 181 (10.7%) of them contained irrelevant or insufficient information, such as “don’t know,” and “its unbranded from the market.” For the remaining 1515 observations, 720 (47.5%) were successfully identified and matched using our initial brand list from 3 sources (Nielsen, Wikipedia, and our unique e-liquid data scraped from web-based stores). Aided by algorithms, we checked the rest of the self-reported brands that were unmatched or unidentified both by humans and by machines and extracted brand information from those observations. Since the same brand can be reported in different ways such as with or without space, we consolidated the brand names by removing the spaces and documented possible variations of a unique brand, such as “geek vape” and “geek vape” in Multimedia Appendix 2. Consequently, 167 new brand names from the ITC 4CV survey were added to our brand database.

Based on our disposable e-cigarette brand data collected from web-based stores in 2022 and the list of e-cigarette brands on Twitter from May 2021 to December 2021 collected by Tang et al [5], we then updated our brand database and in total 138 new e-cigarette brands were added from those 2 data sources.

### Identifying Top Brands

For products sold in brick-and-mortar stores, we used the Nielsen Retail Scanner data to identify the top 5 brands with the greatest sales in 2020. For self-reported brands, we used the frequency counts in the 2020 ITC 4CV US survey to identify the top brands. For products sold in web-based stores, we use the frequencies of offering (number of unique products of a brand multiplied by the number of stores that offer the brand) to identify the top 5 brands for e-liquid and disposable products, respectively.

### Ethical Considerations

In this study, we compiled data on e-cigarette brands using sources including peer-reviewed publications about brand surveillance and brand mentions [5,8], the Nielsen Retail Scanner Data [33], Wikipedia [34], a tobacco use survey, and our unique data scraped from e-cigarette web-based stores. Thus, no human subjects were involved, and the determination of no human subjects was approved by the Ohio State University institutional review board. The survey protocols and all materials including the survey questionnaire for the 2020 (Wave 3) ITC 4CV US survey were cleared for ethics by the Research Ethics Board, University of Waterloo, Canada (REB#20803/30570) and the Medical University of South Carolina (waived due to minimal risk).

### Results

In total, we identified 912 e-cigarette brands available in the United States during 2020-2021 and presented this database in Multimedia Appendix 2. As we compiled the brand database from multiple sources, we observed that e-cigarette manufacturers are creative when naming their product brands; in some cases, generic terms such as “z,” “e s,” “pods,” “something,” “mix,” and “e-hookah” are used as brand names. There are also brand names like “zoom” that could lead to ambiguous search terms and false results. We suggest that for those brands, researchers could use the brand names along with search terms such as “vape” as search terms to conduct market surveillance (eg, social media monitoring) and avoid false results.

The top 5 brands from different retail channels and resources are presented in Textbox 1, which presents the top 5 brands (by sales volume in counts) in the Nielsen Retail Scanner Data during 2020, the top 5 e-liquid brands (by frequency counts) in the scraped data during 2021, the top 5 disposable e-cigarette brands (by frequency counts) in the scraped data during 2022, and the top 5 self-reported brands (by frequency counts) in the ITC 4CV US survey during 2020.
Focused on the sales of e-cigarette products from brick-and-mortar stores, Nielsen Retail Scanner Data captured 82 brand or model names (567 product Universal Product Codes [UPCs]) in the “ELECTRONIC CIGARETTES–SMOKING” product module in 2020. Of the 82 brands or models, 59 sold at least 2 unique e-cigarette products. The top 5 high-level e-cigarette brand names by the number of products (ie, UPCs) in Nielsen Retail Scanner Data were njoy (70 product UPCs), vuse (64 product UPCs), blu (63 product UPCs), logic (31 product UPCs), and JUUL (30 product UPCs), which together accounted for about 45.5% (n=258) of total e-cigarette products observed in the Nielsen data. Product types varied, such as e-liquids, replacement pods, nonreplacement pods, prefilled cartridges or tanks, disposables, devices (eg, pod and mod), and starter kits (both device and pods or cartridges). Furthermore, during 2020 in brick-and-mortar stores, the top 5 brands (in descending order) by sales volume in counts were JUUL, vuse, njoy, blu, and logic, which represented about 137 million (97.9%) out of 140 million total e-cigarette sales volume in Nielsen Retail Scanner Data.

Based on the web-based store data, the top 5 e-liquid brands by frequency counts (reported in parentheses) were blu (7800), king (553), monster (531), twist (488), and air factory (480). Here the frequency count for each top e-liquid brand was calculated as the number of different products from this brand offered by 5 stores, that is, a number of stores offering multiplied by brand variations. The top 5 disposable e-cigarette brands by frequency counts (reported in parentheses) sold by 5 web-based vape shops were hyde (214), pod mesh (126), suorin air bar (125), vaporlax (124), and xtra (115), which were different from the top brands in Nielsen data. Interestingly, self-reported data suggested top brands that differ from both web-based stores and Nielsen data. Based on the ITC 4CV Wave 3 US survey data, during 2020, the top 5 brands by frequency counts (ie, how often were they mentioned by participants, reported in parentheses) were smok (201), aspire (104), vapourco (76), innokin (65), and elfleaf (64).

**Discussion**

 Prior research used keyword searches and identified over 460 e-cigarette brands through January 2014, but the market has grown exponentially since then [35]. Although existing evidence shows that the number of e-cigarette brands has not increased much since 2014, this conclusion relied on limited resources or retail channels [8]. In this study, we used 6 different data sources to consolidate brands identified from multiple retail channels (brick-and-mortar stores, web-based stores, social media mentions, and self-reported data) and identified 912 unique e-cigarette brands as of 2020-2022. This suggests that a large number of e-cigarette brands existed in the market before the enforcement of PMTA, which is in line with the existing assessment that the e-cigarette market is more competitive than cigarettes with many e-cigarette manufacturers [10,13]. However, as tobacco companies increase their shares in the e-cigarette market by producing their own e-cigarette brands or purchasing existing e-cigarette brands, future research could use our database to map brands with tobacco company ownerships to better understand tobacco companies’ interest in e-cigarettes and the changes in the competitiveness and concentration of the e-cigarette market [15].

Regulation on e-cigarettes such as the FDA’s PMTA action might reduce competition in the marketplace, as e-cigarette brands owned by big tobacco companies may have a greater capacity to properly respond to regulatory actions (eg, file a PMTA) and remain in the market [24]. In addition, recent evidence suggests that in response to the FDA regulations that ban flavors other than menthol and tobacco in prefilled cartridges, the marketplace and consumers switch to disposables that provide a spectrum of flavors [3]. Therefore, it is important to track market development, product availability, manufacturers’ market power (ie, market concentration or competitiveness), and market responses to regulations, which this brand database can help.

Since 2021, the FDA has issued marketing denial orders (MDOs) to a number of e-cigarette companies, including prohibiting sales and distribution of all products from JUUL including devices and pods [21,36]. The FDA maintains an updated list of companies (instead of brand names) that have products currently marketed in the United States and have been issued MDOs [36]. It also announces the latest marketing decisions about e-cigarette companies and specific e-cigarette brands and models that have received MDOs [36]. After comparing the brands in the FDA’s marketing decisions from October 12, 2021, to January 24, 2023 [36], with our brand database, we found that all brands in those decisions were included in our database, which supports the validity and completeness of our database, which goes beyond the FDA list with additional brands that are not included in the decision list. This comparison also supports the potential of using this brand database to conduct compliance checks and monitor market product availability that could aid policy making.

Existing evidence from other studies shows that e-cigarette brands owned by tobacco companies typically offer a limited range of e-cigarette products, while brands owned by vape shops are much more likely to have a diverse range of flavor and nicotine options [8]; e-cigarette brands and product types are among the most important factors that influence consumers’
purchasing choices [8,37]. We also show that the most frequently mentioned e-cigarette brands by participants in a tobacco survey, the top brands by sales volume in brick-and-mortar stores based on Nielsen Retail Scanner data, and the top brands by product availability in web-based vape shops are very different, which demonstrates the importance of obtaining information from multiple data sources and purchasing channels when calculating the market share of various e-cigarette brands and the market concentration and power of each brand in the quickly changing e-cigarette marketplace.

We report the top 5 disposable e-cigarette brands (by frequency counts) from our 2022 scraped data as well as the top 5 e-liquid brands (by frequency counts) from our 2021 scraped data. Our database includes e-cigarette brands that are predominately sold in web-based stores; in particular, e-liquid brands are typically not well captured in existing data sources (eg, Nielsen Retail Scanner Data) and our database makes a contribution to the literature by providing those brands. For e-cigarette brands sold on the web, we have also observed that web-based vape shops sell a variety of products with different flavors, nicotine levels, and forms, suggesting the appeals of these brands sold in web-based stores [18,19,32,38,39].

Furthermore, to the best of our knowledge, our database containing brand names from 6 different data sources is the most comprehensive and up-to-date, and could contribute to tobacco regulatory science in multiple ways. In addition to compliance checks, product availability assessments, and market power estimations, the brand database can be used to conduct social media market surveillance using natural language processing and other machine learning techniques [40,41]. Specifically, brand search terms are key to identifying whether e-cigarettes are being mentioned in social media posts or on a website [5,31,42,43]. Therefore, our database not only provides a comprehensive list of e-cigarette brand search terms but also allows for the identification of emerging brands through techniques such as name entity recognition. Future research may also expand this database by linking e-cigarette brands with product characteristics such as flavors, nicotine levels and forms, and device types (disposables vs cartridges vs e-liquid; open vs closed systems) to facilitate rapid market surveillance.

Another potential function of this database is to assist in future tobacco survey development by allowing for a dropdown list that reflects the complexity of the e-cigarette marketplace and brands [2,44-48]. Our experience with the ITC 4CV survey suggests that respondents may have difficulties in self-reporting e-cigarette brands in an open-ended question. A drop-down list could enhance the ability of surveys to capture brands accurately.

Our study has some caveats. The brand names from our web-based vape shop data reflect a snapshot of e-liquid brands sold in popular web-based stores in 2021. As we continue our web scraping efforts, we will be able to extract brand information for a wide range of e-cigarette products sold in the web-based market, such as disposables, devices, and starter kits. Nonetheless, the e-liquid and disposable e-cigarette brand data from web-based vape shops and the list of e-cigarette brands on Twitter complements the brand information in brick-and-mortar stores from the Nielsen data, the existing list of brand names from Wikipedia (accessed on November 29, 2022), and self-reported brands from survey participants during 2020, together making our brand database the most comprehensive and up-to-date to the best of our knowledge. Another limitation is that brands in specialty vape shops are not necessarily captured in our semantic database. Future research is warranted to assess the brand information from specialty vape shops by paying visits to the stores and conducting qualitative interviews with store owners and staff members.

In conclusion, the development of a comprehensive semantic database that contains e-cigarette brands in the US during 2020-2022 demonstrates the competition of the existing e-cigarette market, as well as the use of novel techniques such as name entity recognition to identify emerging brands. It also has broader public health implications on the need to continuously monitor market concentration and manufacturers’ responses to regulations. Our database can be used to facilitate compliance checks and rapid surveillance of product availability to inform policy making.

Acknowledgments
We used data from the Nielsen Company (United States), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. During the preparation of this work, the authors used ChatGPT (version 3.5; OpenAI) to check grammar errors and improve language flow. After using this tool or service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication. We thank Sooa Ahn, a doctoral student at The Ohio State University (OSU) Economics department, for her assistance with the formatting of this paper. This study was supported by The Ohio State University Comprehensive Cancer Center (OSUCCC) Center for Tobacco Research Pilot Grant and by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) and Food and Drug Administration (FDA) Center for Tobacco Products under award (U01CA278695, U54CA228110, and U54CA287392) OSU Tobacco Center of Regulatory Science (TCORS) Marketing Monitoring Core. CS is funded by the NCI under award (R21CA249757). SM and HP are both supported by the Pelotonia Fellowship program at the OSUCCC. The 2020 International Tobacco Control Four-Country Smoking and Vaping US survey was supported by grants from the National Cancer Institute of the United States (P01CA200512) and the Canadian Institutes of Health Research (FDN-148477). Additional support to GTF is
provided by a Senior Investigator Grant from the Ontario Institute for Cancer Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the FDA.

Data Availability
All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1 and 2.

Authors' Contributions
CS and SM contributed to the conceptualization. CS and SM contributed to the methodology. SJ, ZQ, SM, HP, CS, and YH contributed to the software. CS and JC contributed to the validation. SM and CS contributed to the formal analysis. SM, AK, SJ, ZQ, ZX, DL, and HP contributed to the investigation. CS and JC contributed to the resources. SJ, ZQ, SM, AK, ZX, and DL contributed to the data curation. AK and SM contributed to the writing—original draft preparation. SM, AK, RJO, GTF, HP, ZX, DL, and YH contributed to writing—review and editing. CS, JC, RJO, and GTF performed the supervision. SM, SJ, and ZQ contributed to project administration. CS, JC, RJO, and GTF contributed to funding acquisition. All authors have read and agreed to the published version of the study.

Conflicts of Interest
GTF has served as an expert witness or a consultant for governments defending their country’s policies or regulations in litigation. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Additional details of the methodology used to create the brand database.

Multimedia Appendix 2
E-cigarette brand names available in the United States during 2020-2022.

References
12. He Y, Ma S, Yang Q, Shang C. How cigarette excise tax pass-through to prices responds to the uptake and evolution of e-cigarettes (ECs). Tob Control 2023 [FREE Full text] [doi: 10.1136/tc-2023-058078] [Medline: 37640531]


**Abbreviations**

- **FDA**: Food and Drug Administration
- **ITC 4CV**: International Tobacco Control Four-Country Smoking and Vaping
- **MDO**: marketing denial order
- **PMTA**: premarket tobacco market authorization
- **UPC**: Universal Product Code

https://formative.jmir.org/2024/1/e47570

JMIR Form Res 2024 | vol. 8 | e47570 | p.1143

*(page number not for citation purposes)*
Using a Novel Gameplay Intervention to Target Intrusive Memories After Work-Related Trauma: Iterative Qualitative Analysis of Intensive Care Unit Staff Experiences

Priya Patel¹, MSc; Susan Brown²,³, PhD; Boliang Guo¹, PhD; Emily A Holmes⁴, PhD; Lalitha Iyadurai⁵, PhD; Jonathan Kingslake⁵, BSc; Julie Highfield⁶, DClinPsy; Richard Morriss¹,²,³,⁷, MD

¹NIHR ARC East Midlands, University of Nottingham, Nottingham, United Kingdom
²NIHR MindTech MedTech Co-operative, University of Nottingham, Nottingham, United Kingdom
³Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom
⁴Department of Women’s and Children’s Health, Uppsala University, Uppsala, Sweden
⁵P1Vital Products Ltd, Wallingford, Oxfordshire, United Kingdom
⁶Intensive Care Society, London, United Kingdom
⁷NIHR Nottingham Biomedical Research Centre, Nottingham, United Kingdom

Corresponding Author:
Priya Patel, MSc
NIHR ARC East Midlands
University of Nottingham
Innovation Park Jubilee Campus
University of Nottingham Innovation Park, Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 7790988203
Email: Priya.Patel1@nottingham.ac.uk

Abstract

Background: Many intensive care unit (ICU) staff experience intrusive memories following work-related traumatic events, which can lead to long-term mental health outcomes and impact work functioning. There is a need for interventions that target intrusive memories in this population; however, factors such as mental health stigma and difficulty in fitting interventions into busy schedules can pose barriers. The Brief Gameplay Intervention For National Health Service Intensive Care Unit Staff Affected By COVID-19 Trauma (GAINS) study tested a brief, digital imagery-competing task intervention (including computer gameplay) with the aim of reducing the recurrence of intrusive memories, which holds promise for overcoming some of these barriers.

Objective: This substudy aims to explore barriers and facilitators to the uptake and practical use of the intervention by ICU staff, along with its acceptability, and iteratively explore the impact of intervention optimizations to further refine the intervention.

Methods: The GAINS study is a randomized controlled trial comparing access to a brief digital imagery-competing task intervention for 4 weeks with usual care followed by delayed access to the intervention. The participants were ICU staff who worked during the COVID-19 pandemic and experienced intrusive memories. All participants were sent a questionnaire at 4 weeks to gather data about intervention acceptability. Nested within the randomized controlled trial, a subset of 16 participants was interviewed, and data were analyzed using thematic analysis drawing from a framework approach.

Results: Both quantitative and qualitative data indicated high acceptability of the intervention. Intervention use data show that, on average, staff were able to target approximately 73% (3.64/4.88) of their intrusive memories and engaged with the Tetris component for the full 20 minutes per session. Overall, on the acceptability questionnaire, staff found the intervention easy to use, helpful, and highly acceptable. The interviews generated four themes: approach to the intervention, positives of the intervention, negatives of the intervention, and improvements and optimizations. Findings highlighted barriers that ICU staff experienced: stigma, feeling weak for seeking help, not wanting colleagues to know they were struggling, and skepticism. However, they provided suggestions on how barriers could be overcome and discussed the advantages of the intervention when compared with other treatments. Although participants described many positive aspects of the intervention, such as being easy to use, enjoyable, and leading to a reduction in the frequency or intensity of intrusive memories, they also raised practical issues for implementation.
Conclusions: The intervention has the potential to overcome stigma and reduce the frequency of intrusive memories after traumatic events among ICU staff. Further refinement is needed to improve the adoption and reach of this intervention. A limitation is that we could not interview the National Health Service staff who were unable or unwilling to take part in the trial.

**KEYWORDS**

intensive care; posttraumatic stress disorder; PTSD; qualitative research; intervention study; health care professionals; digital intervention; staff well-being; pandemic; intrusive memories; work-related trauma; mobile phone

**Introduction**

**Background**

Following exposure to a psychologically traumatic event (eg, witnessing a severe injury or death) [1], intrusive memories are common, particularly in the first few days and weeks. Intrusive memories are emotional, intrusive, and primarily visual memories of the traumatic event that pop unbidden into the mind [1], that is, it takes the form of mental imagery. When they intrude repeatedly into mind, they comprise a “core clinical feature” of posttraumatic stress disorder (PTSD) [1]. Frontline health care staff are particularly at risk, with 65% of emergency nurses reporting having intrusive memories of work-related traumatic events prepandemic, such as the death of a patient [2]. For some individuals, intrusive memories persist for more than a month and thus become a core symptom of PTSD [3]. It has been known for some time that frontline health care staff experience repeated exposure to potentially traumatic events [4-7], even before the pandemic. This exposure was even worse during the COVID-19 pandemic, with 5 times more UK health care staff reporting PTSD symptoms, such as bothersome intrusive memories, in 2020 than in 2015 [8,9]. In this study, we will focus on intensive care unit (ICU) staff working during the pandemic, although it is assumed that this has wider relevance as trauma exposure and intrusive memories also affect other staff groups, and experiencing trauma was also prevalent prepandemic. A key difference with the pandemic was the increased frequency of trauma exposure for this group.

PTSD symptoms, such as intrusive memories, are associated with poorer long-term physical and mental health outcomes [10]. There is a great cost for patients and society when frontline staff are affected, with 27% of health care staff who reported PTSD symptoms believing that their work functioning was negatively impacted [6]. Furthermore, there are problems with staff shortages and dropouts, with PTSD symptoms causing 20% of staff to consider a job change [6]. Mental health problems remain the leading cause of sickness absence in the National Health Service (NHS) [11]. Owing to these factors, the mental health of frontline health care staff exposed to traumatic events is a major priority internationally [12].

**Prior Work**

A novel approach in this area is the development of a brief mechanism-driven behavioral intervention to reduce intrusive memories [13-15]. This brief imagery-competing task intervention for established intrusive memories after trauma consists of a reminder cue to the traumatic event, followed by playing the computer game Tetris for 20 minutes with instructions to use mental rotation during gameplay [16]. The principles of the intervention are informed by the neuroscience of memory storage and updating (so called consolidation and reconsolidation) [17,18] and cognitive task interference [19]. The hypothesis is that the memory consolidation or reconsolidation process of a traumatic event can be disrupted by engaging in visuospatial demanding tasks, for example, Tetris, and reduce the frequency of the intrusive memories. Randomized controlled trials (RCTs) have shown that this type of intervention approach can reduce the frequency of intrusive memories soon after trauma exposure in women after traumatic childbirth [20] and after a traumatic motor vehicle accident [21,22]. In addition, the intervention has recently been found to reduce established and older intrusive memories in case series studies with patients with chronic PTSD [23], refugees [24], and most recently NHS staff exposed to work-related trauma [16,25]. In particular, ICU staff (those working in intensive care, intensive therapy, and high-dependency units) face repeated exposure to trauma as an inevitable and intrinsic part of the work setting [2]. Now that the adverse effects on staff health and well-being are becoming better recognized [12], it is imperative to find ways to address these needs.

This study is part of a Brief Gameplay Intervention For National Health Service Intensive Care Unit Staff Affected By COVID-19 Trauma (GAINS) study (ClinicalTrials.gov: NCT04992390) for health care staff in the United Kingdom who faced trauma exposure as part of their work during the COVID-19 pandemic. The intervention used a brief imagery-competing task intervention with the aim of reducing and preventing the recurrence of intrusive memories from work-related trauma exposure.

There are many barriers to the implementation of both digital and face-to-face mental well-being interventions in health care staff, specifically owing to the complexity of the role and organization of health care [26]. Personal barriers to uptake by health care staff include a perceived lack of ownership when they feel an intervention is not driven by them; feeling as though they are obliged to participate; and practical barriers to participation, such as cost, time commitments, and age [26,27]. In setting up the GAINS study in collaboration with the Intensive Care Society (ICS) in the United Kingdom, it became clear that time commitment is of particular importance, as the nature of their role as health care staff means they are working in busy and pressurized environments, even more so given staff shortage problems. Staff have indicated that financial barriers have created a perception that spending priorities prioritize patients’ needs over the well-being of the staff [26,27]. The situation is made more complex by organizational barriers,
including changes in senior management, managers being perceived as inaccessible, ongoing organizational changes and restructuring, and the influence of target-driven cultures [26]. A lack of suitable infrastructure to support digital health interventions has also been found to be a barrier to use [28], such as a lack of available computers or internet connection, which could be a potential barrier for staff working in hospital settings.

There are also huge barriers owing to stigma and inclusivity, with many health care professionals experiencing shame for struggling with their mental health [29]. A literature review found that, within the nursing population specifically, many nurses felt that they needed to keep their mental health a secret owing to fear of being judged by their colleagues [30]. This was not only because of fear of what others may think but also because of self-directed stigma, with 21% of nurses struggling with their mental health believing that this was because of a personality weakness or character defect [30]. Therefore, digital health interventions may pose a key strength because they can be accessed independently and from any location, thereby providing users with privacy and anonymity [28]. This is in contrast to, for example, attending mental health services for psychological therapy or medication. In addition, it is vital to ensure that study samples are representative, as intersectionality plays a role in this, with health care professionals from ethnic minority backgrounds experiencing increased workplace discrimination and harassment [31]. Barriers and facilitators to the implementation of mental health interventions for NHS staff will occur at the organizational and individual levels. Therefore, it is necessary to consider the barriers and facilitators that may exist at multiple levels.

This Intervention

The GAINS intervention used a secure web-based mental health and well-being platform (i-spero, P1vital Products Ltd) to allow participants to access the intervention on a smartphone, tablet, or computer. Participants had an initial guided session with a researcher over Microsoft Teams, in which they were provided with step-by-step instructions on how to use the intervention as well as explanatory videos and multiple-choice questions. Participants were assisted in briefly listing their intrusive memories during the initial session (hotspots). Subsequently, they were prompted to recall the image associated with 1 specific intrusive memory. They played the Tetris game for 20 minutes using mental rotation, in which they had to imagine how to rotate the next Tetris piece so that it could fit in the existing structure. The intervention took a total of approximately 25 minutes each time, and they could target different memories on different days. Finally, they were trained in using the i-spero (P1vital Products Ltd) platform to monitor intrusive memories in daily life [25].

The underlying principle of the intervention is its imagery-based nature and that it can be used regardless of the content of the intrusive memory (a motor vehicle accident or witnessing a patient’s death when working in the ICU). The intervention is used once per different intrusive memory, so it can also be used by someone who has experienced a single traumatic event or multiple ongoing traumatic events. It can also be used for new intrusive memories that develop during the trial. Owing to the nature of their roles, participants work in an environment where trauma can be a frequent occurrence. Therefore, choosing an intervention that can address the specific challenge of recurring and frequent trauma is crucial.

This intervention holds particular promise for overcoming some of the mentioned barriers to the implementation of mental well-being interventions after trauma in ICU settings [25]. For example, rather than focusing on a mental health diagnosis such as PTSD, the entry to accessing the intervention is a simpler index problem, namely, having intrusive memories of the traumatic event. It is brief (1 guided intervention session of 1 hour, followed by self-guided use of approximately 25 minutes per session, whereby the aim is 1 session per different intrusive memory). It is digital and can be used flexibly in different locations (eg, on a smartphone during a commute) and may have lower stigma than attending mental health services (as the intervention involves a digital task including a computer game rather than, for example, talking to a trained therapist). As the intervention can be used for new intrusive memories as they arise, it is well suited for health care staff facing repeated or ongoing trauma in their jobs. Finally, as the intervention requires minimal therapist resources, it has the potential to be more cost-effective and scalable than current evidence-based interventions that require more contact.

Aims

This qualitative substudy as part of the GAINS study had the following aims:

- To explore barriers and facilitators to the uptake and use of the imagery-based competing task intervention to reduce intrusive memories of work-related trauma in ICU staff, along with its acceptability.
- To iteratively explore the impact of optimizations made to the intervention to address some of the barriers (and enhance facilitators), allowing us to then further refine the intervention for future use by ICU staff.

Methods

Design

The GAINS study is an RCT comparing immediate access to a brief digital imagery-competing task intervention for 4 weeks (the immediate intervention arm) versus receiving usual care for 4 weeks, followed by delayed access to the intervention for an additional 4 weeks (the delayed intervention arm). This manuscript contains quantitative descriptive data from an acceptability survey completed 4 weeks after the first intervention session as well as data from the intervention itself on uptake and completion of the intervention. This descriptive data are provided as contextual information for the qualitative findings, which explore in more detail barriers and facilitators to the uptake, completion, and overall acceptability of the study. The qualitative analysis draws on 2 sources of data collection: interviews from a maximum variance sampling method in a subset of participants and free-text feedback that was sought from all participants completing the acceptability survey. The narrative feedback was used to triangulate the qualitative
Individuals who met the inclusion criteria were given the participant information sheet again, along with the opportunity to ask questions to the investigator, their general practitioner, or other independent parties to make an informed decision about whether to participate. If they were still interested in participating in the study, a researcher arranged a time to contact them by phone or video call to obtain informed consent. The participant and researcher completed, signed, and dated the consent form using a simple electronic signature via email, which included providing permission to be contacted for the qualitative interview component of the study. The consent form was retained electronically in a secure format, and participants were emailed a copy for their records.

**Study Procedures of the GAINS Study RCT**

After providing informed consent, participants were asked to complete a daily web-based intrusive memory diary for a run-in period of 1 week. Each day, the participants were asked to indicate if they had any intrusive memories and, if so, how many. Participants who met the eligibility criterion of having ≥3 intrusive memories in the run-in week were randomly allocated to either the immediate intervention arm or the delayed intervention arm in a 1:1 overall ratio. Participants were randomized through the P1vital electronic Participant Reported Outcome system (P1vital Products Ltd), and the outcome assessment was completed remotely by the participants (independently of the research team) on this platform. The qualitative team was independent of randomization, delivery of the intervention, and assessment of the quantitative outcomes. The immediate intervention group received the intervention immediately for 4 weeks, whereas the delayed group received usual care for 4 weeks, followed by the 4-week-long intervention.

### Intervention Being Trialed

The brief digital intervention was delivered through a secure web platform that participants accessed on their smartphone or other internet-enabled device (refer to the studies by Iyadurai et al [25] and Ramineni et al [16] for complete details). Participants were provided with an initial session guided by a clinical psychologist or delegated researcher to run through the intervention as well as follow-ups and support available throughout the intervention. During the initial session (approximately 1 hour), participants were asked to briefly list the different intrusive memories they have and choose the one they wish to target first. They were then asked to complete the intervention, which included several key components: (1) the participant was asked to briefly bring to mind the intrusive image as a reminder to the specific memory, (2) they received instructions on how to play the computer game Tetris using “mental rotation,” and (3) they were asked to play Tetris using mental rotation for at least 20 minutes. During the intervention, participants were asked to rate how distressed they are feeling on 3 occasions, to rate the vividness of the image that is brought to mind, and to rate how much they were able to follow mental rotation instructions, to assess adherence to the instructions. After this first session, participants were able to use the intervention as many times as they liked over the next 4 weeks (eg, to target any other intrusive memories on their list or those

**Recruitment for the GAINS Study RCT**

The participants were ICU staff who worked during the COVID-19 pandemic and experienced intrusive memories as a result. Participants were recruited through ICS membership and existing social media followers supplemented by targeted advertisements on social media (eg, Facebook and Twitter) [16,25]. The advertisement email contained a link to the study website, where interested individuals were able to read a study summary, including the participant information sheet (Multimedia Appendix 1), and watch a video explaining intrusive memories in further detail. The study website also included a link to the 10-item prescreening eligibility questionnaire.

**Inclusion and Exclusion Criteria of the GAINS Study RCT**

Potential participants met the inclusion criteria if they were aged ≥18 years; able to read, write, and speak in English; worked in a clinical role in an NHS ICU or equivalent during the COVID-19 pandemic; experienced at least 1 traumatic event related to their work during the COVID-19 pandemic; meeting criterion A of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria for PTSD: “exposure to actual or threatened death, serious injury, or sexual violence” by “directly experiencing the traumatic event(s)” or “witnessing, in person, the event(s) as it occurred to others”; experience intrusive memories of the traumatic event or events; experienced at least 3 intrusive memories in the week before screening; have internet access; were willing and able to provide informed consent and complete study procedures (including briefly listing their intrusive memories [without going into any detail] and playing the computer game Tetris with particular mental rotation instructions and completing a web-based intrusive memory diary); and were willing and able to be contacted by the research team during the study period. Potential participants were excluded if they had <3 intrusive memories during the run-in week.

**Ethical Considerations**

GAINS study part 1 received a favorable opinion from Wales Research Ethics Committee (REC) 6 on May 21, 2021 (REC Reference 21/WA/0173 and Integrated Research Application System project ID number 297063). There were 4 non-substantial amendments made to the Interview Topic Guide - non-substantial amendment 1 on July 21, 2021; non-substantial amendment 3 on Oct 5, 2021; non-substantial amendment 5 on Nov 12, 2021; non-substantial amendment 9 on Jun 8, 2022. The purpose of these amendments was to slightly amend wording of questions, gather information about which NHS Trust the participant worked in and gather thoughts on the optimised version of the intervention.

**Study Procedures of the GAINS Study RCT**

After providing informed consent, participants were asked to complete a daily web-based intrusive memory diary for a run-in period of 1 week. Each day, the participants were asked to indicate if they had any intrusive memories and, if so, how many. Participants who met the eligibility criterion of having ≥3 intrusive memories in the run-in week were randomly allocated to either the immediate intervention arm or the delayed intervention arm in a 1:1 overall ratio. Participants were randomized through the P1vital electronic Participant Reported Outcome system (P1vital Products Ltd), and the outcome assessment was completed remotely by the participants (independently of the research team) on this platform. The qualitative team was independent of randomization, delivery of the intervention, and assessment of the quantitative outcomes. The immediate intervention group received the intervention immediately for 4 weeks, whereas the delayed group received usual care for 4 weeks, followed by the 4-week-long intervention.

**Intervention Being Trialed**

The brief digital intervention was delivered through a secure web platform that participants accessed on their smartphone or other internet-enabled device (refer to the studies by Iyadurai et al [25] and Ramineni et al [16] for complete details). Participants were provided with an initial session guided by a clinical psychologist or delegated researcher to run through the intervention as well as follow-ups and support available throughout the intervention. During the initial session (approximately 1 hour), participants were asked to briefly list the different intrusive memories they have and choose the one they wish to target first. They were then asked to complete the intervention, which included several key components: (1) the participant was asked to briefly bring to mind the intrusive image as a reminder to the specific memory, (2) they received instructions on how to play the computer game Tetris using “mental rotation,” and (3) they were asked to play Tetris using mental rotation for at least 20 minutes. During the intervention, participants were asked to rate how distressed they are feeling on 3 occasions, to rate the vividness of the image that is brought to mind, and to rate how much they were able to follow mental rotation instructions, to assess adherence to the instructions. After this first session, participants were able to use the intervention as many times as they liked over the next 4 weeks (eg, to target any other intrusive memories on their list or those

**Recruitment for the GAINS Study RCT**

The participants were ICU staff who worked during the COVID-19 pandemic and experienced intrusive memories as a result. Participants were recruited through ICS membership and existing social media followers supplemented by targeted advertisements on social media (eg, Facebook and Twitter) [16,25]. The advertisement email contained a link to the study website, where interested individuals were able to read a study summary, including the participant information sheet (Multimedia Appendix 1), and watch a video explaining intrusive memories in further detail. The study website also included a link to the 10-item prescreening eligibility questionnaire.

**Inclusion and Exclusion Criteria of the GAINS Study RCT**

Potential participants met the inclusion criteria if they were aged ≥18 years; able to read, write, and speak in English; worked in a clinical role in an NHS ICU or equivalent during the COVID-19 pandemic; experienced at least 1 traumatic event related to their work during the COVID-19 pandemic; meeting criterion A of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria for PTSD: “exposure to actual or threatened death, serious injury, or sexual violence” by “directly experiencing the traumatic event(s)” or “witnessing, in person, the event(s) as it occurred to others”; experience intrusive memories of the traumatic event or events; experienced at least 3 intrusive memories in the week before screening; have internet access; were willing and able to provide informed consent and complete study procedures (including briefly listing their intrusive memories [without going into any detail] and playing the computer game Tetris with particular mental rotation instructions and completing a web-based intrusive memory diary); and were willing and able to be contacted by the research team during the study period. Potential participants were excluded if they had <3 intrusive memories during the run-in week.

**Ethical Considerations**

GAINS study part 1 received a favorable opinion from Wales Research Ethics Committee (REC) 6 on May 21, 2021 (REC Reference 21/WA/0173 and Integrated Research Application System project ID number 297063). There were 4 non-substantial amendments made to the Interview Topic Guide - non-substantial amendment 1 on July 21, 2021; non-substantial amendment 3 on Oct 5, 2021; non-substantial amendment 5 on Nov 12, 2021; non-substantial amendment 9 on Jun 8, 2022. The purpose of these amendments was to slightly amend wording of questions, gather information about which NHS Trust the participant worked in and gather thoughts on the optimised version of the intervention.

**Study Procedures of the GAINS Study RCT**

After providing informed consent, participants were asked to complete a daily web-based intrusive memory diary for a run-in period of 1 week. Each day, the participants were asked to indicate if they had any intrusive memories and, if so, how many. Participants who met the eligibility criterion of having ≥3 intrusive memories in the run-in week were randomly allocated to either the immediate intervention arm or the delayed intervention arm in a 1:1 overall ratio. Participants were randomized through the P1vital electronic Participant Reported Outcome system (P1vital Products Ltd), and the outcome assessment was completed remotely by the participants (independently of the research team) on this platform. The qualitative team was independent of randomization, delivery of the intervention, and assessment of the quantitative outcomes. The immediate intervention group received the intervention immediately for 4 weeks, whereas the delayed group received usual care for 4 weeks, followed by the 4-week-long intervention.

**Intervention Being Trialed**

The brief digital intervention was delivered through a secure web platform that participants accessed on their smartphone or other internet-enabled device (refer to the studies by Iyadurai et al [25] and Ramineni et al [16] for complete details). Participants were provided with an initial session guided by a clinical psychologist or delegated researcher to run through the intervention as well as follow-ups and support available throughout the intervention. During the initial session (approximately 1 hour), participants were asked to briefly list the different intrusive memories they have and choose the one they wish to target first. They were then asked to complete the intervention, which included several key components: (1) the participant was asked to briefly bring to mind the intrusive image as a reminder to the specific memory, (2) they received instructions on how to play the computer game Tetris using “mental rotation,” and (3) they were asked to play Tetris using mental rotation for at least 20 minutes. During the intervention, participants were asked to rate how distressed they are feeling on 3 occasions, to rate the vividness of the image that is brought to mind, and to rate how much they were able to follow mental rotation instructions, to assess adherence to the instructions. After this first session, participants were able to use the intervention as many times as they liked over the next 4 weeks (eg, to target any other intrusive memories on their list or those
that recur—approximately 20 min/session). The system logged intervention use data, including the number of intrusive memories on the participant’s memory list, number of intrusive memories targeted with the intervention, number of times the intervention was started and used by the participant, and the total time spent playing Tetris.

Data Collection

Intervention Feedback Questionnaire Procedures and Analysis

All participants were sent the Intervention Feedback Questionnaire (IFQ; Multimedia Appendix 2) 4 weeks after the first intervention session to gather information about intervention acceptability. The quantitative data from this questionnaire were analyzed descriptively by an independent and blinded statistician (BG), and the uptake and use data were analyzed by the PIVital data management team. The qualitative research team only received the questionnaire free-text data after completing the qualitative interview analysis. PP categorized any free-text responses that fitted our qualitative themes and subthemes. SB and RM then reviewed these categorizations, and the few very small instances of discrepancies were discussed to create the final structure. These data were analyzed descriptively in terms of frequency of response as a proportion of all participants and also of those who made any response at all. We then examined any data that did not fit any of our themes and considered whether there was enough detail to identify the data as a new theme or subtheme within the overall analysis. If a comment was too vague or general to determine whether it fitted the current thematic or subthematic structure or not, we did not include it in the analysis. As there is a separate trial outcome manuscript that explores the effectiveness of the intervention on a variety of outcomes [16,25], we excluded comments related specifically to the outcomes measured in the trial protocol.

Nested Qualitative Study

Recruitment

Upon completing the 4-week intervention, participants who had previously consented to be contacted for the interview component of the study were asked if they would like to take part in a qualitative interview with a researcher via an audio or a video call. Details of interested participants were stored on a password-protected file, and a researcher used selective sampling to contact participants from diverse backgrounds (age, gender, ethnicity, job role, and location) to schedule an interview.

Interview Schedule

This semistructured interview consisted of several questions designed to gain an in-depth understanding of participants’ experience of using the intervention, including acceptability, improvement suggestions, training or psychoeducation materials, potential barriers or facilitators to recruitment and uptake, and support needed for remote intervention delivery (Multimedia Appendix 3).

Procedures

Before commencing the interview, the researcher confirmed consent to audio record the interviews using a digital voice recorder, and the participants were reminded of the option to withdraw at any point. The interviews lasted approximately 30 minutes, and the audio recordings were immediately transferred to a password-protected laptop and deleted from the voice recorder. The password-protected files were then sent for transcription and anonymized.

Data Analysis

The interview data were analyzed using thematic analysis [32] and drawing from the framework approach by Ritchie et al [33] and Spencer et al [34]. A hybrid approach was used, where themes were generated inductively (from the data) and deductively (based on core areas of interest). This analysis approach was used as the study was exploratory at this stage, so it was important to understand the feasibility and acceptability of the intervention alongside core experiential elements that were not anticipated.

Steps of the analysis included the following:

- Familiarization with data (noting arising concepts and patterns)
- Generating an initial coding framework (iteratively and through team discussion)
- Coding of all transcripts
- Reviewing the content of codes in depth; identifying themes and subthemes; and exploring coherence, variation, consistency, and prevalence
- Creation of mind maps, showing how themes and subthemes fit together and interact, and identifying linkages

In relation to the coding processes, 1 researcher (PP) received interview transcripts, anonymized these transcripts, entered them into NVivo12 (Lumivero), and coded them. A second researcher (SB) coded a sample of transcripts, followed by discussions regarding any discrepancies. Once coding was completed, codes were explored in-depth by PP, who created summaries of coded content to allow the exploration of themes and subthemes. SB repeated the same process on a sample of codes to check for consistency and any discrepancies. PP and SB then worked together to develop and prioritize themes and categorize subthemes according to prevalence.

The analysis initially focused on the barriers and facilitators to using the intervention and how helpful the intervention was for participants, before going on to look at how it could be optimized for future participants and circulated wider.

Results

Overview

In total, 86 participants took part in the GAINS study RCT, 43 (50%) of whom were randomized to the delayed arm and 43 (50%) to the immediate arm [16,25]. Of the 73 participants approached to participate in an interview, 61 (83%) consented to being contacted, 1 (1%) declined, and 11 (15%) did not respond. The mean number of different intrusive memories listed by each participant was 4.88 (SD 2.17), and the mean number of different intrusive memories targeted per participant was 3.64 (SD 2.04). Further intervention use data are presented in Table 1. This shows that participants used the intervention...
an average of 7 times over the 4-week period and spent approximately 20 minutes and 54 seconds playing Tetris per session. They were able to target an average of 73% (3.64 targeted/4.88 total) of the intrusive memories on their list using the intervention.

Table 1. Data on intervention use.

<table>
<thead>
<tr>
<th></th>
<th>Values, median (IQR; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of intrusive memories targeted from list (%)</td>
<td>73 (50-100; 8.33-100)</td>
</tr>
<tr>
<td>Number of times intervention used</td>
<td>7 (5-12.75; 1-44)</td>
</tr>
<tr>
<td>Total time spent playing Tetris per use</td>
<td>20 min 54 s (20 min 22 s-22 min 8 s; 11 min 48 s-31 min 20 s)</td>
</tr>
</tbody>
</table>

The median number of intrusive memories dropped from 14.50 (IQR 10.0-21.50) preintervention to 1.00 (IQR 0.0-3.0) postintervention in the immediate arm group. The median number of intrusive memories dropped from 10.00 (IQR 6.0-17.0) preintervention to 1.00 (IQR 0.0-2.50) postintervention in the delayed arm group. Furthermore, following intervention use, there was a significant reduction in PTSD symptoms ($P<.001$), insomnia ($P<.001$), and anxiety ($P=.02$) and an increase in work functioning ($P<.001$) and well-being ($P<.001$) [16,25].

Of the 86 participants, 84 (98%) completed the IFQ, which contained a mixture of quantitative (scale) and qualitative (free text) response options. The quantitative data will be provided initially, and the qualitative data will be discussed alongside the interview findings. The IFQ respondent demographics are shown in Table 2. Participants’ responses to each quantitative item of the IFQ are presented in Table 3.
Table 2. Demographics of the Intervention Feedback Questionnaire respondents (N=84).

<table>
<thead>
<tr>
<th>Demographic factors</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>69 (82)</td>
</tr>
<tr>
<td>Male</td>
<td>15 (18)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Highest educational qualification</strong></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree or equivalent</td>
<td>47 (56)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>24 (29)</td>
</tr>
<tr>
<td>Doctoral degree</td>
<td>6 (7)</td>
</tr>
<tr>
<td>Sixth form or equivalent (to age 18 y)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Secondary school (up to age 16 y)</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>53 (63)</td>
</tr>
<tr>
<td>Single</td>
<td>26 (31)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Living apart from partner</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Asian (Indian)</td>
<td>5 (6)</td>
</tr>
<tr>
<td>Asian (any other Asian background)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Black (African)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>White (British)</td>
<td>36 (43)</td>
</tr>
<tr>
<td>White (Irish)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>White (any other White background)</td>
<td>23 (27)</td>
</tr>
<tr>
<td>Mixed (any other mixed background)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Other (any other ethnic group)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Unknown (not stated)</td>
<td>12 (14)</td>
</tr>
<tr>
<td><strong>Occupational status</strong></td>
<td></td>
</tr>
<tr>
<td>Working full time</td>
<td>66 (79)</td>
</tr>
<tr>
<td>Working part-time</td>
<td>16 (19)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Sick leave</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>
Table 3. Mean score for each Intervention Feedback Questionnaire (IFQ) item.

<table>
<thead>
<tr>
<th>IFQ item</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFQ0101: How easy did you find it to use the intervention? (0=not at all and 10=very)</td>
<td>8.59 (1.93)</td>
</tr>
<tr>
<td>IFQ0102: How helpful did you find the intervention? (0=not at all and 10=very)</td>
<td>8.23 (2.32)</td>
</tr>
<tr>
<td>IFQ0103: How burdensome did you find the intervention? (0=very and 10=not at all)</td>
<td>6.48 (2.70)</td>
</tr>
<tr>
<td>IFQ0104: How distressing did you find the intervention? (0=very and 10=not at all)</td>
<td>7.07 (2.63)</td>
</tr>
<tr>
<td>IFQ0105: Overall, how acceptable did you find the intervention? (0=not at all and 10=very)</td>
<td>8.50 (1.88)</td>
</tr>
<tr>
<td>IFQ0108: If you were having intrusive memories in the future, how willing would you be to use the intervention if it was offered to you as something that would help? (0=not at all and 10=very)</td>
<td>8.79 (2.11)</td>
</tr>
<tr>
<td>IFQ0109: If a colleague or friend was having intrusive memories, how confident would you be in recommending the intervention to them? (0=not at all and 10=very)</td>
<td>8.43 (2.22)</td>
</tr>
<tr>
<td>IFQ0110: How much do you feel that this intervention could be used within NHS Trusts or health care organizations to support staff who have experienced work-related traumatic events? (0=not at all and 10=very)</td>
<td>8.38 (2.24)</td>
</tr>
<tr>
<td>IFQ0113: Total score (0-80)</td>
<td>64.46 (12.51)</td>
</tr>
</tbody>
</table>

*NHS*: National Health Service.

Table 2 with IQR results shows that the vast majority of participants who took part in the trial were female, working full time, educated to degree level, married or cohabiting, and from a White British or White non-British background. The sample included male, part-time, less educated, single or separated, and ethnic minority staff (from Asian, African, and mixed ethnicity backgrounds). After 4 weeks, most staff reported that the intervention was easy to use, helpful, burdensome, distressing only to a mild degree, and highly acceptable. In the future, they were highly willing to use the intervention again if it was offered to them, highly confident in recommending it to colleagues, and believed strongly that it would help other staff in health care settings who experience trauma.

In the nested qualitative interview study, 16 participants were interviewed. Table 4 shows the demographics of the interview sample. The mean age of the interviewees was 39.4 (SD 8.4) years. Following the first 8 interviews, maximum variance sampling was used to select interviewees from a range of backgrounds. This included a range of ages, genders (predominantly female), ethnicities, job roles (predominantly nurses and consultants), geographical locations, and number of intrusive memories at baseline (range 5-44).
Table 4. Demographics of the interview sample (N=16).

<table>
<thead>
<tr>
<th>Demographic factors</th>
<th>Interview sample, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>0 (0)</td>
</tr>
<tr>
<td>26-35</td>
<td>3 (19)</td>
</tr>
<tr>
<td>36-45</td>
<td>7 (44)</td>
</tr>
<tr>
<td>46-55</td>
<td>4 (25)</td>
</tr>
<tr>
<td>&gt;56</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Number of intrusive memories at baseline</strong></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>1 (6)</td>
</tr>
<tr>
<td>6-10</td>
<td>2 (13)</td>
</tr>
<tr>
<td>11-15</td>
<td>3 (19)</td>
</tr>
<tr>
<td>16-20</td>
<td>4 (25)</td>
</tr>
<tr>
<td>21-25</td>
<td>4 (25)</td>
</tr>
<tr>
<td>26-30</td>
<td>1 (6)</td>
</tr>
<tr>
<td>31-35</td>
<td>0 (0)</td>
</tr>
<tr>
<td>36-40</td>
<td>0 (0)</td>
</tr>
<tr>
<td>41-45</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11 (69)</td>
</tr>
<tr>
<td>Male</td>
<td>4 (25)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>African</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Filipino</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Indian</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Spanish Colombian</td>
<td>1 (6)</td>
</tr>
<tr>
<td>White</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Mixed race</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Job roles</strong></td>
<td></td>
</tr>
<tr>
<td>Nursing</td>
<td>9 (56)</td>
</tr>
<tr>
<td>Consultant</td>
<td>3 (19)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Anesthetist</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Physician</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Dietitian</td>
<td>1 (6)</td>
</tr>
</tbody>
</table>

The final 4 (25%) of the 16 interview participants who participated in the study received an optimized intervention, as this was a Bayesian adaptive RCT. The intervention was optimized as the study progressed over time [16,25] following feedback from previous participants. These optimizations included adding graphs to allow participants to see their own data for each intrusive memory, adding a video at the end of the first guided session to reinforce how to use the intervention independently, and adding an additional reminder cue in the first guided session to ensure that the memory was in their mind just before they played Tetris [16]. The topic guide (Multimedia Appendix 3) was, therefore, amended to include questions about participants’ thoughts on what was required to access the
intervention independently as well as their thoughts on the added graphs and reminders.

The themes and subthemes are outlined in Textbox 1, and example quotes illustrating each theme are provided in the text below.

Textbox 1. Thematic structure.

<table>
<thead>
<tr>
<th>Themes and subthemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitudinal and emotional responses</td>
</tr>
<tr>
<td>Experiencing mental health symptoms as a health care professional</td>
</tr>
<tr>
<td>Stigma and imposterhood</td>
</tr>
<tr>
<td>Value of anonymity</td>
</tr>
<tr>
<td>Using a novel intervention</td>
</tr>
<tr>
<td>Skepticism</td>
</tr>
<tr>
<td>Understanding how it works is helpful</td>
</tr>
<tr>
<td>Positives of the intervention</td>
</tr>
<tr>
<td>Tracking and intervening to reduce intrusive memory frequency and intensity</td>
</tr>
<tr>
<td>Intervention is easy to use and enjoyable</td>
</tr>
<tr>
<td>Intervention is more convenient than psychological therapies</td>
</tr>
<tr>
<td>No need to discuss intrusive memories</td>
</tr>
<tr>
<td>No side effects of the intervention</td>
</tr>
<tr>
<td>Cognitive and emotional coping</td>
</tr>
<tr>
<td>Negatives of the intervention</td>
</tr>
<tr>
<td>No opportunity to discuss intrusive memories in detail</td>
</tr>
<tr>
<td>Unclear whether memories are spontaneous</td>
</tr>
<tr>
<td>Technological issues</td>
</tr>
<tr>
<td>Difficult to find time to use the intervention</td>
</tr>
<tr>
<td>Improvements and optimizations</td>
</tr>
<tr>
<td>Difficult to focus on mental rotation</td>
</tr>
<tr>
<td>How to increase focus on mental rotation</td>
</tr>
<tr>
<td>Researcher support is important</td>
</tr>
<tr>
<td>How to access the intervention independently</td>
</tr>
<tr>
<td>How to aid incorporation of the intervention into participant lifestyle</td>
</tr>
<tr>
<td>Other intervention improvement suggestions</td>
</tr>
</tbody>
</table>

Attitudinal and Emotional Responses

Experiencing Mental Health Symptoms as a Health Care Professional

Stigma and Imposterhood

Choosing whether to participate in the intervention was compounded by barriers, such as stigma surrounding mental health in ICU staff. Many participants described feeling weak for seeking help and not wanting colleagues to know that they were struggling:

I’m a healthcare professional and I do see these things all the time, but then in one way you think these are the things that will never affect you. You help the others, because I’ve always been caring and empathetic with them, but at the same time you don’t think you can be the patient. [001]

But critical care nursing has this kind of almost elitism kind of approach. And the moment you have a little bit of a weakness that’s then seen negatively and your almost- you kind of feel that someone’s going to think that you can’t do your job. [008]
sense of imposter-hood with these. Do I really- Or do I- Are my intrusive memories bad enough? Surely other people have worse ones, therefore- It was helpful getting rid of that feeling. [003]

Value of Anonymity

Overview

In addition, participants stressed how emphasizing the anonymity of this intervention could help reduce the impact of mental health stigma on taking part in the study:

Yes I think highlighting it being anonymous and it doesn’t get reported back to work would probably make people more likely to use it. And the fact that they can – you know if they can access it on their personal emails and not work emails and things like that that would probably make people more likely to use it. [013]

IFQ Data

Despite this, when asked how the intervention could be improved in the feedback questionnaire, 3 participants provided suggestions that required workplace involvement in the intervention:

- Involve managers to support the staff. [IFQ004]
- Time out of work to do it. [IFQ049]
- For it to [be an] option within occupational health because it has impacted on NHS workers that went through the pandemic. [IFQ040]

Using a Novel Intervention

Skepticism

Overview

Participants discussed their initial skepticism when approached with information about this intervention, as its novelty, simplicity, and game design caused them to doubt its effectiveness:

Well when I first heard about it I was very, very sceptical, because yes I know that games focus your mind on something else. But I just wasn’t convinced that it was going to do anything. [005]

IFQ Data

Skepticism toward the intervention was also mentioned by a couple of participants in the feedback questionnaire when asked what they found useful about the intervention:

- Amazing. I genuinely did not expect it to work. [IFQ012]
- The intervention definitely targeted some of those intrusive memories popping up, working much faster than I had expected it to. [IFQ009]

Understanding How It Works Is Helpful

Overview

Therefore, participants described how increasing their understanding about how the intervention works during initial communication is important to reduce the initial skepticism about intervention efficacy:

[Once] you get over people being aware of what it’s trying to do. You know, it’s not, for example, it’s not just a distraction, it’s not trying to distract you from memories and thoughts, but this is- Try this because there is evidence as to how this will work. [010]

People might be a little bit sceptical that it would work. Well especially with nursing... as part of your training you do modules on research so I think if people are shown the research and shown that it works they’re more likely to use it. [013]

Participants suggested providing previous research and testimonials from other ICU staff during initial communication to help normalize intrusive memories and thereby encourage more professionals to seek help:

[Should] have said, look, I also have the same problems, have a personal story like that, and see somebody’s journey. Saying like, you stick with a little work and I think it reassures a little bit and I think it relates a little bit nicer to you mentally than just a- you know a scientific paper which is very objective and impartial. [011]

IFQ Data

This was further highlighted by a couple of participants in the feedback questionnaires, when asked how the intervention could be improved as something that could be offered to health care staff to help reduce intrusive memories after a work-related traumatic event:

- Provide evidence and testimonies from the research participants. [IFQ018]
- To make people aware of it and show success results. [IFQ006]

Positives of the Intervention

Participants discussed various positive effects of the intervention, including its effect on them, and also when compared with other treatments (which was a line of questioning).

Tracking and Intervening to Reduce Intrusive Memory Frequency and Intensity

Overview

Participants described how the intervention was helpful in reducing the frequency or intensity of intrusive memories. Moreover, they found it helpful to track their intrusive memories, as it allowed them to notice reductions and patterns and reinforced intervention use:

It was like being back inside the situation again. I could hear the voices...I was surrounded by the scenario again. And then it became just images of like trying to foreseeing those things again. And now it is just a cloud. There is nothing there. [015]

Because then you can see the trend and then you think oh, actually I haven’t had any memories for three days
or whatever. So, that kind of like spurs you on a bit and then you’re like oh, it’s working, yes that was useful. [006]

Recording the frequency of the intrusive memories and realising that they’re actually going down, it just reinforces that it is a good tool to use for that purpose. [002]

**IFQ Data**

In total, 30% (25/84) of the feedback questionnaire respondents also described how the intervention helped to reduce the frequency and intensity of intrusive memories as well as how some of these memories did not return:

That it actually helped reduce the frequency of the memories. [IFQ011]

The intervention definitely targeted some of those intrusive memories popping up...causing the memories to become less frequent as well as less distressing and vivid when they did appear. [IFQ009]

It helped to reduce my intrusive memories. It gave me something active and engaging to do if I felt distressed by my memories. [IFQ035]

**Intervention Is Easy to Use and Enjoyable**

**Overview**

Participants found the intervention easy to use, in part because of the clear instructions provided, and enjoyable:

I think it was all really straightforward. And for someone of my age, that’s- It must be easy, because I’m of the non-tech generation. [016]

You know, it is a psychological intervention and it is helping you psychologically [laughs]. But it’s also, it is a game and it, there are some enjoyable aspects to it and it’s a good way to help yourself mentally. [006]

**IFQ Data**

These positives were further emphasized in the feedback questionnaire responses, with a couple of respondents mentioning that the intervention was enjoyable and many more saying it was simple, easy to use, and intuitive:

It was enjoyable and very intuitive. [IFQ004]

It’s such a simple tool to use to target these frustrating and upsetting thoughts. [IFQ001 ]

It was straightforward and easy to use. [IFQ002]

However, 1 participant also explained in the feedback questionnaire that their lack of experience with Tetris made it difficult to use the intervention:

Learning to play the game as I had not used before. [IFQ034]

**Intervention Is More Convenient Than Psychological Therapies**

Participants also discussed the benefits of this treatment compared with existing treatment options for intrusive memories, such as psychological therapies. For example, unlike psychological treatment, this treatment is flexible as it does not require an appointment:

You can choose when to target those specific kind of memories rather than being reliant on, well I’ve got an appointment at 2 o’clock on this day. And there were days when I couldn’t- I didn’t have the mental capacity to just focus on kind of day to day let alone do the intervention. So I could pick and choose when I targeted those memories. (008)

It doesn’t take too much time out of the day and the emotional commitment involved is not exhausting, like some other forms of counselling I’ve had before. In fact, I feel better for it, rather than worn out. So, 20 minutes and then you can get on with things really. [003]

It’s because it is more available whenever time you want. That’s the most important thing. [007]

**No Need to Discuss Intrusive Memories**

**Overview**

Some participants also appreciated that this intervention did not require them to relive distressing memories in-depth, unlike in talking therapies:

The fact that you’re actually having a therapy session, but you don’t realise it. So, it can be less distressing as well, especially when you’ve got trauma, some people don’t really want to talk about trauma, they just want trauma to go away. [001]

Because sometimes if you have counselling, you leave the room even worse than when you came in, because you are not thinking about something and then you have to go back to the situation and seeing and then you spend the whole day crying- I’ve had counselling before and after so much crying, you are so exhausted the rest of the day after all that. [015]

**IFQ Data**

There were mixed views on this in the feedback questionnaire responses, with 1 respondent stating that the intervention was nondistressing and 3 respondents explaining that it was difficult to recall the memories as part of the intervention:

Pretty well non-distressing. [IFQ021]

Didn’t want to bring some of the more difficult intrusive memories to mind. [IFQ044 ]

Re living those memories was really difficult sometimes. [IFQ028]

The only difficult thing was facing the intrusive thoughts. [IFQ025]

**No Side Effects of the Intervention**

When compared with medication, the key advantages discussed were the lack of side effects, for example, on sleep or weight, and being able to directly target intrusive memories:

If you’ve got a problem, you still have the problem. Maybe the medication helps you to stay more relaxed,
but it doesn’t really have an impact on intrusive memories, and things like that. [001]
I have my own reservations when it comes to medication...I don’t want to take much of a time, you know in terms of (the drug affecting) sleep or waking or weight loss, things like that. Compared to medication, I was happy to undergo an intervention like this. [009]

In addition to discussing some of the positive effects of the intervention compared with other treatments, I participant also mentioned simply appreciating the availability of an alternative option:

Personally because I’ve had talk therapy and medication and things so, I was looking for something else to try. [006]

Cognitive and Emotional Coping
Overview

The participants described the positive effects of the intervention on their cognition, such as improved concentration and rational thinking at work. They also experienced positive changes in their emotions and other aspects of life, including feeling happier, more in control of their emotions, reduced anxiety, and sleep improvements:

That allowed your brain to then focus more on the important things. And my rational thinking at work seemed to improve as a consequence. [008]

My attention and focusing time came back. Usually, you need focal attention. I can pay attention for extended hours, but I was not able to do that. [009]

My productivity increased to the extent that people started noticing it and I became happier. I think I became my old self. [009]

I was so tearful but now, if I’m tired, I feel more comfortable, more happy, more emotionally controlled. [007]

Obviously, the sleep improvements come with not having the intrusive memory and being troubled by them. [002]

IFQ Data

The feedback questionnaires offered insights into additional positive effects of the intervention, including how it enhanced focus, served as a helpful distraction, aided in relaxation, and allowed individuals to take time out of their day.

Facilitated focus and clarification of thoughts. Organised my mind. [IFQ015]

The game was a great distraction. [IFQ022]

Intervention itself was relaxing, forcing 20 minutes of exclusive concentration on a task providing a break from a busy day. [IFQ032]

Negatives of the Intervention

Participants mentioned some negative effects related to the use of the intervention, some of which were grounded in comparison with other treatments (which was a line of questioning) and some of which were given spontaneously. There is heterogeneity of experience, and these were not discussed by everyone, and these negative effects were grouped as follows.

No Opportunity to Discuss Intrusive Memories in Detail

Overview

Participants discussed some negative effects of this intervention compared with existing treatments. For example, some participants preferred to talk to a clinician about their intrusive memories, which this intervention did not allow:

[If] you are really troubled by one of the memories, I can imagine having direct psychological support to kind of work your way through that thought process, it is just a thought, calming you down, there’s none of that. [002]

IFQ Data

One participant also highlighted in the feedback questionnaire that the inability to discuss intrusive memories in detail meant that the intervention should be accompanied by another support system:

Great as a distraction technique but needs to be accompanied by another support mechanism, being given a safe space to be able to talk through experiences etc. [IFQ017]

Unclear Whether Memories Are Spontaneous

Overview

In addition, participants expressed difficulty in knowing whether memories of the traumatic event were actually spontaneous or simply because of being part of the study:

I think it’s really hard to try and differentiate between, am I putting this thought in my head? Because I know I am doing the GAINS Study, or- So I think like, after the first couple of days, I just had to just try and, like, ignore the research side of it. (004)

IFQ Data

Feedback questionnaire respondents concurred with this; with respondents explaining how processes inherent to the study, such as repeat contacts and reminder texts, could in fact remind them more about the distressing incidents:

Repeat contacts became a bit burdensome. [IFQ003]

The phone calls, texts were far more frequent than I anticipated. The reminder to do the intervention at 7 am and at night reminded me of the distressing incidents and made me more distressed. [IFQ019]

Technological Issues

Overview

Once participants had overcome the barriers to participating in the study (described previously), there were some environmental factors that affected their experience with the intervention. As this is a digital intervention, technological difficulties such as a lack of technological knowledge, reduced access to a device, and device differences were mentioned as issues:
The one thing I did find though, which I didn’t know, is when I occasionally had, I usually used my laptop if I was able, and you know, it was much easier to rotate the blocks. But I found on my mobile phone it was, it was very sensitive. [005]

Some of us are lucky to have laptops, but like, I remember at the initial start, we were told to make sure you have your laptop, because it might not work, like, on your phone or your iPad. [012]

There might be some groups that have potentially worked in different areas in part of the pandemic and they’ll be left with some of these intrusive memories that wouldn’t be so technologically fluent. [002]

IFQ Data
This was echoed by the feedback questionnaire responses, with 1 participant explaining how there were problems owing to the multiple platforms and the inability to access the intervention offline. However, there were mixed views on how well the intervention worked on mobile phones:

Multiple platforms, not an intuitive website...and it would time out if you were on a train/went offline. [IFQ002]

Didn’t work too well on my phone. [IFQ016]

I would not improve the intervention itself as it is very accessible and able to play on phone and computer. [IFQ037]

Difficult to Find Time to Use the Intervention
Overview
In addition to technological difficulties, participants described how it could be difficult to find time to use the intervention because of their busy schedules, shift patterns, and other responsibilities outside of work:

[Found] it a bit difficult to find the time with this- I’ve got a small child, I’m working almost close to full-time. My husband is a [profession name], so he does a lot of out of hours work, so a lot of the childcare and toing and froing, that comes to me. [002]

[Occasionally] I had to do it just before I went to bed after a busy day. And I didn’t like doing that because I try and reduce my mobile phone usage or computer usage late at night. [005]

Particularly about doing it during the day at work, made you think, well that’s fine, I’ve managed to push things, cleared everything up. I’ll be fine for half an hour while I do this, and then inevitably you get a phone call 10 minutes into it. [010]

I suppose compared to medication or whatever, I suppose it’s investing the time, finding the time. Taking a pill is very quick, isn’t it. [002]

IFQ Data
This was also a very common theme in the feedback questionnaire responses, with participants explaining how and why they found it difficult to find time to use the intervention (both inside and outside the workplace):

Finding the best time to do the intervention without interruption from my preschooler Vs being too tired to pay it proper attention to get the most out of it. [IFQ001]

Surprisingly difficult to find 20 mins in busy clinical day or family time daily to do game. [IFQ003]

My concern is that staff wouldn’t find the time during the working day to work on the interventions and with shift patterns may struggle to include it in their free time. [IFQ036]

Three of the feedback questionnaire respondents also highlighted how they found it difficult to remember to record and use the intervention, in part because of the lack of time:

Remembering to do the intervention regularly and record the relevant information. During the day, I don’t have time to work on the intervention so I’ve had to work time into my evening to complete the tasks. [IFQ036]

Improvements and Optimizations
Participants were asked about potential improvements or optimizations to the intervention as well as any challenges experienced while using it. The responses were grouped into the following themes, which were found to be relatively frequent and consistent.

Difficult to Focus on Mental Rotation
Overview
Participants emphasized the importance of recognizing that the mental rotation aspect was the focus rather than the Tetris score:

I had to concentrate quite a bit because when I first started using it I just was having fun playing the game [laughs]. [006]

I think we should highlight that when you do the intervention you shouldn’t really. You should focus more on the arrangement rather than aiming for a score or you should stay away from how you usually play Tetris and focus on the future blocks and the arrangement. [014]

Despite recognizing this, participants often found it difficult to concentrate or focus on mental rotation:

I was thinking about myself, my thoughts were just running away, and I wasn’t really concentrating on the Tetris game. [001]

I remember thinking first few times it [speed] increased I started to panic a bit because- And then you do stop focusing on the mental rotation because you just think, oh they’re all coming so quickly. [005]

I think it was the pre-empting the rotation that’s coming, to rotate it in your head before it actually comes onto the screen, that’s the hard bit. [012]
IFQ Data
This difficulty in concentrating was echoed by participants in the feedback questionnaires:

I think 15 mins is enough—my concentration starts to flag after this. [IFQ005]
I found I was bored quickly. [IFQ038]

How to Increase Focus on Mental Rotation
Overview
Participants provided suggestions on how to improve the intervention for future participants. One of the main suggestions was a pop-up message during gameplay to remind participants to focus on mental rotation, as participants had previously mentioned that their focus drifted from mental rotation. Participants also mentioned that the speed at which the blocks fell made it difficult to focus on mental rotation, and they suggested that capping the speed may be helpful:

I don’t know if during the Tetris game some kind of pop up could be coming up, like a reminder. Yes? A kind of- I don’t know what you say on the reminder, but something like- I don’t know, remember to focus on the next pieces. [001]
And maybe cap the speed. When it gets ridiculously fast I think you’re very aware that you’re not doing mental rotations very well. [002]

IFQ Data
Another participant mentioned capping the speed as an improvement suggestion in the feedback questionnaire:

Cap the speed of the blocks to enable proper rotation planning. [IFQ001]

Researcher Support Is Important
Overview
Researchers provided participants with support during the initial session, in which they showed the participants how to identify intrusive memory images and how to use the intervention with mental rotations very well.

I like the fact that I got contacted midway through, because it was obvious that that one was still bothering me more. [005]
Yes, I think so. I think I had a couple of teething issues, but one of the ladies phoned me up, went through it, face-to-face, got me to play it whilst on the phone with me, so it was fine. [004]

IFQ Data
The importance of the initial session with a researcher was also highlighted by a participant in the feedback questionnaire:

Needs 1-1 at beginning when 1st doing intervention so person using it knows exactly what to do. [IFQ045]

How to Access the Intervention Independently
All participants who received the optimized intervention provided suggestions on how to make it more easily accessible independently. For example, they proposed incorporating video demonstrations showing how to autonomously identify their intrusive memory images and playing the intervention while focusing on mental rotation:

Yes, so see someone playing. And what obviously, because you cannot read their mind, but if you put like a bubble say what is doing in their mind. They are turning, they are not focusing on this, they are doing this, this. [015]
Maybe- I guess, maybe giving some examples of how to break those things [intrusive memory images] down, if that makes sense. And the different ones that are just like an image, or something that’s like a little video that plays, that sort of explanation and how to write that, maybe just some sort of video of how to break that down into something. [016]

How to Aid Incorporation of the Intervention Into Participant Lifestyle: IFQ Data
In the feedback questionnaires, a few participants explained how shorter Tetris sessions could help them incorporate the intervention into their daily lives more easily:

Perhaps if the timing the intervention had to be carried out for was shorter - I'm not sure if it would still be effective but feel 10 minutes twice a day rather than one block of 20 minutes would be easier to fit in. [IFQ009]

Is the 20 minutes a specific time or could it be reduced? Sometimes difficult in a busy working day to get full 20 minutes to spend on it. What effect with say 10 minutes? Or 5...? [IFQ021]
I was sometimes deterred from starting the intervention knowing that it would take up 20mins of time. It may seem more accessible if it only required 10 mins for example. It would be possible to do whilst on break at work etc. [IFQ023]
Other Intervention Improvement Suggestions

Overview
In addition, participants expressed a desire for the intervention to be available as an app rather than solely on the web-based platform:

*I think it would have been really good if you could have had an app. And then every time you have a memory, you just tap the app, or something, and then it. Yes, and otherwise you have to, like- By the time you get home, and log it, you’re like, how many did I have? [004]*

*I think it will help if it’s an App that you can download, like Calm or what’s the other one- Then you can just go and open and do for 15, 20 minutes. [014]*

There were mixed views on the use of graphs to display intrusive memory changes, with some participants not liking graphs, whereas others appreciated the ability to track progress in this way:

*I’m not a great fan of graphs to be honest. [013]*

*And not only to see progress but to see- I guess because I could match as well, like they definitely got worse when I was on nights and I could see that, with my shift pattern and also stuff I was doing at work, just certain things that were going on and then I’d be like, I can see how that happened and where that connection is. So, it kind of makes sense to me. [016]*

In general, participants appreciated receiving brief daily reminders to log their intrusive memories and engage with the intervention, as long as the reminders were not excessive and did not prompt participants who had already completed the task. However, as mentioned in a previous subtheme, some participants did find the reminders distressing as they brought the memories of traumatic incidents to mind:

*A quick reminder with may be just a shortcut to the logon is fine for me. [013]*

*[At] least I had that reminder. And I think if I forgot one day, let’s say today, tomorrow in the morning I could go back and do it...Especially when you have so many things on your mind. [015]*

IFQ Data
In addition to requesting that the intervention be available in an app format, a couple of participants in the feedback questionnaire suggested that the intervention use only 1 platform instead of 2 (1 for the intervention and 1 for logging outcome measures):

*I found it difficult to navigate the website, perhaps an app would have been better. [IFQ022]*

*It might be easier to make it all on one platform. Rather than 2 different places. [IFQ028]*

Although some participants mentioned how they appreciated the text message reminders, 1 participant explained in the feedback questionnaires how reminder messages could be confusing:

*I would suggest keeping up the automatic reminders to record and use the intervention. [IFQ025]*

*Infrequent but relevant text message feedback was useful and not overly intrusive. [IFQ032]*

*I get multiple reminders about completing tasks which sometimes confuses me whether I have completed it or not. [IFQ053]*

Discussion
Principal Findings
This qualitative study explored barriers and facilitators to the adoption of a brief digital imagery-competing task intervention (1 guided intervention session of 1 hour, followed by self-guided use of approximately 20 min/session) to reduce intrusive memories of traumatic events from working in an NHS ICU during the COVID-19 pandemic. Overall, on the acceptability questionnaire, the health care staff found the intervention easy to use, helpful, and highly acceptable. They were highly willing to use the intervention and were confident in recommending it to colleagues and their health care organizations for staff exposed to repeated trauma. In the qualitative data collection, participants described many additional positives of the intervention, such as it being easy to use, enjoyable, and encouraging, as participants were able to track intrusive memories and notice reductions in frequency. They could modify the use of the intervention based on the intrusiveness and frequency of the traumatic memories. Compared with sessions of psychological treatment, it was considered less time consuming, more flexible when it could be used, did not require discussing unpleasant memories, and required less effort. Compared with medication, it was more specific in its effect on intrusive memories of traumatic events and did not have adverse effects on weight, sleep, or alertness. It was seen as complementary to psychological and medication treatments in those who needed them.

Although it has its advantages, participants described how the intervention may not entirely replace the need for psychological therapy to talk about the nature of intrusive memories in those who wish to or the need for medication in some instances. A key finding was that some participants preferred not to access the intervention through their workplace or for colleagues to know that they were using the intervention owing to mental health stigma, a factor that is known to affect mental health help seeking among health care professionals [35], including after witnessing trauma in the workplace [36]. An advantage of the GAINS intervention is that although it could be provided through the workplace and introduced as a normal working practice for staff in the ICU, it could also be accessed independently outside of work.

The intervention use data showed that, on average, staff engaged with the Tetris component for the full 20 minutes per session, approximately 7 times over the 4-week period. They were able to target approximately 73% (3.64/4.88) of their intrusive memories through the intervention, that is, on average, participants were able to target 3.64 intrusive memories and had 4.88 intrusive memories listed. This emphasizes that the intervention was extensively used, indicating its significant...
value. When combined with qualitative findings, it appears feasible and acceptable for staff, particularly in the short term. However, there is a need to further investigate how participants use the intervention for a longer term, particularly whether it can easily fit into their daily lives.

Comparison With Prior Work

The findings highlighted barriers that ICU staff experience when accessing support for their mental health, such as stigma, feeling weak for seeking help [37], questioning if they were bad enough to warrant such help, and not wanting colleagues to know that they were struggling. This is consistent with previous findings investigating mental health in health care professionals [35,36] and a culture of not showing weakness in health care work settings [29]. Participants suggested that these barriers could be partially overcome by normalizing intrusive memories after trauma through testimonials from other ICU staff who participated in the GAINS study. In addition, as discussed in the existing literature [35,38], the anonymity of the intervention was important, as it was completely separated from the health care professionals’ workplace or colleagues. This suggests that staff should have the option to access the intervention through routes other than only the workplace. However, participants in the IFQ suggested that health organizations would benefit from the intervention being endorsed by senior staff members. This endorsement could occur during induction and appraisal meetings involving junior colleagues, especially in environments where staff are repeatedly exposed to trauma. If staff did find it acceptable for the intervention to be used in their work environment, it could even be incorporated into staff induction and colleagues could support another through a “buddy system.”

In a previous meta-synthesis of digital health interventions for mental health [39], one of the key barriers to the initial approach was skepticism about how helpful a remote treatment could really be. This was also the case with the GAINS intervention. The initial skepticism was compounded by it being a simple and novel gameplay intervention, with some participants expecting the intervention to be at best a short-term distraction while they played the game. In fact, many participants went on to report long-lasting effects on the frequency and intrusiveness of their traumatic memories. Publicizing research evidence, discussing the mechanism of action of the intervention, and testimonials from ICU staff were suggested as counters to this possible skepticism. This is consistent with the literature, which highlights the importance of users being on board with digital health interventions’ aims and understanding their purpose [40].

The suggestion to publicize research evidence and provide testimonials from ICU staff is of particular importance, as prior findings emphasize that endorsement from health care professionals is valuable and helps digital health interventions’ aims and understanding their purpose [40]. It highlights the importance of users being on board with digital health interventions and suggests that endorsement could occur during induction and appraisal meetings involving junior colleagues, especially in environments where staff are repeatedly exposed to trauma. If staff did find it acceptable for the intervention to be used in their work environment, it could even be incorporated into staff induction and colleagues could support another through a “buddy system.”

Similar to prior findings [41] that integrating a human component into treatment helps retain engagement and reduce dropout, participants reported that researcher support, both before using the intervention (eg, the initial guided session) and throughout intervention use (eg, booster session), was found to be extremely helpful and important. They valued the continuous support provided to ensure they correctly used the intervention and received guidance on making adjustments when necessary. However, providing this level of support can be difficult when scaling up an intervention [41]. Participants provided suggestions on how the intervention could be more easily accessed independently, which would require fewer therapist and researcher resources and enable the intervention to spread more widely and reach a greater number of ICU staff, for example, by providing video demonstrations of someone identifying the intrusive memory images independently and playing the intervention while focusing on mental rotation. Participants’ suggestions around helping to retain a focus on mental rotation while playing Tetris are helpful to identify, as this may be one of the core aspects to the working of the intervention. They also discussed changing aspects of the game itself and whether more frequent but shorter use of the game might be effective and more feasible for staff to continue using it for a longer term.

Limitations

Limitations of the study include the method of recruitment and sample representativeness and the short duration of use of a novel intervention. The trial was a first trial that is being followed by further trial work to test the robustness of the findings of the first trial. Recruitment of participants through advertising in the ICS, a professional organization, may have recruited participants who would be the most receptive and enthusiastic for such interventions. The sample was geographically drawn from many parts of the United Kingdom and was representative of the NHS at large but appears to have been overrepresentative of ethnic minority staff. There were 44% (7/16) staff from ethnic minority backgrounds in our interview sample compared with 20.7% (248,400/1,200,000) in the NHS workforce [42]. However, selective sampling was used for interview recruitment to capture a broad range of experiences of using the intervention from as diverse a sample as possible, rather than to match the sample to demographic characteristics of the NHS population. Any novel intervention, both in format and purpose, may have a large halo effect in
relation to enthusiasm to take it up and use that may not be sustained over time. There is a need to recruit larger representative samples that use this intervention. The aim of this intervention is for most individuals to only need to use it a few times (once for each distinct intrusive memory of trauma). It is designed to be brief each time it is used, and requiring only a few sessions, rather than for prolonged use like a mindfulness app. However, for people with a very large number of intrusive memories and repeated ongoing traumatic events, it would be useful to consider use over a number of months to obtain more robust data on its likely uptake, use, and acceptability, which could be generalized to the staff experiencing repeated trauma in routine health care settings. Furthermore, we were unable to interview any ICU staff who were not able or chose not to participate in the study. Therefore, it is difficult for us to know how typical our sample and findings are of the wider ICU staff population. We also could not obtain important information about any barriers to participating in these individuals and how these barriers could be overcome, as it is likely that our participants were more open to mental health support in general. Furthermore, participants who took part in the trial but did not consent to participating in an interview may have had a different (more negative) experience of the intervention, and we could not obtain information about their experiences.

Nonetheless, our participants described potential barriers to wider participation, and it appears that the intervention was able to overcome some of these to an extent, such as the anonymity of the intervention helping to reduce the impact of stigma. In addition, we used selective sampling to ensure that our sample was as diverse as possible on factors such as profession, background, ethnicity, age, NHS Trust, and baseline intrusive memory frequencies. Therefore, our sample aimed to be as inclusive as possible of our target population. Saturation of themes was reached through the interview and analysis process, suggesting that further barriers would be unlikely to be present if we had recruited more participants.

We attempted to triangulate our data by comparing feedback from the sample that completed the acceptability questionnaire (IFQ) with our qualitative interview data. Certain topics did not lend themselves to completion on the feedback questionnaire, such as discussion of stigma. However, this was often reported in the qualitative interviews with ICU staff, and the themes resonate with previous literature. The feedback questionnaire, which had a very high rate of completion, confirmed most other barriers and facilitators, identifying a new subtheme around highlighting a number of ways staff improved cognitive and emotional coping with trauma through the intervention. A limitation of our analysis is the inability to delve deeper into certain findings. For example, we could not explore how the intervention might induce relaxation nor whether the distraction and improved focus persisted beyond the gameplay or were solely experienced during the game sessions. Bringing awareness to the intrusive memories could be both a positive and a negative experience, as it might help identify a source of stress; however, some people cope with intrusive memories by suppressing them, whereas others believe it adds to distress. The feedback questionnaire provided many additional suggestions to improve the uptake, feasibility, and acceptability of the intervention that the research team and developers of the intervention could explore and consider. A key strength of the qualitative interviews was the chance to iterate findings to adapt the intervention accordingly while it was still being used in the RCT. We also interviewed some participants who received the optimized intervention to gain feedback about their experience with the optimizations so that we could further improve the intervention.

Data gathered from the use of a survey in addition to interviews demonstrate that even when interviews are repeatedly producing the same subthemes and themes, and despite maximum variance sampling on the basis of characteristics available to us, there might still be important themes that might be missed because of the limits on maximum variance sampling imposed by data protection and trial procedures. We could only be made aware of a limited amount of information without fully consenting individuals for the interview. However, the qualitative interview method delves deeper into extracting information that participants might not readily provide in a feedback survey. In addition, it is an iterative process that builds upon multiple interviews. Therefore, if a theme or subtheme is not supported in the survey feedback, it does not mean that it is unimportant or even uncommon, simply not as immediately obvious to the participants.

Future Research

Other potential issues for us to consider in the next phase are related to the practicalities of the intervention, as participants mentioned that the intervention could be difficult to fit into their extremely busy working lives during the pandemic. Health care demands have remained high; therefore, this ability to fit in may be a continued factor to consider. There was also the issue of lack of privacy when accessing the intervention at work and not having access to a personal device in this setting. As highlighted in a previous literature review [43], most intervention frameworks recognize the importance of understanding how well an intervention fits with existing organizational routines to predict its adoption and implementation on a larger scale. As the intervention can be accessed by ICU staff either at work or outside of work, we must also understand how well the intervention fits into their personal lives. Although this aspect was discussed in our findings, it is crucial to further explore the feasibility of long-term intervention use, especially considering that ICU staff regularly encounter work-related trauma. For the intervention to be beneficial, it must integrate as seamlessly as possible into their lives. An advantage of this intervention is that it can be used at any convenient time (eg, at home or on a commute). A further key issue going forward within work, and especially outside work, is data protection of sensitive information that may require training or other safeguards, for example, if staff members are overlooked while examining graphical outputs (eg, a line graph) of the frequency of traumatic memories. We may need to provide alternative ways of presenting the data to ensure that they are more widely accessible, such as through color chart indicators rather than numerical graphs.
Conclusions
Overall, the data suggest that the intervention to reduce intrusive memories after trauma is highly acceptable to ICU staff and has some unique value compared with other current approaches to staff mental well-being. Through additional refinement and gathering evidence regarding outcomes and implementation, this intervention could potentially present a much-needed approach to address the widespread issue of repeated exposure to trauma, which manifesting as intrusive memories significantly impacts on the mental health and emotional well-being of health care staff.

Acknowledgments
Tetris has been licensed for use within i-spero from The Tetris Company at no cost. The authors would like to thank the Intensive Care Society, in particular Dr Sandy Mather and Alex Day; our Data Monitoring Committee members including professor Andreas Reif, professor Steve Hollon, and professor Ian Penton-Voak; our Trial Steering Committee including Guy Goodwin, Pooyan Bebahani and Rebecca Dias; expert advisors including professor Charlotte Summers, professor Thomas Jaki, Dr Nick Grey, professor Sir Simon Wessely, and professor Jonathon Bisson; statistician Varsha Ramineni and statistical adviser professor Michael Bonsall; members of the study team including Veronika Kubičková, Katie Hurman, Marie Kanstrup, Alfred Markham, and Zunaid Islam for support; and our IT team and data management team at P1vital Products Ltd. This study was funded by The Wellcome Trust (223016/Z/21/Z). The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript. P1vital Products Ltd was the trial sponsor. EAH also received funding from the Swedish Research Council (2020–00873), OAK foundation (OCAY-18-442), AFA Försäkring (200342), The Lupina Foundation, and Rannís—The Icelandic Research Fund. The qualitative study was supported by staff from the National Institute for Health and Care Research (NIHR) Applied Research Collaboration East Midlands, the NIHR Nottingham Biomedical Research Centre Mental Health and Technology Theme, and the NIHR MindTech Medtech Co-operative.

Authors' Contributions
PP carried out the interviews, collected and formally analyzed the data, and drafted, reviewed and edited the paper as first author. SB and RM contributed to methodology, formal analysis, writing, review, and editing. BG contributed to methodology and formal analysis. EAH contributed to conceptualization, funding acquisition, project administration, methodology, review, and editing. LI and JK contributed to conceptualization, project administration, methodology, review, and editing. JH contributed to review and editing.

Conflicts of Interest
JK is a shareholder and director of P1vital Products Ltd, which is the study sponsor and manufacturer of i-spero and the P1vital electronic Participant Reported Outcome system. LI is employed by P1vital Products Ltd. EAH receives funding from The Wellcome Trust (223016/Z/21/Z), the Swedish Research Council (2020–00873), AFA Försäkring (200342), and Rannís—The Icelandic Research Fund. EAH’s salary is partly funded by Wellcome Trust (223016/Z/21/Z) via consultancy to P1vital Products Ltd. EAH is on the Board of Trustees of the MQ Foundation. EAH developed the imagery-competing task intervention for intrusive memories, and know-how in using it over the last 20 years (ANEMONE). EAH receives book royalties from Guildford Press and Oxford University Press and receives occasional honoraria for conference keynotes and clinical workshops. All other authors declare no other conflicts of interest. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health and Care Research, or the Department of Health.

Multimedia Appendix 1
Participant information sheet.
[DOCX File, 29 KB - formative_v8i1e47458_app1.docx ]

Multimedia Appendix 2
Intervention feedback questionnaire.
[DOCX File, 14 KB - formative_v8i1e47458_app2.docx ]

Multimedia Appendix 3
Interview topic guide.
[DOCX File, 15 KB - formative_v8i1e47458_app3.docx ]

References


**Abbreviations**

  - **GAINS**: Brief Gameplay Intervention For National Health Service Intensive Care Unit Staff Affected By COVID-19
  - **Trauma**
  - **ICS**: Intensive Care Society

---

https://formative.jmir.org/2024/1/e47458

JMIR Form Res 2024 | vol. 8 | e47458 | p.1165

*(page number not for citation purposes)*
Edited by G Eysenbach; submitted 21.03.23; peer-reviewed by T Aslanidis, Y Jeem, M Ries; comments to author 12.09.23; revised version received 02.10.23; accepted 20.11.23; published 29.02.24.

Please cite as:
Using a Novel Gameplay Intervention to Target Intrusive Memories After Work-Related Trauma: Iterative Qualitative Analysis of Intensive Care Unit Staff Experiences
JMIR Form Res 2024;8:e47458
URL: https://formative.jmir.org/2024/1/e47458
doi:10.2196/47458
PMID:38421698

©Priya Patel, Susan Brown, Boliang Guo, Emily A Holmes, Lalitha Iyadurai, Jonathan Kingslake, Julie Highfield, Richard Morriss. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
A Mobile Applet for Assessing Medication Adherence and Managing Adverse Drug Reactions Among Patients With Cancer: Usability and Utility Study

Chenxu Ni¹, MM, MS; Yi-fu Wang¹, MM, MS; Yun-ting Zhang¹, MM, MS; Min Yuan¹, MD, PhD; Qing Xu¹, MD, PhD; Fu-ming Shen¹, MD, PhD; Dong-Jie Li¹, MD, PhD; Fang Huang¹, MM, MS
Shanghai Tenth People’s Hospital, Shanghai, China

Corresponding Author:
Fang Huang, MM, MS
Shanghai Tenth People’s Hospital
301 Middle Yanchang Road
Shanghai, 200072
China
Phone: 86 66302570
Email: hazel_huang@126.com

Abstract

Background: Medication adherence and the management of adverse drug reactions (ADRs) are crucial to the efficacy of antitumor drugs. A WeChat applet, also known as a “Mini Program,” is similar to the app but has marked advantages. The development and use of a WeChat applet makes follow-up convenient for patients with cancer.

Objective: This study aimed to assess the usability and utility of a newly developed WeChat applet, “DolphinCare,” among patients with cancer in Shanghai.

Methods: A qualitative methodology was used to obtain an in-depth understanding of the experiences of patients with cancer when using DolphinCare from the usability and utility aspects. The development phase consisted of 2 parts: alpha and beta testing. Alpha testing combined the theory of the Fogg Behavior Model and the usability model. Alpha testing also involved testing the design of DolphinCare using a conceptual framework, which included factors that could affect medication adherence and ADRs. Beta testing was conducted using in-depth interviews. In-depth interviews allowed us to assist the patients in using DolphinCare and understand whether they liked or disliked DolphinCare and found it useful.

Results: We included participants who had an eHealth Literacy Scale (eHEALS) score of ≥50%, and a total of 20 participants were interviewed consecutively. The key positive motivators described by interviewers were to be reminded to take their medications and to alleviate their ADRs. The majority of the patients were able to activate and use DolphinCare by themselves. Most patients indicated that their trigger to follow-up DolphinCare was the recommendation of their known and trusted health care professionals. All participants found that labels containing the generic names of their medication and the medication reminders were useful, including timed pop-up push notifications and text alerts. The applet presented the corresponding information collection forms of ADRs to the patient to fill out. The web-based consultation system enables patients to consult pharmacists or physicians in time when they have doubts about medications or have ADRs. The applet had usabilities and utilities that could improve medication adherence and the management of ADRs among patients with cancer.

Conclusions: This study provides preliminary evidence regarding the usability and utility of this type of WeChat applet among patients with cancer, which is expected to be promoted for managing follow-up among other patients with other chronic disease.

(JMIR Form Res 2024;8:e50528) doi: 10.2196/50528

KEYWORDS
WeChat applet; usability testing; utility testing; cancer patients; patients; cancer; qualitative study
Introduction

Medication adherence is defined by the World Health Organization as the extent to which a person’s medication-taking behavior corresponds with agreed recommendations from health care providers [1]. Medication adherence is crucial to the efficacy of antitumor drugs. Patients with cancer have disproportionately higher burdens of comorbid chronic conditions compared to individuals without a cancer history [2]. For individuals with preexisting chronic conditions, a new cancer diagnosis can lead to tremendous challenges, including the coordination of carers and dependents as well as the management of multiple medications for comorbid conditions alongside cancer treatment [3]. Antitumor drugs are known for their numerous adverse drug reactions (ADRs), which can diminish adherence to treatment and cause medical complications. Several interventions have been developed to improve medication adherence and manage ADRs in patients with cancer [4-6]. However, the effectiveness of these interventions is controversial.

WeChat is a type of social networking software that provides flash messaging services on smart terminals. In 2020, the number of monthly active WeChat users exceeded 1.1 billion, rendering it the most common smartphone app in China. It is no longer a simple social platform but has penetrated into all aspects of people’s lives, including their health [7]. With the constant development of WeChat tools, a new development environment and platform was built for the WeChat applet used by 400 million Chinese users every day [8].

The WeChat applet, also known as the “Mini Program,” is similar to the app but has marked advantages. The steps for using the WeChat applet have been simplified, and the applet can be opened directly without downloading the app package. Interestingly, there is an independent storage space among different WeChat applets. If one no longer uses the applet, one need only close the page without uninstalling the program or clearing the cache, which is convenient for users. In addition, the IT infrastructure of WeChat applets can bring about a rapid transfer of digital data between patients with cancer and doctors or pharmacists, and reduce their burden related to oncotherapy information through real-time communication. All of these make follow-up convenient for patients with cancer by using the WeChat applet.

Norman and Skinner [9] reported that participants who had a high eHealth Literacy Scale (eHEALS) score indicated that they had higher literacy skills in using the internet as a resource to obtain health information. This study selected and interviewed dozens of patients with cancer with high eHEALS scores from a tertiary hospital in Shanghai, who are willing to use an eHealth tool—the WeChat applet. If the WeChat applet can be successfully promoted in Shanghai, it will be spread sequentially in various cities throughout China. At present, the vast majority of WeChat applets are free to use, which can reduce the economic burden of follow-up among patients with cancer. However, the existing WeChat applets were developed without involving relevant stakeholders (such as health care professionals or patients) and were not subjected to mobile health app guidelines [10,11].

Usability is described as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [12]. Utility is defined as the functionality of the app and how useful it is to users [12]. The development phase consisted of 2 parts: alpha and beta testing. Beta testing was conducted using in-depth interviews. Alpha testing combined the theory of the Fogg Behavior Model and the usability model. The Fogg Behavior Model [13] suggests that 3 core elements—motivation, ability, and trigger—must converge at the same time for a desired behavior to take effect. Motivation and ability can be balanced against each other; for example, patients may be willing to perform a difficult task if they are highly motivated by the promise of better health outcomes. Ability refers to the patients’ skill or dexterity with respect to the simplicity (or otherwise) of using the app. A trigger is a stimulus that prompts patients to adopt and use the app (eg, an acute attack or clinician’s suggestion).

To our knowledge, no study has assessed the usability and utility of a WeChat applet for patients with cancer once it has been developed. This study aimed to assess the usability and utility of a newly developed WeChat applet—DolphinCare—among patients with cancer in Shanghai.

Methods

Study Design

A qualitative methodology was used to obtain an in-depth understanding of the experiences of patients with cancer when using DolphinCare from the usability and utility aspects.

Ethical Considerations

Ethics approval was obtained from the ethics committee of Shanghai Tenth People’s Hospital prior to the study (SHSY-IEC-5.0/22K99/PO1). The study was registered in Chinese Clinical Trial Registry (ChiCTR2200058189). Written informed consent was obtained from all study participants.

eHEALS

The modified eHEALS was used to assess participants’ literacy skills in using their smart devices to find health-related information on the internet (Multimedia Appendix 1).

Participants

We included participants who had an eHEALS score of ≥50% and were taking 2 or more prescribed medications for their tumor treatment and chronic conditions. We excluded participants aged <18 years or those who had mental disabilities. Purposive sampling was used to recruit older (≥65 years of age) and younger (<65 years of age) participants, as we required the experiences of older participants who may have more comorbidities but may not be comfortable using mobile apps, as well as younger participants who may have fewer comorbidities (than older patients) but may be more comfortable using mobile apps. The purpose of recruiting participants based
on age was to obtain wider perspectives when using DolphinCare.

**Alpha and Beta Testing in the Development Phase**

The development phase consisted of 2 parts: alpha and beta testing. Alpha testing involved testing the design of DolphinCare using a conceptual framework (Figure 1), which combined the theory of the Fogg Behavior Model of Motivation-Ability-Triggers and the usability model. Our framework also included factors that could affect medication adherence and ADRs (Table 1).

DolphinCare was then used for beta testing (Figure 2). Beta testing was conducted using in-depth interviews from July to September 2022. The first and second rounds of in-depth interviews were both conducted at the resting area of the ward. In-depth interviews were conducted to explore the views of patients with cancer regarding the usability and utility of DolphinCare when using it for the first time. We supplemented this interview process by observing the participants and documenting these observations as field notes. This allowed us to determine whether they encountered any difficulties and whether they liked or disliked the utility of DolphinCare. In-depth interviews allowed us to focus on the individual, assist the individual in using DolphinCare, and create an environment where the individual would be able to express his or her views without being influenced by others.

**Figure 1.** The conceptual framework for the design and development of DolphinCare based on the Fogg Behavior Model's theory of Motivation-Ability-Triggers and the usability model. AI: artificial intelligence.
Table 1. Summary of the preferred features and utilities of DolphinCare.

<table>
<thead>
<tr>
<th>Utility</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individualized monitoring plan</td>
<td>The patients can obtain professional evaluation and guidance by implementing the established process and providing feedback to the medical team. Monitoring plan includes objectives, key indicators, medication list, follow-up plan, and precautions.</td>
</tr>
<tr>
<td>Cultivate patients’ habits through small tasks</td>
<td>After each task is completed, the page assigns the patient a “star” as a reward.</td>
</tr>
<tr>
<td>A pop-up push notification and weekly text alerts</td>
<td>A pop-up push notification, with a “single-click” return to the app, was preset every other day. Apart from a pop-up push notification, there are weekly text alerts as well. Most patients will use the applet after receiving the notification.</td>
</tr>
<tr>
<td>Complex medication regime</td>
<td>Ability to aid patients in managing complex medication regimens, such as a drug combination or changing a medical prescription.</td>
</tr>
<tr>
<td>Timely WeChat web-based communication</td>
<td>In case of emergencies or questions during the treatment, patients can consult the medical team at the WeChat consultation window to solve the problems quickly and effectively.</td>
</tr>
<tr>
<td>The Medical team intelligently manages the patients with cancer</td>
<td>The medical team can see the number of existing patients, their medication situation, adverse drug reactions, examination information, etc. They consider patients at the center for full-dimensional data monitoring and comparison. Adverse drug reactions and inspections are monitored.</td>
</tr>
<tr>
<td>AI²-enabled automatic recognition of characters in pictures in inspection reports</td>
<td>AI automatically recognizes the words and data in the examination documents and inputs them into the patient database.</td>
</tr>
<tr>
<td>Web-based questionnaire on adherence</td>
<td>Users are able to assess their adherence to medications.</td>
</tr>
<tr>
<td>Rehabilitation guidance</td>
<td>Based on the patient’s condition and medication situation, the medical team gives knowledge guidance such as rational medication and nutritional rehabilitation.</td>
</tr>
</tbody>
</table>

\(^{a}\text{AI: artificial intelligence.}\)

Figure 2. The landing page, registration page, home page, and pharmaceutical care page of DolphinCare. ADR: adverse drug reaction.
The Interview Questions
The following questions were asked to the participants during the interviews:

- What is your initial impression of using DolphinCare?
- What is your feeling about using DolphinCare after a period of use?
- What motivates you to use DolphinCare?
- Can you operate DolphinCare yourself, or do you need help from others?
- Will you continue to use DolphinCare?
- What’s the operation interface of DolphinCare that left a deep impression on you?
- What’s function do you expect DolphinCare to further improve on?
- What do you think are the inconveniences of using DolphinCare?

Data Collection Process
Participants used their mobile WeChat app to acquire the official account of DolphinCare free of charge and then filled in the demographic form. Each participant was interviewed twice. During the interview, the “first impression” of participants in using DolphinCare would be captured. The researcher took detailed notes and observed for nonverbal cues during each interview. Facial expressions and body language subconsciously portrayed by the participants were noted down by the researchers. All interviews were audio recorded. The 8-item Morisky Medication Adherence Scale (MMAS-8) [14-16] is a questionnaire designed to facilitate the identification of barriers and behaviors associated with adherence to medication. The possible answers to questions 1 to 7 are “yes” (0 points) or “no” (1 point). Five of the questions are scored in reverse (ie, yes=1 and no=0). The possible answers to question 8 are “Never” (1 point), “Occasionally” (0.75 points), “Sometimes” (0.50 points), “Often” (0.25 points), and “All the time” (0 points) [17].

Data Analysis
All interviews were transcribed verbatim. An interpretive-descriptive approach was used to identify the themes that emerged from the data. This approach was used to obtain a deeper understanding of the usability and utility of the perspectives and experiences of patients with cancer in using the WeChat applet. The researchers reflected on the data and began constructing an interpretive account of what the codes signified from the participants’ perspectives, and its application in clinical practice [14]. The researchers also referred to the field notes for reflections, facial cues, and body languages observed during the interviews. The research team then met to discuss the coding of the transcripts. Any coding discrepancies were resolved through discussion until a consensus was reached.

Results
Participants
A total of 22 participants were recruited (Table 2) for the first interview. Only 20 participants (12 men and 8 women; 16, 80% of them being patients and 4, 20% being carers) were interviewed consecutively, as 2 (P13 and P14) declined to participate. The mean age of the patients and their carers was 62.4 (SD 10.25) years. The average of the number of medications among patients was 2.7.
Table 2. Demographic characteristics of the recruited participants.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Age (years)</th>
<th>Medications, n</th>
<th>Patient or carer</th>
<th>iPhone or Android user</th>
<th>eHEALS(^a) score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Male</td>
<td>64</td>
<td>3</td>
<td>Carer</td>
<td>Android</td>
<td>75</td>
</tr>
<tr>
<td>P2</td>
<td>Female</td>
<td>67</td>
<td>1</td>
<td>Patient</td>
<td>Android</td>
<td>56</td>
</tr>
<tr>
<td>P3</td>
<td>Female</td>
<td>73</td>
<td>2</td>
<td>Patient</td>
<td>iPhone</td>
<td>80</td>
</tr>
<tr>
<td>P4</td>
<td>Female</td>
<td>73</td>
<td>2</td>
<td>Carer</td>
<td>Android</td>
<td>78</td>
</tr>
<tr>
<td>P5</td>
<td>Female</td>
<td>35</td>
<td>4</td>
<td>Patient</td>
<td>Android</td>
<td>57</td>
</tr>
<tr>
<td>P6</td>
<td>Male</td>
<td>72</td>
<td>6</td>
<td>Patient</td>
<td>Android</td>
<td>75</td>
</tr>
<tr>
<td>P7</td>
<td>Male</td>
<td>55</td>
<td>3</td>
<td>Patient</td>
<td>Android</td>
<td>69</td>
</tr>
<tr>
<td>P8</td>
<td>Male</td>
<td>62</td>
<td>4</td>
<td>Patient</td>
<td>Android</td>
<td>69</td>
</tr>
<tr>
<td>P9</td>
<td>Male</td>
<td>47</td>
<td>2</td>
<td>Patient</td>
<td>iPhone</td>
<td>75</td>
</tr>
<tr>
<td>P10</td>
<td>Male</td>
<td>68</td>
<td>2</td>
<td>Carer</td>
<td>iPhone</td>
<td>75</td>
</tr>
<tr>
<td>P11</td>
<td>Male</td>
<td>67</td>
<td>3</td>
<td>Patient</td>
<td>Android</td>
<td>59</td>
</tr>
<tr>
<td>P12</td>
<td>Female</td>
<td>75</td>
<td>4</td>
<td>Patient</td>
<td>Android</td>
<td>57</td>
</tr>
<tr>
<td>P13</td>
<td>Male</td>
<td>65</td>
<td>5</td>
<td>Patient</td>
<td>Android</td>
<td>84</td>
</tr>
<tr>
<td>P14</td>
<td>Male</td>
<td>71</td>
<td>1</td>
<td>Patient</td>
<td>Android</td>
<td>50</td>
</tr>
<tr>
<td>P15</td>
<td>Male</td>
<td>59</td>
<td>1</td>
<td>Patient</td>
<td>Android</td>
<td>62</td>
</tr>
<tr>
<td>P16</td>
<td>Female</td>
<td>64</td>
<td>2</td>
<td>Patient</td>
<td>Android</td>
<td>71</td>
</tr>
<tr>
<td>P17</td>
<td>Male</td>
<td>50</td>
<td>3</td>
<td>Patient</td>
<td>Android</td>
<td>88</td>
</tr>
<tr>
<td>P18</td>
<td>Female</td>
<td>62</td>
<td>3</td>
<td>Patient</td>
<td>Android</td>
<td>65</td>
</tr>
<tr>
<td>P19</td>
<td>Male</td>
<td>73</td>
<td>4</td>
<td>Patient</td>
<td>iPhone</td>
<td>65</td>
</tr>
<tr>
<td>P20</td>
<td>Male</td>
<td>61</td>
<td>2</td>
<td>Carer</td>
<td>Android</td>
<td>80</td>
</tr>
<tr>
<td>P21</td>
<td>Female</td>
<td>53</td>
<td>2</td>
<td>Patient</td>
<td>iPhone</td>
<td>82</td>
</tr>
<tr>
<td>P22</td>
<td>Male</td>
<td>68</td>
<td>1</td>
<td>Patient</td>
<td>Android</td>
<td>73</td>
</tr>
</tbody>
</table>

\(^a\)eHEALS: eHealth Literacy Scale.

Adoption

We used the motivation, ability, and trigger categories of the Fogg Behavior Model to explore why patients adopted the WeChat applet using data from the qualitative interviews and usage data.

Adoption: Motivation

The key positive motivators described by the interviewers were to be reminded to take their medications, to alleviate their ADRs through monitoring, and to aid in medical research. Some patients had more than 1 motivation for adoption.

_The motivation is to observe if I can improve the medication adherence, and manage my ADRs._ [Patient 10, 68 years, male]

_It [taking part in this research] could be helpful for the medical studies or something. So to advance their work._ [Patient 4, 73 years, female]

_My main motivation is to participate in this research for distracting me from my anxiety about the disease._ [Patient 7, 55 years, male]

Adoption: Abilities

Most of the patients were able to activate and use the WeChat applet by themselves. Some (typically older) patients needed assistance from younger family members.

_I’m not very good with these things but somebody younger in my family would help me for handling it._ [Patient 2, 67 years, female]

_I know how to use it … I think it is quite easy and straight forward because the layout is very clear (while showing the Homepage)._ [Patient 5, 35 years, female]

Adoption: Triggers

Most patients indicated that they would receive follow-up through the WeChat applet if it was recommended by their known and trusted health care professionals.

_My trusted physician recommended me to use this ‘wechat applet’ to improve my medication adherence and better manage ADRs, which makes me willing to use it._ [Patient 8, 62 years, male]
Usability Testing
They were challenges encountered when adding a new medication, with regard to patients’ understanding of their complex medication regimen.

Challenges Encountered When Adding a New Medication
Several subthemes emerged under this theme: confusion by terms used when adding medications into DolphinCare, unfamiliarity with the entered generic name of the medication, and patients’ incomprehension of their complex medication regimen.

Unfamiliarity With the Generic Name of the Medication
When entering medication details into the WeChat applet, most patients knew their medication by brand names but not generic names. DolphinCare requires users to enter the generic name of the medication, as the pharmacy label only contains the generic name.

I am not sure of accurately different names of my medications, and I almost always have my pill boxes with me. [Patient 15, 59 years, male]
I’m not too sure what is the precise name of my medication. I only know it by its brand name. So having the indication automatically linked to the medication name is good to have. [Patient 16, 64 years, female]

Utility Testing
Two themes emerged from the utility testing of DolphinCare: utilities that could improve medication adherence and the management of ADRs.

A Medication Reminder System
All participants found the medication reminder useful, including timed pop-up push notifications and text alerts.

Oh yes this was helpful. It prompted me to remind my mum to take her medications. [Patient 20, 61 years, male]
It would like to set timed pop-up push notifications to depend on what time I wake up in the morning and go to bed in the evening. [Patient 18, 62 years, female]

A Medication Adherence Scoring System
Medication adherence among patients was evaluated using the MMAS-8. When a patient uses the WeChat applet for more than 1 month, a questionnaire to evaluate their medication adherence will be provided to the patient to fill in.

I’m sorry that I wasn’t aware of the MMAS-8 questionnaire until you reminded me. Happily, my medication adherence improved after using the applet. [Patient 9, 47 years, male]

The Management of ADRs
When patients input the antitumor medication they are taking, the applet interface automatically matches the corresponding educational information of ADRs. A week later, the applet provides the corresponding collection information forms of ADRs to the patient to fill out. Health care professionals receive messages or provide web-based medication guidance to the patient based on the severity of ADRs. In addition, the WeChat applet can intelligently recognize the inspection and image reports uploaded by the patient in paper-photo versions. The WeChat applet converts the reports into text formats, and organizes and records them in the patient’s file. When there are obvious abnormalities in indicators of the reports related to ADRs, health care professionals can promptly contact the patient.

The Web-Based Consultation System
The web-based consultation system enables patients to consult pharmacists or physicians in time when they have doubts about medications or have ADRs. There will be a professional exclusively in charge of web-based consultation services during working days.

Oh this is good. This saved me the time and efforts to go to the hospital for consultation, and the response I got from online consultations was equally satisfactory. [Patient 17, 50 years, male]

Discussion
Principal Findings
DolphinCare was designed and developed on the basis of the Fogg Behavior Model. This model comprised 3 phases—motivation, ability, and triggers—which suggests that a patient is able to achieve a target behavior if he or she has high motivation, ability, and an effective trigger simultaneously.

The requirement phase was based on utilities that could improve medication adherence and manage ADRs, which led to the design and development of DolphinCare until the medical personnel and patients were satisfied with the prototype. This paper focused on the usability and utility testing of DolphinCare, for which patients with cancer were recruited to use the WeChat applet and provided feedback.

To our knowledge, no previous study has reported the experiences of patients with cancer when using a WeChat applet. The patients also preferred a summary page of medication, which was accessible by tapping on the medication icon on the home page of DolphinCare; this was useful as it provided a brief overview of their medication regimen. This further enhanced the usability of DolphinCare, as it would be more patient-centered and more likely to be adopted. It is worth noting that there was no particularly negative feedback provided by the participants. This is a limitation of the study, and future research needs explore the negative feedback from patients to better improve DolphinCare.

The steps to add a medication were simplified and displayed in a layered order to prevent cognitive overload [18]. However, some participants struggled when adding a new medication due to the complexity of this task. It is challenging to input data into a small device, as it requires the user to navigate the app on a small screen [19]. Despite the challenges encountered by patients, the process of adding a new medication “manually” benefited patients with regard to their medication knowledge.
Participants had to “learn” the generic name of their medications, their administration frequency, and their purpose, which is beneficial for the patient's treatment process.

Several studies have shown that behavioral change is achievable through active reminders, which strengthens the benefits of medication adherence apps [20,21]. A review by Santo et al [22] in 2016 revealed that only 56% of medication adherence apps adjust flexible scheduling for medication reminders. Medication reminders with flexible scheduling (where users may opt for medication reminders on alternate days or on a weekly basis) allows for personalization of the app to suit their individual needs. Participants also reported that they had a better understanding of the frequency and indication of the medication, which appeared on the reminder, thereby improving their medication knowledge. Improved patient knowledge is known to enhance their medication adherence and clinical health outcomes [23]. However, areas involving strategies to improve patients’ medication knowledge require further investigation [24]. DolphinCare offers a pop-up push notification, with a “single-click” return to the WeChat applet, which was preset every 2 days. Besides a pop-up push notification, weekly text alerts were also provided. These allowed users who were unable to take their medications at a specific time point to take them later, which actively prompted the patients to take their medications properly and on time. DolphinCare required users to acknowledge the reminder that, theoretically, would make the patients more conscious of their adherence to medications.

Symptom monitoring is especially important for patients with cancer because they can experience varying acute and chronic side effects from their treatment regimen [25]. Egbring et al [26] conducted a 3-arm randomized controlled trial for 6 weeks with 139 patients with early-stage breast cancer undergoing chemotherapy. The participants were randomly assigned to a control group, an unsupervised group that used a mobile app to record data without physician review, or a supervised group that recorded data in the app with physician review. The results revealed that participants who had physician collaboration when using the health tracking app demonstrated increased reporting of adverse effects of chemotherapy, more precise health data entries, and stabilization of daily functional activities measured using the ECOG (Eastern Cooperative Oncology Group) scale.

Our study confirmed the importance of symptom monitoring among patients with cancer, as these patients can also experience significant side effects from their treatment regimen, which influence their overall quality of life during and after their course of therapy.

The wide demographic range of the participants of this study delineates the experiences of both young and old users. The usability and utility testing of DolphinCare demonstrated the needs of patients with cancer and their caregivers better and helped tailor the WeChat applet to suit their needs. This ensured that DolphinCare would be a more patient-centered WeChat applet and more likely to be used. Our study suggests that DolphinCare can aid patients with cancer and even those with other chronic diseases to improve medication adherence and manage ADRs.

Conclusions
DolphinCare was designed and developed on the basis of the Fogg Behavior Model. This study provides preliminary evidence of the usability and utility of this type of WeChat applet among patients with cancer. This WeChat applet had usabilities and utilities that could improve the patients’ medication adherence and ADR management.

Acknowledgments
This work was supported by a research project designed by the Chinese Pharmaceutical Association Hospital Pharmacy department (CPA-Z05-ZC-2023002) and a program for research-oriented physicians of Shanghai Tenth People’s Hospital (grant 2023LYJFZR002). The MMAS-8 Scale, content, name, and trademarks are protected by US copyright and trademark laws. Permission for use of the scale and its coding is required. A license agreement is available from MMAR, LLC., www.moriskyyscale.com.

Conflicts of Interest
None declared.

Multimedia Appendix 1
eHealth Literacy Scale.
[DOCX File , 26 KB] - formative_v8i1e50528_app1.docx ]

References


8. 2013 Medical Devices. U.S. Food and Drug Administration. URL: https://www.fda.gov/MedicalDevices/ [accessed 2024-02-12]


Abbreviations

ADR: adverse drug reaction  
ECOG: Eastern Cooperative Oncology Group  
eHEALS: eHealth Literacy Scale  
MMAS-8: 8-item Morisky Medication Adherence Scale
Using mHealth to Improve Communication in Adult Day Services Around the Needs of People With Dementia: Mixed Methods Assessment of Acceptability and Feasibility

Amy Zheng¹; Marissa Bergh¹, RN, BSN; Komal Patel Murali¹,², RN, ACNP-BC, PhD; Tina Sadarangani¹,²,³, RN, ANP-C, GNP-BC, PhD

¹New York University Rory Meyers College of Nursing, New York, NY, United States
²Hartford Institute for Geriatric Nursing, New York University Rory Meyers College of Nursing, New York, NY, United States
³New York University Grossman School of Medicine, New York, NY, United States

Corresponding Author:
Marissa Bergh, RN, BSN
New York University Rory Meyers College of Nursing
433 First Avenue
6th Floor
New York, NY, 10010
United States
Phone: 1 212 998 5300
Email: msb7677@nyu.edu

Abstract

Background: Adult day services (ADS) provide community-based health care for older adults with complex chronic conditions but rely on outdated methods for communicating users’ health information with providers. CareMOBI, a novel mobile health (mHealth) app, was developed to address the need for a technological platform to improve bidirectional information exchange and communication between the ADS setting and providers.

Objective: This study aims to examine the feasibility and acceptability of CareMOBI in the ADS setting.

Methods: A concurrent-triangulation mixed methods design was used, and participants were client-facing ADS staff members, including direct care workers (paid caregivers), nurses, and social workers. Interviews were conducted to describe barriers and facilitators to the adoption of the CareMOBI app. The acceptability of the app was measured using an adapted version of the Technology Acceptance Model questionnaire. Data were integrated into 4 themes as anchors of an informational matrix: ease of use, clinical value, fit within workflow, and likelihood of adoption.

Results: A mix of ADS staff (N=22) participated in the study. Participants reported high levels of acceptability across the 4 domains. Qualitative findings corroborated the questionnaire results; participants viewed the app as useful and were likely to implement CareMOBI in their practice. However, participants expressed a need for proper training and technical support throughout the implementation process.

Conclusions: The CareMOBI app has the potential to improve care management in the ADS setting by promoting effective communication through an easy-to-use and portable method. While the integration of CareMOBI is acceptable and feasible, developing role-specific training modules and technical assistance programs is imperative for successful implementation within the ADS setting.

(JMIR Form Res 2024;8:e49492) doi:10.2196/49492

KEYWORDS

adult day services; primary health care; health communication; dementia; mobile health; mHealth; community-based; health care; older adults; older adult; chronic condition; health information; feasibility; acceptability; CareMOBI; mixed methods design; caregivers; caregiver; care workers; nurses; social workers
Introduction

Adult day services (ADS), commonly referred to as adult day care, is a vital but overlooked source of health and social care for the burgeoning population of older adults with complex chronic conditions—particularly those with Alzheimer disease and related dementia (ADRD) [1]. ADS sites are nonresidential, congregate, and community-based facilities that offer interdisciplinary services, including opportunities for socialization, nursing care, and nutritious meals for chronically ill or functionally impaired adults. Each day in the United States, 251,100 adults with complex health and social needs receive care in the ADS setting: 65% have some combination of ADRD, diabetes, depression, heart disease, or other chronic conditions; 72% live below federal poverty lines; and 55% are from a racial or ethnic minoritized group [2]. Clients attend 2 to 5 days per week and may receive assistance with personal care, physical therapy, vital sign monitoring, and medication administration, while also participating in organized group activities [3]. ADS staff are skilled at using their in-depth, serial observations of clients to identify warning signs of acute illness and promote early clinical intervention, which are especially important in persons with ADRD, who may not be able to identify or communicate changes in their health status.

However, ADS sites face numerous barriers in communicating concerning changes in health status to primary care providers [4]. Ineffective communication in health care has been associated with costs exceeding US $10 billion, in addition to adverse outcomes and increased mortality [5,6]. We previously found that communication between free-standing community-based ADS centers is hampered by reliance on antiquated methods of information exchange that challenge information sharing and care coordination [7]. When ADS clients experience acute changes in health status or behavior, information is typically reported to primary care providers through fax or voicemail messages, which often result in nonresponse, delayed diagnosis, referral or treatment, and inadequate follow-up. Primary care providers and ADS staff have agreed that communication between them is infrequent, delayed, incomplete, and unreliable. Primary care providers prefer to communicate using direct messaging systems within their electronic health record; however, 92% of ADS sites in the United States lack the resources for interoperable electronic health record systems that enable e-communication [8]. The lack of resources shifts the burden of communication between ADS sites and primary care to the family caregiver (herein referred to as caregivers) who must deliver the information from the ADS site to the primary care provider at the point of service. Although engaged caregivers are often willing to track medical information and coordinate care, they may lack the necessary time, resources, and health education, resulting in medical errors or delays in care.

We developed CareMOBI (mobile health for organizations to bolster interconnectedness), a mobile health (mHealth) app prototype, to address the consistent need for improved care coordination and communication among care team members supporting care in home and community settings (Figure 1). CareMOBI acts as a centralized hub for families to track and share information about their loved one’s day-to-day health with other care team members (ie, ADS staff, home health aides, and family members). This enables multiple caregivers to support an individual across home and community-based settings. Within CareMOBI, individuals are first invited to a person’s care team. Each member can then record information and share updates about a person’s health, including how they ate, slept, or felt on a given day. They can track vital signs, keep up-to-date medication lists, and track medication administration. They can also track appointments and make notes of observations or questions to ask providers. Most importantly, care team members can report and be alerted when an individual is exhibiting concerning symptoms or experiencing an emergency. Information entered with the app can be summarized and shared via a PDF file to support shared clinical decision-making at appointments or appended to a chart. CareMOBI is a low-cost, portable means of exchanging information between ADS sites, caregivers, and health care providers. It also provides a centralized platform that allows ADS staff, caregivers, and care providers to provide updates and track the health progress of a person who cannot do so independently, as well as share any urgent concerns and observations, such as new or worsening confusion, behavioral changes, or abnormal vital signs. These features are designed to support critical early identification of clinical issues with the goal of reducing costly, traumatic, and avoidable emergency department care or hospitalizations, as well as overall care management for people with complex care needs. The purpose of this mixed methods study was to examine the feasibility and acceptability of CareMOBI in an ADS setting through surveys and interviews with ADS staff.
Methods

We used a mixed methods concurrent triangulation design to (1) assess acceptability and feasibility of the CareMOBI prototype among adult day center staff and (2) identify factors contributing to eventual likelihood of adoption or nonadoption.

Setting and Sample

Participants were eligible if (1) they were paid employees of a participating adult day center and (2) they had a client-facing role that involved daily interaction with persons living with dementia, as these are the target end users of the app in the ADS setting. Examples of client-facing staff are registered nurses, social workers, and program assistants. Individuals were excluded if they had worked in their current position for less than 6 months. Purposive sampling was used to recruit a diverse multistakeholder sample that represented the range of professionals in adult day centers (eg, registered nurses, social workers, and program directors).

Ethical Considerations

Eligible staff members were identified with the help of administrators at participating adult day centers. A research assistant contacted them by email or phone according to their preference, described the study, confirmed participants’ eligibility, and subsequently obtained informed consent. In total, 22 staff members from adult day centers in 3 states (New York, California, and Georgia) enrolled. All enrollees received a US $50 gift card for their participation. The New York University committee on activities involving human subjects provided Institutional Review Board approval for this study (IRB-FY2020-4615).

Procedures

Data collection consisted of one-to-one semistructured interviews and the completion of the Technology Acceptance Model questionnaire, adapted for health care settings [9]. The Technology Acceptance Model was developed 4 decades ago to explore factors that shape workers’ intention to use emerging technology. Rooted in the theory of reasoned action, it explores the beliefs and norms that influence attitudes and expectations that increase the desire to carry out a behavior. Thus, it is a logical and established framework to explore factors determining the likelihood of adoption of CareMOBI in ADS. The Technology Acceptance Model questionnaire provided insight into the acceptability of the app, while subsequent qualitative interviews provided additional insight into concerns around the feasibility of embedding CareMOBI within ADS.

Approximately 1 week prior to the interview, participants received a confirmation email that contained a link to an interactive prototype of CareMOBI. The prototype could be accessed from a smartphone, tablet, or computer. Participants were instructed to watch a 2-minute informational video about the app, also linked in the email, and were then asked to spend approximately 10 minutes navigating through the interactive prototype to complete several relevant tasks, including logging in, adding a new medication, recording day-to-day activities of a typical person living with dementia at their adult day center, as well as any health-related progress notes. They were also asked to use filters within CareMOBI to locate information.
about the person living with dementia in whom they were interested.

**Qualitative Data Collection and Analytic Procedures**

Web-based interviews were scheduled based on participants’ availability and conducted via a secure web-based platform. UX (user experience) and UI (user interface) design professionals, who had extensive knowledge of user-testing, provided input to develop a semistructured interview guide at the product development firm that built the CareMOBI prototype.

Participant interviews lasted 30 minutes on average. Interviews were conducted by either the principal investigator (TS) or a trained research assistant. Both individuals have extensive experience with qualitative interviewing. Open-ended questions allowed participants to elaborate on their reaction to the CareMOBI app and allowed the researchers to elicit information on different factors influencing their perceptions of the app, aspects related to its usability, and potential feasibility issues, including workflow integrations. Sample questions included:

1. What is your overall impression of the app?
2. In what ways do you think this app could help address some of the challenges you face in organizing and communicating information around the needs of person living with dementia?
3. Are there other challenges in caring for person living with dementia that this app doesn’t address? What might those be?
4. What features in the app were most confusing for you to understand/find/utilize?

All interviews were recorded, professionally transcribed, and reviewed for accuracy. Field notes by the interviewer supplemented the tape-recorded interviews. A detailed audit trail documented the rationale for any methodological changes during the interview or analysis (ie, when unique follow-up questions were posed).

Qualitative data were analyzed using a content analysis approach. A preliminary codebook was developed a priori based on the interview guide as a coding scheme for all transcripts. The research team discussed any text that could not be categorized within the codebook to determine if a new category or code needed to be defined or aligned with an existing category or code. The codebook was continuously updated to reflect the iterative process. Initially, 2 coders coded independently in Dedoose, a web-based platform for qualitative and mixed method coding, and met weekly to review coding and resolve any disagreements. To ensure the reliability and consistency of coding, a third independent coder analyzed a subset of 20% (n=5) of the transcripts. The principal investigator addressed any unresolved disagreements, as well as potential new categories or codes, in team meetings. Codes were summarized within cases and then compared across cases to identify emerging themes. Saturation occurred when no new themes emerged. The research team members regularly debriefed to discuss and validate the results of the analysis. Quotations were provided with participant’s initials to maintain anonymity.

**Quantitative Data Collection and Analytic Procedures**

Upon finishing the interview, participants completed a web-based questionnaire. In addition to providing basic demographic information, participants completed an adapted version of the Technology Acceptance Model questionnaire, which was previously validated for use in health care settings [9]. Responses to the 33 survey items enabled further examination of factors that could influence the eventual adoption of CareMOBI, and the anonymous nature reduced the potential for social desirability bias. Respondents rated each item on a 7-point Likert scale ranging from “strongly disagree” to “strongly agree.” Each domain-specific question was averaged to determine scores. Higher scores corresponded to higher perceived acceptability.

Descriptive statistics were used to characterize the sample, with measures of central tendency and spread for continuous measures and frequencies and percentages for dichotomous or categorical variables. Calculations for quantitative statistics were done using Qualtrics software (version May 2023; Qualtrics LLC).

**Integration of Qualitative and Quantitative Data**

Qualitative and quantitative data were integrated in the third and final phase of analysis. We sought to align with the Technology Acceptance Model: perceived ease of use, perceived value in clinical care, fit within existing workflows, and end users’ overall likelihood of adoption. Using the 4 themes as anchors, we developed an informational matrix in which qualitative data were embedded and compared with quantitative data. Using triangulation methods, we sought to understand the overall likelihood of adoption of the app by end users (quantitatively) and factors underpinning this across cases within each stakeholder group (qualitatively).

**Results**

**Overview**

The primary goals of this study were to (1) assess the acceptability of the CareMOBI prototype among adult day center staff and (2) identify factors contributing to the eventual likelihood of adoption or nonadoption. We evaluated the feasibility and acceptability of CareMOBI in an ADS setting quantitatively and qualitatively according to 4 themes: perceived ease of use, perceived value in clinical practice, how the mHealth app fits within existing workflows, and likelihood of adoption.

**Study Sample**

The total sample (N=22) of ADS staff members was majority non-Hispanic (20/22, 91%) and White (16/22, 73%). Most participants (19/22, 86%) were aged between 40 and 69 years. All respondents (100%) identified as women. Among those who responded when asked about their role, the majority (10/18, 56%) identified as direct care workers or professional caregivers (aides) in adult day centers (see Table 1).
Table 1. Demographic characteristics of adult day center staff.

<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>Participants (N=22), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>22 (100)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>African American or Black</td>
<td>5 (23)</td>
</tr>
<tr>
<td>White</td>
<td>16 (73)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>20 (91)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>≤29</td>
<td>2 (9)</td>
</tr>
<tr>
<td>30-39</td>
<td>0 (0)</td>
</tr>
<tr>
<td>40-49</td>
<td>9 (41)</td>
</tr>
<tr>
<td>50-59</td>
<td>5 (23)</td>
</tr>
<tr>
<td>60-69</td>
<td>5 (23)</td>
</tr>
<tr>
<td>≥70</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>High school graduate, or equivalent</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Professional degree</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Some college credit, no degree</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Associate’s degree</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>5 (23)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>5 (23)</td>
</tr>
<tr>
<td>Doctorate degree</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Role at adult day center</strong></td>
<td></td>
</tr>
<tr>
<td>Direct care worker or professional caregiver</td>
<td>10 (44)</td>
</tr>
<tr>
<td>Registered nurse</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Social worker</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Program director</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Activities coordinator</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Preferred not to answer or did not specify</td>
<td>4 (18)</td>
</tr>
</tbody>
</table>

**Perceived Ease of Use**

Perceived ease of use describes the effort level associated with understanding and using CareMOBI. There were 5 questions within the Technology Acceptance Model questionnaire that assessed the perceived ease of use: the overall ease of use, the technological skill required to use the app, and the user’s comfort level with the app (Figure 2). The mean score for this domain was 6.48, indicating a high ease of use for CareMOBI. The question with the highest mean score was “I think that I could easily learn how to use the proposed mHealth app” (mean of 6.77 or strongly agree). The lowest scoring question was “I feel comfortable with information and communication technologies” (mean of 6.00 or agree). The qualitative interviews provided insight into what aspects of CareMOBI facilitated the ease of use for the participants (Table 2). Participants reported that the organized layout was simple and easy to navigate. Many expressed appreciation for how they did not have to spend time searching for what they needed, and that all of the features were “at their fingertips” (ADC-LP). However, the qualitative interviews did reveal that while ADS staff found CareMOBI easy to use, they had concerns about family caregivers’ ability to use it. This included challenges with the functionality of some app features and the lack of non-English language options for caregivers and patients with low English proficiency. One respondent stated:
For those who don’t speak English, that’s another issue—because participants here, we have different languages spoken here, more than five, six, seven languages. For caregivers who don’t understand or don’t speak English well, that’s challenging as well. [ADC-AL]

Furthermore, some participants expressed concern that technology-based solutions may not be appropriate methods of communication for some clients due to low technology proficiency and preferences for traditional methods. While many of the features were easy to use, participants noted a few exceptions that were too complex and potentially confusing for them to navigate, particularly the appointment tracking feature. A staff member noted that while navigating the app was “pretty easy,” it might confuse caregivers who could “get mixed up on appointments, whether it’s outside or they’re coming here” (ADC-NF).

Figure 2. Quantitative survey answer distributions for perceived ease of use. mHealth: mobile health.

Table 2. Qualitative subthemes for perceived ease of use.

<table>
<thead>
<tr>
<th>Qualitative subthemes</th>
<th>Qualitative feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>User centered design</td>
<td>“I liked it. I thought it was well organized. I like where everything’s at your fingertips. You can do it on your phone and share it with who you’d like to share it with. I was pretty impressed.” [ADC-LP]</td>
</tr>
<tr>
<td>Functionality</td>
<td>“It’s not letting me click on event. I think this was the issue before. I had to go all the way out to—and come back in if I wanted to access something other than—so, I’m stuck on mood right now. I can go to vitals. I can go to all, but I can’t go to event.” [ADC-DT]</td>
</tr>
<tr>
<td>Noninclusive Features</td>
<td>“Additional to that, for those who don’t speak English, that’s another—because participant here, we have different languages spoken here, more than five, six, seven languages. For caregiver who don’t understand or don’t speak English well, that’s challenging as well.” [ADC-AL]</td>
</tr>
<tr>
<td>Confusing or overly complex features</td>
<td>“...those boxes are right there in the forefront, so it’s pretty easy, but it’s easy to get mixed up on appointments, whether it’s outside or they’re coming here...” [ADC-NF]</td>
</tr>
</tbody>
</table>

Perceived Value in Clinical Practice

Perceived value in clinical practice is the degree to which CareMOBI could improve patient care and management within ADS. Thirteen questions in the survey assessed this domain (Figure 3). The mean score for this domain was 6.18 out of 7, indicating that participants generally rated the CareMOBI app well in terms of potential value to their clinical practice. The item with the highest average score was “The proposed mHealth app can facilitate the care of my patients/clients/loved ones” (mean of 6.55 or strongly agree). The item with the lowest average score was “The proposed mHealth app can facilitate the care of my patients/clients/loved ones” (mean of 5.70 or agree).

Data from the qualitative interviews suggested that staff felt there was immediate value to implementing the app in the day center (Table 3).

I know how it could help me and my family, so it just lights the torch for me the more. This is so huge because the next wave, they’re gonna be—they’re more technology savvy, so they’re gonna be wanting a lot of this information more so, so we have to keep evolving to be able to help make sure that we are communicating right away and training our staff to use it. Both the people were so excited, like, “Oh, my God, it’s so much easier,” and then they get an alert. I’m like, “Yeah, just to get an alert.” [ADC-NF]
Several features of the CareMOBI app would improve the staff’s ability to manage care for ADS participants. Specifically, study participants appreciated the alert and notifications feature for caregivers and staff and the built-in features that allow easy “communication on a daily basis” for members of the care team. The CareMOBI app has the capacity to track health progress (ie, vital sign trends, medication management, and appointment reminders) beyond ADS through caregiver and health professional portals. These features were viewed as beneficial to augment the traditional emergency communications and to facilitate chronic disease management within the adult day centers. While most features received positive feedback, 1 staff participant noted that the behavioral assessment tool, which used emoticons (or “smiley faces”) to evaluate daily behavior was oversimplified. Generally, both the qualitative and quantitative data indicate that staff participants perceived the CareMOBI app as a potentially valuable clinical tool to augment their practice due to its ability to facilitate communication, consolidate patient information, and organize care management.

**Figure 3.** Quantitative survey answer distributions for perceived value in clinical practice. mHealth: mobile health.
Table 3. Qualitative subthemes for perceived value in clinical care.

<table>
<thead>
<tr>
<th>Subthemes</th>
<th>Qualitative feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient monitoring and management</td>
<td>“I know how it could help me and my family, so it just lights the torch for me the more. Both the people were so excited, like, ‘Oh, my God, it’s so much easier,’ and then they get an alert. I’m like, ‘Yeah, just to get an alert.’” [ADC-NF]</td>
</tr>
<tr>
<td>Real time communication among the care team</td>
<td>“I think that the real-time communication makes it useful. What we’re seeing on a daily basis, if we’re able to communicate on a daily basis as opposed to—right now, we have—we will call a family member if there’s an acute episode. If there’s not an acute episode, we’re not gonna just be like, ‘Hey, this is what happened today. This is what we saw. Blah, blah, blah, blah, blah.”” [ADC-DT]</td>
</tr>
<tr>
<td>Updates on health progress</td>
<td>“Even though they might not be able to communicate on a consistent basis when I call, tell me what was happenin’ this week or this weekend. I could at least gather that information and, like I said, with the app there, I could say, ‘Per a neighbor who’s slash the caregiver, this is what they observed.’” [ADC-LSM]</td>
</tr>
<tr>
<td>Trends</td>
<td>“Then visually having the client on-site, it would be able to give me a better way of communicating, ‘Oh, yeah. This is not just a bump. [Laughter] It’s a big bump from the baseline.’ That’s the whole point, being able to look at the trend.” [ADC-LSM]</td>
</tr>
<tr>
<td>Assist caregivers</td>
<td>“We, at adult daycare center, we take care of not only participant, but the caregiver also. When caregiver’s wellbeing are good, our participant are good. They can have quality of life at home instead of ending up in a placement. It’s not their wishes and their goals. That’s very good that in the app.” [ADC-AL]</td>
</tr>
<tr>
<td>Schedules and reminders</td>
<td>“I think it’s very good to have [appointments] on the calendar. It’s forgetful for those who have cognitive impairment and caregiver busy with the schedule. When you have those in place, it’s very good for the keep ongoing with the appointment, with the medication, make sure they take medication on time, don’t miss any doses. That help both caregiver and participant to do the daily tasks appropriately.” [ADC-AL]</td>
</tr>
<tr>
<td>Patient profile information</td>
<td>“Then they sign up their loved one and their pictures, information about themselves, information about their loved one, the medication they’re on, what they, little bit about themselves, a little bit about their loved one’s background.” [ADC-EK]</td>
</tr>
<tr>
<td>Disliked features</td>
<td>“Sometimes behavior issues where someone is having extraordinary anxiety or—then I think that that might be able to be addressed. It would be challenging to properly do that with just the smiley faces, I think. I thought that was very surface-related, very cursory thing. I’m not sure how incredibly helpful that would be unless there’s some depth later somewhere. That would be my only concern about that.” [ADC-HK]</td>
</tr>
<tr>
<td>Health progress logs</td>
<td>“I found that it had, the activities part of it and how their morning started or how they slept, and things that I deal with hands on here.” [ADC-LR]</td>
</tr>
<tr>
<td>Miscellaneous ways the app enhances care</td>
<td>“That is perfect, because then—in outcomes for us or if we utilize it regularly, it would help us with data collection too.” [ADC-BT]</td>
</tr>
</tbody>
</table>

Fits Within the Workflow

New technology should be compatible with the existing workflow of staff and patients and cause minimal disruptions to optimize adoption. A total of 9 questions assessed the extent to which the CareMOBI app could integrate into the existing health records and workflow of ADS staff (Figure 4). The mean domain score was 5.14 out of 7, indicating that most participants agreed that the CareMOBI app would fit within the workflow. The item with the highest average score was “I often use smartphone apps in my work or daily life” (mean of 6.38 or strongly agree), and the item with the lowest average score was “The use of the proposed app may interfere with the usual follow up of my patients” (mean of 3.36 or slightly disagree). Due to the wording of this item, a low rating has a positive implication for the compatibility of the CareMOBI app with ADS staff workflow.

However, data from the qualitative interviews highlighted several caveats that could preclude the successful integration of the CareMOBI app into the current staff workflow (Table 4). Staff recommended prioritizing the interoperability of CareMOBI with existing electronic health record systems (eg, TurboTAR) to avoid duplicitous documentation and streamline their workflow. Congruent with the quantitative data, staff participants cited that the lack of existing mHealth apps used in their routine workflows meant that staff and caregivers were not familiar with this interface and could be problematic when initially integrating the app into use. To ameliorate this, the staff participants emphasized the need for sufficient training using videos and other methods for both staff and caregivers, as many described being intimidated by the complexity of emerging technologies. Participants (ADC-JL) also voiced concern that documentation in the CareMOBI app would just be “one more thing” that needs to be done and believed it could be challenging to “actually utilize it” given their other responsibilities. Many participants also imparted that the benefits provided by improved e-communication and information sharing could encourage integration into their workflows. One staff member stated that the app was “great [for] communication with physicians” and another remarked that the ability to be able to take pictures of medication bottles and export patient information “clears up so much confusion” for them. Both the qualitative and quantitative data show relative support for using the app in the workflow but the qualitative interview data reveal several barriers that could inhibit integration.
Figure 4. Quantitative survey answer distribution for how CareMOBI fits within the adult day service workflow. mHealth: mobile health.

<table>
<thead>
<tr>
<th>Subtheme</th>
<th>Qualitative feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefits of e-communication</td>
<td>“I actually think it would help a lot, just because it would keep you—it would keep the family more informed, 'cause nowadays everybody uses their cell phones, so it would be right at—right in the palm of their hand. They wouldn’t have to worry about goin’ to their computer. I feel like it would be a great communication with the physicians, letting them know how our members are doing. Also, communication between the staff. Yeah, I thought it was a great tool. I work in another area with home care, and I was thinking how that would work out with that one really well, too. It just kinda would be a great communicator.” [ADC-BW]</td>
</tr>
<tr>
<td>Information sharing</td>
<td>“Also, for the medications where they’re seen, that’s huge ‘cause they don’t know all the time, the majority of the time. Waiting for a doctor’s office to just snap that picture of the bottle, and the RN gets it right away, it clears up so much confusion.” [ADC-NF]</td>
</tr>
<tr>
<td>Suggestions for integrating within existing workflows</td>
<td>“The interoperability, I think that’s the biggie. That we’re entering information one time only so that once you do have your prototype, and it’s down pat, how it maps to the software and, I would say, specifically, TurboTAR ‘cause I think it is the most broadly used in adult day healthcare in California, that whatever terminology you’re using on your app maps to sections within TurboTAR so that that’s automatically populating.” [ADC-DT]</td>
</tr>
<tr>
<td>Training caregivers to use app</td>
<td>“This is so huge because the next wave, they’re gonna be—they’re more technology savvy, so they’re gonna be wanting a lot of this information more so, so we have to keep evolving to be able to help make sure that we are communicating right away and training our staff to use it.” [ADC-NF]</td>
</tr>
<tr>
<td>Feasibility of using app in practice</td>
<td>“I think your issue and hurdle is gonna get people to actually utilize it, especially on the health care side because we already have a thousand things to document and a thousand things to click and a thousand buttons to do. This would just be one more thing. You said something—I—you could put an event in like an urgent situation, but would the caregiver still understand that I’m not sitting here with my cell phone watching for something like that to happen so that they’re still responsible for the care of the member. They’re still responsible for calling 911 in any situation like that.” [ADC-JL]</td>
</tr>
</tbody>
</table>

Table 4. Qualitative subthemes for how CareMOBI fits within the adult day service workflow.

<table>
<thead>
<tr>
<th>Subtheme</th>
<th>Qualitative feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefits of e-communication</td>
<td>“I actually think it would help a lot, just because it would keep you—it would keep the family more informed, ‘cause nowadays everybody uses their cell phones, so it would be right at—right in the palm of their hand. They wouldn’t have to worry about goin’ to their computer. I feel like it would be a great communication with the physicians, letting them know how our members are doing. Also, communication between the staff. Yeah, I thought it was a great tool. I work in another area with home care, and I was thinking how that would work out with that one really well, too. It just kinda would be a great communicator.” [ADC-BW]</td>
</tr>
<tr>
<td>Information sharing</td>
<td>“Also, for the medications where they’re seen, that’s huge ‘cause they don’t know all the time, the majority of the time. Waiting for a doctor’s office to just snap that picture of the bottle, and the RN gets it right away, it clears up so much confusion.” [ADC-NF]</td>
</tr>
<tr>
<td>Suggestions for integrating within existing workflows</td>
<td>“The interoperability, I think that’s the biggie. That we’re entering information one time only so that once you do have your prototype, and it’s down pat, how it maps to the software and, I would say, specifically, TurboTAR ‘cause I think it is the most broadly used in adult day healthcare in California, that whatever terminology you’re using on your app maps to sections within TurboTAR so that that’s automatically populating.” [ADC-DT]</td>
</tr>
<tr>
<td>Training caregivers to use app</td>
<td>“This is so huge because the next wave, they’re gonna be—they’re more technology savvy, so they’re gonna be wanting a lot of this information more so, so we have to keep evolving to be able to help make sure that we are communicating right away and training our staff to use it.” [ADC-NF]</td>
</tr>
<tr>
<td>Feasibility of using app in practice</td>
<td>“I think your issue and hurdle is gonna get people to actually utilize it, especially on the health care side because we already have a thousand things to document and a thousand things to click and a thousand buttons to do. This would just be one more thing. You said something—I—you could put an event in like an urgent situation, but would the caregiver still understand that I’m not sitting here with my cell phone watching for something like that to happen so that they’re still responsible for the care of the member. They’re still responsible for calling 911 in any situation like that.” [ADC-JL]</td>
</tr>
</tbody>
</table>

Likelihood of Adoption

This domain assesses the extent to which staff participants intended to include CareMOBI in their program and use it with clients and caregivers. Five items assessed this domain (Figure 5) and had a mean score of 5.77, a score that suggests most respondents agreed that they intended to use the app after its release. The highest rated item was “I would use the proposed app if I receive appropriate training” (mean of 6.10 or strongly agree) and the lowest rated item was “I have the intention to
use the proposed app routinely for the care of my patients/clients/loved one” (mean of 5.40 or slightly agree).

The qualitative data revealed that many staff participants were motivated to adopt the CareMOBI app (Table 5). One participant said, “we were already talking about how we would utilize it in our program,” reflecting the high quantitative score for the likelihood of adoption. The app was designed to reflect the interdisciplinary nature of ADS, and staff participants were appreciative of this aspect. One participant (ADC-LR) commented that they were “very impressed with it and how it touched on all of the different disciplines that we have to do with in an adult day health center, social work, health, activities.” To increase the likelihood of adoption, staff participants also suggested new features of the app to improve care and user experience, including the option to print reports and collect education-level data from patients to help allocate activities.

Figure 5. Quantitative survey answer distribution for overall likelihood of adoption of CareMOBI. mHealth: mobile health.

![Quantitative survey answer distribution for overall likelihood of adoption of CareMOBI.](image)

Table 5. Qualitative subthemes for overall likelihood of adoption of CareMOBI.

<table>
<thead>
<tr>
<th>Subthemes</th>
<th>Qualitative feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall high likelihood of adoption</td>
<td>“Yep, yep. Very much. I definitely see for those that will——and I...my co-director, so we were already talking about how we would utilize it in our program. By creating direction and maybe even assistance with family members and setting it up.” [ADC-BT]</td>
</tr>
<tr>
<td>Suggested additional features</td>
<td>“I think also, education level. That would help because we do a lot of brain work. If we need to know their education level as to which kind of work to give them. I think education level would be a good question too.” [ADC-EK]</td>
</tr>
<tr>
<td>General impression of the app</td>
<td>“I was very impressed with it and how it touched on all of the different disciplines that we have to do with in an adult day health center, social work, health, activities. Yeah, I touched on everything, so I was really impressed with that.” [ADC-LR]</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

The results of this study show that the CareMOBI app is highly acceptable to ADS staff, but there were questions about the feasibility of implementation. The high scores in all 4 domains of the survey—perceived ease of use, perceived value in clinical care, fit within existing workflows, and end users’ overall likelihood of adoption—demonstrate that ADS staff members generally considered that the app was easy to use; added clinical value; fit into their workflow; and that if given the option, would likely adapt the app into their system.

Prior to developing the app, members of the study team conducted qualitative interviews nationwide to identify facilitators of effective communication between ADS sites and primary care providers. Results indicated that ADS staff wanted to exchange information bidirectionally with other care team members. They sought technology that would reduce their reliance on fax and voicemail while enabling them to share serial observations of a person living with dementia’s day-to-day health with all members of the care team (eg, ADS staff, health care providers, and caregivers) [7].

In this study, qualitative interviews largely qualified the high scores from survey data; however, interviewees expressed practical concerns about the app’s feasibility, particularly about its fit within the existing workflow of ADS. The feasibility of CareMOBI was threatened by the added workload and documentation burden it has the potential to create, especially for overburdened providers who “already have a thousand things
to document and a thousand things to click and a thousand buttons to do.” This concern is not limited to the ADS setting; other research has shown that increased documentation burden in hospitals is a contributing factor to clinician burnout [10,11]. Notably, cost was not mentioned by any of the participants as a concern or barrier to adoption. This is likely because, with low-cost subscriptions (<$5 US $/month) and little to no investment overhead, mHealth apps are reputed as a cost-effective solution to meet the needs of complex patients.

The potential workflow disruption and incompatibility with existing technology infrastructures have been cited as one of the major barriers to successful eHealth implementation in a systematic review, and our study was no exception [9,12]. As another participant pointed out, “The interoperability, I think that’s the biggie. That we’re entering information one time only.” This could be mitigated by making CareMOBI interoperable with ADS billing software which is what most sites rely on for day-to-day practice management. In a future study, a back-end dashboard for CareMOBI will be designed to integrate with existing ADS practice management systems, like TurboTAR. This will avoid what 1 participant referred to as “double documentation.”

Another key priority for future development will be the translation of CareMOBI into different languages. ADS is the most diverse subset of long-term care, with nearly 60% of participants identifying as a racial or ethnic minority or individual [13]. Studies show that ADS can deliver relevant, person-centered care to diverse populations [3]. Adaptations of CareMOBI into other languages will be necessary to improve adoption and inclusivity and ensure the technology reflects the population it is designed to serve.

Successful adoption of new technology in health care settings requires buy-in from staff. It is, therefore, imperative that technology adds clinical value and does not hinder efficiency. While barriers and facilitators of app use in certain health care settings have been well documented, our study is the first to explore the unique perspectives of staff in ADS settings. Resourcing this setting with low-cost mobile technology is especially innovative because ADS sites provide essential care to persons living with dementia but are often resource-constrained and rarely embedded within health care systems.

CareMOBI was developed as a reaction to the overwhelming sentiment that methods of communication in ADS need to be modernized and streamlined to provide optimal care for their clients. Overall, our results from prototype testing show a high level of acceptability among ADS staff with a need for greater attention to feasibility. Recognized by many of the study’s participants, CareMOBI’s biggest strength is its ability to bring all the members of the care team—ADS staff, caregivers, patients, and health care providers—to the same digital table, managing the complex care of older adults as a true team. Previous research by the Agency for Health Care Research and Quality found that care models emphasizing collaborative care were the only effective, evidence-based care models for persons with dementia [14]. By emphasizing effective communication, CareMOBI, with ongoing refinement and enhancement, has significant potential to help streamline and improve care for participants with dementia and other chronic conditions.

**Limitations**

While this study is the first to our knowledge to document the perceived feasibility and acceptability of an mHealth app in the ADS setting, these results should be interpreted with some limitations. The participants of this study were only furnished with a prototype of the CareMOBI app with a limited set of capabilities. At times, this elicited frustration from the participants due to certain functions not working correctly. Conversely, testing with only a prototype precludes assessment of real-world functionality, particularly in the assessment of integration within the workflow. This study consisted of a relatively small, sample of exclusively women participants representing 3 states. Regulations and documentation requirements for ADS and other community-based services differ greatly between states, and preferences around the design and use of technology can be influenced by gender [15]. These findings, therefore, may therefore not be generalizable. A larger, more representative study sample would have potentially offered more robust results. However, these findings—the emphasis on minimizing disruptions to workflow, training, and institutional support—are consistent with other previous implementation studies in a broad array of health care settings [9,12,14].

**Future Directions**

The feedback gleaned in this study will guide future improvements to the CareMOBI app and on-site implementation of the app. Prior to implementation, the CareMOBI team will aim to develop interactive training modules personalized for each type of end user (ie, a caregiver-specific training module vs an ADS staff or nurse training module). Small-scale implementation studies will help develop a robust framework for successful CareMOBI adoption in ADS settings. Further technological development will be pursued to integrate CareMOBI’s interface with other prominent technologies, like TurboTAR, that are already required in ADS settings.

**Conclusions**

This mixed methods study assessed the feasibility and acceptability of ADS staff toward adopting the CareMOBI app in ADS settings. Overall, staff participants were motivated to integrate CareMOBI into their clinical settings because of its potential to ameliorate long-standing communication challenges between ADS staff and other key members of the care team. However, the qualitative interviews highlighted many important potential barriers to successful implementation that researchers will have to consider as they move forward to the next phase of CareMOBI’s development. The CareMOBI team will continue to leverage this important feedback to develop an app that is optimized to fit the needs of all end users and help bring ADS communication into the 21st Century.
Acknowledgments

Research reported in this publication is supported by the National Institute on Aging (NIA) of the National Institutes of Health (NIH) under awards (1) R21AG069801; (2) U54AG063546, which funds the NIA Imbedded Pragmatic Alzheimer and AD-Related Dementias Clinical Trials Collaboratory (NIA IMPACT Collaboratory); and (3) K23AG071948. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References


Abbreviations

ADRD: Alzheimer disease and related dementia
ADS: adult day services
mHealth: mobile health
UI: user interface
UX: user experience
The Impact of a Web-Based Restorative Dentistry Course on the Learning Outcomes of Dental Graduates: Pre-Experimental Study

Rasha Al-Sbei1,2, MEd; Jawdat Ataya1,2, MEd; Issam Jamous1,2, PhD; Mayssoon Dashash1,2, PhD

1Medical Education Program, Syrian Virtual University, Damascus, Syrian Arab Republic
2Faculty of Dental Medicine, Damascus University, Damascus, Syrian Arab Republic

Corresponding Author:
Jawdat Ataya, MEd
Medical Education Program
Syrian Virtual University
Al-Mazzeh Street
Damascus, 35329
Syrian Arab Republic
Phone: 963 992287487
Email: dr.jawdat.ataya@gmail.com

Abstract

Background: Restorative dentistry plays a crucial role in dental practice, necessitating professionals to stay abreast with the latest advancements in the field. The advancement of technology has made web-based learning a widely used method of education delivery in dentistry, providing learners with extensive information and flexibility.

Objective: This study aims to evaluate how effective an online educational course in restorative dentistry is for dental graduates in Syria.

Methods: This study used a pre-experimental study design, with pretest and posttest assessments to measure changes in participants’ knowledge and skills. A total of 21 dental graduates completed the online course in restorative dentistry, which was hosted on Moodle, using the learning management system of the Syrian Virtual University. Participants were provided with a suggested learning sequence and had the flexibility to navigate the course on their own and at their own pace. The course was developed based on the principles of web course design and web-based course development using the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) general instructional design model. The pretest and posttest assessments consisted of 50 multiple-choice questions with a single correct answer, aligning with the course content. Furthermore, participants were asked to complete a course acceptance survey upon finishing the course.

Results: The results showed a significant improvement in the participants’ knowledge of restorative dentistry, supported by a statistically significant $P$ value of less than .05. The effect size of the difference between the pre and posttest indicated that the effect size, as indicated by $\omega^2$, demonstrated a significant 62.1% difference between the pre and posttest, indicating a high and statistically significant effect. Furthermore, the value derived from the Haridy obtained work ratio formula indicated that the educational program was effective, with an effectiveness amount of 3.36%. Additionally, 93% (n=19) of respondents expressed confidence in having gained the expected benefits from the educational course upon its completion.

Conclusions: The findings indicated a notable enhancement in the participants’ understanding of restorative dentistry. The participants’ high satisfaction rate and positive feedback from the course acceptance survey further emphasize the favorable reception of the web-based learning approach. This study highlights the potential of web-based learning in dental education, opening the door for future research in this area. The findings of this study carry important implications for the design and implementation of web-based educational programs in dentistry, suggesting that such programs can serve as an effective tool for continuous professional development in the field.

(JMIR Form Res 2024;8:e51141) doi:10.2196/51141

KEYWORDS
restorative dentistry; online learning; dental education; dental graduates; Syria; education; dental; dentistry; dental practice; effectiveness; educational program; survey
**Introduction**

Restorative dentistry is a crucial aspect of dental practice, involving the use of various materials and techniques to restore the function and aesthetic appearance of teeth [1]. Continuous education, particularly in restorative dentistry, is essential for dental graduates and practitioners to stay updated with the latest developments in the field [2-4].

In recent years, web-based learning, also known as e-learning, has gained popularity as a method of delivering education in dentistry, thanks to technological advancements [5,6]. This approach provides learners with access to a wealth of information while offering the flexibility to learn at their own pace, overcoming the limitations of time and space [7]. In response to this trend, the dental profession has developed web-based dental courses and web-based continuing dental education programs [8]. The global implementation of web-based learning in dentistry has been further accelerated by the COVID-19 pandemic [9,10].

The transition to new education delivery models, such as web-based education, has reshaped teaching methods into a blended learning methods, particularly in dental education [11,12]. Studies have shown that web-based teaching is perceived as equally as effective as traditional classroom approaches in terms of knowledge acquisition and academic performance [1,13]. Web-based dental education programs and courses have been developed to enhance the knowledge of dental students and practitioners, providing them with easy access to a wide range of information and fostering empowerment [14-16].

Several universities and dental schools worldwide have integrated web-based education into their dental programs, acknowledging its advantages and potential to enhance patient health care outcomes through a patient-centered approach and continuous learning [17-19]. However, while web-based learning offers numerous advantages, careful consideration of technical and pedagogical factors is essential to ensure its effectiveness. Hands-on training with manikins during preclinical education remains crucial [19-21].

This study aims to evaluate the effectiveness of a web-based educational course in restorative dentistry in improving the learning outcomes for Syrian dental graduates. Through assessing the influence of this web-based course, we aim to contribute to the understanding of the benefits and effectiveness of web-based learning in dental education.

**Methods**

**Ethical Considerations**

The ethical approval of this study was obtained from the Syrian Virtual University Ethics Committee (237/0) on January 7, 2023. This study involved a pre-experimental web-based course with human participants. The course did not involve any harmful interventions or interactions that could harm the participants. There is a preprint version of this study [22].

Before their participation, informed consent was obtained from all participants. The consent form, written in Arabic, the first language of all participants, provided a clear description of the research objectives and procedures. Participants were assured that their information would be securely stored and used solely for this research. To ensure confidentiality and privacy, the study data collected from participants were anonymized and protected with appropriate security measures. Access to the participants’ identities was restricted to only one of the researchers involved in the study.

The participation questionnaire was indirectly distributed to approximately 50 graduates, and a voluntary and random selection process was conducted to include participants. Out of the 50 participants contacted individually, a total of 21 participants willingly agreed to take part in the study. The selection of participants was not influenced by any specific criteria or biases, ensuring a representative sample for the research. A total of 21 dental graduates completed the online course in restorative dentistry (OCRD) which used a pre-experimental study design with pretest and posttest assessments.

**Overview**

The course was asynchronous and hosted on Moodle, using the learning management system (LMS) of the Syrian Virtual University. Each participant had an account and full access to the course at any time. The course duration was 3 weeks, and it was available on LMS during this duration with full access for participants, it was divided into 4 thematic units with resources including documents, downloadable articles, presentations, short educational videos, examinations, and surveys.

The OCRD was developed by the principal researcher, RA, based on the principles of web course design [23,24], and web-based course development [25,26], using the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) general instructional design model [27]. The course and its contents were reviewed by several experts from the Faculty of Dentistry at Damascus University.

The course primarily focused on theoretical aspects related to clinical topics. The course provided participants with a comprehensive understanding of various clinical topics within the field of dentistry. It aimed to enhance their theoretical knowledge and conceptual understanding of the principles and practices relevant to restorative dentistry. The course covered a wide range of theoretical aspects, including but not limited to dental materials, treatment planning, occlusion, tooth preparation, and aesthetic considerations. Participants were exposed to evidence-based theories, research findings, and best practices in the field. The course was theory about clinical topics and consisted of four main themes: (1) composite restoration including anterior and posterior composite restoration, (2) diastema closure, (3) direct composite veneer, and (4) tooth whitening.

In addition, we supported the course with clinical applications and clinical cases. The participants were given a suggested sequence for optimal learning but were allowed to access and navigate the course at their own pace. Participant contributions were reviewed daily to guide them in their studies and provide...
feedback. The daily feedback was given by the course tutor. Discussion of any issues related to the revised unit was also allowed, and any doubts were addressed. The study outcomes were based on the structure of the course, which was organized into units, lectures, and educational videos. Tools, methods, and resources were selected for use in the course, which included a set of presentations and related educational short videos.

The participants were informed about the pretest and posttest assessments that would be conducted as part of the study, and they were required to complete both assessments. The pretest was administered before the start of the course, while the posttest was conducted immediately after the course ended.

The study population comprised participants who had graduated from the Faculty of Dentistry. The invitation to the course was sent to a group of graduate students. A total of 21 participants agreed and accepted the invitation to participate in the course, and informed consent was obtained from those who accepted the invitation. The participants have been recruited according to the inclusion criteria including participants who had completed all restorative dentistry courses in their undergraduate studies, had access to a laptop, iPad, or smartphone, had good internet access, and had good knowledge of both Arabic and English, including medical terminology. Participants who did not complete the pretest were excluded from the study.

The pretest and posttest assessments were designed to measure the participants’ knowledge of restorative dentistry. The assessments were based on the course content. The assessments consisted of 50 multiple-choice questions with a single correct answer. The questions were designed to cover the 4 main themes of the course. Each multiple-choice question had 1 correct answer and 3 distractor options. A correct response received 2 points toward a total score of 100, while an incorrect response received zero points. The test was web-based, with limited time access. The duration of the test was 40 minutes.

In addition, participants were also asked to complete a course acceptance survey at the end of the course. The survey consisted of Likert scale questions and open-ended questions. The Likert scale questions were designed to measure the participant’s satisfaction with the course, the quality of the course content, and the effectiveness of the course in improving their knowledge of restorative dentistry. The open-ended questions allowed participants to provide additional feedback and comments about the course.

Data collected from the assessments were analyzed using SPSS (version 25.0; IBM). Descriptive statistics, including means and SDs, were calculated to describe the participants’ demographic characteristics and their pretest and posttest scores. The paired-sample 2-tailed t test was used to compare the mean scores of the pretest and posttest assessments. A P value of less than .05 was considered statistically significant.

Results
Overview
The study included 21 participants, of which 12 (57%) participants were female and 9 (43%) participants were male, with ages ranging from 23 to 30 years. Sixteen (76%) participants graduated from public universities and 5 (24%) participants graduated from private universities. The participants were distributed as follows: 12 (57%) participants were master candidates, 8 (38%) participants were general dentists, and 1 (5%) participant was a specialist. All participants had different years of experience in dental practice with 7 (33%) dentists having only 1 year of experience, 6 (29%) dentists having 2 years of experience, 6 (29%) dentists having 3 years of experience, and only 2 (10%) dentists having more than 3 years of experience.

A total of 21 participants completed the course, with course completion tracked by the LMS, and the time taken to complete the course ranged from 2 to 7 days. The pretest and posttest results were collected and stored in Google Forms and analyzed using SPSS. All 21 participants completed both tests without any significant lack of information. The change in knowledge between pretest and posttest results was analyzed using the paired-sample 2-tailed t test. A P value of less than .05 was considered significant. The P values, mean scores, and SDs for the precourse and postcourse tests for the entire test and each section are shown in Table 1.
### Table 1. Results of paired-sample 2-tailed t test to test knowledge level change.

<table>
<thead>
<tr>
<th>Unit and test</th>
<th>Score, mean (SD)</th>
<th>P value</th>
<th>t test (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Entire test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precourse</td>
<td>53.05 (6.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcourse</td>
<td>77.24 (11.8)</td>
<td>&lt;.001</td>
<td>9.592 (21)</td>
</tr>
<tr>
<td><strong>Composite restoration</strong></td>
<td></td>
<td>&lt;.001</td>
<td>4.468 (21)</td>
</tr>
<tr>
<td>Precourse</td>
<td>17.8 (1.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcourse</td>
<td>21.52 (2.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diastema management</strong></td>
<td></td>
<td>&lt;.001</td>
<td>6.664 (21)</td>
</tr>
<tr>
<td>Precourse</td>
<td>11.33 (3.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcourse</td>
<td>17.14 (4.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Direct composite veneer</strong></td>
<td></td>
<td>&lt;.001</td>
<td>8.087 (21)</td>
</tr>
<tr>
<td>Precourse</td>
<td>14.38 (3.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcourse</td>
<td>20.3 (2.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Teeth whitening</strong></td>
<td></td>
<td>&lt;.001</td>
<td>9.118 (21)</td>
</tr>
<tr>
<td>Precourse</td>
<td>9.52 (3.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcourse</td>
<td>18.3 (4.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The normal distribution was tested using the 1-sample Kolmogorov-Smirnov test, which showed that the distribution of the sample was standard. The paired-sample 2-tailed t test was then applied, revealing no statistically significant difference in the post-course test results of the participants, regardless of their age group, gender, academic status, type of university, or prior experience with web-based courses.

**Acceptance Questionnaire**

The internal consistency of the user acceptance questionnaire items, as determined by Cronbach α, was demonstrated to be 0.924, indicating high reliability. In terms of user acceptance, all survey respondents (N=21) reported satisfaction with the web-based course when using a 5-point Likert scale. Moreover, the majority of survey respondents 93% (n=19) believed that they gained the expected benefit from the educational course after its completion.

After analyzing the data, we found that there was no statistically significant difference (with a P value>.05) in the postcourse test results of the participants, regardless of their age group, gender, academic status, type of university, or prior experience with web-based courses.
Table 2. Results of statistical analysis of the user acceptance questionnaire.

<table>
<thead>
<tr>
<th>Number</th>
<th>Questionnaire</th>
<th>Results, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Level of knowledge at the beginning of the course</td>
<td>2.86 (0.727)</td>
</tr>
<tr>
<td>2</td>
<td>Level of knowledge at the end of the course</td>
<td>3.86 (0.478)</td>
</tr>
<tr>
<td>3</td>
<td>The learning process was easy and clear</td>
<td>3.67 (0.658)</td>
</tr>
<tr>
<td>4</td>
<td>I had full control of the learning process</td>
<td>3.81 (0.512)</td>
</tr>
<tr>
<td>5</td>
<td>Time assumed for the learning process was sufficient</td>
<td>3.86 (0.478)</td>
</tr>
<tr>
<td>6</td>
<td>I had full decisions in the sequences of the learning process</td>
<td>3.95 (0.218)</td>
</tr>
<tr>
<td>7</td>
<td>The learning objectives of the course were listed</td>
<td>4.38 (0.590)</td>
</tr>
<tr>
<td>8</td>
<td>The e-content was well organized</td>
<td>4.57 (0.507)</td>
</tr>
<tr>
<td>9</td>
<td>Loads of the educational process were affordable</td>
<td>4.14 (0.727)</td>
</tr>
<tr>
<td>10</td>
<td>The course was organized to allow participants to participate adequately</td>
<td>4.29 (0.784)</td>
</tr>
<tr>
<td>11</td>
<td>The e-content was useful and gave added value</td>
<td>4.52 (0.512)</td>
</tr>
<tr>
<td>12</td>
<td>The e-content was obvious and clear</td>
<td>4.52 (0.512)</td>
</tr>
<tr>
<td>13</td>
<td>I think e-Learning is interesting</td>
<td>4.19 (0.512)</td>
</tr>
<tr>
<td>14</td>
<td>I believe that e-Learning is useful and gives an added value</td>
<td>4.14 (0.727)</td>
</tr>
<tr>
<td>15</td>
<td>I believe that the e-Learning process is easy and feasible</td>
<td>4.19 (0.814)</td>
</tr>
<tr>
<td>16</td>
<td>Log-in to the platform was easy and clear</td>
<td>4.38 (0.669)</td>
</tr>
<tr>
<td>17</td>
<td>Navigation with platform was easy and clear</td>
<td>4.48 (0.512)</td>
</tr>
<tr>
<td>18</td>
<td>The e-content on the platform was well-organized</td>
<td>4.62 (0.498)</td>
</tr>
<tr>
<td>19</td>
<td>The course instructor was active</td>
<td>4.43 (0.507)</td>
</tr>
<tr>
<td>20</td>
<td>The presentations were well-organized and clear</td>
<td>4.62 (0.498)</td>
</tr>
<tr>
<td>21</td>
<td>The recorded presentations were clear and understandable</td>
<td>4.19 (0.512)</td>
</tr>
<tr>
<td>22</td>
<td>The suggested educational videos were useful and related</td>
<td>4.33 (0.577)</td>
</tr>
<tr>
<td>23</td>
<td>The course instructor was available and useful</td>
<td>4.52 (0.512)</td>
</tr>
<tr>
<td>24</td>
<td>The feedback was direct and constructive</td>
<td>4.43 (0.507)</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

This study aimed to evaluate the effectiveness of a web-based educational program in restorative dentistry for increasing the knowledge of Syrian dental graduates. The course was designed based on the fundamental principles of web course design and web-based course development. A total of 21 participants completed the entire course. The results of the study showed a significant increase in the participants’ levels of restorative dentistry knowledge after completing the web-based course, as demonstrated by the postcourse test scores.

The course was also well-received by the participants, as indicated by their positive evaluations. These findings are consistent with previous studies that have investigated the effectiveness of web-based learning in enhancing participants’ knowledge and skills. This success of the course could be attributed to the well-structured presentation and content of the course, as well as the ease of use of the web-based platform (LMS-Moodle) and its flexibility in navigating the course content [29,30].

The highest score in the students’ precourse tests was obtained about composite restoration (first unit), which may be attributed to the fact that composite restoration is widely used and considered the most popular restorative material in dentistry. In contrast, the lowest score in the precourse test was obtained for tooth whitening (fourth unit), because tooth whitening lectures are only given in advanced dental practice programs and postgraduate studies. However, the students’ scores on specific knowledge areas in the postcourse tests showed a notable improvement in all units, including tooth whitening, despite the widespread use of tooth whitening in dental clinics to restore the aesthetic appearance of teeth [31]. This result ensures that the course has been focused on new concepts and interesting ideas that the participants had not been taught before, as well as influencing their daily practice.

The result of this study is consistent with those of Morales - Pérez et al [29], Absi et al [32], and Rosenberg et al [33] regarding the effectiveness of web-based learning and web-based courses in increasing the knowledge of participants.

The current situation in Syria makes attending traditional courses challenging, due to, for example, the lack of transportation, fuel,
and electricity; lack of sufficient and qualified places to hold the courses; and shortage of human resources and staff [34-36].

However, the course offered flexibility in terms of delivery and content review, allowing participants access to the platform 24 hours a day. This approach is similar to previous studies by Murphy et al [37] and Rosenberg et al [33]. Asynchronous web-based learning was also an option that enabled participants to revisit the course materials as needed, based on their commitments and social life as noted by Ruiz et al [38].

Asynchronous web-based learning is considered an attractive, flexible, and convenient option for learners, with lower costs and easy access to information. Similarly, Kenjrawi and Dashash [39] found that asynchronous electronic medical education is an effective and feasible approach for improving the knowledge and attitude of Syrian clinical practitioners [39].

The web-based course proved a valuable option for continuing medical education in Syria given the current circumstances in Syria and after the impact of the COVID-19 pandemic [40]. In dental education, computerized sources of information and virtual reality are increasingly being used as educational tools and have shown promise in training dental students [41]. These technological advancements have the potential to revolutionize dental education and enhance the learning experience of students.

The results of this study indicate that the participants expressed high levels of acceptance and satisfaction with the restorative dentistry course, as evidenced by the questionnaire results. The majority of the cohort (19/21, 93%) reported that the course provided the desired benefit, and all participants reported enjoying it. The well-designed course materials and the instructor’s enthusiasm for delivering the web-based course were also praised by participants. However, there were some areas of the course that needed improvement, such as technical resources and communication methods required to improve the web-based course.

To further assess the effectiveness of the web-based course, future research using larger sample sizes would be beneficial, to evaluate both knowledge and skills in each unit of restorative dentistry separately. Nevertheless, implementing web-based courses of restorative dentistry in continuous medical education programs in Syria is advisable, as it may help dentists stay up to date with minimal requirements.

Limitations
The study did not assess the long-term retention of knowledge or the impact of this web-based course on the participants’ clinical practice. Furthermore, the study focused only on an OCRD. The findings might not apply to other areas of dentistry or other forms of web-based learning.

Conclusion
This study provides evidence that web-based learning can be an effective tool in improving the knowledge and skills of dental graduates about restorative dentistry. The high satisfaction rate expressed by the participants and the positive feedback received through the course acceptance survey indicate that the web-based educational program was well-received by the participants. These findings support the potential of web-based learning in dental education and suggest that it can be a valuable tool for providing continuous education to dental professionals, as it can help dental professionals stay up to date with the latest advancements in their field, which can ultimately benefit the patients they serve.

Acknowledgments
The authors would like to thank Syrian Virtual University for its assistance and all the participants who agreed to take part in this study. All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided APF support for the publication of this article. Artificial intelligence was not used in writing this manuscript.

Data Availability
The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


26. Shattuck M. What we’re learning from quality matters-focused research: research, practice, continuous improvement. Qual Matters 2012;1-29 [FREE Full text]
32. Absi EG, Drage NA, Thomas HS, Nash ES, Newcombe RG. The effectiveness of dental postgraduate courses--are we doing the right thing? Br Dent J 2006 Sep;Suppl:19-23. [doi: 10.1038/sj.bdj.4814068] [Medline: 16964275]

Abbreviations
ADDIE: Analysis, Design, Development, Implementation, and Evaluation
LMS: learning management system
OCRD: online course in restorative dentistry

©Rasha Al-Sbei, Jawdat Ataya, Issam Jamous, Mayssoon Dashash. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete
bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Comparison of the Discrimination Performance of AI Scoring and the Brixia Score in Predicting COVID-19 Severity on Chest X-Ray Imaging: Diagnostic Accuracy Study

Eric Daniel Tenda¹*, MD, DIC, PhD; Reyhan Eddy Yunus²*, MD, MSc; Benny Zulkarnaen³, MD; Muhammad Reynalzi Yugo⁴, MD; Ceva Wicaksono Pitoyo⁵, MD; Moses Mazmur Asaf⁶, MD; Tiara Nur Islamiyati¹, MD; Arierta Pujitresnani³, BSc, MSc; Andry Setiadharma⁴, MD; Joshua Henrina⁴, MD; Cleopas Martin Rumende⁴, MD, Prof Dr; Vally Wulanii⁷, MD; Kuntjoro Harimurti⁸, MSc, MD, PhD; Aida Lydii⁹, MD, PhD; Hamzah Shattii⁶, MEPid, MD, PhD; Pradana Soewondo⁴, MD, Prof Dr; Prasandhya Astagiri Yusuf⁵, BSc, MT, PhD

¹Department of Internal Medicine, Pulmonology and Critical Care Division, Faculty of Medicine Universitas Indonesia, RSUPN Dr. Cipto Mangunkusumo, Universitas Indonesia, Jakarta, Indonesia
²Department of Radiology, RSUPN Dr. Cipto Mangunkusumo, Universitas Indonesia, Jakarta, Indonesia
³Department of Medical Physiology and Biophysics/ Medical Technology Cluster IMERI, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia
⁴Department of Internal Medicine, Geriatric Division, Faculty of Medicine Universitas Indonesia, RSUPN Dr. Cipto Mangunkusumo, Universitas Indonesia, Jakarta, Indonesia
⁵Department of Internal Medicine, Nephrology and Hypertension Division, Faculty of Medicine Universitas Indonesia, RSUPN Dr. Cipto Mangunkusumo, Universitas Indonesia, Jakarta, Indonesia
⁶Department of Internal Medicine, Psychosomatic Division, Faculty of Medicine Universitas Indonesia, RSUPN Dr. Cipto Mangunkusumo, Universitas Indonesia, Jakarta, Indonesia
⁷Department of Internal Medicine, Endocrinology – Metabolism – Diabetes division, Faculty of Medicine Universitas Indonesia, RSUPN Dr. Cipto Mangunkusumo, Universitas Indonesia, Jakarta, Indonesia
*these authors contributed equally

Corresponding Author:
Prasandhya Astagiri Yusuf, BSc, MT, PhD
Department of Medical Physiology and Biophysics/ Medical Technology Cluster IMERI, Faculty of Medicine Universitas Indonesia
Jalan Salemba Raya No.6
Jakarta, 10430
Indonesia
Phone: 62 812 8459 4272
Email: prasandhya.a.yusuf@ui.ac.id

Abstract

Background: The artificial intelligence (AI) analysis of chest x-rays can increase the precision of binary COVID-19 diagnosis. However, it is unknown if AI-based chest x-rays can predict who will develop severe COVID-19, especially in low- and middle-income countries.

Objective: The study aims to compare the performance of human radiologist Brixia scores versus 2 AI scoring systems in predicting the severity of COVID-19 pneumonia.

Methods: We performed a cross-sectional study of 300 patients suspected with and with confirmed COVID-19 infection in Jakarta, Indonesia. A total of 2 AI scores were generated using CAD4COVID x-ray software.

Results: The AI probability score had slightly lower discrimination (area under the curve [AUC] 0.787, 95% CI 0.722-0.852). The AI score for the affected lung area (AUC 0.857, 95% CI 0.809-0.905) was almost as good as the human Brixia score (AUC 0.863, 95% CI 0.818-0.908).

Conclusions: The AI score for the affected lung area and the human radiologist Brixia score had similar and good discrimination performance in predicting COVID-19 severity. Our study demonstrated that using AI-based diagnostic tools is possible, even in low-resource settings. However, before it is widely adopted in daily practice, more studies with a larger scale and that are prospective in nature are needed to confirm our findings.
Introduction

Background

Humans have been learning to adapt to the COVID-19 pandemic. While vaccine development has mitigated the spread, mortality, and morbidity associated with COVID-19, waves of COVID-19 cases are still reported. The main driver of these cases is viral mutation, with the latest mutant, the XBB omicron subvariant, reported to be more virulent and responsible for another COVID-19 wave in Singapore [1]. As of November 1, 2022, a total of 4,707 new cases have been reported throughout Indonesia [2]. This is the highest number of new cases reported since September 1, 2022, and might be attributed to SARS-CoV-2 variants. Importantly, at the time this study was conducted, the number of new daily cases reached 14,518 confirmed cases per day at its peak (January 30, 2021) [3].

Currently, the World Health Organization (WHO) endorses the use of the nucleic acid amplification test, including reverse transcription–polymerase chain reaction (RT-PCR), as the gold-standard diagnostic method for COVID-19 cases [4]. Nonetheless, at the height of the pandemic, the weaknesses associated with the test are accentuated, thus increasing the false-negative rates [5,6]. This is similar to the situation we experienced in Indonesia during our study period.

Therefore, another alternative is needed to help with the COVID-19 triage process. Imaging modalities, primarily chest computed tomography (CT) scan and chest x-ray (CXR), are widely available in most health care facilities. Lung CT is the most effective and sensitive method for diagnosing lung lesions during early disease progression [7]. While CXR is less sensitive than lung CT, it is easier to perform, is more cost-effective, is faster, is more portable, has less radiation exposure, has simpler decontamination, and is more widely distributed. Hence, the latter is the initial radiographic modality of choice amid the COVID-19 pandemic [8].

Several scoring systems have been developed to increase CXR’s diagnostic accuracy and reliability in diagnosing COVID-19. The Brixia scoring system (BSS), one of the most commonly used scoring systems, is a semiquantitative CXR scoring system for COVID-19 diagnosis [9]. However, this system is complicated because of the issue of ground truth, which is influenced by rater experience, interobserver agreement, CXR quality, the facility, and the environment surrounding the scoring process. Moreover, additional work burden is imposed on the radiologists as the system relied on manual, subjective scoring and became a less-interesting option at the height of the pandemic [10].

Fortunately, artificial intelligence (AI) is available to ease the workforce burden amid the COVID-19 pandemic. In the past decade, AI has been advancing rapidly, especially in radiology, with applications mainly to diagnose respiratory diseases such as tuberculosis [11]. With the help of AI and machine learning, diagnostic precision can be optimized through computing algorithms for image identification and analysis, resulting in quantitative image scoring [12]. Another important AI role is to determine COVID-19 severity [13], especially in the setting of limited medical resources, equipment, and hospital beds. The correct identification of disease severity can facilitate efficient, adequate, and prompt treatment for those who need it the most.

One of the most used AI software in the COVID-19 pandemic is CAD4COVID x-ray, which detects and scores COVID-19 pneumonia through the color heat map method. This software has been shown to be significantly superior in diagnosing COVID-19 pneumonia through CXR in 454 participants compared with 6 radiologists with an excellent area under the curve (AUC) [14]. However, studies that examined this software utility for disease severity classification remained scarce, especially in low- and middle-income countries such as Indonesia.

Similar to the study mentioned earlier, our study aimed to compare AI performance against that of radiologists. However, in our study, the radiologists used the BSS.

Objectives

The research questions are two-fold: (1) How does the AI scoring system, using the color heat map methodology, compare to the BSS when assessing CXR in correlation with SARS-CoV-2 RT-PCR results among participants suspected of having COVID-19 pneumonia? (2) What is the effectiveness of the AI scoring system in comparison with the BSS for classifying disease severity in participants suspected of having COVID-19 pneumonia?

The rationale for our research questions is in alignment with the WHO’s guidelines, which recommend the use of chest imaging for the diagnostic evaluation of COVID-19 in scenarios where (1) RT-PCR testing is available but results are delayed or in cases where (2) initial RT-PCR testing returns negative results but there is a strong clinical suspicion of COVID-19 [15]. Our practical experience indicates that these delays in RT-PCR results can extend up to a maximum of 2 weeks. In addition, during periods of high COVID-19 prevalence, the occurrence of false negatives in RT-PCR tests can be notably elevated. A meta-analysis revealed that under conditions of a 50% disease prevalence rate, the rate of misdiagnosis reached 290 out of 1000 participants [16].

Moreover, the wait for a positive RT-PCR result can significantly disrupt the triage system and the clinical flow for patients suspected with COVID-19 infection, consequently leading to delays in the allocation of appropriate treatments. To the best of our knowledge, this is the first study to examine these questions in an Indonesian population.
Methods

Study Design
This retrospective cross-sectional diagnostic study used secondary data from medical records and picture archiving and communication system chest radiography repositories. This study was conducted at the Rumah sakit Dr. Cipto Mangunkusumo (RSCM) National Referral Hospital, Jakarta. This study included adults (aged ≥18 years) hospitalized with suspected COVID-19 and RT-PCR–confirmed COVID-19 infection, with or without comorbidities, from April 2020 to April 2021. This study excluded cases with substandard chest radiography qualities, large lung cavities on CXR, concurrent mediastinal or lung mass, and an interval between RT-PCR and CXR acquisition of >7 days. Data were extracted from inpatient medical records from the RSCM department of internal medicine from April 2020 to April 2021 that met the inclusion criteria. Sampling was performed consecutively until the minimum number of samples was obtained.

Ethical Considerations
This study protocol was reviewed and approved by the Faculty of Medicine, University of Indonesia’s Ethical Board (approval number Nomor KET-588/UN2. FI/ETIK/PPM.00.02/2020). Written informed consent was waived because of the retrospective nature of the study and amid the COVID-19 pandemic. All medical records and CXRs were deidentified and anonymized to ensure patient confidentiality and compliance with privacy standards. No compensations were given to the participants because of the nature of the study.

Operational Definition
Vaccination data were ascertained from history taking and medical records and ordinarily stratified into not vaccinated, vaccinated once, and vaccinated twice. COVID-19 disease severity was determined on hospital admission by emergency medical doctors and was stratified according to the local Indonesian guideline, which adopted the WHO COVID-19 disease severity stratification (Table S1 in Multimedia Appendix 1) [17]. Oxygen saturation data were measured using the transmittance pulse oximeter. They were ordinarily stratified according to normal oxygen saturation (94%-100% on room air), mild to moderate hypoxia (90%-93% on room air), and severe hypoxia (<90% on room air). The RT-PCR data were ascertained from the medical records and used naso- oropharyngeal specimens. Specimen handling and processing for RT-PCR have been described elsewhere [18].

Diabetes was defined according to the American Diabetes Association and the Indonesian Guidelines for the Management and Prevention of Diabetes [19]. Hypertension was defined according to the Indonesian Society of Hypertension Guidelines and the Eighth Joint National Committee [20,21]. Chronic obstructive pulmonary disease and asthma were determined according to the Global Initiative for Obstructive Lung Disease and the Global Initiative for Asthma guidelines, respectively [22,23]. Finally, pulmonary tuberculosis was defined according to the WHO guideline. It was deemed positive if there was a previous history of tuberculosis or active pulmonary tuberculosis [24]. Acute respiratory distress syndrome was defined according to the Berlin criteria [25].

BSS Measurement
The BSS is a semiquantitative method used to measure the severity of lung lesions on CXR. The method has been described in detail elsewhere [9]. Briefly, the lung image on CXR is divided into 6 zones, and each zone can have a score of 0 to 3 with a total maximum score of 18. A total of 2 observers, both board-certified radiologists, measured the score. The third radiologist acted as the ground truth, with >20 years of experience. Each CXR was anonymized before the scoring was performed, and radiologists were blinded to the clinical data. In this study, we focused on the overall CXR score domain of BSS for comparison. We did not establish a predefined BSS cutoff for the classification of positive SARS-CoV-2 RT-PCR results and for the disease severity classification in advance.

AI System for CXR Interpretation
For AI-based CXR interpretation, we used the CAD4COVID x-ray software. It is an advanced AI system built upon deep learning techniques designed for the identification of COVID-19 indicators in frontal CXRs. This cutting-edge system, an extension of the commercially available CAD4TB software (version 6; Thirona), primarily developed for tuberculosis detection in chest radiographs, undergoes initial preprocessing steps, including image normalization and lung segmentation via U-net software. Subsequently, the system uses a patch-based analysis with a convolutional neural network and concludes with image-level classification using an ensemble of networks [14].

The following steps were performed on the software:

1. Digital CXR (Digital Imaging and Communications in Medicine) of patients suspected with COVID-19 infection was pseudoanonymized before image upload with the picture archiving and communication system INFINITT software (INFINITT Healthcare).

2. CXR scoring was performed in 4 consecutive steps, which are as follows:
   - Normalization: this step was to normalize the CXR scale from CXRs with larger or smaller sizes and to be generalized by AI so that it can be processed uniformly.
   - Lung fields segmentation: the AI automatically did this step to delineate the lungs and distinguish them from the rest of the image.
   - Texture analysis: this step was to determine relevant abnormalities in lung segments.
   - Finally, area analysis was done to estimate the percentage of involved lung parenchyma.

3. All filter weights were calculated. The average filter weight was used as a mask on the CXR image to generate a color heat map, which was visualized only in the lung area previously segmented by the previously trained model.

4. The color heat map produced different colors corresponding to its weight. Red, yellow, green, and blue correspond to high,
medium, low, and very low probability of abnormality on the CXR, respectively.

5. The digital CXR was uploaded to the CAD4COVID software to generate 2 AI scorings. First, the affected lung area (ALA) score, with a scale from 0 to 100, was determined according to the total lung volume with abnormalities detected on the CXR. A higher value indicates more lung area that is affected. Second, the COVID-19 probability score, with a scale from 0 to 100, was determined according to the average final weight of all layers. A higher value indicates a higher probability of COVID-19.

The CAD4COVID cutoff for a positive SARS-CoV-2 RT-PCR result and disease severity classification were determined during the study.

Data Analysis

Descriptive Statistics

We summarized baseline characteristics, presenting categorical variables as frequencies (n) and proportions (%). Continuous variables were described as means with SDs for normally distributed data and as medians with IQRs for nonnormally distributed data. The normality of continuous data was assessed using the Kolmogorov-Smirnov test.

Interobserver Reliability

Interobserver reliability was evaluated using the intraclass correlation coefficient (ICC) with a 2-way mixed-effect model (k=2) and consistency. We categorized ICC values as poor, moderate, good, or excellent reliability. Estimated means with their respective 95% CIs were reported for each respective ICC. ICC values of <0.50, 0.50 to 0.75, 0.75 to 0.90, and >0.90 were interpreted as poor, moderate, good, and excellent reliability, respectively. In addition, we also evaluated interobserver reliability based on CXR projections, that is, posteroanterior (PA) and anteroposterior (AP).

Receiver Operating Characteristic Analysis

To assess the diagnostic performance of both AI scoring and Brixia scoring, we used receiver operating characteristic curves and AUC analyses. AUC values of <0.60, 0.60 to 0.70, 0.70 to 0.80, 0.80 to 0.90, and 0.90 to 1 were classified as failure, poor, fair, good, and excellent, respectively. The results were calibrated and internally validated using the Hosmer-Lemeshow test and bootstrapping. A comparison between the AUCs was performed using the DeLong test. A P value of <.05 indicates significant difference.

Optimum Cutoff Values

We determined the optimum cutoff values for AI scoring (probability score and affected lung score) and the BSS to distinguish the RT-PCR results and classify disease severity. The Youden Index method guided our selection process, aiming for the highest sensitivity with a specificity of ≥50% (Tables S2 to S4 in Multimedia Appendix 1).

Diagnostic Performance

We calculated sensitivity, specificity, accuracy, positive predictive value, and negative predictive value based on the chosen cutoff values. The reference standard for COVID-19 diagnosis was RT-PCR, as it is the diagnostic modality needed to confirm COVID-19. The reference standard for disease severity classification was the BSS.

Agreement Tests

The agreement between AI and Brixia scoring in relation to RT-PCR results and disease severity was assessed using the kappa statistic. Kappa values of 0 to 0.20, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80, and 0.81 to 1 were classified as slight, fair, moderate, substantial, and near-perfect agreement, respectively.

We adhered to the Standards for Reporting of Diagnostic Accuracy statement in our reporting of results. Statistical analyses were performed using SPSS for Macintosh (version 27; IBM Corp), MedCalc for Windows (version 20.114; MedCalc Software Ltd), and Stata Statistical Software for Macintosh (version 14; StataCorp LP)

Results

From April 2020 to April 2021, there were 1145 hospitalized patients with COVID-19 in RSCM National Referral Hospital, Jakarta, with complete clinical data, CXR, and RT-PCR. Only 26.2% (300/1145) of the participants met the inclusion and exclusion criteria. The study outline is presented in Figure 1.
Baseline Characteristics
A total of 300 participants (refer to the Methods section) were enrolled in this study. Demographics, comorbidities, clinical data, and radiology scoring are presented in Table 1.

In this study, most hospitalized patients with COVID-19 were aged <60 years (211/300, 70.3%), with a median population age of 52 (IQR 39.0-61.0) years and male gender preponderance (159/300, 53%). In total, >two-thirds of the patients had a negative RT-PCR result (203/300, 67.7%). Moderate disease severity dominated COVID-19 disease severity. Moreover, the 3 most common comorbidities were hypertension, diabetes, and pulmonary tuberculosis.
## Table 1. Baseline characteristics of study participants (N=300).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), median (IQR)</td>
<td>52.0 (39.0-61.0)</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
</tr>
<tr>
<td>&gt;60</td>
<td>89 (29.7)</td>
</tr>
<tr>
<td>&lt;60</td>
<td>211 (70.3)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>159 (53)</td>
</tr>
<tr>
<td>Female</td>
<td>141 (47)</td>
</tr>
<tr>
<td>Oxygen saturation b, n (%)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>132 (44)</td>
</tr>
<tr>
<td>Mild to moderate hypoxia</td>
<td>18 (6)</td>
</tr>
<tr>
<td>Severe hypoxia</td>
<td>25 (8.3)</td>
</tr>
<tr>
<td>RT-PCR c, n (%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>97 (32.3)</td>
</tr>
<tr>
<td>Negative</td>
<td>203 (67.7)</td>
</tr>
<tr>
<td>Disease severity, n (%)</td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>42 (14)</td>
</tr>
<tr>
<td>Moderate</td>
<td>178 (59.3)</td>
</tr>
<tr>
<td>Severe</td>
<td>27 (9)</td>
</tr>
<tr>
<td>Critical</td>
<td>53 (17.7)</td>
</tr>
<tr>
<td>Comorbidities, n (%)</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>77 (25.7)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>96 (32)</td>
</tr>
<tr>
<td>COPD d</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Asthma</td>
<td>4 (1.3)</td>
</tr>
<tr>
<td>Pulmonary tuberculosis</td>
<td>16 (5.3)</td>
</tr>
<tr>
<td>ARDS e</td>
<td>45 (15)</td>
</tr>
<tr>
<td>Length of stay (days), median (IQR)</td>
<td>8.0 (4.0-13.3)</td>
</tr>
<tr>
<td>Overall CXR f score-Brixia score, median (IQR)</td>
<td>3.00 (0.0-9.5)</td>
</tr>
<tr>
<td>AI g-CAD4COVID, median (IQR)</td>
<td></td>
</tr>
<tr>
<td>Probability score h</td>
<td>62.0 (35.75-83.25)</td>
</tr>
<tr>
<td>ALA i score j</td>
<td>6.5 (1.0-27.0)</td>
</tr>
</tbody>
</table>

a Normally distributed data are presented as mean (SD). Otherwise, it is presented as median (IQR).
b The sum of participants falls short of 300 since room air peripheral oxygen saturation data was missing for 125 patients.
c RT-PCR: reverse transcription–polymerase chain reaction.
d COPD: chronic obstructive pulmonary disease.
e ARDS: acute respiratory distress syndrome.
f CXR: chest x-ray.
g AI: artificial intelligence.
h Higher AI probability scores are commensurate with a higher COVID-19 probability.
i ALA: affected lung area.
j Higher AI scores of affected lung area are commensurate with a larger affected lung area.
Interobserver Reliability of Lung Lesion Severity on CXR With the BSS

The analysis showed no statistically significant difference for every lung zone evaluation and the total score in Brixia scoring between the 2 observers (Table 2).

The ICC score for the BSS for every lung zone was >0.75, with good to excellent reliability for zone A (right upper lobe [RUL]) and zone D (left upper lobe [LUL]). Excellent reliability was noted for zone B (right middle lobe), zone C (right lower lobe), zone E (left middle lobe), zone F (left lower lobe), and for the overall CXR score. The Brixia score for each lung zone and the overall CXR score had similar proportions for both AP (187/300, 62.3%) and PA (101/300, 33.7%) CXR projections. The AP had a lower ICC score with a wider 95% CI (Figure S1 in Multimedia Appendix 1).

### Table 2. Difference for every lung zone evaluation and the total score in Brixia scoring between the 2 observers.

<table>
<thead>
<tr>
<th>Scoring parameter</th>
<th>Observer, median (IQR)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Zone A (RUL&lt;sup&gt;b&lt;/sup&gt;)</td>
<td>0.0 (0.0-1.0)</td>
<td>0.0 (0.0-1.0)</td>
</tr>
<tr>
<td>Zone B (RML&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>0.0 (0.0-2.0)</td>
<td>0.0 (0.0-2.0)</td>
</tr>
<tr>
<td>Zone C (RLL&lt;sup&gt;d&lt;/sup&gt;)</td>
<td>1.0 (0.0-2.0)</td>
<td>1.0 (0.0-2.0)</td>
</tr>
<tr>
<td>Zone D (LUL&lt;sup&gt;e&lt;/sup&gt;)</td>
<td>0.0 (0.0-0.0)</td>
<td>0.0 (0.0-1.0)</td>
</tr>
<tr>
<td>Zone E (LML&lt;sup&gt;f&lt;/sup&gt;)</td>
<td>0.0 (0.0-2.0)</td>
<td>0.0 (0.0-2.0)</td>
</tr>
<tr>
<td>Zone F (LLL&lt;sup&gt;g&lt;/sup&gt;)</td>
<td>1.0 (0.0-2.0)</td>
<td>1.0 (0.0-2.0)</td>
</tr>
<tr>
<td>Overall CXR&lt;sup&gt;h&lt;/sup&gt; score</td>
<td>3.0 (0.0-9.25)</td>
<td>3.0 (0.0-10.0)</td>
</tr>
<tr>
<td>Δ Overall CXR score</td>
<td>0.0 (0.0-1.0)</td>
<td>N/A&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>P<.05 is considered statistically significant.  
<sup>b</sup>RUL: right upper lobe.  
<sup>c</sup>RML: right middle lobe.  
<sup>d</sup>RLL: right lower lobe.  
<sup>e</sup>LUL: left upper lobe.  
<sup>f</sup>LML: left middle lobe.  
<sup>g</sup>LLL: left lower lobe.  
<sup>h</sup>CXR: chest x-ray.  
<sup>i</sup>N/A: not applicable.

Performance Comparison Between Color Heat Map–Based AI Scoring Performance and the BSS on CXR Against SARS-CoV-2 RT-PCR Results of Patients Suspected With COVID-19 Infection

Of the 300 participants suspected with COVID-19 infection, only 32.3% (97/300) had a positive RT-PCR result. Owing to the small number of RT-PCR–positive cases and its large measurement error, no scoring system was able to statistically discriminate between patients who had a positive RT-PCR result and those who had a negative RT-PCR result.

Performance Comparison Between Color Heat Map–Based AI Scoring and the BSS on CXR Against COVID-19 Disease Severity

All scores were higher among 86% (258/300) of cases of moderate to critical disease, when compared with 14% (42/300) of cases of mild disease (P<.001). The receiver operating characteristic analysis showed that the AI probability score, AI ALA score, and BSS had excellent discrimination against COVID-19 disease severity (Table 3; Figure 2).

Compared with the performance of BSS to discriminate disease severity (sensitivity 75.7% and accuracy 79.3%), the AI ALA score had better sensitivity and accuracy (Sn 84.5% and accuracy 83.0%), while the AI probability score did not (Sn 68.2% and accuracy 69.7%). The kappa statistic showed there were moderate agreements between AI probability score (κ=0.271±0.050; P<.001), AI ALA score (κ=0.452±0.063; P<.001), and BSS (κ=0.456±0.053; P<.001) against COVID-19 disease severity (Table 3). However, there was no significant difference between the AUC for the AI probability score (AUC 0.787) and the BSS (AUC=0.863), with a difference of 0.076 (SD 0.034, 95% CI 0.010-0.142; P=.04). Similarly, no significant difference was observed between the AI ALA score (AUC 0.857) and the BSS, with a negligible difference of 0.006 (SD 0.023, 95% CI –0.039 to 0.052; P=.76), indicating that both AI scores were comparable to BSS in discriminating disease severity (Table 4).
Table 3. AUC\textsuperscript{a}, optimum cutoff, Sn\textsuperscript{b}, Sp\textsuperscript{c}, positive predictive value, negative predictive value, and diagnostic accuracy for AI\textsuperscript{d} scores (probability and ALA\textsuperscript{e}) and BSS\textsuperscript{f} in discriminating 86% (258/300) of cases of moderate to critical disease from 14% (42/300) of cases of mild disease\textsuperscript{g}.

<table>
<thead>
<tr>
<th></th>
<th>AUC, mean (SD)</th>
<th>P value\textsuperscript{h}</th>
<th>Value, 95% CI</th>
<th>Cutoff</th>
<th>Sn, %</th>
<th>Sp, %</th>
<th>Acc\textsuperscript{i}, %</th>
<th>PPV\textsuperscript{j}, %</th>
<th>NPV\textsuperscript{k}, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI probability score</td>
<td>0.787 (0.033)</td>
<td>&lt;.001</td>
<td>0.722-0.852</td>
<td>≥56</td>
<td>68.2</td>
<td>78.6</td>
<td>69.7</td>
<td>95.1</td>
<td>28.7</td>
</tr>
<tr>
<td>AI ALA score</td>
<td>0.857 (0.024)</td>
<td>&lt;.001</td>
<td>0.809-0.905</td>
<td>≥1</td>
<td>84.5</td>
<td>73.8</td>
<td>83.0</td>
<td>95.2</td>
<td>43.7</td>
</tr>
<tr>
<td>BSS</td>
<td>0.863 (0.023)</td>
<td>&lt;.001</td>
<td>0.818-0.908</td>
<td>≥1</td>
<td>76.7</td>
<td>95.2</td>
<td>79.3</td>
<td>99.0</td>
<td>40.0</td>
</tr>
</tbody>
</table>

\textsuperscript{a}AUC: area under the curve.
\textsuperscript{b}Sn: sensitivity.
\textsuperscript{c}Sp: specificity.
\textsuperscript{d}AI: artificial intelligence.
\textsuperscript{e}ALA: affected lung area.
\textsuperscript{f}BSS: Brixia scoring system.
\textsuperscript{g}Interpretation: <0.60: fail; 0.60 to 0.70: poor classification; 0.70 to 0.80: fair classification; 0.80 to 0.90: good classification; 0.9 to 1: excellent classification.
\textsuperscript{h}P<.05 was considered statistically significant and emphasized by bold texts.
\textsuperscript{i}Acc: accuracy.
\textsuperscript{j}PPV: positive predictive value.
\textsuperscript{k}NPV: negative predictive value.

Figure 2. Receiver operating characteristic curves of artificial intelligence and Brixia scoring systems against disease severity of suspected and confirmed patients with COVID-19. CXR: chest x-ray.

Table 4. The comparison between AI\textsuperscript{a} and Brixia scoring AUC\textsuperscript{b} values against COVID-19 disease severity.

<table>
<thead>
<tr>
<th></th>
<th>AUC, mean difference (SD)</th>
<th>95% CI</th>
<th>P value\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI probability score vs BSS\textsuperscript{d}</td>
<td>0.076 (0.034)</td>
<td>0.010 to 0.142</td>
<td>.04</td>
</tr>
<tr>
<td>AI ALA\textsuperscript{e} score vs BSS</td>
<td>0.006 (0.023)</td>
<td>−0.039 to 0.052</td>
<td>.76</td>
</tr>
</tbody>
</table>

\textsuperscript{a}AI: artificial intelligence.
\textsuperscript{b}AUC: area under the curve.
\textsuperscript{c}P<.05 is considered statistically significant.
\textsuperscript{d}BSS: Brixia scoring system.
\textsuperscript{e}ALA: affected lung area.
The distribution of the 2 binarized groups for all 3 scores is illustrated in a histogram (Figure S2 in Multimedia Appendix 1).

The Hosmer-Lemeshow test shows good calibration and internal validation for the AI probability score ($P=.90$) and AI ALA score ($P=.99$). Calibration and internal validation of the BSS could not be performed.

**Discussion**

**Principal Findings**

In this study, we demonstrated that the BSS had excellent interobserver reliability (ICC $>0.75$) to determine the severity of lung lesions on CXR (ICC $0.950$, 95% CI $0.937$-$0.960$). Our findings are in agreement with previous studies by Mruk et al [26] (ICC $=0.847$, 95% CI $0.816$-$0.848$) and Chaudhari et al [27] (ICC $=0.920$, 95% CI $0.880$-$0.950$). Our study also aligns with that of Mruk et al [26], who reported that PA projection had higher interobserver agreement and reliability than AP projection. Thus, it can be assumed that PA CXR projection has better image quality and thus can affect the interpretation.

In the context of lung anatomy, zone A (RUL) and zone D (LUL) had lower ICC values than the other zones. These findings were similar to those of a previous study by Monaco et al [28], who also reported that LUL had the lowest ICC, followed by the right lower lobe and the RUL. A hypothesis explaining these findings is that the upper zone sparing phenomenon seen in COVID-19 pneumonia, that is, infiltrates rarely develop on the LUL and the RUL. Moreover, anatomical structure juxtaposition with the clavicles and scapula influences the scoring subjectivity of these zones.

Although the interobserver reliability was excellent and the mean difference in the overall CXR score was low, the interobserver score had a very wide range ($Δ$ Overall CXR score $=0.0$-$16.0$). This difference was influenced by a myriad of factors that were not studied in this research. Similarly, van Assen et al [29] reported poor interobserver agreement when assessing disease severity. The disease severity classification was different in 82% of CXR, with 59% showing a 1° difference in disease severity. The differences were primarily observed in the intermediate group (mild and moderate severity), which can greatly affect clinical decision-making for patient management.

While the semiquantitative method may be seen as more convenient, the subjective nature of this scoring system can give rise to multiple problems, including how to determine the ground truth. Thus, supporting data is needed to justify its use [10]. Moreover, the BSS can contribute to health care workers’ burnout, especially radiologists, as the system relies on them to manually sum up the score.

No scoring system was able to statistically discriminate between which patients had a positive RT-PCR result and which ones had a negative RT-PCR result. These findings might be explained by the low prevalence of positive RT-PCR results ($97/300$, 32.3%), which might have been caused by a high false-negative rate, considering the true COVID-19 prevalence during the participant enrollment was 47.6% ($143/300$).

False-negative RT-PCR results can be explained by numerous factors, including exposure time, symptom onset, SARS-CoV-2 virulence, and specimen handling and processing [30-35]. Lai and Lam [31] showed that the interval between the day of exposure and the day of RT-PCR specimen sampling contributed to false-negative rates. False-negative rates of RT-PCR for the 0-, 5-, 8-, and 21-day interval were 100%, 35%, 20%, and 66%, respectively. Viral virulence also contributed to a false-negative rate, as reported in studies by Alteri et al [34] and Petrillo et al [35]. They showed that the false-negative rates of participants infected with low viral virulence were approaching 20% to 30%.

Another possible contributor to an increased false-negative rate was specimen transportation, considering that the viral specimen should be kept at a minimum of $–70$ °C to maintain viral isolation and viability [36]. This factor is crucial owing to the unavailability of an in-house RT-PCR facility in our hospital because it was centralized in the early days of the COVID-19 pandemic. Our hospital is a national referral hospital, which might have led to referral bias. Therefore, most patients referred from lower-tier health care facilities with COVID-19 may have negative RT-PCR results as they may have passed the virulence period and many days have passed since the first onset of symptoms. Finally, according to the Indonesian COVID-19 guidelines, the RT-PCR test did not have to be repeated, contrary to the WHO and European Center for Disease Control guidelines, which state that before discharging patients, 2 negative RT-PCRs are needed [4,37]. Therefore, in our circumstances, solely relying on RT-PCR can lead to an underestimation of the COVID-19 diagnosis, which will impact the decision-making and clinical management of COVID-19. In retrospect, as the RT-PCR turnaround time is high with a high operating cost, at the height of the COVID-19 pandemic, another screening modality that is fast, inexpensive, practical, reliable, and noninvasive is needed for triaging, diagnosing, and quarantining suspected COVID-19 cases as measures to curb the pandemic. Nonetheless, this modality will act as an adjunct to conventional CXR and should be incorporated while waiting for the RT-PCR results [38]. Therefore, we preferred CXR rather than CT scan as the modality that we researched because it is widely available, inexpensive, and has low radiation exposure.

Although our findings found that AI and Brixia scoring were poor COVID-19 diagnostic modalities, they were similar to previous studies that also reported that AI and manual radiologist scoring had similar performance. Murphy et al [14] showed that the AI system (AUC $0.810$) gave similar scoring results and was even superior compared with 6 radiologists ($P<.001$). This was likely because of the heterogeneous lung lesions seen on CXR in patients with COVID-19, including peripheral and diffuse opacities, which made distinguishing COVID-19 from other pulmonary diseases more challenging. In contrast, Chamberlin et al [39] stated that radiologists had a superior diagnostic ability for COVID-19 (AUC $0.936$, 95% CI $0.918$-$0.960$) compared to AI (AUC $0.890$, 95% CI $0.861$-$0.920$), despite similar discriminatory abilities between them.

However, the AI scoring system has several advantages over manual evaluation for CXR interpretation. First, an automated
and quantified AI scoring system will decrease radiologists’ overall work burden. Second, with the help of the AI scoring system, radiologists can increase their accuracy up to 99.05%, which is comparable to that of RT-PCR [40,41]. Moreover, as nonradiologists, especially medical doctors who work on the front line, are the first ones to see the CXR before interpretation by the radiologists, AI use can increase the diagnostic agreement. Hwang et al [42] showed that AI scoring gave similar results to radiologists (AUC 0.714 vs 0.712) but was superior to nonradiologists (AUC 0.714 vs 0.584). Furthermore, hybrid AI use increased diagnostic agreement significantly for both groups (radiologists’ Fleiss κ=0.688, 95% CI 0.665-0.710; nonradiologists’ Fleiss =0.510, 95% CI 0.488-0.533) [42].

In our assessment of AI and BSS performance against disease severity in patients suspected with COVID-19 infection participants, the AI ALA score exhibited higher sensitivity and diagnostic accuracy compared with the BSS, although the difference was not statistically significant. This can be explained by the fact that CAD4COVID can eliminate the technical limitations of conventional CXR, such as its quality. Furthermore, before generating scores, the software normalized the CXR image and segmented lung fields, which optimized CXR image quality [14]. Considering the limited size of our data set, it is not possible to definitively assert the superiority of the AI system over human Brixia. Nonetheless, our analysis indicates a 95% CI that the AI system’s performance is not statistically distinct from that of human Brixia. This finding serves as a promising safety signal, warranting additional comprehensive testing and assessment of AI scoring systems in larger subsequent studies aimed at real-world implementation. Regardless, both had excellent discrimination without significant differences in AUC.

In the context of evaluating AI and Brixia scoring in assessing disease severity among patients suspected with COVID-19 infection participants, both AI scoring systems demonstrated notably higher sensitivity and diagnostic accuracy when compared with the BSS. This superiority can be attributed to the capacity of CAD4COVID to overcome the inherent technical limitations often associated with conventional CXR imaging, including variations in image quality. Notably, CAD4COVID effectively normalized CXR images and meticulously segmented lung fields before generating probability and ALA scores, thereby enhancing overall image quality. Notably, both AI and human Brixia scoring exhibited excellent discrimination without significant disparities in the AUC.

To the best of our knowledge, this is the first study to compare the AI scoring performance using the CAD4COVID software based on AI probability and AI ALA score with the BSS against disease severity in patients suspected with COVID-19 infection participants. Our findings align with those of Guiot et al [43], albeit with a different modality. In their research, CAD4COVID-CT, through the ALA and CT severity score (CT-SS), was able to predict the length of stay, the odds of intensive care unit admission, the odds of mechanical ventilation, and the odds of in-hospital mortality [43]. The cutoff value chosen for odds of intensive care unit admission was CT-SS 14 with an AUC of 0.84 (95% CI 0.79-0.90) and, for odds of mechanical ventilation, it was CT-SS 16 with an AUC of 0.71 (95% CI 0.63-0.78).

The sensitivity, specificity, and accuracy of the BSS were 75.6%, 100%, and 78.4%, respectively. These values were lower than reported by Abo-Hedibah et al [44]. In their study, the BSS sensitivity, specificity, and accuracy in diagnosing moderate disease were 90.4%, 100%, and 94.6%, respectively, whereas in diagnosing severe disease, they were 100%, 84.5%, and 86.7%, respectively. Nevertheless, they referred to disease severity stratified by the WHO, whereas our study referred to the Indonesian national guidelines.

The clinical implications of these findings are 2-fold. First, with the help of AI scoring, a clinician can more confidently exclude moderately to critically ill patients with COVID-19. However, the AI scoring system is image based and not clinically applicable. Thus, it will generate conflicting results when the disease severity classification relies on clinical criteria, as seen in the WHO and Indonesian guidelines [17,45]. In contrast, the US National Institutes of Health COVID-19 guidelines do not solely rely on clinical data [46]. For example, moderate COVID-19 infection can be diagnosed if there are pulmonary infiltrates on imaging. As COVID-19 pneumonia can be asymptomatic or without typical signs and symptoms of pneumonia, we argue that disease severity should be stratified according to the National Institutes of Health guideline [46].

Second, minimal pulmonary lesions on the CXR that are sometimes undetected by manual readers could be identified by AI. The AI probability and the ALA score had higher sensitivities than the BSS to rule out moderate to critical disease. The AI probability and the ALA score can mitigate drawbacks when relying on clinical judgment and conventional radiographs. These findings are significant, as the downstream effect will include patient management, that is, outpatient or inpatient treatment.

According to our study results, we propose the incorporation and clinical application of AI use on CXR as an ancillary diagnostic tool for patients suspected with COVID-19 infection in a structured algorithm. We hope that for future COVID-19 outbreaks, this algorithm can shorten triage and diagnostic time and shorten clinical decision-making as to whether the patients need to be quarantined or hospitalized. As AI and Brixia scoring did not have discriminatory ability against RT-PCR results in suspected COVID-19 cases, we hope that with the addition of clinical and laboratory data, a more precise diagnostic model can be developed. In contrast to the RT-PCR results, the AI and Brixia scoring had an excellent ability to discriminate disease severity in patients suspected with COVID-19 infection, with superior sensitivity and accuracy observed for the former. Thus, AI scoring can be considered for CXR interpretation because of the clinical–radiological incompatibility that can sometimes be observed in patients with COVID-19 pneumonia.

Our study showed that the AI scoring system has the potential to become a disease severity classifier for patients with suspected COVID-19 infection. As the AI scoring system was generated through machine learning, with more data available to train the system, the AI scoring accuracy will continue to increase. We proposed an algorithm for AI incorporation and...
AI application on CXR as an ancillary diagnostic test for patients with COVID-19 (Figure 3).

Figure 3. Proposed algorithm of artificial intelligence (AI) incorporation and AI application on chest x-ray as an ancillary diagnostic test on patients with COVID-19. ALA: affected lung area; CXR: chest x-ray; Prob: probability; RT-PCR: reverse transcription–polymerase chain reaction.

Limitations
Our study has several limitations. First, the study population is relatively small compared with the COVID-19 prevalence during the participant’s enrollment, owing to missing data and the study’s exclusion criteria. Second, as our hospital is a national referral hospital, most patients came with moderate to critical disease and presented with other pulmonary lesions. Furthermore, several study variables could not be retrieved as they were not incorporated into the patient’s medical record. Finally, in the early days of the pandemic, our hospital did not have an in-house RT-PCR facility because of the centralized specimen processing, so the specimen had to be delivered to another facility, which could further compromise RT-PCR results.

Conclusions
The AI score for the ALA and the human radiologist Brixia score had similar and good discrimination performance in predicting COVID-19 severity. Our study demonstrated that using AI-based diagnostic tools is possible, even in low-resource settings. However, before it is widely adopted in daily practice, more studies with a larger scale and prospective in nature are needed to confirm our findings.

Acknowledgments
The authors thank Delft Imaging for providing CAD4COVID free of charge. The research team institutions have also signed a joint research arrangement between Universitas Indonesia, Dr. Cipto Mangunkusumo National Referral Hospital, and Delft Imaging System, encompassing rights and obligations for each party on data confidentiality during and after the research period. This study was supported by Universitas Indonesia PUTI Q1 2022 Research Grant NKB-429/UN2. RST/HKP:05.00/2022. The code for CAD4COVID is not publicly available as it is proprietary software. Any questions that arise regarding the CAD4COVID software can be directed to the Delft Imaging website [47]. The authors are also grateful for the valuable assistance provided by Dr Kahlil Gibran in acquiring data. All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided APF support for the publication of this article.
Data Availability
The data that support the findings of this study are available from the corresponding author on reasonable request.

Authors’ Contributions
EDT, REY, and BZ were instrumental in the conceptualization, formal analysis, investigation, methodology, and writing of the original draft. MRY contributed to data curation and played a significant role in reviewing and editing the manuscript. CWP provided supervision and was involved in the manuscript review and editing process. MMA also contributed to data curation and manuscript editing. TNI was responsible for formal analysis and contributed to the writing by reviewing and editing. AP provided resources and data curation. AS and JH were involved in data curation, formal analysis, and enhanced the manuscript through their reviewing and editing efforts. CMR, VW, KH, AL, HS, and PS provided supervision and lent their expertise to the reviewing and editing of the writing. Finally, PAY was pivotal in project administration and resources, and provided supervision in addition to reviewing and editing the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables and figures.

References


Abbreviations

AI: artificial intelligence
ALA: affected lung area
AP: anteroposterior
AUC: area under the curve
BSS: Brixia scoring system
CT: computed tomography
CT-SS: computed tomography severity score
CXR: chest X-ray
ICC: intraclass correlation coefficient
LUL: left upper lobe
PA: posteroanterior
RSCM: Rumah sakit Dr. Cipto Mangunkusumo
RT-PCR: reverse transcription–polymerase chain reaction
RUL: right upper lobe
WHO: World Health Organization

Edited by A Mavragani; submitted 27.02.23; peer-reviewed by Z Zrubka, T Tillmann; comments to author 24.08.23; revised version received 28.09.23; accepted 29.12.23; published 07.03.24.

Please cite as:
Comparison of the Discrimination Performance of AI Scoring and the Brixia Score in Predicting COVID-19 Severity on Chest X-Ray Imaging: Diagnostic Accuracy Study
JMIR Form Res 2024;8:e46817
URL: https://formative.jmir.org/2024/1/e46817
doi:10.2196/46817
PMID:38451633

https://formative.jmir.org/2024/1/e46817 JMIR Form Res 2024 | vol. 8 | e46817 | p.1212
(page number not for citation purposes)
Original Paper

Developing a Novel Web-Based Self-Management Support Intervention for Polycystic Ovary Syndrome: Mixed Methods Study With Patients and Health Care Professionals

Carol Percy1*, BSc, PhD; Andrew Turner1*, BSc, PhD; Charys Orr2, BSc, MSc

1Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom
2Harris Church of England Academy, Diocese of Coventry Multi Academy Trust, Coventry, United Kingdom
*these authors contributed equally

Corresponding Author:
Carol Percy, BSc, PhD
Centre for Intelligent Healthcare
Coventry University
Priory Street
Coventry, CV1 5FB
United Kingdom
Phone: 44 2477 659 337
Email: c.percy@coventry.ac.uk

Abstract

Background: Polycystic ovary syndrome (PCOS) represents a significant global health burden requiring urgent attention. This common chronic endocrine and cardiometabolic condition affects around 1 in 10 women and individuals assigned female at birth, with significant adverse effects on well-being, quality of life, and mental health, as well as serious and complex long-term health consequences. International guidelines for best health care practice recommend the provision of comprehensive cognitive behavioral interventions to support self-management and improve health outcomes for those living with PCOS. Web-based health interventions have the potential to meet this need in an accessible and scalable way.

Objective: We aim to identify barriers to self-management and psychological well-being in women with PCOS and adapt a web-based self-management program to provide a prototype digital support intervention for them.

Methods: We adapted an existing support program (HOPE) for PCOS using the antecedent target measure approach. We conducted qualitative interviews with 13 adult women living with PCOS, 3 trustees of a patients with PCOS advocacy charity, and 4 endocrinologists to identify “antecedents” (barriers) to self-management and psychological well-being. Framework analysis was used to identify potentially modifiable antecedents to be targeted by the novel intervention. At a national conference, 58 key stakeholders (patients and health professionals) voted for the antecedents they felt were most important to address. We used research evidence and relevant theory to design a prototype for the PCOS intervention.

Results: Voting identified 32 potentially modifiable antecedents, relating to knowledge, understanding, emotions, motivation, and behaviors, as priorities to be targeted in the new intervention. A modular, web-based prototype HOPE PCOS intervention was developed to address these, covering six broad topic areas (instilling HOPE for PCOS; managing the stress of PCOS; feeding your mind and body well; body image, intimacy, and close relationships; staying healthy with PCOS; and keeping PCOS in its place).

Conclusions: We identified barriers to self-management and psychological well-being in women with PCOS and used these to adapt a web-based self-management program, tailoring it for PCOS, which is a comprehensive group intervention combining education, empowerment, lifestyle management, peer support with cognitive behavioral tools, and goal-setting (to be delivered by peers or codelivered with health care professionals). The modular structure offers flexibility to adapt the program further as new clinical recommendations emerge. The intervention has the potential to be delivered, evaluated for feasibility, and, if effective, integrated into health care services. Self-management interventions are not designed to replace clinical care; rather, they serve as an additional source of support. The HOPE PCOS program conveys this message in its content and activities. Future research should evaluate the prototype intervention using primary outcomes such as measures of psychological well-being, self-management self-efficacy, depression, anxiety, and PCOS-related quality of life. They should also assess the intervention’s acceptability, scalability, and cost-effectiveness.
Introduction

A Complex Condition With Serious Physical and Mental Health Consequences

Polycystic ovary syndrome (PCOS) represents a significant global health burden requiring urgent attention [1]. It is a common chronic cardiometabolic condition affecting around 1 in 10 women and individuals assigned female at birth. PCOS has diverse symptoms, for example, acne, alopecia, hirsutism, obesity, impaired glucose metabolism, menstrual disturbances, and subfertility [2-4]. It has significant adverse effects on well-being, quality of life, and mental health. For example, compared with healthy controls, women with PCOS have lower health-related quality of life and higher scores on symptoms of depression and anxiety [5-9], and appear to be at risk for disordered eating [10-14] and unhealthy weight management practices [15]. There is also evidence that women with PCOS may experience impaired sexual satisfaction [16], which may impact relationship satisfaction for those in intimate relationships [17]. The adverse mental health impact of PCOS may have been worsened during the COVID-19 pandemic, as patients faced an uncertain risk of contracting COVID-19, more limited access to health care services, and barriers to using their normal support strategies [18].

PCOS has serious and complex long-term health consequences [19,20]. Women with PCOS are at increased risk of sleep disturbances [21] and obstructive sleep apnea [22,23] and appear to have an elevated risk for postpartum depression [24]. They also have an increased risk of developing type 2 diabetes and cardiovascular complications [4]. Hospital admission rates for patients with PCOS have been reported to be twice as high as those without the condition [25].

The Need for Self-Management Support

PCOS is a long-term condition with no cure, having a wide range of treatment options but no single treatment that targets every symptom or concern. Treatments are often complex and multifaceted and often include recommendations for lifestyle change, for example, altering diet and increasing physical activity [4]. Considerable self-management is required on the part of those living with PCOS, including adherence to treatment. There are wide disparities in adherence, with 1 review suggesting adherence rates ranging from 21.7% to 86% [26]. A large-scale survey suggests women with PCOS are open to adjusting their diet and physical activity to improve their health, but few achieve their goals [27].

Recent qualitative research [28-33] has identified that women with PCOS face a range of barriers to following health professionals’ lifestyle advice, including limited access to credible information, time, cost, and lack of access to safe exercise spaces, with some women lacking adequate social support, having unsupportive partners, or struggling to prioritize their own health because of responsibilities for feeding children [29]. In addition, women’s motivation and capability to adhere to recommended lifestyle changes may be impacted by the complex and multifaceted nature of PCOS, including fatigue, anxiety, depression, difficult emotions, disordered eating, the impact of stress, sleep disturbances, and a lack of critical health literacy [28,29,31-33]. Features of PCOS itself may add to the burden of following lifestyle advice, especially around weight management. For example, altered regulation of gut hormones and energy expenditure may affect women’s capability to follow diet and exercise regimes [28].

The Need for Comprehensive, Evidence-Based, Coproduced Interventions

Patients with PCOS value specialized, integrated health care services [34], and it has been argued that the ideal model of care would be “evidence-based, patient-centered, codeveloped by consumers and health professionals” [35]. International guidelines advise (in addition to appropriate medical treatment) “Comprehensive health behavioral or cognitive behavioral interventions could be considered to increase support, engagement, retention, adherence, and maintenance of a healthy lifestyle and improve health outcomes in women with PCOS.” [4]. There is therefore scope for developing new interventions that combine education and cognitive behavioral approaches with peer support to promote physical activity, healthy eating, and emotional well-being in women with PCOS. This multifaceted approach reflects the tradition of long-term self-management that is well established in other endocrine conditions, such as diabetes [36]. Self-management is arguably broader than lifestyle change, referring to what patients must do to manage their own health in collaboration with health professionals [37,38]. This may include behavior change but also requires emotional regulation, coping, and maintaining general well-being, such as self-worth, a positive outlook, and hope [39].

Objectives for This Study

The first aim of the study was to identify barriers to self-management and psychological well-being in women with PCOS. The second aim was to combine insights from these data to cocreatively adapt an existing web-based program to offer a prototype peer-delivered, self-management intervention to empower and support women with PCOS to enhance their psychological well-being and, if they chose to, to make appropriate lifestyle changes to self-manage their condition.

To develop our novel self-management intervention for PCOS, we selected an existing intervention that has been researched and evaluated with multiple patient groups. The HOPE program [40] is a complex intervention for long-term condition self-management with a distinctive theory and evidence base from positive psychology [41]. The HOPE program has reduced anxiety and depression and increased positive well-being for...
people with a range of health conditions and support needs, for example, people living with multiple sclerosis [42], parents of children with developmental disorders [43], and people living with and after cancer [44].

Methods

Approach to Intervention Development

The UK Medical Research Council’s (MRC) guidance on developing complex interventions [45,46] recommends that intervention development should be “a dynamic iterative process, involving stakeholders, reviewing published research evidence, drawing on existing theories, articulating program theory, undertaking primary data collection, understanding context, paying attention to future implementation in the real world, and designing and refining an intervention using iterative cycles of development with stakeholder input throughout” [46]. The most recent framework commissioned by the National Institute of Health Research and the MRC [47] further emphasizes the importance of developing interventions that are “implementable, cost-effective, transferable, and scalable in real-world conditions.” This may include adapting existing interventions and using existing, for example, digital infrastructure, rather than developing wholly new interventions ab initio. We chose to adapt and tailor an existing intervention to build on its previous successes and use its proven, scalable digital infrastructure to offer a PCOS intervention that would be as implementable, transferable, and scalable as possible.

The HOPE cancer program from which this intervention was adapted [44] was developed following the MRC recommended processes and continues to undergo refinements as each cohort or variant of the program is completed. In adapting HOPE for PCOS, we conducted an initial needs assessment with key stakeholders, reviewed published research evidence from primary sources, review papers, and international clinical guidelines, and drew on established and emerging theories about psychological well-being to produce a prototype digital intervention [48] that could be subjected to feasibility testing.

Underpinning Theory and Evidence: Self-Management Versus Behavior Change

Interventions in the field of health behavior are more likely to be successful if they are theory-based [49]. A number of different approaches were considered for adapting the existing HOPE program, including the behavior change wheel [50] and intervention mapping [51]. However, both of these approaches require very clear specification at the outset of the intervention design of what the target behavior is that the intervention seeks to increase, decrease, promote, or modify. “Lifestyle change” is recommended as a first-line treatment in relation to the management of PCOS, and this often refers to changes to diet and physical activity. However, there is as yet no international consensus as to what precise optimal self-management behaviors might be in PCOS, and it was felt that the groundwork was lacking for the development of a behavior change intervention per se. Given the diversity of the PCOS population, for example some are lean, some are living with overweight and obesity, some are trying to conceive, others are not, some are physically active, some are more sedentary, some have issues around eating, and others are not, the development of a behavior change intervention to meet the needs of all these was beyond the scope of this study. As our objective was to adapt the existing HOPE program to support self-management, reduce depression and anxiety, and promote psychological well-being rather than develop a completely new intervention from scratch, we used the design process outlined in the following sections.

Cocreation Process

An iterative process of cocreation was used to develop the intervention, using a mixed methodology. An initial needs assessment was conducted with key stakeholders, following the antecedent target measure (ATM) approach [52,53]. This is a flexible model that has been used in the development of previous self-management interventions, for example, for people living with cancer or dementia [54,55].

Initial Data Collection

Semistructured interviews were conducted with women living with PCOS, trustees of the Verity UK PCOS charity who themselves had PCOS, and endocrinologists who provided services to patients with PCOS at a large publicly funded teaching hospital.

Participant Characteristics

Those with lived experience of PCOS were 13 adult women recruited through purposive sampling from social media. All had received a formal diagnosis of PCOS from health professionals at least 6 months before the study and took part in individual interviews. The Verity trustees were 3 adult women, all diagnosed with PCOS by National Health Service (NHS) clinicians, who volunteered to be interviewed to assist with the “cocreation” of the new intervention. The health professionals interviewed in December 2016 were 4 endocrinologists working for a large UK NHS teaching hospital, which has a specialist clinic for PCOS.

Procedure

Interviewees were presented with the problem statement “Some women with PCOS struggle to cope and self-manage effectively” and asked if they agreed and, if so, to give all the reasons why. The reasons given (“antecedents” in the terminology of ATM) were typically difficulties women experience with PCOS and in self-managing their condition. Interviews continued until participants felt they had exhausted all the antecedents they viewed as relevant for PCOS. Antecedents and the conversation through which they were generated were audio recorded and transcribed verbatim for the patient interviews. Antecedents were transcribed contemporaneously for the trustee and health professional interviews.

Analysis

All the antecedents generated across interviews were collated and analyzed thematically following the framework analysis approach to theme development [56] to identify potential targets for intervention. When targets were chosen for intervention, only those considered to be potentially amenable to change by a self-management intervention were selected. For example,
the following antecedents were excluded: endocrine characteristics of PCOS itself, social and environmental factors such as stigmatizing behavior from others, and issues with the health care system such as general practitioners’ awareness of PCOS. Once antecedents judged to be unsuitable targets for the intervention were excluded, 36 potentially modifiable antecedents remained.

**Stakeholder Voting on Priorities for Intervention**

A web-based survey was created listing the 36 modifiable antecedents, and delegates at a national public engagement conference were asked to rate each of these on a 5-point scale of how important they would be to target in a self-management intervention (0=not at all important, 1=somewhat important, 2=important, 3=extremely important, and 4=crucial). Conference delegates included patients with PCOS and advocates and some health professionals treating PCOS, of whom 58 voted in the survey. All modifiable antecedents scoring 2.5 or above were retained as intervention targets; antecedents scoring below 2.5 were discarded.

**Literature Searches for Intervention Targets**

Literature searches were conducted, and clinical guidelines were consulted to generate further intervention targets. Review and systematic review papers [2,16,57-68] were reviewed alongside an international clinical guideline for the management of PCOS [4]. No new targets were generated from the literature, which confirmed the findings from the needs assessment with key stakeholders.

**Adaptation of Existing Digital Intervention and Consultation With Stakeholders**

The existing HOPE program has previously been specified using the Practical Reviews in Self-Management Support (PRISMS) self-management intervention taxonomy [69], and new intervention components were mapped onto or added to the program specification. New program content specific to the prototype PCOS program was created to complement the existing HOPE program materials. A paper prototype was presented to key stakeholders from the PCOS support charity, and following discussion, the new program materials were created and uploaded into the secure web-based platform that hosts the HOPE program. This produced a 6-week web-based intervention ready for feasibility testing.

**Ethical Considerations**

This study received clearance from the Coventry University Research and Ethics Governance Committee (approvals P40730, P44631, and P45355). For the data collection from health care professionals, NHS Research Ethics Committee approval was not required. A letter of access for research was provided by the Research, Development, and Innovation office of the University Hospitals Coventry and Warwickshire NHS Trust. Participants provided written, informed consent before being interviewed. Patients with PCOS and health care professionals at the public engagement event completed a digital informed consent statement before completing the web-based voting survey. Interview transcripts were anonymized for analysis. No personally identifiable data were collected in the voting survey. No compensation was offered to any participants.

**Results**

**Stakeholder Needs Assessment and Voting**

Table 1 shows the list of modifiable antecedents presented in the next stage of the intervention design. Figure 1 shows a diagrammatic representation of the logic model developed from the PCOS self-management needs assessment done with stakeholders.
Table 1. Modifiable antecedents of polycystic ovary syndrome (PCOS) self-management and psychological well-being in order of priority, as voted on by patient and health care professional delegates at the public engagement event.

<table>
<thead>
<tr>
<th>Potentially modifiable antecedent</th>
<th>Mean score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge and understanding: long-term implications of PCOS</td>
<td>3.34</td>
</tr>
<tr>
<td>Knowledge and understanding: how to get the most from the health care system</td>
<td>3.31</td>
</tr>
<tr>
<td>Knowledge and understanding: how to eat well in PCOS</td>
<td>3.26</td>
</tr>
<tr>
<td>Emotions: low self esteem</td>
<td>3.21</td>
</tr>
<tr>
<td>Knowledge and understanding: what self-management of PCOS entails</td>
<td>3.19</td>
</tr>
<tr>
<td>Knowledge and understanding: how to use physical activity to manage PCOS</td>
<td>3.12</td>
</tr>
<tr>
<td>Knowledge and understanding: which information sources to trust</td>
<td>3.12</td>
</tr>
<tr>
<td>Motivation: women may find it hard to maintain motivation long-term</td>
<td>3.12</td>
</tr>
<tr>
<td>Emotions: feeling not properly feminine</td>
<td>3.09</td>
</tr>
<tr>
<td>Knowledge and understanding: the basic biology of PCOS</td>
<td>3.09</td>
</tr>
<tr>
<td>Emotions: depression and low mood</td>
<td>3.07</td>
</tr>
<tr>
<td>Behaviors: women may not follow a PCOS-healthy diet</td>
<td>3.02</td>
</tr>
<tr>
<td>Emotions: struggling to make healthy choices</td>
<td>3.02</td>
</tr>
<tr>
<td>Behaviors: women may eat in disordered ways, for example, binge eating and overeating</td>
<td>3.00</td>
</tr>
<tr>
<td>Motivation: women may lack motivation to self-manage</td>
<td>3.00</td>
</tr>
<tr>
<td>Emotions: embarrassment and shame</td>
<td>2.95</td>
</tr>
<tr>
<td>Motivation: women may have low confidence to self-manage</td>
<td>2.95</td>
</tr>
<tr>
<td>Behaviors: women may not use physical activity to manage their PCOS effectively</td>
<td>2.91</td>
</tr>
<tr>
<td>Emotions: lack of trust in health professionals</td>
<td>2.91</td>
</tr>
<tr>
<td>Emotions: anxiety and fear</td>
<td>2.90</td>
</tr>
<tr>
<td>Behaviors: women may not prioritize their own needs effectively</td>
<td>2.88</td>
</tr>
<tr>
<td>Motivation: women may lack clear self-management goals</td>
<td>2.79</td>
</tr>
<tr>
<td>Behaviors: women may avoid or give up on self-management</td>
<td>2.76</td>
</tr>
<tr>
<td>Emotions: frustration and anger</td>
<td>2.71</td>
</tr>
<tr>
<td>Emotions: loneliness and isolation</td>
<td>2.71</td>
</tr>
<tr>
<td>Motivation: women may not find physical activity enjoyable</td>
<td>2.67</td>
</tr>
<tr>
<td>Skills: women need better skills to communicate needs to health professionals</td>
<td>2.62</td>
</tr>
<tr>
<td>Behaviors: women may not self-monitor their menstrual cycle effectively</td>
<td>2.60</td>
</tr>
<tr>
<td>Motivation: women may be in denial and avoid trying to self-manage</td>
<td>2.60</td>
</tr>
<tr>
<td>Motivation: women may have unrealistic goals</td>
<td>2.59</td>
</tr>
<tr>
<td>Knowledge and understanding: women may have unrealistic expectations of the female body</td>
<td>2.57</td>
</tr>
<tr>
<td>Behaviors: women may isolate themselves from social contact</td>
<td>2.52</td>
</tr>
<tr>
<td>Behaviors: women may hide their self-management activity from others^</td>
<td>2.47</td>
</tr>
<tr>
<td>Motivation: women may not be ready yet to start self-managing^</td>
<td>2.41</td>
</tr>
<tr>
<td>Behaviors: women may not take medications as prescribed^</td>
<td>2.21</td>
</tr>
<tr>
<td>Behaviors: women may self-harm^</td>
<td>2.17</td>
</tr>
</tbody>
</table>

^Potential antecedents scoring below 2.5 were discarded as intervention priorities. This left 32 antecedents that were targeted by the new intervention, 10 of which were thematically categorized in the framework analysis under behaviors, 9 under emotions, 8 under knowledge and understanding, 8 under motivation, and 1 under skills.
In a typical ATM design, logic models are often presented as linear processes in which 1 antecedent, such as knowledge or understanding, is represented as preceding another, such as motivation or behavior. We have selected the metaphor of interlocking gears for our schematic logic model on the assumption that knowledge, understanding, skills, motivation, emotions, and behavior interlock and may drive or impede each other, at times facilitating or hampering effective self-management. For example, a person might gain helpful knowledge about their condition, but without developing skills to apply this to practice, they might be unable to engage in effective self-management behaviors. Similarly, a person might have adequate knowledge, understanding, and skills to manage their condition, but difficult emotions might adversely impact their motivation and prevent effective self-management. This logic model was used to inform the adaptation of the existing HOPE program framework to produce a specific version for PCOS.

Mapping Modifiable Antecedents Onto Intervention Content

Table 2 shows how the antecedent “problems” derived from the logic model were mapped onto solution-focused content across 6 sessions of the prototype intervention.
### Table 2. Mapping of antecedents of polycystic ovary syndrome (PCOS) self-management and psychological well-being onto new PCOS self-management intervention session content.

<table>
<thead>
<tr>
<th>Potentially modifiable antecedent</th>
<th>Course sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Long-term implications of PCOS</td>
<td>✓</td>
</tr>
<tr>
<td>Getting the most from the health care system</td>
<td>✓</td>
</tr>
<tr>
<td>How to eat well in PCOS</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with low self esteem</td>
<td>✓</td>
</tr>
<tr>
<td>What self-management of PCOS entails</td>
<td>✓</td>
</tr>
<tr>
<td>How to use physical activity to manage PCOS</td>
<td>✓</td>
</tr>
<tr>
<td>Which information sources to trust</td>
<td>✓</td>
</tr>
<tr>
<td>Maintaining motivation longer term</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with feeling “not properly feminine”</td>
<td>✓</td>
</tr>
<tr>
<td>Basic biology of PCOS</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with depression and low mood</td>
<td>✓</td>
</tr>
<tr>
<td>Following a PCOS-health diet</td>
<td>✓</td>
</tr>
<tr>
<td>Making healthy choices</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with eating distress, for example, binge eating and overeating</td>
<td></td>
</tr>
<tr>
<td>Motivation to self-manage</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with embarrassment and shame</td>
<td>✓</td>
</tr>
<tr>
<td>Confidence to self-manage</td>
<td>✓</td>
</tr>
<tr>
<td>Using physical activity to manage PCOS</td>
<td>✓</td>
</tr>
<tr>
<td>Trust in health professionals</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with anxiety and fear</td>
<td>✓</td>
</tr>
<tr>
<td>Prioritizing your own needs</td>
<td>✓</td>
</tr>
<tr>
<td>Setting clear self-management goals</td>
<td>✓</td>
</tr>
<tr>
<td>Not giving up on self-management</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with frustration and anger</td>
<td>✓</td>
</tr>
<tr>
<td>Coping with loneliness and isolation</td>
<td>✓</td>
</tr>
<tr>
<td>Finding a physical activity that is enjoyable</td>
<td></td>
</tr>
<tr>
<td>Skills to communicate with health professionals</td>
<td>✓</td>
</tr>
<tr>
<td>Monitoring your menstrual cycle</td>
<td>✓</td>
</tr>
<tr>
<td>How to keep trying to self-manage</td>
<td>✓</td>
</tr>
<tr>
<td>Setting realistic goals</td>
<td>✓</td>
</tr>
<tr>
<td>Realistic expectations of the body</td>
<td>✓</td>
</tr>
<tr>
<td>Getting social support</td>
<td>✓</td>
</tr>
</tbody>
</table>

#### Choosing Theory- and Evidence-Based Intervention Components to Address Modifiable Antecedents of Self-Management

The HOPE program framework used to develop the PCOS intervention includes the following self-management intervention components: information about available resources, safety netting, training and rehearsal to communicate with health care professionals, training and rehearsal in psychological strategies, social support, and lifestyle advice and support [69].

Below, we outline how specific intervention content was selected to address the antecedents identified in the logic model for PCOS self-management.

#### Information and Activities to Target Knowledge and Understanding for Self-Management

Women with PCOS have unmet information needs [63,70-72] and may lack knowledge of how to manage their condition, for example, how to eat well to manage insulin resistance [33,73]. Unmet information needs may adversely impact...
self-management, including adherence to treatment [26,28,32,74,75]. PCOS-specific information is valued by patients with PCOS and may address unmet psychosocial needs [76-78], potentially enhancing understanding, reducing anxiety, and promoting quality of life [79]. There is a plethora of information sources available on the web, of varying quality, making it difficult for women with PCOS to select trustworthy material [71,80-83]. Signposting to trustworthy information sources may support patient activation and self-management. In the prototype intervention, factual information is provided about PCOS, including androgen excess and food intake, blood sugar and insulin resistance, eating well, and managing weight.

The needs assessment and previous research [84,85] indicate women may have concerns about the long-term health sequelae of PCOS. Motivation to adopt healthy lifestyles may be impacted by perceptions of the health risks associated with PCOS [28]. In the prototype intervention, content is provided on focusing on future health and ways to manage emerging health concerns and worries. Information about specialist health professionals, the range of support, and some potential treatment options is designed to empower participants to be active partners in their health care. The intervention seeks to empower participants to engage in health protective behaviors, for example, exercise, healthy eating, self-monitoring of their health, managing worries to alleviate health anxiety, and future help-seeking as any new health issues arise. Links to further information and resources are provided, for example, Verity (a UK PCOS charity), the NHS Choices website, the NHS Improving Access to Psychological Therapies service, and the AskPCOS app [86].

**Goal Setting and Feedback to Target Self-Management**

**Motivation and Behavior**

Women with PCOS may struggle to achieve the goals they set for lifestyle change [27], may lack the ability to identify and resolve barriers [32], experience tiredness, feel unrewarded, or have depressive and defeating thoughts that act as barriers to achieving their goals [29]. Goal setting is a widely used and effective component of self-management support [87,88]. Goal-oriented care, emphasizing patient priorities and values, may be particularly appropriate for patients with chronic or multifaceted health concerns [89]. Goal-setting theory and research recommend paying attention to multiple goal factors and multiple phases in the goal process [90].

Goal setting and solution-focused goal feedback are included in every session of the prototype intervention, with weekly and long-term goals chosen by participants themselves based on their personal priorities and values rather than set or recommended by program facilitators. The goal-setting and feedback processes are facilitated by peer facilitators, with attention to goal difficulty, goal specificity, goal proximity, goal commitment, and solution-focused feedback.

**Psychoeducation and Activities to Target Emotions**

Women with PCOS are prone to higher perceived stress [91], and stress may also impact women with PCOS differently and have an impact on the physiology of the condition [92,93]. Stress may be a barrier to lifestyle change [32]. Mindfulness-based stress reduction activities have been effective in reducing stress, depressive, and anxiety symptoms in PCOS [94] and in increasing self-efficacy for physical activity and nutrition behaviors [95]. Physical activity has well-recognized benefits for mood and perceived stress [96,97], including in women with PCOS [98,99]. In the prototype intervention, content is provided on managing stress, including mindfulness, soothing rhythm breathing, guided imagery, and “get active, feel good.” Because interviews with health professionals during the needs assessment suggested that some women with PCOS may exercise in dysfunctional ways in order to lose weight, the PCOS prototype intervention does not promote exercise for weight loss per se but rather for stress management and general well-being.

Psychological comorbidities may be a barrier to lifestyle change [32]. Interventions using aspects of cognitive behavioral therapy (CBT) have shown promise for improving fatigue, quality of life, weight loss, and depressive symptoms in women with PCOS [100-103]. In the prototype intervention, CBT-informed content is provided on managing common unhelpful thinking patterns. Compassionate mind approaches (sometimes referred to as part of a “third wave” of CBT) are particularly relevant to disordered eating, emotional eating, and weight management [104-107]. In the prototype intervention, a compassionate mind approach is taken to eating well, including eating mindfully, overeating, binge eating, and self-soothing.

Self-compassion activities, compassionate mind training, and compassion-focused therapy have been demonstrated to be particularly beneficial in improving anxiety, depression, shame, and self-criticism [106,108-111]. These issues have been identified as common in populations with PCOS. In the prototype intervention, content is provided on self-compassion, including, for example, compassion for perceived flaws, toward feelings of embarrassment or shame, and “getting to know your inner critic.”

Women with PCOS may have reduced relational and marital satisfaction [17,112], sexual satisfaction, function, and sense of sexual attractiveness [16,113-119]. Body dissatisfaction may be a barrier to lifestyle change [28]. The prototype intervention discusses and seeks to normalize the topic of common body changes and associated distress, for example, acne, alopecia, hirsutism, obesity, and associated low self-esteem, frustration, and depression, which have been identified in previous research [84,120-123], and in the needs assessment. Content is provided on common body changes in PCOS and associated difficult emotions. The potential impact of PCOS on intimate relationships and sexuality is explored, along with suggestions for ways to cope and adjust.

The population with PCOS is diverse, including women who will be more or less concerned about visible or invisible body changes. Interventions may empower them to respond and adjust in ways that are personally appropriate, including accepting changes, making adaptations, or seeking treatment, for example, for obesity or hirsutism. Previous research, including the needs assessment, indicates that embarrassment or shame may lead women to conceal the extent of their bodily changes, even from health professionals [124-127]. In the prototype intervention,
content is provided on responding to change, including acceptance, treatments and adaptations, and overcoming embarrassment to get help for body changes. The PCOS intervention takes an autonomy or body acceptance approach [128,129]. Unrealistic female body standards and body-positive activism are discussed in the program materials.

Women may struggle to manage and control their PCOS symptoms, especially if they have comorbidities, sometimes feeling controlled by their condition [84]. Motivational imagery interventions are widely used in sports psychology, with benefits for motor performance, motivational outcomes, and affective outcomes [130] and found to be motivating by a number of different populations, including people trying to manage their weight or type 2 diabetes [131,132]. Emotional mental imagery interventions have shown promise for reducing anxiety and depression [133,134]. Prospective (future-focused) mental imagery interventions are particularly relevant to fostering optimism in people with depression [135]. In the prototype intervention, content is provided on life priorities and motivational imagery—“keeping PCOS in its place”—and future-focused mental imagery.

Positive psychology interventions, such as character strengths and gratitude diaries, which focus on function rather than dysfunction, have shown promise in treating anxiety, depression, low mood, and low self-esteem and in promoting positive well-being [136-138]. In the prototype intervention, content is provided on maximizing psychological resources, for example, by focusing on character strengths and gratitude diaries rather than focusing purely on psychological deficits and dysfunctions.

Activities for Development of Self-Management Communication Skills

Being able to communicate the need for practical and social support is essential for effective self-management. Communicating with friends, family, peers, and health professionals may be particularly difficult for women with PCOS because of its sometimes visible, sometimes invisible nature and associated taboos and stigma [139-142]. In the prototype intervention, content is provided on communication, including communicating concerns with friends, family, peers, and health professionals.

The intervention explicitly acknowledges the issue that some women with PCOS may lack trust in, or feel dismissed or stigmatized by, health professionals [70,71,143-147]. Some health professionals treating PCOS recognize that “lifestyle change,” especially around weight, is a sensitive topic [148], but some patients report a lack of trust in and perceived lack of empathy from health professionals. Intervention materials in the prototype explore and deconstruct health communication around weight to foster shared understanding and trust.

People with long-term health conditions need a range of knowledge and skills to navigate health care systems and participate actively in consultations and care [149]. In the prototype intervention, content is provided on maximizing support from health services and health professionals, including health care specialists who treat PCOS, treatments, “why it may sometimes sound like health professionals are being preachy or judgemental,” summarizing concerns and requesting a referral, shared agenda setting at health care appointments, and communicating clearly and assertively with health professionals.

Building on Therapeutic Peer Group Factors and Collective Advocacy

Women with PCOS are interested in group educational interventions [150]. Group psychosocial and psychoeducational interventions provide particular therapeutic factors, for example, instillation of hope, universality, imparting of information, and opportunities for altruism and catharsis [151]. A lack of social support may be a barrier to lifestyle change [32,33]. Group and peer support, provided face-to-face or on the web, have been beneficial for those living with PCOS [76,78,152-154]. Peers delivering interventions may be well placed to express empathy and to act as realistic role models, supporting self-management [155-157].

Peer group factors are emphasized and facilitated throughout the prototype PCOS intervention, including, for example, self-management, “your PCOS journey,” getting peer support, open space forums, support from health professionals, peers, and the UK PCOS charity, and signposting to ongoing participation with PCOS-related groups and networks. The intervention is delivered in a web-based group format with secure digital social sharing features, for example, a gratitude wall, hopes and dreams, goal sharing, and delivery by trained peer facilitators who themselves had PCOS. Group delivery allows the intervention to be made available to multiple participants, saving time and resources for facilitators.

Improving health care access and outcomes for PCOS is not simply a matter of individual patient assertiveness. Collective collaborative action between patients, health professionals, and powerful others is needed to bring about change. In the prototype intervention, content acknowledges the limits of personal assertiveness and emphasizes the value of collective action. For example, we signpost to activism by Verity [158] and other women’s health organizations.

Modular Web-Based Format for Accessibility And Scalability

Digital health interventions offer considerable scope for accessibility and scalability [159]. The prototype PCOS intervention is designed to be delivered digitally through a secure social platform developed by the social enterprise community interest company Hope for the Community (H4C) [160]. The social enterprise company hosts the intervention and has a track record of scaling up the existing HOPE interventions for a range of commissioners. The H4C web-based intervention platform allows peers to support each other and engage in synchronous or asynchronous therapeutic activities in a web-based space without the logistical and practical difficulties of travel and finding venues and times suitable for all.

Program Specification

Textbox 1 shows the content of the 6-week prototype intervention program. Solution-focused goal setting and feedback, open forum discussions, and further resources are provided in every session.
The H4C team uploaded the additional content and created a working prototype intervention that was checked and tested, ready for a pilot study with women living with PCOS.
Textbox 1. Content of polycystic ovary syndrome (PCOS) intervention.

**Session 1: Instilling HOPE for PCOS**
- Welcome and introductions
- Responsibilities and ground rules
- Instilling HOPE
- Diaphragmatic breathing
- Gratitude diary
- Your PCOS journey
- Support from health professionals, peers, and Verity (UK PCOS charity)
- PCOS basics: androgen excess & insulin resistance
- Test your PCOS basics: quiz
- Factual information about PCOS
- Self-management
- Self-compassion

**Session 2: Managing the stress of PCOS**
- Gratitude diary
- Managing stress
- Mindfulness
  - Physical activity for stress management: “get active, feel good”
  - Managing common unhelpful thinking patterns (cognitive behavioral therapy [CBT])
  - Mindfulness: and am I doing this “right”??
  - Why self-compassionate mindfulness?
- Compassion-focused therapy

**Session 3: Feeding your mind and body well**
- Gratitude diary
- Eating well in PCOS: role of food intake, blood sugar, and insulin
- Why being insulin resistant is a problem for the whole body
- Why it may sometimes sound like health professionals are being preachy or judgmental
- Eating well without depriving yourself, losing weight in PCOS
- Eating mindfully to eat well in PCOS
- Overeating and binge eating in PCOS: some helpful tips for reducing the chances of overeating
- Feeding your mind and body well
  - Ways to soothe yourself without over- or undereating
  - Getting to know your inner critic, developing a compassionate ideal

**Session 4: Body image, intimacy, and close relationships**
- Gratitude diary
- Body changes, sexuality, and intimacy
- Communication
- Common body changes in PCOS
- Difficult emotions that can come with body changes
- Getting ready to be self-compassionate
- Compassion for your perceived flaws
Discussion

Principal Results
We cocreatively identified and prioritized barriers to self-management and psychological well-being to adapt a successful web-based self-management program and tailor it for the needs of adult women with PCOS. We developed a web-based prototype program for the PCOS program, ready for testing with this population.

Comparison With Previous Work
Women with PCOS value specialized, integrated, evidence-based, patient-centered health care services, ideally codeveloped with consumers and health professionals [34,35]. The new HOPE PCOS intervention has been co-developed by combining input from patients, patients with PCOS’ advocates, and health care professionals. It is a comprehensive intervention integrating education and empowerment with lifestyle management, cognitive behavioral tools, and peer support, ready to be delivered by trained peer facilitators or codelivered by peers and health care professionals.

Developed following MRC guidance for the development of complex interventions [45,47], the intervention is evidence-based, underpinned by relevant theory, and designed to support a collaborative approach to care. Subject to evaluation, the new intervention could be integrated as part of PCOS services to support self-management and the model of collaborative, patient-centered health care advocated by international guidelines [4,161].

The intervention deploys an existing secure web-based social platform that has been used successfully to scale up other self-management support programs for other populations and conditions [40,44]. The multicomponent and modular structure of the program offers flexibility in adapting it further as new clinical recommendations and patient information emerge. The intervention is therefore well placed to be delivered and evaluated at scale.

The HOPE PCOS intervention is designed to support self-management and promote psychological well-being, rather than as a behavior change intervention per se. Since we started the process of cocreating the HOPE PCOS intervention, a different team has undertaken development work for a behavior change lifestyle intervention using a different intervention design.
model [28,31-33]. That work has not yet, to our knowledge, resulted in a prototype intervention. However, the findings from the needs assessment work conducted by that team for their lifestyle intervention concurred considerably with our own needs assessment results, showing that women living with PCOS may often experience barriers of capability, opportunity, and motivation, for example, a lack of credible information, difficulty managing multiple health conditions, limited access to resources, adequate social support, and issues such as health expectancies and emotional eating affecting their motivation to engage in lifestyle change [33].

Limitations
The current HOPE PCOS intervention is based on a needs assessment conducted with adult women. As part of the iterative cocreation process, further needs assessments may be required to adapt and tailor future versions of the intervention for other populations, for example, adolescents with PCOS, as concerns, distress, and intervention format preferences may differ [162]. There is scope to develop variants of the intervention for specific adult populations, for example, perimenopausal, menopausal, or postmenopausal women with PCOS who may have specific concerns [163-165]. The intervention might also be tailored to support trans and nonbinary people with PCOS, those with particular concerns such as weight management, comorbid conditions, or those actively trying to conceive. More cocreation and development work could be done to ensure the intervention is culturally appropriate. The intervention is intended to provide extra support and not to bypass what is recommended by a clinician. Women with PCOS are not advised to rely fully on this self-management intervention and are advised to consult a clinician as appropriate.

Cocreative intervention development is an iterative process, and, in addition to tailoring the HOPE PCOS intervention for different populations, future iterations may need to develop new content or place greater emphasis on some components, depending on how participants and health professionals evaluate the program. If consensus emerges about specific target behaviors for self-management in PCOS, it may be necessary in the future to adapt the intervention to promote specific behavior change. For example, the current content promoting physical activity could be modified and refined to promote a target level of physical activity, or the current material on healthy eating could be modified to promote specific changes in eating behavior.

Although the prototype intervention has been collaboratively co-designed to support self-management and psychological well-being, further feasibility research is needed with patients with PCOS using the program. This should evaluate the program’s acceptability as well as its impact on key outcomes such as measures of self-management, depression, anxiety, and psychological well-being. Future work is also needed to assess the intervention’s scalability, cost-effectiveness, and suitability for integration with standard care.

Conclusions
The development of the novel HOPE PCOS intervention contributes to ongoing efforts to support patients with long-term conditions to self-manage effectively and is, to our knowledge, the first such program for patients with PCOS. This study has demonstrated that comprehensive intervention programs can be co-developed with patients with PCOS, patient advocates, and health care professionals to address multiple barriers to self-management and psychological well-being. A person-centered, holistic approach, organized around the self-reported needs and priorities of patients, may produce interventions that complement services provided by health care professionals and patient advocacy and support organizations. This study demonstrates that the use of a remote digital platform with preloaded, evidence-based intervention content may offer economies of scale in that complex self-management support interventions might be delivered to many patients in parallel, potentially saving clinic resources for one-to-one care. A key message for those working to develop lifestyle or self-management support interventions is the potential value of adapting an existing program and reusing existing digital infrastructure to provide a tailored intervention without the additional resource implications of developing an entire intervention from scratch.

Acknowledgments
The HOPE Program is delivered by Hope for the Community, a community interest company of Coventry University. The intervention development work for this polycystic ovary syndrome intervention was supported by a short-term temporary research sabbatical for CP funded by Coventry University. The patient interviews were conducted by CO as part of her master’s study in Health Psychology at Coventry University.

Data Availability
The data sets generated or analyzed during this study are not publicly available due to participants’ consent being given for anonymized extracts only (not whole transcripts) to be used in published reports, but data are available from the corresponding author upon reasonable request.

Conflicts of Interest
AT is the co-inventor of the original HOPE Program, co-founder, and nonexecutive director of Hope For The Community (H4C).

References


84. Williams S, Sheffield D, Knibb RC. 'Everything's from the inside out with PCOS': exploring women's experiences of living with polycystic ovary syndrome and co-morbidities through Skype™ interviews. Health Psychol Open 2015;2(2):2055102915603051 [FREE Full text] [doi: 10.1177/2055102915603051] [Medline: 28070371]


160. Hope for the Community—Community Interest Company. URL: https://www.h4c.org.uk/ [accessed 2024-01-31]


**Abbreviations**

ATM: antecedent target measure  
CBT: cognitive behavioral therapy  
H4C: Hope for the Community  
MRC: Medical Research Council  
NHS: National Health Service  
PCOS: polycystic ovary syndrome  
PRISMS: Practical Reviews in Self-Management Support
A Deep Learning–Based Approach for Prediction of Vancomycin Treatment Monitoring: Retrospective Study Among Patients With Critical Illness

Dohyun Kim\textsuperscript{1*}, MSc; Hyun-Soo Choi\textsuperscript{1,2*}, PhD; DongHoon Lee\textsuperscript{1}, MSc; Minkyu Kim\textsuperscript{1}, MSc; Yoon Kim\textsuperscript{1,3}, PhD; Seon-Sook Han\textsuperscript{4}, MD, PhD; Yeonjeong Heo\textsuperscript{4}, MD; Ju-Hee Park\textsuperscript{5}, MD; Jinkyeong Park\textsuperscript{6}, MD, PhD

\textsuperscript{1}Department of Research and Development, ZIOVISION Co, Ltd, Chuncheon, Republic of Korea
\textsuperscript{2}Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul, Republic of Korea
\textsuperscript{3}Department of Computer Science and Engineering, Kangwon National University, Chuncheon, Republic of Korea
\textsuperscript{4}Department of Internal Medicine, Kangwon National University, Chuncheon, Republic of Korea
\textsuperscript{5}Department of Internal Medicine, Dongguk University Ilsan Hospital, Goyang, Republic of Korea
\textsuperscript{6}Department of Pulmonary, Allergy and Critical Care Medicine, School of Medicine, Kyung Hee University Hospital at Gangdong, Seoul, Republic of Korea

*these authors contributed equally

Corresponding Author:
Jinkyeong Park, MD, PhD
Department of Pulmonary, Allergy and Critical Care Medicine, School of Medicine
Kyung Hee University Hospital at Gangdong
892, Dongnam-ro, Gangdong-gu
Seoul, 05278
Republic of Korea
Phone: 82 1027747808
Email: pjk3318@gmail.com

Abstract

Background: Vancomycin pharmacokinetics are highly variable in patients with critical illnesses, and clinicians commonly use population pharmacokinetic (PPK) models based on a Bayesian approach to dose. However, these models are population-dependent, may only sometimes meet the needs of individual patients, and are only used by experienced clinicians as a reference for making treatment decisions. To assist real-world clinicians, we developed a deep learning–based decision-making system that predicts vancomycin therapeutic drug monitoring (TDM) levels in patients in intensive care unit.

Objective: This study aimed to establish joint multilayer perceptron (JointMLP), a new deep-learning model for predicting vancomycin TDM levels, and compare its performance with the PPK models, extreme gradient boosting (XGBoost), and TabNet.

Methods: We used a 977-case data set split into training and testing groups in a 9:1 ratio. We performed external validation of the model using 1429 cases from Kangwon National University Hospital and 2394 cases from the Medical Information Mart for Intensive Care–IV (MIMIC-IV). In addition, we performed 10-fold cross-validation on the internal training data set and calculated the 95% CIs using the metric. Finally, we evaluated the generalization ability of the JointMLP model using the MIMIC-IV data set.

Results: Our JointMLP model outperformed other models in predicting vancomycin TDM levels in internal and external data sets. Compared to PPK, the JointMLP model improved predictive power by up to 31% (mean absolute error [MAE] 6.68 vs 5.11) on the internal data set and 81% (MAE 11.87 vs 6.56) on the external data set. In addition, the JointMLP model significantly outperforms XGBoost and TabNet, with a 13% (MAE 5.75 vs 5.11) and 14% (MAE 5.85 vs 5.11) improvement in predictive accuracy on the inner data set, respectively. On both the internal and external data sets, our JointMLP model performed well compared to XGBoost and TabNet, achieving prediction accuracy improvements of 34% and 14%, respectively. Additionally, our JointMLP model showed higher robustness to outlier data than the other models, as evidenced by its higher root mean squared error performance across all data sets. The mean errors and variances of the JointMLP model were close to zero and smaller than those of the PPK model in internal and external data sets.
Conclusions: Our JointMLP approach can help optimize treatment outcomes in patients with critical illnesses in an intensive care unit setting, reducing side effects associated with suboptimal vancomycin administration. These include increased risk of bacterial resistance, extended hospital stays, and increased health care costs. In addition, the superior performance of our model compared to existing models highlights its potential to help real-world clinicians.

(JMIR Form Res 2024;8:e45202) doi: 10.2196/45202

KEYWORDS

critically ill; deep learning; inflammation; machine learning; pharmacokinetic; therapeutic drug monitoring; vancomycin

Introduction

Vancomycin is frequently used for severe infections caused by gram-positive bacteria (including methicillin-resistant Staphylococcus aureus), such as pneumonia, skin and soft tissue infections, and other sepsis or septic shock in patients with critical illnesses. For vancomycin, because of the narrow therapeutic range and individual differences in pharmacokinetic parameters, therapeutic drug monitoring (TDM) is recommended to minimize toxicity and improve therapeutic efficacy. Although various pharmacokinetic indicators related to the antibacterial action of vancomycin have been suggested, the most recent guideline [1] recommends that a ratio of the area under the curve (AUC) over 24 hours to the minimum inhibitory concentration (MIC) of ≥400 should be the primary pharmacokinetic and pharmacodynamic (PK/PD) predictor of vancomycin activity [2]. However, in the real world, especially in intensive care units (ICUs), a more practical method of vancomycin monitoring is evaluating with trough concentrations since calculating the AUC requires multiple blood samples per patient [3-5].

There are different methods to determine the vancomycin AUC, such as the Bayesian approach or equation-based methodologies such as the trapezoidal model. The Bayesian approach uses population data to estimate individual patient pharmacokinetic parameters, called Bayesian priors. However, this method requires complex mathematical calculations to estimate the posterior distribution of model parameters and relies on certain assumptions, such as the normality of errors and data independence. Violating these assumptions can result in biased results, and using the Bayesian approach requires expensive commercial software programs [6,7]. On the other hand, equation-based approaches, like the trapezoidal model, do not require specialized software but need multiple blood sampling to achieve 2 steady-state levels. Since it is done under steady-state conditions, it cannot account for potential changes in AUC due to ongoing acute physiological changes. The pharmacokinetic profile of vancomycin exhibits linear pharmacokinetics over a range of therapeutic doses, is highly variable, especially in older people or people with critical illnesses, and is different for adults with normal renal function [8,9].

Population pharmacokinetic (PPK) models of vancomycin required to determine the AUC have very limited robustness beyond the specific population studied. The training and experience of the clinician or clinical pharmacist are still vital in achieving adequate vancomycin TDM, especially in patients with critical illnesses with high heterogeneity. A more multidimensional and consistent decision-making system is needed in such a clinical decision-making system because differences inevitably influence the results of individual clinicians’ abilities. Machine learning algorithms have emerged as a promising approach for improving decision-making on specific questions related to rich multidimensional data and for medical research and clinical care [10-12]. In previous studies, machine learning techniques such as decision trees [13] and extreme gradient boosting (XGBoost) [14,15] were used to make better predictions on vancomycin concentrations. In addition, various deep learning models have had great success in various fields [16], including clinical practice. These might help solve more complex clinical problems. In this study, we aimed to establish an ideal model by comparing and integrating methods that have been successful so far in the decision-making system that predicts the vancomycin TDM level in patients in the ICU.

Methods

Study Population

This retrospective study involved 2406 patients with critical illnesses admitted to medical ICUs, including 1 each at Dongguk University Ilsan Hospital (DUIH) and Kangwon National University Hospital (KNUH) between January 1, 2010, and February 28, 2022. We used the data collected from DUIH and KNUH as the internal and external validation data sets. The internal validation data set contained 977 patients, while the external validation data set contained 1429 patients.

Patients with critical illnesses (>18 years) with intravenous vancomycin treatment history and who had at least 1 test for vancomycin TDM were eligible for this study. Vancomycin TDM was investigated for trough-level information. If the TDM was checked multiple times, only the first TDM value was selected. However, in patients with normal renal function, if the interval between stopping vancomycin and readministering it was 2 weeks or more, it was considered an independent TDM value and used for analyses. Patients who received oral vancomycin or were aged 18 years or younger were excluded. The total serum vancomycin concentrations were determined using a fluorescence immunoassay (VANC3, Cobas c 702; Roche Diagnostics).

The data set was split into training and testing groups in a 9:1 ratio, with 879 and 98 cases representing the internal data set, respectively. For external validation, 1429 cases from KNUH were used. To evaluate the model's generalization and the range of expected errors of the classifiers, 10-fold cross-validation was performed on the internal training data set, with a 9:1 ratio and 791 and 88 cases, respectively. The model was tested on...
both the internal and external test data sets for each fold and the metrics obtained were used to calculate the 95% CI over the 10 folds. The internal and external data sets remained constant across all folds. To evaluate the generalization ability of the joint multilayer perceptron (JointMLP) model, we used 2394 cases from the Medical Information Mart for Intensive Care–IV (MIMIC-IV) [17] as a US critical care database used for large-scale multi-institutional research for external validation.

Experimental Model

We proposed a JointMLP to predict the vancomycin TDM level in patients in the ICU. The JointMLP model is an ensemble model, meaning it combines the predictions of multiple individual MLPs (multilayer perceptrons) to generate a final prediction. The MLP is an artificial neural network. The MLP consists of multiple layers of interconnected nodes (also known as neurons) that perform mathematical operations on the input data to generate an output. In the case of the JointMLP model proposed for vancomycin TDM prediction, each MLP in the ensemble consists of 3 hidden layers with a hidden unit size of 32, and the LeakyReLU activation function is used to introduce nonlinearity into the model (Figure 1). Using an ensemble helps reduce the risk of overfitting and improves the accuracy and robustness of the model. In this case, the JointMLP model comprises 100 different MLPs, each with its own set of weights learned during the training process. To find the best set of hyperparameters for the JointMLP model, grid search is used, which systematically tests different combinations of hyperparameters to find the combination that yields the best performance on a validation data set. In this case, the tuned hyperparameters included the number of MLPs, the number of hidden layers, and the size of the hidden units. In a JointMLP model, the MLPs are trained independently in parallel. During training, each MLP in the ensemble receives a randomly selected subset of the training data and learns to make predictions based on that subset. The individual MLPs are then combined to generate a final prediction, typically by taking the average of their outputs or using a weighted combination. The MLPs are trained independently of each other and in parallel. Our JointMLP model is available on the websites [18] for predicting TDM levels and recommending the frequency and dose.

Figure 1. The architecture of joint multilayer perceptron (JointMLP). The JointMLP model was built with 100 different multilayer perceptrons (MLPs) consisting of 3 hidden layers with a hidden unit size of 32 and LeakyReLU as an activation function.
We compared 4 other models: the PPK model [19], XGBoost, TabNet, and 300-layer MLP. The PPK is a pharmacokinetic model used to assess the relationship between drug concentration and time in the body. XGBoost is an open-source library for decision tree–based gradient-boosting machine learning. The XGBoost model especially works well on tabular data learning. It is a vital model for distributed training or regularization. TabNet is a deep learning model specialized for tabular data set analysis. TabNet could be optimized by automatically transforming input variables and using additional validation data to select the required variables.

For XGBoost, the hyperparameters are eta (learning rate), gamma (minimum loss reduction required to make a further partition on a leaf node of the tree), and max_depth (maximum depth of a tree). The grid search was performed to find the best combination of these hyperparameters. The eta, gamma, and max_depth values were 0.2, 0.001, and 6, respectively.

For TabNet, the hyperparameters are n_d (width of the decision prediction layer), n_a (width of the attention embedding for each mask), n_steps (number of steps in the architecture), gamma (coefficient for feature reuse in the masks), and verbose (verbosity for notebook plots). The grid search was performed to find the best combination of these hyperparameters. The values chosen for n_d, n_a, n_steps, gamma, and verbose were 8, 8, 3, 1.3, and 0, respectively.

The 300-layer MLP model was constructed to verify the performance of the JointMLP model. The hyperparameters for the MLP model were chosen to have the same depth as the JointMLP model, which is 100 horizontal MLPs times 3 hidden layers. The number of hidden layers was set to 300, and the number of hidden units was set to 32.

Data Processing

Variables that are associated with vancomycin PK/PD were selected. These variables included the total dose of vancomycin from start to end, the usage of the loading dose at the first use, the total number of vancomycin infusions, the dose of vancomycin per infusion, the mean infusing interval of vancomycin, the interval from the start of vancomycin administration to the measurement of vancomycin serum levels, age, sex, height, weight, serum creatinine levels, serum vancomycin levels, dialysis, and the total volume of transfusion from the start of vancomycin administration to the measurement of vancomycin levels. Patients with impaired kidney function may have issues with the metabolism and elimination of vancomycin. The estimated glomerular filtration rate (eGFR), which represents renal function, was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula and used as a variable. For the normalization of the data, for both the input and target variables, all values were scaled in the range (−1 to 1) using the minimum-maximum scaler. Moreover, we only had binary categorical variables in the model. Therefore, no special categorical variable encoding was performed in the model but a simple change to 0 and 1. To select significant input variables, we grouped variables into 5 categories. Each variable was classified as (1) baseline demographics—age, sex, body weight, and height; (2) drug administration-related variables; (3) variables related to the volume of distribution; (4) eGFRs using serum creatinine levels; and (5) variables related to the amount of transfusion. With the above-grouped categories, first, we have selected the default variables that would be used throughout all scenarios: the baseline demographics and the drug administration–related variables. Then, from the remaining 3 categories, with the clinician’s consultation, we tried to minimize the redundancy of the variables that would have the same meaning. Moreover, in accordance with the feature importance of each variable, we were able to find the best combination of model input variable sets. Furthermore, the contribution of the input data to the model in the test set was determined through the Shapley value of Shapley Additive exPlanations (SHAP) [19]. A high mean absolute estimated Shapley value indicates that the variable has a stronger impact on the output value.

Statistical Analysis

The primary end point was the predictive performance of the vancomycin level within the therapeutic range. The baseline variables and patient characteristics of internal and external data sets were presented as frequencies with percentages or mean values with SDs. Between–data set comparisons were performed using the paired 2-tailed t test for continuous variables or the chi-square test for categorical variables. The measured serum vancomycin levels were used as true values. The predictive abilities of 4 kinds of models for vancomycin trough concentrations, using the PPK model on the website of the infectious diseases management program at the University of California, San Francisco [20], machine learning models (XGBoost) [21], and deep learning models (TabNet [22] and JointMLP), were evaluated for bias and precision by calculating the mean absolute error (MAE), root-mean-square error (RMSE), R², and adjusted R². The paired t test on MAE was used [23,24] to confirm significant differences in prediction performance between models. P values of <.05 were considered statistically significant. All analyses were performed with R software (version 3.4.4; R Foundation for Statistical Computing) and Python (version 3.8.12; The Python Software Foundation) using the SciPy 1.9.1, PyTorch 1.10.2, and PyTorch Lightning 1.6.4 libraries.

Ethical Considerations

This study involved a retrospective analysis of medical records and did not involve collecting prospectively obtained data. As such, the study was considered to pose minimal risks and did not directly affect patients’ rights, welfare, or clinical care while offering recognized social benefits. To protect patient confidentiality during data collection and record review, all investigative records were securely stored in an encrypted database, and published results did not include any participants’ names or identifiers. The study received ethical approval from the institutional review board of Dongguk University Hospital (DUH 2021-08-015) and Kangwon National University Hospital’s Institutional Review Board (KNUH-2022-03-020). This study was performed per the principles of the Declaration of Helsinki.

https://formative.jmir.org/2024/1/e45202
Results

Overview

Out of the 977 patients included in the DUIH data set, the testing data set was composed of 98 (10%) patients, with the other 879 patients belonging to the training data set. All 1429 patients included in KNUH were for external validation. The baseline characteristics did not differ significantly between the testing and training groups (Table 1). However, there were more older patients and shorter heights in the external validation set. The estimated renal function by the CKD-EPI method was worse in the patients for external validation. The patients included in MIMIC-IV showed significant differences in sex distribution compared to internal data sets or patients in other ICUs in Korea, and the proportion of non-Hispanic White patients was over 66.4% (n=1591).
Table 1. Baseline characteristics of all enrolled patients.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Internal data set</th>
<th>External data set</th>
<th>External data set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test (n=98)</td>
<td>Train or validation (n=879)</td>
<td>From KNUH&lt;sup&gt;a&lt;/sup&gt; (n=1429)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>31 (31.6)</td>
<td>328 (37.3)</td>
<td>555 (38.8)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>67.65 (15.49)</td>
<td>68.38 (15.56)</td>
<td>70.48 (14.39)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>163.56 (9.04)</td>
<td>162.42 (9.61)</td>
<td>161.36 (9.51)</td>
</tr>
<tr>
<td>Body weight (kg), mean (SD)</td>
<td>58.54 (15.80)</td>
<td>57.02 (13.27)</td>
<td>58.36 (13.60)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>98 (100)</td>
<td>979 (100)</td>
<td>1429 (100)</td>
</tr>
<tr>
<td>African-American</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Serum trough value (mg/L), mean (SD)</td>
<td>14.49 (10.33)</td>
<td>15.01 (10.29)</td>
<td>13.37 (9.49)</td>
</tr>
<tr>
<td>Serum creatinine (µmol/L), mean (SD)</td>
<td>1.10 (1.16)</td>
<td>1.11 (1.32)</td>
<td>1.06 (1.34)</td>
</tr>
<tr>
<td>Dialysis, mean (SD)</td>
<td>3 (3.1)</td>
<td>41 (4.7)</td>
<td>55 (3.8)</td>
</tr>
<tr>
<td>The total dose of vancomycin before TDM&lt;sup&gt;e&lt;/sup&gt; (mg), mean (SD)</td>
<td>3961.28 (2812.91)</td>
<td>4289.52 (3504.59)</td>
<td>4446.94 (4933.77)</td>
</tr>
<tr>
<td>Number of vancomycin injections, n (%)</td>
<td>4.42 (2.72)</td>
<td>4.80 (4.03)</td>
<td>4.77 (5.17)</td>
</tr>
<tr>
<td>The average dose of vancomycin before TDM (mg), mean (SD)</td>
<td>882.40 (207.66)</td>
<td>891.90 (186.89)</td>
<td>931.31 (184.90)</td>
</tr>
<tr>
<td>Time between administration of vancomycin and the TDM (day), mean (SD)</td>
<td>3.81 (1.37)</td>
<td>4.01 (2.70)</td>
<td>2.86 (2.62)</td>
</tr>
<tr>
<td>Interval among each dose of vancomycin (hours), mean (SD)</td>
<td>21.43 (11.59)</td>
<td>21.12 (11.75)</td>
<td>16.90 (10.58)</td>
</tr>
<tr>
<td>Vancomycin loading, mean (SD)</td>
<td>8 (8.2)</td>
<td>54 (6.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Volume distribution (L), mean (SD)</td>
<td>40.98 (11.06)</td>
<td>39.92 (9.29)</td>
<td>40.86 (9.52)</td>
</tr>
<tr>
<td>Adjusted volume distribution (L), mean (SD)</td>
<td>39.38 (3.81)</td>
<td>39.17 (3.34)</td>
<td>39.82 (3.25)</td>
</tr>
<tr>
<td>CrCl&lt;sup&gt;f&lt;/sup&gt; (minutes), mean (SD)</td>
<td>91.49 (73.26)</td>
<td>84.06 (58.85)</td>
<td>90.36 (82.44)</td>
</tr>
<tr>
<td>CrCl (hours), mean (SD)</td>
<td>0.08 (0.06)</td>
<td>0.07 (0.05)</td>
<td>0.08 (0.07)</td>
</tr>
<tr>
<td>The elimination rate constant at infusion time (Kt), mean (SD)</td>
<td>1.70 (1.26)</td>
<td>1.59 (1.21)</td>
<td>1.17 (1.07)</td>
</tr>
<tr>
<td>Trough levels of vancomycin (mcg/mL), mean (SD)</td>
<td>10.07 (9.15)</td>
<td>12.19 (14.62)</td>
<td>20.87 (20.79)</td>
</tr>
<tr>
<td>eGFR&lt;sup&gt;g&lt;/sup&gt; MDRD&lt;sup&gt;h&lt;/sup&gt;, mean (SD)</td>
<td>116.94 (100.14)</td>
<td>114.23 (90.33)</td>
<td>122.70 (123.49)</td>
</tr>
<tr>
<td>eGFR CKD-EPI&lt;sup&gt;i&lt;/sup&gt;, mean (SD)</td>
<td>82.33 (39.87)</td>
<td>82.00 (38.47)</td>
<td>84.35 (36.96)</td>
</tr>
<tr>
<td>Vancomycin clearance, mean (SD)</td>
<td>4.00 (3.03)</td>
<td>3.69 (2.43)</td>
<td>3.96 (3.41)</td>
</tr>
</tbody>
</table>

<sup>a</sup>KNUH: Kangwon National University Hospital.
<sup>b</sup>P value: comparison of the internal data set with the external data set from KNU hospital in Korea.
<sup>c</sup>MIMIC-IV: Medical Information Mart for Intensive Care–IV.
<sup>d</sup>P value: comparison of the internal data set with the external data set from MIMIC-IV.
<sup>e</sup>TDM: therapeutic drug monitoring.
<sup>f</sup>CrCl: creatinine clearance.
<sup>g</sup>eGFR: estimated glomerular filtration rate.
<sup>h</sup>MDRD: Modification of Diet in Renal Disease.
<sup>i</sup>eGFR CKD-EPI: Chronic Kidney Disease-Epidemiology Collaboration.
Compared to the PPK model, the XGBoost model only significantly improved the predictive performance of external data sets (20.43 vs 11.59, a 76% change by RMSE and 11.87 vs 8.75, a 36% change by MAE; Table 2) when evaluated by RMSE and MAE. Although statistical significance could not be determined, the predictive power of internal data sets also improved by 8% (10.38 vs 9.58 by RMSE) and 16% (6.68 vs 5.75 by MAE).

**Table 2. Performances of different models with bootstrapping.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Internal, RMSEA (95% CI)</th>
<th>External, RMSEA (95% CI)</th>
<th>Internal, MAEB (95% CI)</th>
<th>External, MAEB (95% CI)</th>
<th>Internal, R² (95% CI)</th>
<th>External, R² (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPKc</td>
<td>10.38 (7.38 to 13.42)</td>
<td>20.43 (18.15 to 22.64)</td>
<td>6.68 (5.29 to 8.45)</td>
<td>11.87 (11.05 to 12.75)</td>
<td>-0.02 (-0.44 to 0.22)</td>
<td>-3.64 (-5.16 to -2.48)</td>
</tr>
<tr>
<td>XGBoostd</td>
<td>9.58 (6.31 to 12.6)</td>
<td>11.59 (10.88 to 12.17)</td>
<td>5.75 (4.37 to 7.48)</td>
<td>8.75 (8.34 to 9.13)</td>
<td>0.13 (-0.63 to 0.48)</td>
<td>-0.49 (-0.81 to -0.24)</td>
</tr>
<tr>
<td>TabNet</td>
<td>8.81 (6.33 to 11.29)</td>
<td>13.89 (11.01 to 17.71)</td>
<td>5.85 (4.53 to 7.25)</td>
<td>7.50 (6.90 to 8.13)</td>
<td>0.26 (-0.15 to 0.51)</td>
<td>-1.15 (-2.53 to -0.38)</td>
</tr>
<tr>
<td>300-layer MLPf</td>
<td>10.17 (7.06 to 13.09)</td>
<td>9.94 (8.84 to 11.04)</td>
<td>6.98 (5.61 to 8.63)</td>
<td>7.45 (6.55 to 7.28)</td>
<td>0.021 (-0.086 to 0.056)</td>
<td>-0.098 (-0.26 to -0.023)</td>
</tr>
<tr>
<td>JointMLPg</td>
<td>8.27 (5.33 to 11.19)</td>
<td>9.50 (8.72 to 10.30)</td>
<td>5.11 (3.92 to 6.58)</td>
<td>6.56 (6.18 to 6.90)</td>
<td>0.35 (-0.03 to 0.59)</td>
<td>-0.005 (-0.17 to 0.13)</td>
</tr>
</tbody>
</table>

aRMSE: root mean squared error.  
bMAE: mean absolute error.  
cPPK: population pharmacokinetic.  
dXGBoost: extreme gradient boosting.  
eMLP: multilayer perceptron.  
fJointMLP: joint multilayer perceptron.  

The proposed JointMLP model significantly outperformed XGBoost and TabNet with a 13% (5.75 vs 5.11) and 14% (5.85 vs 5.11) improvement in predictive accuracy under MAE on the internal data set, respectively. Additionally, our trained JointMLP showed better performance not only on the internal data set but also on the external data set that consisted of 1429 patients with critical illnesses by 34% (8.75 vs 6.56) and 14% (7.50 vs 6.56) when compared with XGBoost and TabNet. Furthermore, JointMLP’s RMSE performance is higher than that of other models on all data sets, which means that the proposed model is more robust than theirs against outlier data.

Within the external data set, we clearly see that all models are statistically significant among themselves (Figure 2). However, the P value of JointMLP within the internal data set expressed a trend more robust than all other models while showing statistical significance only when compared with PPK with MAE. The mean error and variance of the JointMLP model were nearer to zero and smaller in both internal and external data sets than in the PPK model (Figure 3).
Figure 2. $P$ value heat map by paired t test on all the combinations of prediction models. The $P$ value for each comparison of the mean absolute error for predictive performance on (A) the internal validation data set (B) and the external validation data set. 300MLP: 300-layer multilayer perceptron; JointMLP: joint multilayer perceptron; PPK: population pharmacokinetic; XGB: extreme gradient boosting.

Figure 3. Comparisons of the population pharmacokinetic (PPK) and joint multilayer perceptron (JointMLP) models on the internal and external validation sets. Scatter plots of (A) the internal data set and (B) the external data set with the predicted value against a target value. The error histograms of (C) the internal data set and (D) the external data set with a predicted value against a target value. TDM: therapeutic drug monitoring.

The most influential variables in TDM prediction were eGFR CKD-EPI, the average dose of vancomycin, and the average volume of distribution in both internal and external data sets. These variables consistently had high SHA $P$ values in both data sets (Table 3).
Table 3. The mean values of Shapley Additive exPlanations on the internal and external validation sets.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Internal validation, mean (SD)</th>
<th>External validation, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFR\textsuperscript{a} CKD-EPI\textsuperscript{b}</td>
<td>0.076 (0.09)</td>
<td>0.063 (0.088)</td>
</tr>
<tr>
<td>Average dose of vancomycin</td>
<td>0.045 (0.05)</td>
<td>0.04 (0.073)</td>
</tr>
<tr>
<td>Average volume of distribution</td>
<td>0.037 (0.029)</td>
<td>0.032 (0.04)</td>
</tr>
<tr>
<td>Interval of vancomycin</td>
<td>0.028 (0.014)</td>
<td>0.031 (0.03)</td>
</tr>
<tr>
<td>Body weight</td>
<td>0.027 (0.038)</td>
<td>0.024 (0.044)</td>
</tr>
<tr>
<td>Total dose</td>
<td>0.014 (0.028)</td>
<td>0.02 (0.066)</td>
</tr>
<tr>
<td>Height</td>
<td>0.013 (0.023)</td>
<td>0.015 (0.026)</td>
</tr>
<tr>
<td>Sex</td>
<td>0.01 (0.018)</td>
<td>0.011 (0.004)</td>
</tr>
<tr>
<td>Number of vancomycin injection</td>
<td>0.01 (0.012)</td>
<td>0.015 (0.031)</td>
</tr>
<tr>
<td>Creatinine</td>
<td>0.01 (0.007)</td>
<td>0.01 (0.011)</td>
</tr>
<tr>
<td>Loading</td>
<td>0.007 (0.012)</td>
<td>0.003 (0.01)</td>
</tr>
<tr>
<td>Age</td>
<td>0.006 (0.006)</td>
<td>0.006 (0.003)</td>
</tr>
<tr>
<td>Dialysis</td>
<td>0.005 (0.018)</td>
<td>0.005 (0.028)</td>
</tr>
<tr>
<td>The elimination rate constant at infusion time</td>
<td>0.005 (0.022)</td>
<td>0.005 (0.022)</td>
</tr>
<tr>
<td>Time between vancomycin injection and TDM\textsuperscript{c}</td>
<td>0.003 (0.002)</td>
<td>0.007 (0.007)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}eGFR: estimated glomerular filtration rate.  
\textsuperscript{b}CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration.  
\textsuperscript{c}TDM: therapeutic drug monitoring.

**JointMLP Versus 300-Layer MLP**

The purpose of Table 2 was to compare the predictive performances of the 300-layer MLP model and the proposed JointMLP model. The results showed that JointMLP outperformed the 300-layer MLP model in both the internal (8.27 vs 9.94, a 23% improvement by RMSE, and 6.98 vs 5.11, a 37% improvement by MAE) and external (9.50 vs 9.94, a 5% improvement by RMSE, and 7.45 vs 6.56, a 14% improvement by MAE) data sets, indicating the effectiveness of the boosting ensemble approach used in JointMLP. Figure 2 shows the $P$ value analysis of the comparison between JointMLP and the 300-layer MLP model. The analysis revealed that the difference between the 2 models was statistically significant ($P<.001$) within the external data set. However, compared to other models, the difference between JointMLP and the 300-layer MLP model was not statistically significant ($P=.09$) within the internal data set.

**Model Comparisons in the MIMIC-IV Data Set**

The findings presented in Table 4 clearly indicated that the JointMLP model performed better than both the PPK and 300-layer MLP models by a significant margin of 160% and 25%, respectively, when evaluated using the RMSE metric. This improvement in predictive performance was observed even in the MIMIC-IV data set, which included patients from multiple centers and races. The JointMLP model also demonstrated superior performance compared to the XGBoost and TabNet models, with improvements of 5% and 4%, respectively. When examining the MAE metric, the JointMLP model significantly outperformed the PPK model by 90%, with further improvements of 8%, 6%, and 5% over the XGBoost, 300-layer MLP, and TabNet models, respectively.
Table 4. Performances of models with Medical Information Mart for Intensive Care–IV data set.

<table>
<thead>
<tr>
<th>Model</th>
<th>MIMIC&lt;sup&gt;a&lt;/sup&gt;, RMSE&lt;sup&gt;b&lt;/sup&gt; (95% CI)</th>
<th>MIMIC, MAE&lt;sup&gt;c&lt;/sup&gt; (95% CI)</th>
<th>MIMIC, $R^2$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPK&lt;sup&gt;d&lt;/sup&gt;</td>
<td>22.22 (18.88 to 25.72)</td>
<td>12.13 (11.42 to 12.87)</td>
<td>−5.93 (−8.52 to −3.99)</td>
</tr>
<tr>
<td>XGBoost&lt;sup&gt;e&lt;/sup&gt;</td>
<td>8.93 (8.60 to 9.23)</td>
<td>6.93 (6.69 to 7.15)</td>
<td>−0.12 (−0.21 to 0.04)</td>
</tr>
<tr>
<td>TabNet</td>
<td>8.91 (8.53 to 9.25)</td>
<td>6.69 (6.45 to 6.93)</td>
<td>−0.11 (−0.17 to −0.06)</td>
</tr>
<tr>
<td>300-layer MLP&lt;sup&gt;f&lt;/sup&gt;</td>
<td>10.65 (8.88 to 12.72)</td>
<td>6.80 (6.48 to 7.13)</td>
<td>−0.59 (−1.30 to −0.12)</td>
</tr>
<tr>
<td>JointMLP&lt;sup&gt;g&lt;/sup&gt; (proposed)</td>
<td>8.53 (8.17 to 8.88)</td>
<td>6.40 (6.17 to 6.64)</td>
<td>−0.02 (−0.08 to 0.03)</td>
</tr>
</tbody>
</table>

<sup>a</sup>MIMIC: Medical Information Mart for Intensive Care.
<sup>b</sup>RMSE: root mean squared error.
<sup>c</sup>MAE: mean absolute error.
<sup>d</sup>PPK: population pharmacokinetic.
<sup>e</sup>XGBoost: extreme gradient boosting.
<sup>f</sup>MLP: multilayer perceptron.
<sup>g</sup>JointMLP: joint multilayer perceptron.

All models, except for the 300-layer MLP, were statistically significant in the MIMIC-IV data set (Figure 4A). The most influential variables in TDM prediction using the JointMLP model were consistent with eGFR CKD-EPI, average vancomycin dose, and average distribution volume, which were variables identified in the internal data set (Figure 4B). The 300-layer MLP was statistically significant ($P<.001$) compared to the PPK model. Additionally, the JointMLP model had a mean error and variance that were close to zero and smaller than those of the PPK model in the MIMIC-IV data set (Figure 4C and D).
Figure 4. Result figures from the Medical Information Mart for Intensive Care (MIMIC) validation set. (A) The $P$ value for each comparison of the mean absolute error for predictive performance on the MIMIC validation data set; (B) the Shapley Additive exPlanations (SHAP) values of the joint multilayer perceptron (JointMLP) model on the MIMIC validation set; (C) scatter plots of the MIMIC data set with the JointMLP predicted values against target values; and (D) the error histograms of the MIMIC data set with the JointMLP predicted values against target values. 300MLP: 300-layer multilayer perceptron; CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration; eGFR: estimated glomerular filtration rate; PPK: population pharmacokinetic; TDM: therapeutic drug monitoring; XGB: extreme gradient boosting.

Discussion

Principal Findings

The PPK method is widely used for predicting vancomycin TDM levels. However, it is population-dependent, leading to inappropriate results for some patients. Therefore, this study proposes a consistent decision-making system that is population-independent and unaffected by clinicians’ abilities by applying a joint MLP model that integrates various successful deep learning models and decision trees. The proposed JointMLP showed the best performance for predicting vancomycin trough concentration in all data sets compared to other models. Significantly, in the MAE metric, the JointMLP model improved performance by 90% over the PPK model. Results mean that the proposed model can be applied to predict appropriate vancomycin trough concentrations in patients with critical illnesses who require vancomycin treatment in various situations.

Comparison to Machine Learning Model

In our study, we compared the performance of 3 machine learning models for predicting vancomycin levels: JointMLP, XGBoost, and TabNet. XGBoost is a popular gradient-boosting algorithm known for its high accuracy and flexibility, while TabNet is a newer deep-learning model specifically designed for tabular data. Compared to XGBoost and TabNet, ourJointMLP model outperformed both in terms of predictive accuracy. One of the main reasons for this is that our model was designed to handle noisy data samples, a common characteristic of medical data sets. A deep neural network architecture accomplishes this with multiple layers, which allows the model to learn complex relationships between the input features and the output variable. In contrast, XGBoost and TabNet are both tree-based models, which may need to be more effective at handling noisy data.

Compared to previous studies [15,24-26], our proposed JointMLP model performs better in accurately predicting vancomycin levels. Specifically, our model outperforms XGBoost and TabNet, 2 well-known models in analyzing tabular data and predicting drug concentrations. Moreover, our JointMLP model is designed to handle noisy data samples effectively, a crucial feature for real-world applications in the medical domain where data quality can be highly variable. Our model’s robustness is demonstrated by comparing it to a
300-layer MLP, confirming the effectiveness of our model structure in handling vancomycin TDM predictions. Importantly, our model is trained on a larger and more diverse data set than previous studies and tested on an unseen data set, demonstrating its generalizability and suitability for predicting drug concentrations. Overall, our findings highlight the potential of deep learning models in improving the accuracy and reliability of drug concentration prediction, which can lead to better treatment outcomes and improved patient health. Overall, our findings suggest that the JointMLP model is a promising approach for predicting vancomycin levels, particularly in the presence of noisy data. However, the choice of which model to use may depend on the specific characteristics of the data set being analyzed and the goals of the analysis.

Limitations and Future Work
When considering implementing our approach in clinical settings, there are several factors to consider. First, although we extended our model to validate across diverse ethnic groups by including patients from Asia and other regions in the MIMIC-IV data set, we acknowledge that there is still room for improvement in this regard. We used a larger data set than previous studies, with 977 samples as the internal data set and 1429 samples as the external data set. Additionally, we sought to develop a generalized model by applying the MIMIC-IV data set, which includes 2394 samples. However, additional validation with external data sets is necessary to enhance further the generalizability of our model for predicting vancomycin levels. Second, our comparison of the predictive performance with the PPK method was based on the trough-based target method, which is easily used in the treatment of patients with critical illnesses, rather than the AUC/MIC-specific range that is currently recommended. Therefore, further research is needed to compare the AUC/MIC range with our approach. Despite these limitations, our deep learning model, if implemented in actual clinical practice, could significantly improve the treatment outcomes of ICUs by supporting clinical decision-making in a more standardized and consistent manner. Our proposed JointMLP model will continue to evolve and update its performance to mimic the human brain better and determine the optimal vancomycin dose.

Conclusions
In this study, proposing the JointMLP approach in clinical settings has significant implications for public health, as it can help optimize treatment outcomes for patients with critical illnesses in ICUs. By providing a more accurate and consistent method for predicting vancomycin levels, the model can reduce adverse events associated with suboptimal vancomycin dosings, such as the increased risk of bacterial resistance, longer hospital stays, and higher health care costs. Furthermore, the ongoing development and improvement of the JointMLP model through continuous learning and updating can lead to more effective treatments and better health outcomes for patients, improving the accuracy and reliability of other clinical prediction models in the future. Overall, the implementation of the JointMLP approach has the potential to improve public health outcomes and benefit patients worldwide.

Acknowledgments
We would like to thank Jihye Han for helping us with data mining. This study was supported by the National Research Foundation of Korea Grant and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Korean Government (HI21C1074).

Data Availability
The data used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
JP contributed to the conceptualization. HSC, DHK, DHL, and MGK were involved in the methodology. DHK, DHL, and MGK were responsible for the software. HSC was responsible for validation. JP, HSC, and DHL contributed to the formal analysis. HSC and DHK were involved in the investigation. YK and MGK were responsible for resources. JP, JHP, YH, and SSH were involved in data curation. JP, DHK, DHL, and MGK wrote and prepared the original draft. JP, HSC, DHK, DHL, and MGK were involved in writing, reviewing, and editing. HSC, DHK, DHL, and MGK were responsible for visualization. YK, JP, HSC, and DHL contributed to supervision and project administration. YK, JP, and HSC were involved in funding acquisitions. All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

References
2. Lodise TP, Drusano G. Vancomycin area under the curve-guided dosing and monitoring for adults and pediatric patients with suspected or documented serious methicillin-resistant staphylococcus aureus infections: putting the safety of our patients first. Clin Infect Dis 2021;72(9):1497-1501 [FREE Full text] [doi: 10.1093/cid/ciaa1744] [Medline: 33740042]


20. Vancomycin IV. Infectious Diseases Management Program at UCSF. URL: https://idmp.ucsf.edu/content/vancomycin-iv [accessed 2023-12-29]


Abbreviations

AUC: area under the curve
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration
DUIH: Dongguk University Ilsan Hospital
eGFR: estimated glomerular filtration rate
ICU: intensive care unit
JointMLP: joint multilayer perceptron
KNUH: Kangwon National University Hospital
MAE: mean absolute error
MIC: minimum inhibitory concentration
MIMIC-IV: Medical Information Mart for Intensive Care–IV
MLP: multilayer perceptron
PK/PD: pharmacokinetic/pharmacodynamic
PPK: population pharmacokinetic
RMSE: root-mean-square error
SHAP: Shapley Additive exPlanations
TDM: therapeutic drug monitoring
XGBoost: extreme gradient boosting

Edited by T Leung, G Eysenbach; submitted 21.12.22; peer-reviewed by JY Hong, G Lim, S Kim, Y Zhang; comments to author 22.02.23; revised version received 28.03.23; accepted 27.12.23; published 08.03.24.

Please cite as:
A Deep Learning–Based Approach for Prediction of Vancomycin Treatment Monitoring: Retrospective Study Among Patients With Critical Illness
JMIR Form Res 2024;8:e45202
URL: https://formative.jmir.org/2024/1/e45202
doi:10.2196/45202
PMID:38152042

©Dohyun Kim, Hyun-Soo Choi, DongHoon Lee, Minkyu Kim, Yoon Kim, Seon-Sook Han, Yeonjeong Heo, Ju-Hee Park, Jinkyeong Park. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Optimization of Using Multiple Machine Learning Approaches in Atrial Fibrillation Detection Based on a Large-Scale Data Set of 12-Lead Electrocardiograms: Cross-Sectional Study

Beau Bo-Sheng Chuang1*; Albert C Yang1,2,3*, MD, PhD

1School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan
2Digital Medicine and Smart Healthcare Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan
3Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan

*all authors contributed equally

Corresponding Author:
Albert C Yang, MD, PhD
Digital Medicine and Smart Healthcare Research Center
National Yang Ming Chiao Tung University
No 155, Li-Nong St, Sec.2
Beitou District
Taipei, 112304
Taiwan
Phone: 886 228267995
Fax: 886 228267389
Email: accyang@nycu.edu.tw

Abstract

Background: Atrial fibrillation (AF) represents a hazardous cardiac arrhythmia that significantly elevates the risk of stroke and heart failure. Despite its severity, its diagnosis largely relies on the proficiency of health care professionals. At present, the real-time identification of paroxysmal AF is hindered by the lack of automated techniques. Consequently, a highly effective machine learning algorithm specifically designed for AF detection could offer substantial clinical benefits. We hypothesized that machine learning algorithms have the potential to identify and extract features of AF with a high degree of accuracy, given the intricate and distinctive patterns present in electrocardiogram (ECG) recordings of AF.

Objective: This study aims to develop a clinically valuable machine learning algorithm that can accurately detect AF and compare different leads' performances of AF detection.

Methods: We used 12-lead ECG recordings sourced from the 2020 PhysioNet Challenge data sets. The Welch method was used to extract power spectral features of the 12-lead ECGs within a frequency range of 0.083 to 24.92 Hz. Subsequently, various machine learning techniques were evaluated and optimized to classify sinus rhythm (SR) and AF based on these power spectral features. Furthermore, we compared the effects of different frequency subbands and different lead selections on machine learning performances.

Results: The light gradient boosting machine (LightGBM) was found to be the most effective in classifying AF and SR, achieving an average $F_1$-score of 0.988 across all ECG leads. Among the frequency subbands, the 0.083 to 4.92 Hz range yielded the highest $F_1$-score of 0.985. In interlead comparisons, aVR had the highest performance ($F_1=0.993$), with minimal differences observed between leads.

Conclusions: In conclusion, this study successfully used machine learning methodologies, particularly the LightGBM model, to differentiate SR and AF based on power spectral features derived from 12-lead ECGs. The performance marked by an average $F_1$-score of 0.988 and minimal interlead variation underscores the potential of machine learning algorithms to bolster real-time AF detection. This advancement could significantly improve patient care in intensive care units as well as facilitate remote monitoring through wearable devices, ultimately enhancing clinical outcomes.

(JMIR Form Res 2024;8:e47803) doi: 10.2196/47803
KEYWORDS
machine learning; atrial fibrillation; light gradient boosting machine; power spectral density; digital health; electrocardiogram; machine learning algorithm; atrial fibrillation detection; real-time; detection; electrocardiography leads; clinical outcome

Introduction
Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia, impacting an estimated 33.5 million people worldwide [1]. This severe cardiac condition heightens the risks of stroke and heart failure [2]. Clinicians typically detect and diagnose AF through the noninvasive electrocardiogram (ECG) method. However, ECG interpretation relies heavily on the expertise of the medical professional, creating a need for automated ECG classification to support clinicians. Machine learning, a subset of artificial intelligence, has shown great potential in improving the detection and management of AF through automated ECG analysis [3], risk stratification [4], or treatment planning [5].

The 2020 PhysioNet Challenge data sets offer 12-lead ECG recordings that can be used to evaluate machine learning techniques for ECG interpretation [6]. Compared to machine learning approaches based on small and homogeneous data sets, algorithms using PhysioNet data are likely to be more representative of realistic clinical scenarios, thereby making them better suited for practical implementation. Various machine learning techniques have been used to detect AF using ECG data. Some of the widely explored methods include support vector machines, decision trees, random forests, and deep learning approaches like convolutional neural networks and recurrent neural networks. These techniques have demonstrated promising results in classifying normal sinus rhythm (SR) and AF, with fair accuracy and $F_1$-scores [4].

The performance of machine learning algorithms depends on the quality of the input features. Common feature extraction methods in AF detection include time-domain analysis, frequency-domain analysis, and wavelet transform. Power spectral density (PSD) is a popular frequency-domain feature that has been used to differentiate SR from AF. PSD analysis reveals information about the distribution of a signal’s power and frequency components. An ECG signal comprises various frequency components, including those related to the sinus heartbeat as well as atrial and ventricular activity. The PSD distribution of these frequency components may alter heart conditions that affect the cardiac contraction cycle, making it a potential indicator for identifying cardiac arrhythmias. Consequently, we developed an automated machine learning algorithm based on PSD to differentiate normal SR from AF based on a large-scale data set from PhysioNet 2020.

Methods

ECG Data

We used the 2020 PhysioNet Challenge data sets (Table 1) [6], comprising the China Physiological Signal Challenge (CPSC) Database (men: n=3699, women: n=3178, total: n=6877, and sampling rate: 500 Hz), CPSC-Extra Database (men: n=1843, women: n=1610, total: n=3453, and sampling rate: 500 Hz), St Petersburg Institute of Cardiological Technics INCART 12-lead Arrhythmia Database (total: n=72 and sampling rate: 257 Hz), Physikalisch Technische Bundesanstalt (PTB) Diagnostic ECG Database (men: n=377, women: n=139, total: n=516, and sampling rate: 1000 Hz), Physikalisch Technische Bundesanstalt extra large (PTB-XL) electrocardiography Database (men: n=11,379, women: n=10,458, total: n=21,837, and sampling rate: 500 Hz), and Georgia 12-lead ECG Challenge Database (men: n=5551, women: n=4793, total: n=10,344, and sampling rate: 500 Hz). The PhysioNet Challenge data sets can be accessed publicly [7], and data access is licensed under the Creative Commons Attribution 4.0 International Public License [8].

Table 1. 2020 PhysioNet challenge data sets.

<table>
<thead>
<tr>
<th>Properties</th>
<th>Data set</th>
<th>Recording time</th>
<th>Sampling frequency (Hz)</th>
<th>Participants, n</th>
<th>Participants, n</th>
<th>Participants, n</th>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPSC² Database</td>
<td>6–60 seconds</td>
<td>500</td>
<td>Male</td>
<td>3699</td>
<td>N/A ³</td>
<td>377</td>
</tr>
<tr>
<td></td>
<td>CPSC-Extra Database</td>
<td>6–60 seconds</td>
<td>500</td>
<td>Female</td>
<td>1843</td>
<td>139</td>
<td>11,379</td>
</tr>
<tr>
<td></td>
<td>St Petersburg INCART² Database</td>
<td>30 minutes</td>
<td>257</td>
<td>Total</td>
<td>6877</td>
<td>1610</td>
<td>5551</td>
</tr>
<tr>
<td></td>
<td>PTB² Diagnostic</td>
<td>Unknown</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td>4793</td>
</tr>
<tr>
<td></td>
<td>PTB-XL²</td>
<td>10 seconds</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td>5551</td>
</tr>
<tr>
<td></td>
<td>Georgia</td>
<td>10 seconds</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td>10,344</td>
</tr>
</tbody>
</table>

²CPSC: China Physiological Signal Challenge.
³INCART: Institute of Cardiological Technics.
⁴PTB: Physikalisch Technische Bundesanstalt.
⁵PTB-XL: Physikalisch Technische Bundesanstalt extra large.
⁶N/A: not applicable.

https://formative.jmir.org/2024/1/e47803
From these data sets, we initially selected data featuring a 500 Hz sampling rate. Next, we identified ECG diagnoses using the Systematized Nomenclature of Medicine—Clinical Terminology codes 426783006 and 164889003 for SR and AF, respectively. This led to the identification of 20,766 SR and 3458 AF ECG recordings. After excluding recordings corresponding to multiple diagnoses, the final data set comprised 9102 SR and 1088 AF ECG recordings.

**Ethical Considerations**

This study was approved by the institutional review board of Taipei Veterans General Hospital (2022-04-004BC).

**ECG Preprocessing and Feature Extraction**

The ECG patterns associated with AF exhibit greater variability compared to those of SR. To explore this distinction, we used the Welch method [9,10] to estimate PSD and determine the frequency composition of the ECG signals. The recording durations of the ECG data ranged from 6 to 60 seconds, with 10 seconds being the most common. This variability could lead to differences in PSD resolution. To address this issue, we used the frequency bin method to ensure consistent frequency resolution across our PSD features. We set the frequency resolution to 1/6 Hz and calculated the arithmetic mean of the PSD frequencies within this resolution to represent the midpoint of each envelope. The data we used had a sampling rate of 500 Hz. Based on Nyquist-Shannon sampling theorem, the highest frequency information we can get is 250 Hz. After observing several data plots of the original PSD graph, we can see that most of the power of the signal is within 0 to 25 Hz. Thus, for a 10-second ECG recording, we selected a frequency of 0 to 25 Hz which was converted to 0.083 to 24.92 Hz. Figure 1 presents an example of original PSD plots and explains why we chose the PSD from 0 to 25 Hz. For further segmentation, it was just reasonable to separate the whole frequency segment into 5 smaller segments since the whole frequency ranges from 0 to 25 Hz. Figure 2 presents example ECG signals and PSD plots, highlighting notable differences in PSD between SR and AF; narrower harmonic frequency components can be observed in SR, while AF patterns display more erratic frequency components. We used these PSD features as inputs for our machine learning algorithm.

Figure 1. Initial results of PSD for SR and AF in lead 1. The original frequency information we got from PSD is 0 to 250 Hz. From the figure, we can see that the power of the signal concentrates within 0 to 25 Hz. Thus, we selected 0 to 25 Hz as our main frequency band. AF: atrial fibrillation; PSD: power spectral density; SR: sinus rhythm.
Figure 2. A comparison of the ECG and PSD for SR and AF in lead V4. While there is a discernible difference between SR and AF in the ECG representation, it remains challenging to distinguish them in this format. In contrast, the PSD figure demonstrates the harmonic nature of SR and the chaotic nature of AF, making it much easier to differentiate between the two. As a result of these observations, we have chosen to use PSD as the primary feature for our machine learning algorithms. This decision is based on the enhanced clarity and distinction provided by the PSD representation, which allows for more accurate and effective differentiation between SR and AF in our analysis. AF: atrial fibrillation; ECG: electrocardiogram; PSD: power spectral density; SR: sinus rhythm.

Machine Learning Methods
Supervised machine learning is a type of machine learning where the algorithm is trained on a labeled data set, and the goal is to make predictions for new, unseen data. There are several types of supervised machine learning methods used in this study, including (1) discriminant analysis: a method for modeling the relationship between a dependent variable and one or more independent variables (used for continuous target variables); (2) logistic regression: a method for modeling the probability of a binary outcome based on one or more independent variables; (3) decision trees: a method for making predictions by creating a tree-like model of decisions and their potential consequences; (4) random forests: an ensemble method that combines multiple decision trees to make a prediction; (5) support vector machines: a method for classifying data by finding the best boundary (or “hyperplane”) that separates the classes; (6) neural networks: a method inspired by the structure and function of the human brain (can be used for a wide range of tasks, including classification and regression); and (7) naive Bayes: a probabilistic method for making predictions based on Bayes’ theorem (often used for text classification).

Additionally, we used a gradient boosting decision tree algorithm, namely light gradient boosting machine (LightGBM) [11]. It is one of the most efficient algorithms in recent years, known for its speed and accuracy. Boosting is an approach for combining multiple base models into a composite one. One of the ensemble tree algorithms of LightGBM is a “leaf-wise” tree algorithm, wherein the tree grows vertically; most of the others are “level-wise” tree algorithms, wherein the trees grow horizontally. The leaf-wise structure reduces time complexity and offers a favorable balance of accuracy and efficiency, especially for large-scale, high-dimensional data. Thus, it is useful in classification problems.

Perceptrons are single-neuron models and the precursors to larger neural networks; now, they are the building blocks of neural networks. A deep neural network (DNN) is a supplement of a feed-forward neural network and consists of 3 types of layers: the input layer, the output layer, and the hidden layer. The input layer receives the input signal to be processed. The output layer performs the required task, such as prediction or classification. The true processing engine is the hidden layers, placed between the input and output layers. In a DNN, data flow from the input layer to the output layer, just as in feed-forward neural networks. The neurons are trained through back propagation. DNNs are widely used in various applications, including automated diagnosis using ECG data [12].

Statistical Analysis
Data Imbalance
We used the synthetic minority over-sampling technique [13] to filter the data and reduce the imbalance between the numbers of SR and AF recordings (n=9102 and n=1088, respectively).

F1-Score
To mitigate bias resulting from data imbalance, we used the F1-score as the primary metric for evaluating machine learning models. F1-scores treat false-positive and false-negative errors equally and are more useful than accuracy in cases of class
imbalance. The $F_1$-score is the harmonic mean of precision and recall, which measures the errors contributed by false positives and false negatives, respectively.

**Training, Validation, and Testing Data Sets**

We divided our data sets into training, validation, and testing sets. The training set was used to develop the models, the validation set for model tuning, and the testing set for model assessment. In this study, 60% of the data comprised the training set, 20% the validation set, and 20% the testing set. For model tuning, we used ensemble learning to identify optimal parameters for each model based on the validation set.

**Results**

**Comparison of the Performance of Various Machine Learning Algorithms**

We evaluated various machine learning methods, including extra tree, LightGBM (Microsoft Corporation), CatBoost (Yandex), XGBoost (The XGBoost Contributors), decision tree, k-nearest neighbors, stochastic gradient descent, gradient boosting, random forest, naive Bayesian, logistic regression, and DNN. Our analysis reveals that machine learning algorithms with boosting methods outperform those without boosting. In particular, LightGBM outperformed the other methods, achieving the highest average $F_1$-score of 0.988 across all 12 ECG leads and the lowest computation time for features within the entire frequency range (0.083-24.92 Hz). Conversely, our results indicate that the naive Bayesian algorithm performs poorly in classifying SR and AF. Despite its simplicity, naive Bayesian appears to be less suited for this specific task, and alternative approaches should be considered for better accuracy and reliability. Thus, we focused solely on LightGBM for subsequent analyses.

**Effect of Frequency Band on Model Performance**

We examined the contributions of various frequency subbands of ECG PSD features to model performance in detecting AF (Table 2). The entire frequency range was divided into 5 subbands (0.083-4.92, 5.083-9.92, 10.083-14.92, 15.083-19.92, and 20.083-24.92 Hz). The highest overall $F_1$-score was achieved by the model using the full frequency range (0.083-24.92 Hz), and the frequency subband of 0.083 to 4.92 Hz yielded the highest $F_1$-score among the subbands. Generally, the $F_1$-score for each subband was around 0.9, indicating that every subband contains valuable information for ECG classification.

**Table 2. $F_1$-scores for various frequency bands.**

<table>
<thead>
<tr>
<th>Performance</th>
<th>Frequency range</th>
<th>0.083-4.92 Hz</th>
<th>5.083-9.92 Hz</th>
<th>10.083-14.92 Hz</th>
<th>15.083-19.92 Hz</th>
<th>20.083-24.92 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$-score</td>
<td>Whole range</td>
<td>0.988</td>
<td>0.985</td>
<td>0.956</td>
<td>0.925</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td>0.083-4.92 Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.083-9.92 Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.083-14.92 Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.083-19.92 Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20.083-24.92 Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Interlead Differences in Model Performance**

In lead comparisons, aVR demonstrated the highest performance across all frequency subbands, although the differences between leads were minimal (Table 3). Specifically, for the entire frequency range, using aVL resulted in the lowest classification performance ($F_1$-score=0.980), while using aVR achieved the highest performance ($F_1$-score=0.993).

**Table 3. $F_1$-scores for various leads and frequency bands.**

<table>
<thead>
<tr>
<th>Frequency range</th>
<th>Different leads</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>aVR</th>
<th>aVL</th>
<th>aVF</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole range</td>
<td></td>
<td>0.985</td>
<td>0.990</td>
<td>0.981</td>
<td>0.993</td>
<td>0.980</td>
<td>0.984</td>
<td>0.988</td>
<td>0.988</td>
<td>0.991</td>
<td>0.991</td>
<td>0.990</td>
<td>0.989</td>
</tr>
<tr>
<td>0.083-4.92 Hz</td>
<td></td>
<td>0.980</td>
<td>0.989</td>
<td>0.978</td>
<td>0.990</td>
<td>0.980</td>
<td>0.980</td>
<td>0.986</td>
<td>0.988</td>
<td>0.987</td>
<td>0.990</td>
<td>0.990</td>
<td>0.987</td>
</tr>
<tr>
<td>5.083-9.92 Hz</td>
<td></td>
<td>0.961</td>
<td>0.968</td>
<td>0.930</td>
<td>0.978</td>
<td>0.936</td>
<td>0.957</td>
<td>0.961</td>
<td>0.947</td>
<td>0.959</td>
<td>0.958</td>
<td>0.963</td>
<td>0.958</td>
</tr>
<tr>
<td>10.083-14.92 Hz</td>
<td></td>
<td>0.920</td>
<td>0.928</td>
<td>0.900</td>
<td>0.949</td>
<td>0.917</td>
<td>0.921</td>
<td>0.926</td>
<td>0.923</td>
<td>0.936</td>
<td>0.940</td>
<td>0.912</td>
<td>0.937</td>
</tr>
<tr>
<td>15.083-19.92 Hz</td>
<td></td>
<td>0.878</td>
<td>0.915</td>
<td>0.854</td>
<td>0.948</td>
<td>0.900</td>
<td>0.909</td>
<td>0.920</td>
<td>0.907</td>
<td>0.923</td>
<td>0.927</td>
<td>0.938</td>
<td>0.924</td>
</tr>
<tr>
<td>20.083-24.92 Hz</td>
<td></td>
<td>0.868</td>
<td>0.915</td>
<td>0.857</td>
<td>0.939</td>
<td>0.871</td>
<td>0.885</td>
<td>0.877</td>
<td>0.882</td>
<td>0.920</td>
<td>0.916</td>
<td>0.940</td>
<td>0.925</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

In this study, we found that the LightGBM algorithm was the most effective machine learning model for our purposes. Notably, all frequency subbands contained distinct information, achieving independent $F_1$-scores of approximately 0.9. Consequently, using the entire frequency range (0.083-24.92 Hz) provided the most comprehensive features and yielded the highest $F_1$-score of 0.988. Interestingly, among the limb leads, aVR produced the best result, which was marginally superior to the outcomes of chest leads and lead II. This finding is unexpected, as lead aVR is rarely used in real-world scenarios for detecting arrhythmia. The contributions of different frequency bands to classification performance showed that even though each subband contains unique features, the dominant frequency subband (0.083-4.92 Hz) is sufficient for distinguishing AF from SR.
In terms of the contributions of frequency bands to classification performance, it is important to note that the normal human heart beats between 60 to 100 times per minute (1-1.67 Hz), whereas the dominant frequency in AF ranges from 3.8 to 8 Hz [14]. Given the differences in frequency components between SR and AF, they can be used to distinguish one from the other. As demonstrated in Figure 1, multiple frequency peaks are observed within the 0 to 25 Hz range for both SR and AF. Consequently, frequency subbands containing distinct features may yield different results.

Initially, we hypothesized that the subband containing the dominant frequency (0.083-4.92 Hz) would result in the optimal classification of SR and AF, similar to the whole frequency range (0.083-24.92 Hz), while other subbands would struggle to classify SR and AF effectively. Indeed, models using the dominant frequency subband performed well, with results closely resembling those of models using the whole frequency range. However, other subbands also demonstrated high performance independently (Table 2), with $F_1$-scores exceeding 0.854. The dominant frequency subband (0.083-4.92 Hz) yielded an $F_1$-score of 0.985, which is comparable to that of the whole frequency range ($F_1$-score=0.988). These findings suggest that although each subband contains unique features and cannot be replaced by other subbands, the dominant frequency subband alone is sufficient for differentiating AF from SR.

In this study, we explore the variations in AF classification performance across different ECG leads. Each ECG lead has unique physiological applications, which in turn influence the preference of medical professionals in identifying specific types of arrhythmias. For instance, lead 1, aVF, or V1 are frequently used in diagnosing right ventricular hypertrophy. Routine clinical practice suggests that lead 2, which is typically the focus of medical professionals when examining AF, would yield the most accurate results. Surprisingly, our findings revealed that all leads, within the frequency range of 0.083-24.92 Hz, generated $F_1$-scores in close proximity to 0.985. Among these, lead aVR achieved the highest $F_1$-score at 0.993. We believe this can be attributed to the visibility of AF in the majority of leads and the comprehensive features encompassed by the entire frequency range.

Contrary to our initial expectations, lead aVR, which is not a common choice in clinical settings, produced the most optimal results when using subbands [15]. This finding may potentially shed light on the reentry mechanism that underlies AF [16,17]. It is worth noting that lead aVR captures data from the right ventricle outflow tract and the basal portion of the septum, whereas AF predominantly manifests in the left ventricle. Consequently, the reentry mechanism emerges as the most plausible explanation for this observed phenomenon.

In a prior study [18], the 2020 PhysioNet Challenge data sets were used alongside the LightGBM machine learning algorithm. However, the approaches to signal extraction and processing in that study differed from those used in our research. The previous study implemented filtering techniques and wavelet multiresolution analysis, while our investigation used PSD and the Welch method. Another key distinction between the 2 studies is the scope of the classification task. While our research focused solely on differentiating AF from SR, the previous study attempted to identify 24 different diseases. Due to the varied physiological information provided by each ECG lead, we opted to compare the performance of LightGBM for each individual lead, rather than inputting data from all 12 leads into the LightGBM simultaneously. Furthermore, our study identified the most critical frequency subband within the total frequency range. We determined the dominant frequency subband to be 0.083-4.92 Hz, which played a significant role in our analysis.

There are several limitations to be addressed in this study. One notable constraint is our selection of 10-second ECG recordings, which may not accurately represent the full spectrum of real-world clinical scenarios. The data sets used in this study were predominantly clean, while actual clinical situations may involve data contaminated by noise, potentially affecting the results. Moreover, our study focused exclusively on single-diagnosis data, representing either AF or SR. This approach, however, does not account for the possibility of a patient experiencing multiple arrhythmias concurrently. Consequently, our findings were derived under simplified conditions that may not fully reflect the complexity of real-life cases. In light of these limitations, it is essential to interpret our results with caution and consider further research that incorporates a broader range of ECG recordings, addresses potential noise-related challenges, and examines cases with multiple coexisting arrhythmias to enhance the generalizability and applicability of our findings to real-world clinical settings.

In summary, LightGBM proves to be a highly effective algorithm for distinguishing AF from SR. Our study demonstrated that when dividing the entire ECG frequency range into subbands with fewer features, lead aVR delivered the best performance. This outcome could potentially be associated with the underlying pathological mechanisms of AF. By incorporating adequate frequency band features, lead 2 ECG data can achieve an $F_1$-score of approximately 0.99. This level of accuracy and efficiency of the LightGBM model renders the algorithm suitable for implementation in clinical practice and integration into commercialized electronic devices, such as the smartwatch with the ECG functionality, for a range of health care applications. This study’s findings suggest that leveraging the power of LightGBM could enhance arrhythmia detection and monitoring, ultimately improving patient care and outcomes.

Literature Review

In this review, we will talk about studies that introduce techniques for AF detection, analyzing their methods, results, and potential implications in the field of cardiology. First, one study [19] proposed a feature extraction method based on a gradient set for AF detection, which features simplicity, noise tolerance, and adaptability to various classifiers. However, we can still improve the feature extraction method in the future, such as improving AF detection performance by proposing a more representative feature set or combining it with other types of features. This approach shows the potential for streamlined, efficient feature extraction, providing a solid foundation for further research in the domain.
Next, the Transposed Projection–Convolutional Neural Network (TP-CNN) method was introduced in another study [20], aiming to use “compressed ECG signals” for AF detection. The approach demonstrates promising results in accurately detecting AF in wearable application scenarios while addressing energy consumption concerns. It shows the effective use of compressed signals and a highly accurate algorithm. Yet, its data sources only contain 25 patients. More patients should be included in further studies.

Another study [21] demonstrates a unique method using multiple parameters, including the average number of f waves in a TQ interval, showing robust real-time AF detection capabilities. This approach not only differentiates between AF and normal ECGs but also outperforms distinguishing AF from other arrhythmias. However, they can discuss more about how their algorithms will be implemented in clinical situations.

There is another study [22] focusing on low-complexity algorithms for AF detection, a method based on RR interval features. This approach demonstrated reasonable feature selection while maintaining a low computational cost, making it suitable for low-power devices. However, in future studies, it should focus on getting a better $F_1$-score.

In addition to ECG AF detection, there is also a signal called photoplethysmography common in wearable devices for AF detection, while we usually use ECG in clinical situations. This Fitbit Heart study [23] provides an algorithm that exhibited a high positive predictive value for concurrent AF, highlighting the potential of consumer wearables for large-scale AF identification. Though the result is nice, we hope we can know more about photoplethysmography comparison with ECG.

In conclusion, these studies collectively contribute to the evolving landscape of AF detection methods. While each approach offers unique advantages, further research should focus on refining these methods, exploring real-time applications, and enhancing the overall accuracy and efficiency of AF detection algorithms. Potential limitations should also be emphasized in future studies.

Acknowledgments
This study was supported by the National Science and Technology Council of Taiwan with grants to ACY (grant MOST 111-2622-B-002-103-MY2, MOST 111-2628-B-033-003-MY3, MOST 111-2728-B-033-001, MOST 111-2628-B-033-006, MOST 112-2628-B-033-001, MOST 112-2728-B-033-001, MOST 112-2634-F-033-001, MOST 112-2634-F-033-002, MOST 112-2634-F-033-003, MOST 112-2634-F-033-004, MOST 112-2634-F-033-005, MOST 112-2634-F-033-006, MOST 112-2634-F-033-007, MOST 112-2634-F-033-008, MOST 112-2634-F-033-009, MOST 112-2634-F-033-010, MOST 112-2634-F-033-011, MOST 112-2634-F-033-012, MOST 112-2634-F-033-013, MOST 112-2634-F-033-014). ACY was also supported by the Mt Jade Young Scholarship Award from the Ministry of Education, Taiwan, as well as Brain Research Center, National Yang Ming Chiao Tung University, and the Ministry of Education (Aim for the Top University Plan), Taipei, Taiwan. All the funding resources provided financial supports only and had no other role in study design, data collection, analysis, or interpretations.

Data Availability
The data sets generated and analyzed during the current study are available in the 2020 PhysioNet Challenge data sets repository [7].

Conflicts of Interest
None declared.

References
8. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-01-22]


Abbreviations

- **AF**: atrial fibrillation
- **CPSC**: China Physiological Signal Challenge
- **DNN**: deep neural network
- **ECG**: electrocardiogram
- **INCART**: Institute of Cardiological Technics
- **LightGBM**: light gradient boosting machine
- **PSD**: power spectral density
- **PTB**: Physikalisch Technische Bundesanstalt
- **PTB-XL**: Physikalisch Technische Bundesanstalt extra large
- **SR**: sinus rhythm
- **TP-CNN**: Transposed Projection–Convolutional Neural Network
Designing an App to Support Measurement-Based Peer Supervision of Frontline Health Workers Delivering Brief Psychosocial Interventions in Texas: Multimethod Study

Anubhuti Poudyal1, MPH; Delta-Marie Lewis2, BA; Sarah Taha3, SM; Alyssa J Martinez3, MS; Lauren Magoun2, MPH; Y Xian Ho2, PhD; Natali Carmio3, PsyD; John A Naslund3, PhD; Katherine Sanchez3, PhD; Neal Lesh2, PhD; Vikram Patel3, PhD

1Department of Sociomedical Sciences, Columbia Mailman School of Public Health, New York, NY, United States
2Dimagi, Inc, Cambridge, MA, United States
3Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, United States
4Patient and Community Engaged Research Center, Baylor Scott and White Research Institute, Dallas, TX, United States

Corresponding Author:
Anubhuti Poudyal, MPH
Department of Sociomedical Sciences
Columbia Mailman School of Public Health
722 W 168th St
New York, NY, 10032
United States
Phone: 1 (212) 305 2862
Email: ap4150@cumc.columbia.edu

Abstract

Background: The unmet need for mental health care affects millions of Americans. A growing body of evidence in implementation science supports the effectiveness of task sharing in the delivery of brief psychosocial interventions. The digitization of training and processes supporting supervision can rapidly scale up task-shared interventions and enable frontline health workers (FLWs) to learn, master, and deliver interventions with quality and support.

Objective: We aimed to assess the perceived feasibility and acceptability of a novel mobile and web app designed and adapted to support the supervision, training, and quality assurance of FLWs delivering brief psychosocial interventions.

Methods: We followed human-centered design principles to adapt a prototype app for FLWs delivering brief psychosocial interventions for depression, drawing from an app previously designed for use in rural India. Using a multimethod approach, we conducted focus group sessions comprising usability testing and group interviews with FLWs recruited from a large health system in Texas to assess the feasibility and acceptability of the app. The positive System Usability Scale was used to determine the app’s overall usability. We also calculated the participants’ likelihood of recommending the app to others using ratings of 0 to 10 from least to most likely (net promoter score). Focus group transcripts were coded and analyzed thematically, and recommendations were summarized across 4 key domains.

Results: A total of 18 FLWs varying in role and experience with client care participated in the study. Participants found the app to be usable, with an average System Usability Scale score of 72.5 (SD 18.1), consistent with the industry benchmark of 68. Participants’ likelihood of recommending the app ranged from 5 to 10, yielding a net promoter score of 0, indicating medium acceptability. Overall impressions of the app from participants were positive. Most participants (15/18, 83%) found the app easy to access and navigate. The app was considered important to support FLWs in delivering high-quality mental health care services. Participants felt that the app could provide more structure to FLW training and supervision processes through the systematic collection and facilitation of performance-related feedback. Key concerns included privacy-related and time constraints regarding implementing a separate peer supervision mechanism that may add to FLWs’ workloads.

Conclusions: We designed, built, and tested a usable, functional mobile and web app prototype that supports FLW-delivered psychosocial interventions in the United States through a structured supervision mechanism and systematic collection and review of performance measures. The app has the potential to scale the work of FLWs tasked with delivering these interventions to the
hardest-to-reach communities they serve. The results of this project will inform future work to evaluate the app’s use and efficacy in real-world settings to support task-shared mental health programs across the United States.

(JMIR Form Res 2024;8:e55205) doi:10.2196/55205

KEYWORDS
digital technology; mental health; depression; task sharing; nonspecialist providers; peer supervision; therapy quality

Introduction

Background

In the United States, 22.8% of adults have a mental illness, and 5.5% have a serious mental illness [1]. Following the COVID-19 pandemic, the percentage of adults with symptoms of anxiety or depressive disorders has increased substantially, with preliminary data reporting a prevalence of 20% to 32.3% [2,3], although the extent of the increase is still unclear [4]. In 2021, in the United States, only 65.4% of adults with serious mental illness and 47.2% with any mental illness received mental health services [1], and the proportion receiving “minimally adequate” treatment [5], defined as treatment minimally sufficient for common mental disorders, is low [6,7]. In addition to the direct impact on health and well-being, the economic cost of mental illness is high [8]. Worldwide, 418 million disability-adjusted life years and an economic burden of US $5 trillion are associated with mental health disorders [8]. In the United States, the federal government spent US $280 billion on mental health services in 2020 [9]. Despite significant efforts, access to mental health services remains highly limited owing to barriers such as cost and availability of services [9].

The rate of untreated mental health conditions is a major concern. Although psychosocial interventions have shown remarkable effectiveness in supporting individuals with mental conditions [10,11], barriers to accessing services limit the number of people benefiting from such interventions [12]. Access to mental health support varies widely across the country. According to a 2018 study, over 60% of rural counties reported lacking access to a psychiatrist or psychologist [13]. Studies have reported long delays in appointment scheduling [14], with clients sometimes having to wait many months before they are seen; this is especially true for scheduling first-time appointments [15] and has thus become a common issue for mental health care [16,17]. The COVID-19 pandemic has exacerbated the gaps between a need for and access to mental health services [18], with renewed calls to address the shortage of mental health professionals in the United States [19].

To address the gap in providing adequate services for mental health conditions, the World Health Organization and other institutions have advocated for task-sharing approaches, including screening and treatment protocols that frontline health workers (FLWs) can follow to extend the role of mental health specialists wherever possible, particularly in areas where specialist support is nonexistent or inaccessible [20]. Worldwide, there is growing evidence of the effectiveness of psychosocial interventions delivered by FLWs who are not behavioral health specialists [21-24]; frontline interventions have demonstrated effectiveness and cost-effectiveness in addressing mental health in high-income and low-income countries worldwide [25-29].

Task-sharing programs with trained FLWs were initially pioneered and have since been broadly evaluated across various low- and middle-income countries, many with <1 psychiatrist per 100,000 people, rendering specialist access a near impossibility for much of the world’s population [30-33]. Following the concept of “reciprocal innovation,” in which mutual benefit and learnings are derived from innovations demonstrated to be effective in low-income settings and then applied to high-income settings and vice versa [34], task-sharing programs are now being advocated for use in high-income countries experiencing critical shortages of mental health providers and a need for equitable, culturally competent, and value-based care [21,25,35]. Task-sharing programs show great promise for increasing the health system’s capacity to meet the overwhelming demand for mental health care in the United States, particularly in areas with already limited access to specialized care. They can further allow specialists to focus on more complicated cases, potentially reducing delays for those who require more specialized care [36,37]. Scaling up such task-shared programs requires innovative strategies to ensure quality delivery of interventions and support FLWs’ skill enhancement. Using digital tools for training and supervision of FLWs is a promising next step [21] to enhance FLW skills, deliver psychosocial interventions effectively, and provide supportive supervision. Here, we define an FLW as any health worker without specialized training in mental health care who may be tasked with providing mental health support to clients in their communities, which in the United States can include primary care providers, physician assistants, community health workers (CHWs), lay therapists, peer navigators, nurses, midwives, and health coaches [38]. FLW-delivered interventions often have brief core elements (eg, behavioral activation [BA] [21,39]) and are intended to be conducted during FLW visits with clients in the community or primary care settings.

We built an app prototype guided by the principles of measurement-based care and supportive supervision based on an earlier app designed to facilitate supervision for CHWs in rural India [40] to facilitate skill enhancement of FLWs and effective delivery of FLW-led interventions in Texas. Measurement-based care includes methods of monitoring and improving the quality of psychosocial intervention delivery through objective, measurable metrics generated through systematic data collection to monitor client progress and directly inform care decisions [41-43]. Supportive supervision shifts the focus from the supervisor-supervisee dyad to the entire workforce by including self-assessments and assessments by peers, community members, and supervisors [44]. By digitally supporting measurement-based care and supportive supervision, the app can provide scalable access to measurable, evidence-based metrics that can guide assessments by peers as well as oneself and one’s supervisor. The app is designed to
support FLWs trained to implement a psychosocial intervention and who can benefit from a metrics-driven, supportive supervision approach to enhance individual performance and ensure intervention fidelity.

**Goal of This Study**

In this study, we aimed to assess the feasibility and acceptability of an adapted version of this prototype app in preparation for potential implementation in the United States following the technology acceptance model [45] and learnings from prior work using similar technology in India and other global settings [46].

**Methods**

We used human-centered design principles to adapt an app for FLWs in the United States and used a multimethod approach combining usability testing and group interviews to yield collective interpretations of the feasibility and acceptability of the app among potential target end users [47].

**Prototype Design**

The design of the app was based on learnings from implementing an existing app in the Promoting Effective Mental Healthcare Through Peer Supervision (PEERS) project built on the CommCare platform and developed as part of a large-scale randomized controlled trial in India to understand whether the quality of mental health care offered by FLWs can be improved through a scalable measurement-based peer supervision model [46,48]. This study builds on evidence that an allied peer supervision model can provide a scalable approach to improve the quality of mental health care available to populations in low-resource settings [46,49]. As of late 2023, over 200 non-specialist providers or lay counselors in Madhya Pradesh and Goa with appropriate training and supervision in mental health care but no formal mental health qualifications continue to use the app to support the peer supervision component of their work.

In this study, following the PEERS model, we adapted an Android and web app using an iterative prototype design with feedback from an advisory panel of stakeholders and subject matter experts with backgrounds in clinical professional development education, mental health policy, clinical social work, and development and implementation of CHW programs in Texas. Members of the study team held 4 rounds of discussions with 5 advisory panelists to assess face validity and help ensure that potential target users’ input was directly incorporated throughout the design and development process. Through a series of in-depth interviews, advisory panel members initially provided commentary on the original PEERS app, current supervision practices in the United States, and the app’s potential to support the supervision process. Subsequent in-depth interviews probed feedback on the usability of the prototype and considerations for its implementation within the United States. The advisory panel’s feedback included inputs for the design and development of the app and considerations for future implementation.

After incorporating preliminary advisory panel feedback, a final app prototype and study materials were created and prepared for usability testing with target end users. The final app was designed to support the following FLW activities: (1) registering clients, conducting and audio recording intervention sessions, and entering session-related notes; (2) reviewing and rating audio-recorded intervention sessions they conducted themselves using a validated scale (eg, the Behavioral Activation Quality Scale [44]); (3) reviewing and rating audio-recorded intervention sessions conducted by peers; (4) reviewing data visualizations of self-ratings as well as those given by their peers and supervisors for reviewed sessions; and (5) accessing related training materials. A screenshot of the app’s menu and the details of each module are shown in Figure 1.

**Figure 1.** Screenshot of the peer supervision web application menu with descriptions of each menu item.
Previous Mental Health Training
To assess the feasibility and acceptability of this prototype app, we identified participants who had completed a 2-part training on foundational skills and BA interventions for depression and received access to measurement-based peer supervision training developed by a nonprofit program, EMPOWER, that leverages digital tools for training and supervising FLWs worldwide [46]. The FLWs in this study had been trained as part of efforts to build workforce capacity and scale up access to brief psychosocial interventions for depression in primary care and community settings across Texas [46].

Recruitment
FLWs from a large health system in Texas (ie, Baylor Scott & White Health) were recruited via email by an experienced human subjects–trained member of the study team to participate in remote focus group sessions. Candidates who met the following eligibility criteria were invited to participate: (1) aged ≥18 years, (2) fluent in English, and (3) having completed training on foundational skills and BA interventions for depression and having access to the measurement-based peer supervision training course developed by EMPOWER. Informed consent was obtained before participation.

Onboarding to the App
After completing the aforementioned training courses, which included a brief introduction to the app and brief video tutorials on how to use it, participants were invited to join a remote focus group session comprising usability testing and a group interview. All sessions were conducted via Zoom (Zoom Video Communications) and lasted approximately 1.5 hours. Participants who did not own an Android device were mailed a study tablet with the app preinstalled for use during the usability testing portion of the focus group session. Study team members trained in user experience and qualitative methods moderated the sessions (YXH and DML).

Usability Testing
To assess the app’s usability, a set of 11 usability tasks and related scenarios were created based on tasks we anticipated FLWs would commonly perform if they were to use the app in practice (see the task list in Multimedia Appendix 1). One moderator (DML) introduced each task to the participants, who then performed each task independently at their own pace. The next task was introduced once everyone had completed the preceding one unless the task took much longer than anticipated, in which case the moderator then moved participants to the next task. Moderators observed and noted participants’ task performance while providing minimal assistance with task instructions when needed.

Surveys
In addition to usability testing, each participant independently completed a series of surveys via Qualtrics (Qualtrics International Inc) to collect information on participant demographics; experience with technology; and perceived ease of use, usefulness, and user satisfaction. We used the positive System Usability Scale (SUS) [50], a 10-item Likert scale survey to assess general usability, with possible scores ranging from 0 to 100, where higher scores indicate higher usability. We also calculated the net promoter score (NPS) [51], a single-item user-reported questionnaire to rate the likelihood of recommending the app to others to assess market viability. To generate an NPS, respondents provide scores between 0 (not at all likely) and 10 (extremely likely); scores of 9 or 10 are considered “promoters,” scores of 7 or 8 are considered “passives,” and scores of 0 to 6 are considered “detectors.” The NPS is then derived by subtracting detractors from promoters (both calculated as the percentage of total respondents), with possible NPS scores ranging from −100 to 100.

Group Interviews
All participants engaged in semistructured interviews following usability testing and survey completion. An interview guide was developed and used to collect feedback on user experience with probes on the overall impressions of the app, current supervision mechanisms, and barriers to and facilitators of the broader implementation of the app. A trained and experienced moderator (YXH) followed the guide to probe feedback on the app’s perceived feasibility, acceptability, and usefulness.

Data Analysis
The focus group sessions were audio recorded, transcribed, and anonymized before coding and analysis. The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist is included in Multimedia Appendix 2 [52].

We used both deductive and inductive approaches to coding. We first defined usability, feasibility, and acceptability following definitions published by Ginsburg et al [53]. In total, 2 coauthors (YXH and DML) who moderated the focus group sessions independently read 2 transcripts and generated codes from the data. YXH, AP, and DML then finalized the codebook, which was used to direct the coding process. We used a systematic thematic coding approach [54] to generate key themes based on a priori codes and emergent codes: (1) perceived feasibility of the app, (2) acceptability of the app, (3) perceived usefulness of the app, and (4) barriers to and facilitators of scalability. Textbox 1 summarizes the definitions of the themes as used in this paper. Following thematic coding, we developed code summaries or preliminary narratives describing each theme. YXH, AP, and DML then discussed and refined the code summaries to ensure depth and breadth across all participants. The final code summaries were used to describe the findings in this paper. An intercoder agreement was calculated between 2 coders (YXH and DML) on a portion of coded data until a κ value of 0.61 was achieved. All interview transcripts were double coded and analyzed using the qualitative research software Dedoose (version 9.0.107; SocioCultural Research Consultants) [55]. SUS and NPS scores were calculated following the focus group sessions.
Textbox 1. Key themes and definitions.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived feasibility</td>
<td>Perceived structural factors that could influence the introduction of the app, including administrative infrastructure and operational capabilities of the health systems</td>
</tr>
<tr>
<td>Acceptability</td>
<td>Frontline health workers’ willingness to use the app during client interactions</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>Includes the ability to use the app for enhanced functioning and improvement of the current workflow</td>
</tr>
<tr>
<td>Barriers to and facilitators of scalability</td>
<td>Any challenges that could limit scaling up or wider use of the app</td>
</tr>
<tr>
<td></td>
<td>Improvements to the current app, along with changes in the current systems that could result in wider adoption of the app across current and wider health systems</td>
</tr>
</tbody>
</table>

Ethical Considerations

The institutional review boards at Baylor Scott & White Health (388136) and Harvard Medical School (IRB22-0692) approved the study.

Results

Overview

A total of 18 FLWs—including clinical managers, CHWs, social workers or social work students, research assistants, nurses, and medical assistants—participated in 1 of 6 focus group sessions. Participants were aged between 21 and 70 years. Most identified as female (15/18, 83%), were not Hispanic/Latino (11/18, 61%), and had bachelor’s degrees (8/18, 44%). Table 1 summarizes the demographic characteristics of the participants.
Table 1. Participant characteristics (N=18).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Female</td>
<td>15 (83)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>8 (44)</td>
</tr>
<tr>
<td>30-39</td>
<td>3 (17)</td>
</tr>
<tr>
<td>40-49</td>
<td>3 (17)</td>
</tr>
<tr>
<td>50-59</td>
<td>2 (11)</td>
</tr>
<tr>
<td>60-69</td>
<td>1 (6)</td>
</tr>
<tr>
<td>70-79</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>7 (39)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>1 (6)</td>
</tr>
<tr>
<td>White</td>
<td>7 (39)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>7 (39)</td>
</tr>
<tr>
<td>Not Hispanic/Latino</td>
<td>11 (61)</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
</tr>
<tr>
<td>High school graduate or equivalent</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Some college or certificate program</td>
<td>7 (39)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>8 (44)</td>
</tr>
<tr>
<td>Doctorate degree</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Role</strong></td>
<td></td>
</tr>
<tr>
<td>Clinical manager</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Social worker</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Nurse</td>
<td>3 (17)</td>
</tr>
<tr>
<td>Social work student</td>
<td>4 (22)</td>
</tr>
<tr>
<td>Research assistant</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Medical assistant</td>
<td>3 (17)</td>
</tr>
<tr>
<td>Community health worker</td>
<td>4 (22)</td>
</tr>
</tbody>
</table>

Most participants (11/18, 61%) did not have mental health training before the EMPOWER training but interacted with individuals with mental illness regularly during their work (Table 2). None of the participants had experience delivering structured BA therapy.
Table 2. Mental health (MH) training (N=18).

<table>
<thead>
<tr>
<th>Previous MH training</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>7 (39)</td>
</tr>
<tr>
<td>No</td>
<td>11 (61)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years of MH training</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>13 (72)</td>
</tr>
<tr>
<td>1</td>
<td>3 (17)</td>
</tr>
<tr>
<td>2</td>
<td>2 (11)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experience delivering structured psychosocial therapies</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No</td>
<td>18 (100)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average number of clients with MH issues interacted with per month</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7 (39)</td>
</tr>
<tr>
<td>1-9</td>
<td>6 (33)</td>
</tr>
<tr>
<td>10-19</td>
<td>2 (11)</td>
</tr>
<tr>
<td>20-29</td>
<td>2 (11)</td>
</tr>
<tr>
<td>30-39</td>
<td>0 (0)</td>
</tr>
<tr>
<td>40-49</td>
<td>1 (6)</td>
</tr>
</tbody>
</table>

Table 3 summarizes the participants’ current technology use. Less than half (7/18, 39%) of the participants owned Android-supported devices. Participants were provided with Android tablets if they reported not having or being able to use an Android device during the focus group session; 28% (5/18) of the participants installed the app on their own devices, and the remaining participants were provided with Android tablets for this study.

Table 3. Technology use (N=18).

<table>
<thead>
<tr>
<th>Number of hours spent on a computing device (daily)</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-3</td>
<td>0 (0)</td>
</tr>
<tr>
<td>4-6</td>
<td>4 (22)</td>
</tr>
<tr>
<td>7-9</td>
<td>9 (50)</td>
</tr>
<tr>
<td>10-12</td>
<td>3 (17)</td>
</tr>
<tr>
<td>≥13</td>
<td>2 (11)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency of internet use</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least once a day</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Multiple times a day</td>
<td>5 (28)</td>
</tr>
<tr>
<td>Several times a day</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Most of the day</td>
<td>10 (56)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confidence using electronic devices</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somewhat confident</td>
<td>4 (22)</td>
</tr>
<tr>
<td>Very confident</td>
<td>14 (78)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Owns an Android device</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>7 (39)</td>
</tr>
<tr>
<td>No</td>
<td>11 (61)</td>
</tr>
</tbody>
</table>
Overall, participants found the app usable, with an average SUS of 72.5 (SD 18.1) compared with the industry benchmark of 68, which has been found to be the average SUS for digital health apps and used in previous digital health studies [56,57]. The average SUS scores from older participants (aged >50 years) were lower than those from younger participants (Table 4). The NPS was 0 (neutral), with an equal distribution of promoters, detractors, and passives (6/18, 33% per NPS category).

Table 4. System Usability Scale score (N=18).

<table>
<thead>
<tr>
<th>Participants, n (%)</th>
<th>Scores, mean (SD)</th>
<th>Scores, median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>18 (100)</td>
<td>72.5 (18.1)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>8 (44)</td>
<td>74.7 (18.4)</td>
</tr>
<tr>
<td>30-39</td>
<td>3 (17)</td>
<td>79.1 (20.2)</td>
</tr>
<tr>
<td>40-49</td>
<td>3 (17)</td>
<td>73.3 (11.3)</td>
</tr>
<tr>
<td>≥50</td>
<td>4 (22)</td>
<td>62.5 (22.7)</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or some college</td>
<td>9 (50)</td>
<td>74.4 (20.7)</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>9 (50)</td>
<td>70.5 (16.0)</td>
</tr>
</tbody>
</table>

Perceived Feasibility of the App

In this section, we discuss the perceived feasibility of the app, which includes structural factors such as the current administrative and operational capabilities of health systems that could influence the introduction of the app. The participants discussed their current client-provider engagements and supervision mechanisms that could affect the potential introduction and incorporation of the app into current workflows and across the existing health system.

**Client-Provider Engagement**

Participants described how they worked with clients, which revealed potential considerations of how the app may need to be flexibly designed to support the ways in which FLWs meet with clients today. First, although in-person visits were still common, many participants shared that they engaged with their clients via telehealth. Furthermore, when meeting remotely with clients, participants shared that they may not necessarily have a quiet, private space to speak with clients and connectivity may be less reliable. Thus, obtaining high-fidelity recordings of sessions conducted in settings with frequent disruptions is an important consideration. Second, FLWs have limited time to interact with clients, and some participants shared that additional time would be needed to deliver a brief psychosocial intervention. Third, in some cases, screenings such as the Patient Health Questionnaire–9 for depression are currently administered and updated in electronic health record (EHR) systems such as Epic, and it may not be the FLW’s role to access, collect, or update these data even though such screening data may be needed for the delivery of a psychosocial intervention such as BA. When discussing client engagement, one participant noted the following:

...the social worker is the first person that meets with the patient because she talks to the patient to see what barriers they have to their personal care, which would involve depression screening, and any kind of behavioral health barriers that they may have going on. So that’s how we get our information about what the patient is kind of going through and what they need as far as behavioral health management. [P04; nurse]

The integration of the app into existing EHR systems was reported as another important consideration.

Finally, although we did not specifically collect feedback on supervisors’ perceptions of the app, participants in supervisory roles suggested that additional time would be needed not only to review all the recordings but also to actually deliver the intervention—time that FLWs may not have:

From my role, like I said before, I don’t know that I would have the time to incorporate something like this into a patient encounter just because, like I said, we’ve got about 15, 20 minutes to see a patient for whatever issue is happening. And generally, when I see a patient for depression, anxiety, or something mental health, we certainly talk about it, talk about the role of treatment, maybe some options for counseling, not necessarily just medication. But I feel like when they get to me, it’s more about treatment and not so much about talking through things just because they’re not going to have as much time with a physician or a nurse practitioner because of their schedules. [P08; nurse practitioner]

**Supervision Structure**

According to the participants, the current supervision structure includes structured supervisor-supervisee meetings scheduled as a part of annual performance reviews. In most settings, although they are not perceived as part of supervision and are not standardized, peer support groups are also available. Although the existing peer support mechanisms are informal and do not follow a structure similar to the supervisor-supervisee meetings, they function as a good support mechanism for team members in which “a doctor can support a doctor, a manager can support a manager, or even their employees” (P06; clinical manager). A participant noted the following:
We do huddles where we kind of discuss different things that are going on and changes within our settings and our clinical practice that are similar, but we don’t necessarily do any ratings. [P16; certified medical assistant]

Thus, introducing the app would mean integrating these traditionally siloed supervisor-supervisee and peer support mechanisms and potentially introducing a new workflow that allows additional time and resources to be invested in supporting FLWs.

Acceptability of the App

On the basis of feedback from the focus group sessions, participants’ overall impressions of the app were mainly positive. Participants reported that they found the app “easy,” “useful,” and overall “important” to their current workflows. The app could provide a much needed structured platform to support peer supervision and overall training of FLWs. A participant with no training in mental health care delivery before the BA training program noted the following:

I don’t have to be a licensed professional counselor, and I can actually use it myself to help and...get supervision from my other peers to help me along the way. [P07; certified medical assistant]

Another participant described the app and process as an “objective way to actually hear the session and hear how much each person reacted, how they delivered the information to the person, how the patient received it” (P08; nurse practitioner). She added that, as a supervisor, she could instruct the trainee to “go ahead and do it, and then we’ll listen, and we can talk about feedback together.”

Even though most participants did not necessarily review or remember information from the app-related training materials provided to them before the focus group sessions, they considered the app easy to use. Participants who accessed the app on their devices found it easy to install. Most respondents also found the app navigation to be easy and self-explanatory. Nonetheless, it was noted that it could be a bit confusing at first “because you don’t know where to click,” (P02; social worker or social work student) and there were a few features within the app that the participants found more challenging to navigate. Participants needed time to familiarize themselves with the app and use it comfortably as they had never used a similar app before. One FLW reported the following:

I watched the (training) videos a long time ago and didn’t remember exactly everything that was talked about in the videos. And so, there’s a little bit of difficulty at first, but it’s pretty intuitive [as to] where everything is and how to use it. Or at least for me, it was, and so I could be able to grasp onto it pretty easily. [P03; social worker or social work student]

The features that participants needed the most time to navigate included accessing recordings, graphs, and self-ratings.

Perceived Usefulness of the App

Participants generally perceived the app to be potentially useful for a few reasons. One respondent shared that such an app would be useful to have now to address gaps in care because of limited access to and availability of mental health specialists:

For me, it would be a great thing to refer someone to if this was—if this was available because I think right now our psychiatrists are scheduling out next year. I think it’s a year before you can get the next appointment. So obviously, mental health is a big problem. [P08; nurse practitioner]

The app could also help enhance FLWs’ job performance, highlight their areas for growth, and ensure overall intervention fidelity. The app’s ability to record and allow users to review and rate sessions can allow for more tailored feedback on performance enhancement, highlighting areas of growth:

I think as a supervisor, being able to record sessions could be really impactful if you have group supervision that you’re doing. Because if you have the consent from the patient and you’re able to kind of play this recording for everybody, you could kind of talk about what things went good [sic], areas of growth, basically. And so being able to use a recording, I think, could be helpful, as well as peer ratings and getting the ratings from your supervisor to kind of also see what areas of growth that you might have for behavioral activation or any other skills that you’re using at your place of employment. [P03; social worker or social work student]

In addition to individual performance enhancement, the app could ensure intervention fidelity. A CHW shared that counseling sessions were often interrupted in the hospital setting; therefore, audio recording or rating could verify that the sessions were comprehensive despite disruptions. Reviewing the recording asynchronously could also help FLWs ensure that their sessions adhere to the intervention protocol.

Finally, most participants mentioned how their current peer support groups lacked structure and did not use systematic guidance or frameworks such as ratings for peer supervision. Respondents highlighted the potential of this app to fill that gap; however, the challenge would be in how to systematically incorporate the use of the app across the health systems currently implementing or planning to implement task-shared mental health interventions. As aptly summarized by a CHW, “to say that we need something like this, we definitely do. It’s just ‘where-do-we-put-it type’ of situation” (P17; CHW).

Barriers to and Facilitators of Scalability

Following discussions on the app’s perceived feasibility, usefulness, and acceptability, we probed considerations that could help the app’s broader implementation, including its fit into the FLWs’ current workflows and recommendations for future implementation.

Addressing Privacy Concerns

Protecting client privacy was expressed as a top priority for most respondents. Although feedback was mixed about clients’
apprehension toward audio-recorded sessions, it was generally agreed that obtaining proper client consent before recording was critical and necessary. Participants shared that consent from both FLWs and clients would be essential to ensure transparency and that all parties were informed of and agreed to how recordings would be used. Suggestions for increasing client comfort and obtaining adequate consent included building a trusted relationship with the client before asking to record sessions. By doing this, clients may be less hesitant to be recorded and more comfortable sharing their potential concerns.

Finally, regarding physical infrastructure, as noted previously, there is not always a private, physical space for FLWs to engage with clients and conduct peer supervision activities. One participant shared the following:

...we would probably talk about a lot of stuff that people would really want to be kept private with no real chance of anyone else hearing it, but that space is a big issue. We have a lot of shared offices—a lot of people in the same room at the same time...it might be better for us to, say, record a session so that we can make sure we’re getting everything, make sure I’m not missing something because that person’s over here, this person’s doing this, this person interrupted me, that person bumped into me. Yeah, we would just need to find a way to do it where we could minimize distractions, interruptions, and that kind of stuff. [P17; CHW]

Another participant agreed, stating the following:

I’m in a room that can accommodate 25 people. And I have people all around me, and we’re all talking about different things. And it’s really hard. [P18; CHW]

Ensuring the App Fits Workflow and Time Capacity

Although the app was perceived as an important addition to support their work, some participants raised concerns about finding time to deliver brief interventions and reviewing their sessions. It was suggested that the ability to review and rate recordings anytime and anywhere is an advantageous feature of an app accessible from any device and that FLWs may prefer different device types depending on their environment or workflow. For example, one FLW noted the following:

I think mostly it would be done at a computer because we have very set clinic hours where I’m at. But in some other environments, I could see somebody trying to access it from a car or something like that. [P16; certified medical assistant]

Another commented that, for her role, “...these (tasks) would be more flexible in very small devices, like iPad or phone, rather than recording it and everything, doing it in a computer” (P10; nurse). Asynchronous review would also allow for more flexibility in completing these follow-up activities:

I think being able to kind of reference the sessions that you’ve recorded will kind of allow for being a little more flexible in the schedule. Right now, you just have to go with each patient as it is, but this will allow...you to refer back. So, you don’t always have to just do it all at one time. [P16; certified medical assistant]

Collecting and managing data offline would also be critical for FLWs working in areas with limited connectivity.

Providing Structured, Supportive Supervision

Respondents noted how the app could help supervisors give structured feedback to their staff, identify and highlight areas of growth, and delegate appropriate tasks to FLWs based on their performance and expertise. Similarly, peer supervision could allow for more structured feedback from peers in workplaces that currently have informal peer support groups. Respondents confirmed that structured supervisor feedback could help with performance enhancement and quality assurance, particularly during the training phase, when workers need the most support. One participant noted the following:

...it would be good to be able to listen to how we interact with the patients. And it would be good to listen to it with the provider so...the nursing staff would know what we could do different, if we did well, just to get an idea of how we’re doing and how the patients are reacting to what we do. [P09; medical assistant]

Participants felt that current app-supported supervision could allow for a more consistent feedback mechanism, systematic tracking of gradual performance improvement, and a constant feedback loop to enhance learning opportunities for FLWs.

Similarly, although respondents reported that informal peer feedback is available in some settings, it is largely unstructured. Most participants agreed that having their peers listen to their sessions and provide feedback is beneficial. Respondents were generally receptive to their peers being involved in supervisory processes, primarily as the participants viewed peer supervision more as “feedback” rather than a performance appraisal. However, it was suggested that it would be important, if not critical, to ensure that peers provide constructive and positive feedback to each other; concerns were raised about how the care team culture may be affected by peers giving and receiving negative ratings and critiques on performance. When asked how they would feel about their peers listening and rating their sessions, one participant noted that it is “beneficial to have another set of ears as long as it’s constructive criticism or constructive help” (P18; CHW). Regarding constructive criticism, “I think it’s all in bettering us so that we can better our patients, so that’s good” (P16; certified medical assistant). For some, ratings were only associated with yearly performance reviews with a supervisor. Respondents emphasized the importance of a nurturing and supportive peer group rather than a competitive one. Finally, participants agreed that the app had the potential to help with more structured self-assessments as well. The ability to record and review sessions could allow for more flexible self-evaluation and help with self-improvement and retrospective performance assessment.
Discussion

Principal Findings

We conducted this study to assess the feasibility and acceptability of a prototype app developed for FLW supervision using objective, metric-based performance evaluation and supportive feedback mechanisms in the United States. Participants found the app easy to use and acceptable. It was considered useful for supporting task sharing of mental health care services, enhancing FLWs’ skills, ensuring intervention fidelity, and providing structure to the traditionally unstructured peer support groups by introducing peer ratings and structured feedback mechanisms. Some key feasibility issues revealed in discussions with FLWs and stakeholders included (1) how the ability to record sessions with high fidelity may depend largely on the nature of the client encounter, (2) the potential need to integrate the app into an existing EHR system (eg, Epic), and (3) the introduction and integration of a new app-supported peer supervision workflow into existing supervisor-supervisee processes and peer support mechanisms. Among areas of improvement for scalability were privacy considerations both on recording and rating the session and ensuring the availability of private space in the health facility to conduct the mental health intervention successfully. Other areas of improvement included focusing on the additional time commitment required of FLWs and supervisors when introducing the app, including time to record and rate the sessions. Findings from our study highlight the app’s potential to catalyze the existing traditional supervisor-supervisee dyad model by shifting focus to a supportive supervision model that incorporates feedback not just from supervisors but also from peers and through self-assessments [44].

We would like to highlight 2 key findings that affect the scalability of the current app and similar digital tools that aim to support the monitoring of psychosocial intervention delivery through objective, measurable metrics. The first is the utility of the current app for supportive supervision of FLWs in the United States. Given the current focus on the scalability of task-shared mental health interventions and the need to ensure the quality delivery of task-shared psychosocial interventions, digital solutions such as the one proposed in this study are crucial [46,58]. Worldwide, digital interventions have shown great promise in supporting FLWs [59], and studies are now exploring digital supportive supervision more extensively [60,61]. The current app can rapidly scale this process, standardize performance measures, and account for evolving training needs with great flexibility. In this study, we demonstrated that an app based on work with nonspecialist providers in Madhya Pradesh and Goa, India, was found to be usable and potentially useful for FLWs in Texas and shows promise for supporting the scale-up of services delivered by FLWs in similar settings across the United States.

Another key finding of this study is the potential of the app to help merge traditionally siloed supervisor-supervisee and ad hoc peer feedback approaches to build a more unified, structured supervision process across existing health systems. This requires effort beyond adding the app to the existing workflow; rather, it may be necessary to develop a new workflow entirely. Participants expressed concerns about the additional time commitment required to use the app as a part of their job. However, it is important to note that the time constraints mentioned were not necessarily related to the use of the app itself but to adding objective, measurable metrics to assess FLWs’ performance. Introducing measurable metrics to evaluate performance requires recording, rating, and discussing sessions; additional time; and skill sets often beyond the FLWs’ job description. The time commitment required can constrain effective supervision in task-shared projects [29]. Therefore, the app’s scalability depends greatly on creating a workflow that benefits FLWs and communicating its need to support FLWs and task-sharing mental health care across health systems. Creating protocols that outline how to record sessions safely and securely for quality assurance is also critical, particularly in light of the privacy concerns associated with recording internet-based sessions [62]. Ensuring readiness for a systematized supportive supervision process guided by infrastructure and protocols contextualized for a particular type of setting is crucial for the success of task-shared mental health care delivery [63]. More importantly, the current app must flexibly align rather than conflict with existing supervision models and diverse FLW workflows to ensure that it truly supports FLW skill enhancement, effective feedback mechanisms, and adequate fidelity in task-shared interventions [64-66].

Recommendations for Future Implementation

On the basis of the feedback collected in this study, there are several key considerations, challenges, and proposed solutions for the future implementation of this app. Table 5 summarizes the key considerations raised during the preliminary prototype design of the app, including advisory panel discussions, usability testing, and focus group session interviews. We propose solutions for each consideration that can inform future studies aimed at introducing similar digital tools to assist FLWs. These recommendations will also facilitate organizations’ development of supportive and structured supervision mechanisms.
Table 5. Key considerations, challenges, and proposed solutions for improvement.

<table>
<thead>
<tr>
<th>Key considerations</th>
<th>Challenges</th>
<th>Proposed solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target users</td>
<td>• App must support a variety of FLW roles with different workflows and role requirements as well as their supervisors</td>
<td>• Design the app to include role-specific features that can be customized to support site-specific needs as allowed (eg, inclusion of depression screening)</td>
</tr>
<tr>
<td>Integration into the current systems</td>
<td>• Potential to increase workload associated with integrating measurement-based peer supervision into current workflows</td>
<td>• Integrate the new app with existing EHRs such as Epic, which can reduce data redundancy and support more comprehensive client care</td>
</tr>
<tr>
<td>Interoperability across operating systems</td>
<td>• Prototype is currently only supported as a web application or Android mobile app</td>
<td>• Ensure that the app is easily accessible and can support offline data collection on other commonly used devices (eg, iPhones)</td>
</tr>
<tr>
<td><strong>Organization level</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Available resources in the implemen ting health facilities | • Inadequate physical space for conducting private and secure counseling sessions  
• Lack of current protocols on supportive supervision to guide FLWs  
• Need for structured peer supervision to implement the app | • Assess the availability and accessibility of facilities’ resources before implementing the app  
• Facilitate the development of protocols or updates to current protocols to promote the use of the app  
• Set up or use the existing structured peer supervision mechanisms to support app integration |
| Peer supervision model              | • More structured peer supervision could be cumbersome and time-consuming, especially for FLWs who are used to feedback from supervisors and are more comfortable with informal peer feedback mechanisms | • Include specific training and guidance on delivering and receiving peer feedback  
• Ensure supervisor buy-in and system support |
| Hybrid workflow                     | • Both in-person and remote (telehealth) client visits are common            | • Ensure the app can flexibly support recordings either directly in-app or via integration with other tools that may already be used |
| **App related**                     |                                                                             |                                                                                                                                                    |
| Peer ratings                        | • Introducing a mechanism for peers to rate and critique each other may feel uncomfortable | • Include positive and encouraging language and features (eg, badges) to support improvement in performance  
• Include space for qualitative feedback to provide more context for ratings |
| Navigation                          | • Navigation through parts of the app remains cumbersome                   | • Enhance the app to support the following:  
  • Unnested features (eg, client or patient lists)  
  • An easy way to view completed self-ratings  
  • A way to flag incomplete self- and peer ratings  
  • locating a session recording via multiple search parameters, such as client or patient ID, peer, or date |
| Data visualization                  | • Graphs and self-ratings are not intuitive                               | • Simplify current features such that graphs are more easily accessed and usable across device screen sizes  
• Highlight and indicate change in performance over time, noting specific areas that may need improvement |

aFLW: frontline health worker.
Conclusions

In summary, we demonstrated the feasibility and acceptability of a usable and functional mobile app prototype to support FLW-delivered psychosocial interventions in the United States. However, the success of future implementation hinges on further development of the app to ensure its adoption by both FLWs and their supervisors and overall readiness to adopt a novel peer supervision model of care. Feedback from stakeholders and target end users (FLWs) suggests that the app can help structure supervision mechanisms and systematize the collection and review of performance metrics in task-shared mental health interventions. With appropriate organization- and system-level support coupled with strong FLW training programs, the proposed app has the potential to scale up the delivery of high-quality, evidence-based mental health interventions in the United States through the critical role of FLWs who are already connected with and best positioned to serve some of the hardest-to-reach communities with the greatest needs for psychosocial support and care.

Acknowledgments

The authors would like to thank their colleagues at Dimagi, Akshita Sharma, and Surabhi Dubey, who assisted with app building and provided helpful inputs throughout this study. They would also like to thank Dr Daisy Singla, who has been leading the related study in rural India on which this study is based. Finally, they express their deepest gratitude to their advisory panelists from the Meadows Mental Health Policy Institute, Baylor Scott & White Health, the University of Texas at Arlington, and all the study participants for their time and contributions to this work. This project was supported by the National Institute of Mental Health award R43MH130305-01A1. AP has received support from the National Institute of Mental Health T32 on Social Determinants of HIV (T32MH128395-01). The development and testing of the app previously deployed in rural India were supported by Grand Challenges Canada under grant TTS-2109-47591 in collaboration with Sangath and the Sinai Health System.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

YXH, DML, LM, and NL are employed by Dimagi, Inc, whose revenue is derived from the open-source platform on which the digital app described in this paper was built and studied. VP is a co-founder of the EMPOWER At Scale company in the US. He has zero income from this company.

Multimedia Appendix 1

Focus group discussion task list.
[DOCX File, 18 KB - formative_v8i1e55205_app1.docx ]

Multimedia Appendix 2

COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File, 18 KB - formative_v8i1e55205_app2.docx ]

References


10. Institute of Medicine, Board on Health Services Policy, Committee on Developing Evidence-Based Standards for Psychosocial Interventions for Mental Disorders, England MJ, Butler AS, Gonzalez ML. Psychosocial Interventions for Mental and Substance Use Disorders: A Framework for Establishing Evidence-Based Standards. Washington, DC: National Academies Press (US); Sep 18, 2015.


16. Chatterjee R. Psychologists say they can’t meet the growing demand for mental health care. National Public Radio. 2023 Dec 06. URL: https://tinyurl.com/5n7cvvh [accessed 2023-12-06]

17. Dampier C. Mental health care appointments often come with a long wait. 3 ways to cope while help is delayed. Chicago Tribune. 2018 Oct 25. URL: https://tinyurl.com/4c3erxs9 [accessed 2023-08-02]


19. Weiner S. A growing psychiatrist shortage and an enormous demand for mental health services. Association of American Medical Colleges. 2022 Aug 09. URL: https://tinyurl.com/54s9kjw [accessed 2023-08-02]


37. HHS announces $226.5 million to launch community health worker training program. U.S. Department of Health and Human Services. 2022 Apr 15. URL: https://tinyurl.com/5d37arzr [accessed 2023-08-02]


48. PEERS (promoting effective mental healthcare through peer supervision). Sangath. URL: https://sangath.in/peers/ [accessed 2023-11-03]


55. Dedoose. URL: https://www.dedoose.com/ [accessed 2023-08-08]


Abbreviations

BA: behavioral activation
CHW: community health worker
COREQ: Consolidated Criteria for Reporting Qualitative Research
EHR: electronic health record
FLW: frontline health worker
NPS: net promoter score
PEERS: Promoting Effective Mental Healthcare Through Peer Supervision
SUS: System Usability Scale
©Anubhuti Poudyal, Delta-Marie Lewis, Sarah Taha, Alyssa J Martinez, Lauren Magoun, Y Xian Ho, Natali Carmio, John A Naslund, Katherine Sanchez, Neal Lesh, Vikram Patel. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Improving Behavioral-Based Safety Training in Using Verbal Commands Through a Theory-Driven and Feedback-Based Nonimmersive Virtual Reality Game: Development and Usability Study

Chukwudiebube Atagbuzia1*, MS; Ean H Ng1*, PhD; Ganapathy Natarajan2*, PhD

1Oregon State University, Corvallis, OR, United States
2University of Wisconsin-Platteville, Platteville, WI, United States
*all authors contributed equally

Corresponding Author:
Ean H Ng, PhD
Oregon State University
204 Rogers Hall
Corvallis, OR, 97331
United States
Phone: 1 15417370570
Fax: 1 15417372600
Email: ean.ng@oregonstate.edu

Abstract

Background: The construction, chemical, aviation, medical, and health care industries have used serious games for safety training. To our knowledge, serious games have not been developed focusing on behavioral change to improve safety through the use of verbal commands and instilling players with heightened awareness of their spatial proximity to other people in their surroundings.

Objective: We aimed to develop a theory-driven serious game for improving safety behavior using verbal commands and validate the implementation of the theoretical frameworks used for game development. The game developed, KitchenSpeak, was a first-person character (FPC) game where users respond to in-game prompts to use loud verbal commands when they are approaching another employee’s blind spot.

Methods: In addition to using the SERES framework in guiding the general game design and development, and the Reflection, Engagement, Choice, Information, Play, Exposition (RECIPE) framework to inform the design of the game mechanics, we also applied gestalt laws of perception for graphic design to guide the design of the game’s user interface. We conducted 2 evaluative tests (alpha and beta) to collect end user and stakeholder feedback on the implementation of the theoretical frameworks, as well as to collect relevant information for full-scale implementation and a future validation study.

Results: The alpha and beta tests had 8 and 40 participants, respectively. The alpha test results revealed that the theoretical frameworks were adequately applied; however, suggestions were also made to modify and improve the game. The beta test results suggested further improvements for the game design and found no differences in the perception of ease of play between participants with and without previous FPC gaming experience (P=.47; Kruskal-Wallis). Results suggested that the game met its design and theoretical requirements, and it would be easily playable by all players regardless of their previous experience in FPC games.

Conclusions: A theory-driven and evidence-based FPC game titled KitchenSpeak was developed to teach the use of kitchen-speak terms in commercial kitchens. Evaluative tests were conducted to validate the implementation of the theoretical frameworks. Our main contributions are creating and validating game-based training to improve behavioral-based safety in the workplace and the incorporation of gestalt laws of perception for graphic design in the game’s user interface.

(JMIR Form Res 2024;8:e48080) doi:10.2196/48080
KEYWORDS
behavioral safety training; SERES framework; Reflection, Engagement, Choice, Information, Play, Exposition framework; gamification; gestalt laws of perception

Introduction
Overview
In recent times, serious games have been used in industries and various occupations as training tools for employee safety. Serious games combine elements such as visual and auditory effects and the euphoria of overcoming challenges to facilitate user engagement to motivate the user to transfer gained knowledge to real-life experiences [1]. Serious games are effective tools to improve safety knowledge in the construction [2-8], chemical [9-11], medical and health care [12-15], and aviation [16-18] industries. However, to our knowledge, serious games have not been used in improving behavioral-based safety. In addition to acquiring safety knowledge, behavioral-based safety requires employees to repeat safe behavior over a period of time. Using verbal commands to alert fellow employees of their proximity is a safety practice that is important in high-traffic workplaces, where struck-by injuries can occur. Consistently using verbal commands is a behavior-based safety practice.

Verbal commands are used in various work settings, and this study used a commercial kitchen as a representative workplace to develop and validate our serious game. Verbal commands given in a commercial kitchen environment are generally called “kitchen-speak,” and they are widely used in commercial kitchens as a safety behavior aimed at reducing the risk of accidents due to collisions between employees. Existing methods to train employees to use verbal commands in the workplace using traditional training methods (lectures or videos with quizzes) have been ineffective. This study developed a theory-driven and feedback-based serious game for teaching and training employees to use verbal safety commands within a commercial kitchen.

KitchenSpeak Game Overview
KitchenSpeak is a first-person character (FPC) game where users respond to in-game prompts to use loud verbal commands when they are approaching another employee’s blind spot. These verbal commands to alert other employees of their proximity are called kitchen-speak, the namesake of this game.

Commonly used verbal commands in a commercial kitchen include “corner,” “behind,” “hot,” “sharp,” and combinations of these commands, such as “hot corner” and “sharp behind.” In our game, the player will use the keyboard and mouse to navigate the game environment. Whenever a scenario to use a verbal command occurs, a pop-up text message will inform the player of the verbal command required to be yelled. Upon successful yelling of the command, the game will display another pop-up to inform the player. The player will not be able to proceed through the scenario until the proper command is verbally completed.

KitchenSpeak game consists of 4 stages, where players will progressively learn a new command in every stage while applying the commands learned in previous stages. The last stage is a free-roam stage, where players will have to use a combination of the commands they have learned without any pop-up window to indicate the place of use. The players should now be able to identify the safety situation and act without a prompt. In each stage, a minimum of 4 verbal commands are required for completion, and the players are allowed 1 manual override per stage to minimize frustration and game abandonment.

Prior to starting the game, the player will have to enable the microphone on their computer. The voice recognition module used in this game is Google’s speech-to-text application programming interface which is trained with Google’s speech recognition models. The players do not calibrate or train the speech recognition module.

Significance of the Research
This study contributes to the body of knowledge on serious games by focusing on behavioral change to improve safety through the use of verbal commands and instilling the player with a heightened awareness of their spatial proximity to other people in their surroundings. Existing research in using serious games to improve safety has mainly focused on knowledge such as the proper use of tools [4,5] and proper procedures to perform high-risk tasks [19]. None of the existing serious games for safety focus on behavioral change and using verbal commands. The development and beta testing of the game from this study can be adapted to other training scenarios where behavioral change is needed and incorporate verbal commands as part of the gameplay.

Methods
Ethical Considerations
This study was reviewed by Oregon State University (OSU) institutional review board (IRB-2021-1174) and was determined to be exempt from review.

Methods Overview
The SERES framework [20] for developing serious games was used to guide KitchenSpeak’s development. Nicholson’s [21] Reflection, Engagement, Choice, Information, Play, Exposition (RECIPE) framework was used to guide the design of the game mechanics within the design foundation stage of the SERES framework. KitchenSpeak’s user interface (UI) was designed according to the 6 gestalt laws of perception as proposed by Smith-Gratto and Fisher [22]. An effective UI is important for game development as it ensures that the game players engage with the game in the intended manner [23].

SERES Framework
Overview
The five stages in the SERES framework are (1) scientific foundations, (2) design foundations, (3) development, (4)
validation, and (5) implementation. This study applies the first 3 stages of the framework to KitchenSpeak game development. The last 2 stages of the framework involve evaluating the effectiveness of the training tool, which is planned for a follow-up study. The following subsections present a summary of how the first 3 stages of the SERES framework were applied, as well as the plan for implementing the last 2 stages.

**Scientific Foundations**

There were 4 aspects of this stage. First, the target audience was Marketplace West Dining Center (MWDC) student employees (84% of the MWDC employee population). Second, the outcome objectives were to increase the use and knowledge of kitchen-speak among student employees. Third, the theoretical basis was our hypothesis that kitchen-speak use and knowledge will increase among people who play the game. Fourth, the tool evaluation will be performed in future work prior to implementation.

**Design Foundations**

**Game Mechanics**

We designed the game mechanics according to the RECIPE framework for gamification. Please see the RECIPE Framework for Game Mechanics section for further details.

**Design Requirements**

We modeled the virtual environment after the real-world MWDC environment to motivate reflection and foster immersion. A plug-in was used to transcribe verbal inputs from the players. For easy access, KitchenSpeak was designed for a web browser platform. MWDC management advised that the game’s completion time should be between 5 and 10 minutes.

**Game Authoring Tools**

Based on the design requirements, the following software and services were used to develop the game: (1) Unity-3D game engine for WebGL, (2) Autodesk 3ds Max for object modeling, (3) Google’s speech-to-text plug-in for speech transcription, and (4) PHP database hosted on OSU servers for game data storage.

**Trial Design**

Two validation tests (alpha and beta tests) were used to validate the implementation of the theoretical frameworks.

**Development**

**Genre**

KitchenSpeak was modeled to be a FPC game. The FPC perspective camera model was used since it encourages user immersion more than games in the third-person camera perspective [24].

**Content**

KitchenSpeak had 4 game stages: Corner, Behind, Hot Corner, and Free Roam stages. Each user must use kitchen-speak terms at least 3 times in the Corner and Behind stages, and at least 5 times in the Hot Corner and Free Roam stages. Each stage has a simple narrative where users are asked to complete tasks assigned by their supervisor. While navigating the virtual kitchen to complete tasks, the player is prompted to use kitchen-speak in blind areas. Figure 1 shows the “yell corner” prompt shown to a user as they approach a blind area. Learned kitchen-speak words cascade in each stage, that is, the Behind stage includes “corner” and “behind” kitchen-speak, the Hot Corner stage includes “corner,” “behind,” and “hot-corner,” and Free Roam includes all previously used kitchen-speak, with an additional “sharp-corner” kitchen-speak.

![Figure 1. Prompt to use kitchen-speak in the Corner stage.](https://formative.jmir.org/2024/1/e48080)
Rules
There are three general rules: (1) All game stages must be completed to complete the game. (2) Within the Corner, Behind, and Hot Corner stages, a player must use kitchen-speak terms when prompted; otherwise they do not advance through the blind area and cannot complete their assigned task (“enforced-use rule”; the Free Roam stage does not have this rule). (3) In the Corner, Behind, and Hot Corner stages, users are allowed to override the prompt to use kitchen-speak if they are stuck in the blind area for over 30 seconds; users are allowed a maximum of 3 overrides throughout the 3 game stages (override rule).

Visuals and User Experience
Gestalt laws of perception for graphic design were applied to guide the design of the UI elements. User experience was tested and modified through the alpha test.

Figure 2. KitchenSpeak virtual environment and Marketplace West Dining Center (MWDC) kitchen.

Validation and Implementation
Validation and implementation, the last 2 elements of the SERES framework, will be performed in future work.

RECIPE Framework for Game Mechanics
The six elements of the RECIPE framework were implemented in KitchenSpeak game mechanics as follows.

Reflection
This was implemented by modeling KitchenSpeak environment and activities after the MWDC kitchen environment, as well as the activities that occur within the MWDC kitchen. Figure 2 shows a scene of the virtual environment from KitchenSpeak compared to the MWDC environment.

Exposition
This was implemented by including simple narratives, such as “Your colleague John put some Teriyaki chicken into the oven, and it is now ready to be checked and served. Go and pick up the chicken from the oven.”

Choice
This was implemented by (1) allowing users to choose different navigating routes to complete an activity and (2) giving users the ability to choose when they want to complete a task. They may decide to free-roam within the game before completing the task.

Information
This was implemented by creating 3 tutorial stages that allow the user to practice the game controls before the main game stages. Informational texts highlighting the objectives of each stage are displayed at the start of each stage. A quick message is also shown to alert the player if they correctly use kitchen-speak, and if they have started, completed, or are in the process of completing a task.

Play
This was implemented by allowing players to restart each stage from within the game. If a user fails to complete certain tasks within a stage, the user is permitted to restart that stage an infinite number of times until they have completed the stage or feel satisfied with their progress. A player is always informed on their progress through the informational text during task progress and completion. Allowing the user to fail and restart within the learning environment creates an opportunity for the user to self-reflect and discuss their experiences with other players, thus improving long-term learning [25].

Engagement
This was implemented through interaction by including activities within the game that require the user’s response. The user is made to provide verbal input as a form of response to the kitchen-speak prompts and is asked to provide keyboard input while completing activities.

Gestalt Laws of Perception for Graphics Design
Gestalt laws of perception for graphics design leverage the laws of perception to develop a comprehensible screen design that is aimed at aiding viewers to interpret and remember the
presented materials [22, 23]. The gestalt laws were applied to the UI elements within KitchenSpeak; however, the Controls menu and Stage Select menu in Figure 3 are used as examples to show how the laws were applied. The 6 gestalt laws of perception for graphics design applied in this game are as follows.

**Figure 3. Controls and stage select menus.**

**Figure-Ground Contrast**
This law states that the figure and background must be distinct so that viewers can easily distinguish the figure from the background. In the Controls and Stage Select menus, the primary background color is dark orange (#FF8000) while the text color on the primary background is white (#FFFFFF). This color combination creates high contrast, which improves its legibility.

**Simplicity**
This law states that people simplify their perceptions based on previous experiences. The UI should be designed with simple and familiar designs so that viewers may easily understand it. In KitchenSpeak, unrelated information is shown in separate UI objects, while related information is shown in a single UI object.

**Proximity**
This law states that objects that are close together in space are generally perceived to be grouped together. Similar game UI objects were placed together to reduce the mental load that a player uses to identify game objects. In the Stage Select menu, the game stages and tutorial stages buttons are arranged in close proximity so that the player can easily identify that those buttons represent the game stages and training stages, respectively.

**Similarity**
This law states that objects that possess similar characteristics are generally perceived as a whole. KitchenSpeak game objects that were similar in function were designed to have similar appearances. In the Controls menu, the same background color was used for the subheadings Controls and Stages to signify that these 2 texts were subheadings and not parts of the instructions. In the Stage Select menu, the game stage buttons are made the same color apart from the Free Roam stage button. The Free Roam stage button turns green after the player successfully completes the 3 other stages.

**Symmetry**
This law states that objects with an unbalanced symmetry are perceived to be incomplete and could be distracting to viewers. In the Controls menu, the white vertical line between the Controls and Stage subheadings was designed to create a symmetrical design in the UI. Symmetry was also applied in the Stage Select menu to guide the arrangement of the game stage buttons.

**Closure**
This law states that open shapes are perceived as incomplete. Important figures should have closed shapes so that viewers do not have to be distracted from trying to close the shape. In the Controls menu, the message “Successfully complete all stages to finish the game” conveys information that is different from the information that the other UI objects convey. As a result, its UI object was designed to have a closed red rectangular figure that distinguished it from the other UI objects, thus reducing ambiguity and making the information easier to read.

**Trial Design**
Two evaluation tests (alpha and beta tests) were conducted. The alpha test was conducted as a qualitative test in the form of an open-ended discussion where participants played KitchenSpeak version 1.0 and were asked to provide their verbal feedback on their gameplay experiences. Our focus for the alpha test was the game content, navigation, UI, speech recognition, realism of environment, narrative, choice, information, and the restart and override rules.

The beta test was conducted by releasing KitchenSpeak version 2.0 to MWDC staff and collecting their (1) game completion times, (2) perception of ease of play, (3) gameplay engagement, and (4) experience with bugs and glitches during gameplay. A 5-point Likert scale survey was used to collect user perception of ease of play, while a survey with predetermined questions was used to collect users’ experience with bugs and glitches.

**Results**

**Alpha Test Results**
We recruited 8 participants, including 5 MWDC management staff and 3 nonmanagement staff with varying gaming experience, who were recruited by MWDC management.
Participants played the game one after another with the debugging window opened to observe how the speech recognition model transcribed the participant’s speech input. The results from this qualitative test are provided in Table 1.

Table 1. Results from the alpha test.

<table>
<thead>
<tr>
<th>Aspects of the game</th>
<th>Results</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realism of game environment</td>
<td>Participants deemed the game environment design was similar in appearance to the MWDC\textsuperscript{a} kitchen.</td>
<td>Include more details to reflect real-world execution of activities, for example, while a player is picking up a hot pot from the gas oven top, they should turn off the gas and use heat-resistant gloves.</td>
</tr>
<tr>
<td>Content and narrative</td>
<td>Game narratives were perceived as short, but it was explained that they were intentionally kept short to avoid confounding the narrative with the game’s purpose. All participants agreed that the content and narrative were sufficient.</td>
<td>No suggestions for improvement.</td>
</tr>
<tr>
<td>Navigation</td>
<td>Difficulty in navigating for participants with no previous FPC\textsuperscript{b} gaming experience.</td>
<td>Limit the vertical camera viewing angle such that it cannot go above the game character’s horizontal line of sight. That is, FPC cannot look up, but only downwards.</td>
</tr>
<tr>
<td>Information</td>
<td>Instructions in the Stage Select menu were not easy to understand, and participants did not know how to start playing a stage.</td>
<td>Include descriptive instructions on how participants could access game stages.</td>
</tr>
<tr>
<td>UI\textsuperscript{c}</td>
<td>UI elements were without clutter, simple, and easy to understand.</td>
<td>No suggestions.</td>
</tr>
<tr>
<td>Speech recognition usability</td>
<td>Single-word kitchen-speak (ie, “corner” and “behind”) was transcribed correctly, but the “hot corner” phrase was often transcribed as “hot burner” and “hot porter.”</td>
<td>Include “hot burner” and “hot porter” phrases into a similar keyword dictionary such that they will be accepted as correct alternatives to “hot corner” when transcribed.</td>
</tr>
<tr>
<td>Choice</td>
<td>Participants enjoyed the liberty to explore the game environment while completing tasks. One participant mentioned that exploring the kitchen environment reminded them of the MWDC kitchen.</td>
<td>No suggestions.</td>
</tr>
<tr>
<td>Game rules</td>
<td>Two overrides per game session were too little given the accuracy of the speech recognition module. Participants agreed that the restart rule was good as it would allow users to practice within each stage.</td>
<td>Increase the allowable number of overrides per session to 3.</td>
</tr>
</tbody>
</table>

\textsuperscript{a}MWDC: Marketplace West Dining Center.  
\textsuperscript{b}FPC: first-person character.  
\textsuperscript{c}UI: user interface.

Beta Test Results

Overview

Feedback from the alpha test was used to release version 2.0, which was used for the beta test. The beta test had 40 participants; 34 of them played and responded to the FPC experience and the ease of play surveys, and 32 responses were valid. A total of 21 participants indicated they were in the 15-to-24-year age group, while 11 participants did not respond to the age question; 7 participants were male, 11 were female, 1 was nonbinary, and 13 did not indicate their gender. The age group and gender proportions were representative of the student population in MWDC. A total of 13 participants responded to the Bugs and Glitches questionnaire.

Ease of Play

We looked at the difference in perception of ease of play between participants with previous FPC experience and those without previous FPC experience. Table 2 shows the summary of the ease of play results. The average ease of play of all 32 participants was 3.91 (SD 1.12). There was no significant difference in the user’s perception of ease of play between players who had previous FPC experience and players who had no previous FPC experience at a .05 significance level.

Table 2. Examining perception of ease of play.

<table>
<thead>
<tr>
<th>Group</th>
<th>Ease of play score\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants, mean (SD)</td>
<td>3.91 (1.12)</td>
</tr>
<tr>
<td>Had previous FPC experience, mean (SD)</td>
<td>3.63 (1.30)</td>
</tr>
<tr>
<td>No previous FPC experience, mean (SD)</td>
<td>3.91 (1.12)</td>
</tr>
<tr>
<td>Kruskal-Wallis P value</td>
<td>.47</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Measured on a 5-point Likert scale.
**Game Completion Time**

We evaluated the average time it took a player to complete game stages and the difference in completion time between participants with FPC gaming experience and no FPC gaming experience. Table 3 shows a summary of the game completion time test. The Kruskal-Wallis test for game completion times indicated that there was a significant difference in the median time it took to complete the game between users with FPC experience and users without FPC experience.

### Table 3. Statistical test results for completion times.

<table>
<thead>
<tr>
<th>Group</th>
<th>Game completion time (seconds), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants (n=32)</td>
<td>369.94 (164.33)</td>
</tr>
<tr>
<td>Had previous FPC experience (n=24)</td>
<td>325.95 (129.41)</td>
</tr>
<tr>
<td>No previous FPC experience (n=8)</td>
<td>501.93 (195)</td>
</tr>
<tr>
<td>Kruskal-Wallis P value</td>
<td>.005(^a)</td>
</tr>
</tbody>
</table>

\(^a\)Significant at 5% significance level.

**Gameplay Engagement**

At least 160 prompts were shown to all 32 players (minimum of 5 kitchen-speak prompts per player). We measured disengagement by the ratio of the number of times a user did not use kitchen-speak when prompted to the minimum number of times a prompt to use kitchen-speak was shown in the Free Roam stage (5 prompts). The results revealed that out of more than 160 prompts that were shown to all 32 players, 11 (6.9%) prompts did not receive valid user responses, while at least 149 (93.1%) prompts received valid user responses.

**Bugs and Glitches**

A total of 13 participants completed the Bugs and Glitches questionnaire. A total of 8 reported no issues, 5 had bugs in the Free Roam stage, and 2 had speech recognition issues. One participant found the game realistic. While the speech recognition module was adjusted to reduce recognition errors after the alpha test, issues with speech recognition were still expected.

**Discussion**

**Overview**

As part of this discussion, we will examine implications of this study and implications for future game developers based on our results. The major topics of discussion will be the frameworks used for game development, the results of the alpha and beta tests, and the development of serious games for behavioral safety.

**Frameworks for Game Development**

The frameworks used for our game design were SERES and RECIPE frameworks. In addition, we used gestalt laws of perception to design an effective UI for a serious game. In evaluating our implementation of the RECIPE framework, we found that evaluating player engagement was challenging. We measured engagement through the lack of disengagement, and we found our approach to be lacking. Engagement is a multiconstruct measurement [26] and as such must use a triangulation method using more than 1 input. For future game development, developers should consider including multiple measurements for engagement during the design phase.

Effective UI design can improve the gaming experience for inexperienced players, and the gestalt laws of perception can bridge the gap for game developers in designing an effective UI. Previous gaming experience affects user experience as inexperienced players report less positive experiences than experienced players [27]; however, effective UI design can be used to improve user experience and aid effective learning in electronic learning platforms [23]. Gestalt laws of perception for UI design are effective because they provide a holistic perspective that considers humans’ perception and interpretation of information [23]. We applied gestalt laws of perception to design the UI and evaluated the UI’s usability. The alpha test results showed that the UI design was without clutter, simple, and the information was easy to understand. We encourage serious game developers to incorporate the laws of perception to improve their game design by reducing clutter and improving information retention with players.

**Alpha and Beta Testing**

The alpha tests showed us the importance of receiving feedback to improve the game. Including the players, the most important stakeholders, is key to successful game deployment. In our case, we used a qualitative method that provided rich feedback for version 2. Our methods here show that a full testing regime that includes quantitative methods such as A/B testing is not always required. Receiving qualitative feedback to make quick improvements is an effective and agile methodology we suggest for serious game developers.

Our analysis of game completion time showed players without FPC gaming experience completed the game significantly more slowly than those with prior experience. Even though we had a tutorial for inexperienced players, the difference in completion time results was significant. This shows that additional training is required. Players with no prior FPC experience must be eased into gameplay by observing player behavior and receiving real-time feedback. Failure to ease players into the game may cause frustration and game abandonment.

In terms of bugs and glitches, voice recognition was an area of interest. There was variability due to factors such as background noise levels and user-specific characteristics. User accent and voice pitch are factors that will inherently affect the accuracy...
of a speech recognition module. Accent recognition is a problem of particular interest that requires further exploration.

KitchenSpeak game used the Unity3D WebGL. We used Unity3D’s profiling tool to review game assets and memory usage. The results showed that CPU memory was mostly used by rendering game object textures and audio data. To improve overall game performance, similar textures were applied to different game objects to reduce the number of times a type of texture was rendered. We encourage game developers to consider this aspect while developing a game and to use this technique to promote game performance.

**Serious Game for Safety Behavior**

Prior work in serious games for safety mainly focused on safety knowledge training. There are a few challenges in developing a serious game for behavior change from the safety behavior theoretical perspective. Safety behaviors are defined as acts that can be observed by other employees. Existing theories on changing safety behaviors are rooted in changing values, attitudes, and motivations, and emphasize providing consequences for not practicing the behavior [28].

The first challenge we encountered was finding the balance between providing sufficient exposition in the game to affect the value, attitude, and motivations of the players and managing the length of the gameplay. Our current approach to exposition is a pop-up text box to provide context. If the pop-up window is text-heavy or if we use too many consecutive pop-up windows, based on our own gameplay experience, it can lead to players skipping the pop-up windows or abandoning the game. We have opted to minimize the text in exposition, but have the player repeat the same behavior (verbal commands) multiple times for each play, each with a different exposition.

Second, the literature on safety behavior states that behavior must be observable by other employees; we included nonplayer characters (NPCs) into the scenes within our game. For future work, we plan to collect the gameplay information of actual players and create NPC movements based on actual players’ gameplay data to increase the realism of the NPCs. In either case, we are still looking into methods to assess whether the NPCs create the effect of observers.

Third, in terms of consequences to players when they did not perform the safety behavior, we considered adding consequences, and we evaluated two potential approaches. One is to create and display virtual scenes of accidents such as the spilling of hot food content after the collision and another is to use a pop-up window to inform the player that an accident had happened due to them not providing verbal commands. We decided against the pop-up window as we used it for exposition and did not want to overload the players and have them ignore it. Due to time constraints, we were not able to implement the actual scene of the accident for this study.

For safety behavior, using verbal commands consistently throughout an entire work shift is a relatively new concept to many employees who have never worked in a commercial kitchen. The behavior change we initiated through this study can serve as the baseline for future work to assess whether a change in value or attitude or motivations, a need for observers, and a need for consequences are necessary in a serious game to initiate behavior change.

**Conclusion and Future Work**

Serious games and other e-learning modules could be used as a cost-effective method for training as employers do not have to deal with the costs of face-to-face in-person training [29]. A serious game is a game that achieves an additional goal that is not for entertainment purposes [30]. By this definition, KitchenSpeak is a serious game. This study aimed to advance the serious games research field by developing a game aimed at changing behavior, specifically, instilling the behavior of consistently using verbal commands in blind areas within high-traffic workplaces.

KitchenSpeak development was successful; however, a new study should evaluate its effectiveness in improving the use of verbal commands and its effect on accidents that lead to struck-by injuries. In order to assess how effective KitchenSpeak is in enhancing the use of verbal commands, we need to compare the frequency of verbal command terms used in a commercial kitchen where the game is deployed to the frequency of kitchen-specific terms used in a commercial kitchen where the game is not deployed. To evaluate KitchenSpeak’s effect on accidents that lead to struck-by injuries, we will compare safety records on struck-by injury frequency before and after the game’s deployment in a commercial kitchen.

In this study, we discussed and validated the implementation of 3 theoretical frameworks that were used to develop the KitchenSpeak serious game for commercial kitchen safety. A serious game’s effectiveness is improved by incorporating theoretical concepts and principles that intend to achieve the desired results [30]. While theoretical concepts and principles serve as relevant tools for guiding serious game development, their effectiveness is linked to their success in application and implementation. In addition to testing the effectiveness of a serious game in achieving its desired outcomes, developers may also evaluate the implementation of their foundational theoretical principles, as they may provide relevant information on the cause of their game’s effectiveness or ineffectiveness.

**Acknowledgments**

We are grateful to Christopher Anderson, Kent Hildebrandt, Suzi Kritikos, and Michael Meeker of University Housing and Dining Services at Oregon State University for their assistance and input in the game’s development and validation. We are also grateful to Dan Kermoyan, MPH, and Tom Doyle, MBA, of Environmental Health and Safety at Oregon State University. There was no funding for this study. We used Google’s speech-to-text plug-in for transcribing users’ speech input to text, as well as a PHP database hosted on OSU’s servers for gathering all game-related data. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this study.
Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
EHN was responsible for the conceptualization of this study. CA wrote the study, developed the game, and provided data for all tables in the study. EHN and GN provided administration and supervision of the project. All authors reviewed the final study.

Conflicts of Interest
None declared.

References


Abbreviations

FPC: first-person character  
MWDC: Marketplace West Dining Center  
NPC: nonplayer character  
OSU: Oregon State University  
RECIPE: Reflection, Engagement, Choice, Information, Play, Exposition  
UI: user interface

©Chukwudiebube Atagbuzia, Ean H Ng, Ganapathy Natarajan. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Implementing a Sodium-Glucose Cotransporter 2 Inhibitor Module With a Software Tool (Future Health Today): Qualitative Study

Matthew Suen¹, BBiomed, MD; Jo-Anne Manski-Nankervis¹, MBBS, PhD; Caroline McBride¹, MD; Natalie Lumsden¹, PhD; Barbara Hunter¹, PhD

Department of General Practice and Primary Care, University of Melbourne, Parkville, Australia

Corresponding Author:
Matthew Suen, BBiomed, MD
Department of General Practice and Primary Care
University of Melbourne
Level 3, North Wing, Building
181 Grattan Street, Medical Building
Parkville, 3010
Australia
Phone: 61 383443369
Email: matthew.suen@unimelb.edu.au

Abstract

Background: Primary care plays a key role in the management of type 2 diabetes. Sodium-glucose cotransporter 2 (SGLT2) inhibitors have been demonstrated to reduce hospitalization and cardiac and renal complications. Tools that optimize management, including appropriate prescribing, are a priority for treating chronic diseases. Future Health Today (FHT) is software that facilitates clinical decision support and quality improvement. FHT applies algorithms to data stored in electronic medical records in general practice to identify patients who are at risk of a chronic disease or who have a chronic disease that may benefit from intensification of management. The platform continues to evolve because of rigorous evaluation, continuous improvement, and expansion of the conditions hosted on the platform. FHT currently displays recommendations for the identification and management of chronic kidney disease, cardiovascular disease, type 2 diabetes, and cancer risk. A new module will be introduced to FHT focusing on SGLT2 inhibitors in patients with type 2 diabetes who have chronic kidney diseases, cardiovascular diseases, or risk factors for cardiovascular disease.

Objective: The study aims to explore the barriers and enablers to the implementation of an SGLT2 inhibitor module within the Future Health Today software.

Methods: Clinic staff were recruited to participate in interviews on their experience in their use of a tool to improve prescribing behavior for SGLT2 inhibitors. Thematic analysis was guided by Clinical Performance Feedback Intervention Theory.

Results: In total, 16 interviews were completed. Identified enablers of use included workflow alignment, clinical appropriateness, and active delivery of the module. Key barriers to use were competing priorities, staff engagement, and knowledge of the clinical topic.

Conclusions: There is a recognized benefit to the use of a clinical decision support tool to support type 2 diabetes management, but barriers were identified that impeded the usability and actionability of the module. Successful and effective implementation of this tool could support the optimization of patient management of type 2 diabetes in primary care.

(JMIR Form Res 2024;8:e50737) doi:10.2196/50737

KEYWORDS
type 2 diabetes; CP-FTT; electronic health; clinical decision support tool; primary care; SGLT2 inhibitor; complication; tool; digital health intervention; thematic analysis; decision support; diabetes management

Introduction

Type 2 diabetes places a significant burden on both people with this condition and the Australian health system. An estimated 1.3 million Australians older than the age of 15 years have diabetes [1], with an associated Aus $14 billion (US $12.9 billion at 2010 rates) [2] of health spending. This creates an enormous social and economic burden. General practitioners

https://formative.jmir.org/2024/1/e50737
(GPs) play a vital role in reducing the impact of diabetes as most people receive their medical care in general practice. Guidelines produced by the Australian Diabetes Society [3], The Royal Australian College of General Practitioners [4], and the Living Evidence for Diabetes Consortium [5] provide support to GPs to inform both pharmacological and nonpharmacological management decisions.

Sodium-glucose cotransporter 2 (SGLT2) inhibitors have been included in Australian guidelines for years, but their place in therapy has evolved as recent evidence demonstrates secondary prevention benefits for cardiovascular disease (CVD) and chronic kidney disease (CKD) irrespective of glycated hemoglobin [6-11]. Prescribing of SGLT2 inhibitors remains low [12], and reported barriers include a lack of knowledge of its nonglycemic benefits and concerns about side effects [13]. This can be mitigated if appropriate and relevant education is available [14]. However, education alone is not likely to have a significant impact on changing prescribing behavior.

Future Health Today (FHT) is a co-designed quality improvement and clinical decision support technology platform aimed at the detection and management of conditions in general practice. FHT integrates with the electronic medical record to provide two components: (1) a clinical decision support tool active at the point of care (PoC) and (2) a web-based dashboard that facilitates practice-wide audit and quality improvement activities, including patient recall. Wraparound activities are also included with links to the latest guidelines and access to educational resources. Following initial optimization in 12 general practices, using guidelines for CKD, CVD, and cancer prevention, it is now in use in 55 general practices across Australia. A new module was introduced to FHT practices in July 2022 recommending SGLT2 inhibitor prescribing in patients with type 2 diabetes who also have CKD, CVD, or risk factors for CVD consistent with the Australian Evidence-Based Clinical Guidelines for Diabetes [5]. If a patient is suitable for prescribing, the statement “Consider initiation of SGLT2 inhibitor to reduce CVD and CKD risk” will appear in the PoC. The objective of this paper is to report on the evaluation of the implementation of this new SGLT2 inhibitor module in FHT.

**Methods**

Study Design
Fifty-two practices in Victoria (n=51) and Tasmania (n=1) had access to the SGLT2 module between July 2022 and January 2023. A qualitative evaluation exploring the use of the module was undertaken using the Clinical Performance Feedback Intervention Theory (CP-FIT), a theory for designing, implementing, and evaluating feedback [15]. CP-FIT was chosen as it incorporates and builds on 30 pre-existing theories and was developed specifically for the health care context. The theory proposes 42 variables that influence a feedback cycle with each step vital for successful feedback to occur. The theory postulates that the cycle is affected by 3 variables: feedback, recipient, and context. In the context of FHT, CP-FIT outlines the steps that users move through when guideline-based recommendations are communicated to users. Algorithms are applied to electronic medical records (data collection and analysis); recommendations are delivered to users (feedback); and they are received (interaction), interpreted (perception), and interrogated (verification). If users find the recommendations appropriate (acceptance), they will respond to them (intention and behavior) and ultimately leads to changes in patient care (clinical performance improvement).

Participants were recruited through expression of interest to participate in semistructured interviews, from those who had at least 1 month access to the SGLT2 inhibitor module. Expressions of interest were sent via email to practices encouraging participation in an interview. Advertisements in the department and practice-based research network newsletters were also used to promote the study. Interviews were conducted between September and December 2022. All participants had access to the FHT software for between 6 and 18 months prior to the use of the SGLT2 inhibitor module in clinical practice.

Data Collection
Semistructured interviews using CP-FIT as a guide were conducted with GPs, general practice nurses (GPNs), and clinical health assistants (CHAs) to explore their perspectives on using the module and recommendations for improving the tool. CHA are administrative staff who use patient data for health research and quality improvement. The aim was to gather as many perspectives as possible until saturation was reached. Participants participated in one-to-one interviews that occurred via telephone or Zoom (Zoom Technologies). Interviews were conducted by MS, a male academic GP registrar with the University of Melbourne. An interview guide was developed to focus on questions regarding the clinical usefulness of the recommendations, the impact on clinical workflow, and perceived changes in clinical performance. Data were uploaded to NVivo (version 12; QSR International) and coded by MS to identify themes using CP-FIT as a coding framework. A second researcher (CM) reviewed the data to ensure the reliability of the analysis.

Ethical Considerations
This study was approved (23269) by the Faculty of Medicine, Dentistry and Health Sciences Human Ethics Subcommittee at The University of Melbourne. Consent was taken through a signed consent form prior to the interview, and verbal consent was also gained to record the audio component of the interview. Participation was voluntary, and participants were compensated with an Aus $50 (US $32) voucher for their time. Participants’ names and practice locations were deidentified and replaced by a pseudonym only known to MS and CM to ensure the protection of their privacy.

**Results**

Overview
Invitations for interviews were sent out to clinics that had at least 1 month of experience with the SGLT2 inhibitor module. There was interest from 14 general practice clinics, and 16 interviews were completed with participants from 11 clinics (Table 1). The sex of the interviewee was confirmed during the interview. A further 3 interviews were not completed as the
The interviewees used FHT and previous modules but not the SGLT2 inhibitor module. Interviews ranged from 15 to 42 minutes. The themes were mapped across the 3 variables: context, feedback, and recipient. Barriers and facilitators related to these variables are summarized in Figure 1.

Table 1. Characteristics of study participants (N=16).

<table>
<thead>
<tr>
<th>Role</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General practitioner</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Practice nurse</td>
<td>6 (38)</td>
</tr>
<tr>
<td>Practice manager and practice nurse</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Clinical health assistant</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11 (69)</td>
</tr>
<tr>
<td>Male</td>
<td>5 (31)</td>
</tr>
<tr>
<td>Rurality of practice</td>
<td></td>
</tr>
<tr>
<td>Metro</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Regional or rural or remote</td>
<td>8 (50)</td>
</tr>
</tbody>
</table>

Figure 1. How CP-FIT explains the effectiveness of the FHT intervention. CP-FIT: Clinical Performance Feedback Intervention Theory; FHT: Future Health Today; PoC: point of care; SGLT2: sodium-glucose cotransporter 2.
Feedback Variables

Goal: Evidence-Based Nature, Relevance, and Importance

Common across all interviews were goals relating to the importance, relevance, and evidence-based nature of the SGLT2 recommendations. Users had confidence that the recommendations were accurate knowing they were based on the latest guidelines. Some incorporated FHT and the module in their conversation with patients by highlighting and explaining to patients the background and concept of FHT to instill patient confidence and acceptance of SGLT2 inhibitor prescribing.

> I can say to people that the reason I’m suggesting the change in medicine or the increase in dose is because there is a little point of care prompt, it’s telling me there are advantages in doing this, in making a change. [GP2]

Feedback Delivery: Active Delivery, Frequency, and Function

Active delivery was demonstrated through the PoC but not the dashboard. Participants found the PoC user friendly as it would automatically pop up once the patient file was opened. They also enjoyed the flexibility of being able to minimize and scroll through the recommendations in the PoC. In contrast, a majority indicated the need to log in to the dashboard to access it was a barrier to its use.

> Well it [PoC] comes up automatically every time we enter a patient where it’s relevant and it’s a very good prompt to double check that we’re following the guidelines. [GP8]

Practices using the SGLT2 inhibitor module also had access to over 20 recommendations within FHT relating to other chronic diseases. With the multiple modules available, prompt fatigue was highlighted as a barrier, especially with time constraints during consultations. Several participants highlighted the possibility of a behavioral norm of not checking the PoC recommendations when they pop up. Despite this, most still believed the recommendations were helpful in identifying patients who would not have been highlighted if it was not for the prompt, reiterating the tool as an additional safety net.

> It comes up on every patient when we open the file but because we’re so busy during a consultation...and it comes up so often. [GP2]

Context Variables

Organization and Team Characteristics: Workflow Fit, Intraorganizational Networks, and Competing Priorities

Many users reported using the PoC function before or during the consultation to guide clinical decision-making when talking broadly about FHT. They indicated that they referred to it when the consultation was diabetes focused but would otherwise flag it for future reference or disregard it. “I have a look at the point of care and see what kind of recommendations are popping up” (GP1).

The SGLT2 module encouraged collaboration between the staff. With the use of the cohort function, CHA and GPNs were able to generate recall list and book appointments for GPs to review patient suitability. This reduced the workload for GPs. In addition, GPNs felt comfortable initiating a conversation about the benefits of the medications, although they also acknowledged the final decision lay with the GP.

> The autonomy is given to be able to do recalls without having to have the GP absolutely check through every little detail of the patient. [CHA1]

Competing priorities were identified as barriers to fully use FHT with a flow on effect to the SGLT2 inhibitor module, especially with the use of the dashboard. Multiple GPs commented that they used the PoC as the recommendations came to them once the patient file was open. This is in contrast with the dashboard which involved creating patient lists, recalling patients, and filling quality improvement forms. Time pressure was highlighted as a barrier, with many GPs commenting about having time constraints and an increased in workload due to workforce shortages, particularly in rural and regional clinics.

> As with all of those things in a busy practice, the catch is putting time aside [to use FHT]. I never get around to doing so. [GP8]

Participants believed a lack of familiarity and knowledge with SGLT2 inhibitors hindered their use of the module. This stemmed from a lack of time for users to upskill. They were aware of the education resources and support offered through FHT, but time pressures did not allow them to review these resources. For the feedback cycle to be completed, GPs must show leadership in engaging with the module as they are the ones that make the decision on whether the recommendation is clinically appropriate. Practice staff have commented that due to time constraints and competing priorities, GPs were unable to prioritize the use of FHT.

> I think our GPs, you need to actually show them and go through it with them [FHT], then there’s always that “I’m time poor, not right now, can’t do it too much.” [CHA1]

Patient Population: Clinical Appropriateness and Choice Alignment

Participants recognized that the recommendations were a guide only and still required the use of clinical acumen. Although they acknowledged the accuracy, they may not be clinically appropriate at the individual patient level due to factors such as age, health literacy, comorbidities, and likely compliance.

> They’re [GPs] not willing to introduce stricter control for a 70 year old that’s got diabetes and he’s well controlled and everything else.... It’s just a little bit reassessment and the individualisation of the knowing your patients. [Practice nurse 7]

Implementation Process: Cost, Training, and Support

The main cost attributed to the use of FHT was time. For GPs, the time spent on the use of PoC before and during consultations was appropriate for the benefit given as the recommendations...
could be actionable at the time or delayed. However, the dashboard was unused by GPs as they did not have time to use it. In contrast, practice nurses and the CHA found the dashboard function useful to create recall lists and as a quality improvement tool.

Would I love to have more of an opportunity to look at the online module and setting up cohorts and doing things? I’d love to have more time doing that. [GP2]

Participants reported that more support and training would increase their familiarity of confidence in discussing and prescribing SGLT2 inhibitors. They were aware of the educational materials offered to them through FHT, but time constraints prevented them from using them. Practice nurses noted that the resources provided specifically for the SGLT2 inhibitor module were tailored more to GPs for prescribing with broader education on the topic likely to be more beneficial for them. While feedback for these resources was positive, participants believed other modes of education would be beneficial for the broader use of FHT.

I suppose maybe just some how to put it into practice, maybe. Like some scenarios or that type of thing. See while these patients come in, this has popped up, this is what we should be talking about and discussing. [Practice nurse 6]

Recipient Variables

Many found it useful to prompt and raise awareness of the benefits of prescribing SGLT2 inhibitors but found the overcapture of patients an issue. While commenting they were confident with the accuracy of the recommendations, the guideline-based recommendations did not take the patient’s personal history and suitability into account. Participants believed that to maximize the potential of the module, they are still needed to use their own clinical judgment to interrogate and decide if the recommendations were appropriate. Some commented that by following the recommendations blindly, it would reduce the “patient centredness” of the clinical decision and remove shared decision-making with the patient. “I wouldn’t have thought of it without the popup.... It’s making medicine more recipe-like” (GP3).

After seeing the recommendations, users identified their own need for further education on SGLT2 inhibitors and diabetic management in general. This stemmed from the need for more knowledge on the topics for them to confidently discuss these medications with patients. While they were aware of the resources offered through FHT, time constraints for education were again noted.

It’s improved my ability...my little areas of where I need to hone my education skills... I probably didn’t do it that well until we had the popup that forced me to think about why am I telling this patient. [Practice nurse 2]

The recommendation is shown every time the patient file is opened independent of the reason for presentation. Users have commented that it would be difficult to introduce the recommendations if the patients presented for a nondiabetes-related consultation and to steer the conversation toward diabetes. Some users would take note of it and often suggest making a subsequent appointment to discuss SGLT2 inhibitor prescribing. Others used the recommendation as a prompt to change their consultation style to incorporate the module in their consultations.

If I was to see a diabetic who wasn’t on a SGLT2 and had cardiovascular risks, then I would - I would in time change my plan. But as you would well know, best practice is hard to change because you’ve got to learn a new mental routine. [GP6]

Discussion

Principal Findings

The overall response to the use of the SGLT2 inhibitor module in FHT was positive. Enablers and barriers to the use of this module were explored. Enablers of the module’s use included users’ confidence about the accuracy of the recommendations and that they found it easy to incorporate into appropriate consultations and comfortable introducing the topic to patients. They felt the recommendations were proactive and a useful prompt for users to consider prescribing the medication. Unfamiliarity with SGLT2 inhibitors is a known barrier to its prescription [16], and GPs’ awareness of the nonglycemic benefits is low [14]. The inclusion and presentation of SGLT2 inhibitors in guidelines are relatively new, and they have also encouraged users to upskill on knowledge for them to feel more comfortable discussing it with patients.

The role of the GP and GPN and both vital to the management of diabetes in the primary care setting [17], and the module has helped define these roles. GPNs used the module to create lists from the dashboard to recall patients to initiate a conversation about possible SGLT2 inhibitor prescribing and optimization of diabetes management in general with the final decision made by the GP to determine suitability. The main barrier to the use of the module was time pressure and competing priorities during the consultations. Many users stated that the heavy workload prevented them from using the full function of FHT.

By mapping against the CP-FIT framework, the PoC function allowed users to move through the feedback cycle successfully. It has led to better recordkeeping (data collection and analysis) to maximize the efficacy of the algorithms to produce accurate recommendations. They still required to use their clinical acumen to decide whether the recommendation was appropriate for the patient (feedback, interaction, and perception). If appropriate (acceptance), they would initiate a conversation with the patient about prescribing SGLT2 inhibitors and discussing the benefits of improved glycemic control and better patient outcomes (clinical performance improvement). The expanding number of modules has led to prompt fatigue causing users to reduce their interaction with PoC and not initiating the feedback cycle.

There were several limitations to this study. Approximately half of the users of this module were the practice champions for FHT in their workplace. This may lead to responder bias with their enthusiasm for FHT making them more likely to participate and have a better understanding of the program. The study was

https://formative.jmir.org/2024/1/e50737

Suen et al JMIR FORMATIVE RESEARCH

(422x43)
conducted during the COVID-19 pandemic, which may prevent potential users from engaging with the program as the effects of the pandemic caused disruption to workflow, especially with staff and resource shortages. We aim to conduct follow-up interviews in the future to see if the pandemic had any significant effect on the use of FHT.

The use of clinical decision support systems (CDSSs) in diabetes has previously been shown to improve patient outcomes [18], although studies are focused on the hospital setting and glycemic control [19,20]. Literature has consistently reported a gap between optimal diabetes care practice and recommended care standards. Previous CDSS studies have found high clinician satisfaction with its use to facilitate the intensification of glucose control [21]. There is strong evidence that users of CDSS consistently recommend its use to others [22] and is appropriate for general practice use [23].

Previous barriers with CDSSs previously reported include increased workload and time constraints [24], which align with the findings of this study. A previous study showed that the implementation of a CDSS alone did not improve quality of care but required multifaceted strategies including continuing education and feedback mechanisms, organizational changes, and patient-orientated strategies [25]. While these strategies were made available to participants, engagement and uptake of the SGLT2 inhibitor module recommendations were variable.

With the increasing burden of type 2 diabetes on the Australian health care system, early diagnosis and treatment will no doubt reduce the risk of developing and delaying comorbidities and improve the quality of life for diabetics. CDSS can assist clinicians in diagnosing and optimizing the management of chronic disease beyond type 2 diabetes.

Conclusions

This study highlights the benefit of a clinical decision support tool to improve appropriate prescribing and increase clinician awareness of SGLT2 inhibitors for their diabetic and cardiorenal effects. Successful implementation of this module could be used to detect patients who will benefit from the effects of SGLT2 inhibitors in primary care and assist in reducing all-cause mortality and morbidity with guideline-concordant prescribing.

Acknowledgments

The authors would like to acknowledge the contributions of the Future Health Today project team and investigators. We would also like to acknowledge the time and commitment of all participants. The project was supported by the Australian Government’s Targeted Translation Research Accelerator program and the Paul Ramsey Foundation.

Data Availability

The data sets generated during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References


Abbreviations

CDSS: clinical decision support system
CHA: clinical health assistant
CKD: chronic kidney disease
CP-FIT: Clinical Performance Feedback Intervention Theory
FHT: Future Health Today
GP: general practitioner
GPN: general practice nurse
PoC: point of care

https://formative.jmir.org/2024/1/e50737

JMIR Formative Research 2024 | vol. 8 | e50737 | p.1292
(page number not for citation purposes)
SGLT2: sodium-glucose cotransporter 2
Background: Hot flashes are associated with a lower quality of life and sleep disturbances. Given the many consequences of hot flashes, it is important to find treatments to reduce them. Hypnotherapy, the use of hypnosis for a medical disorder or concern, has been shown in clinical trials to be effective in reducing hot flashes, but it is not routinely used in clinical practice. One solution to close this implementation gap is to administer hypnotherapy for hot flashes via a smartphone app. Evia is a smartphone app that delivers hypnotherapy for hot flashes. Evia has made hypnotherapy more widely accessible for women who are experiencing hot flashes; however, the app has yet to undergo empirical testing. Additionally, research on user characteristics is lacking.

Objective: This study aims to (1) determine the average age, stage of menopause, and length of menopause symptoms for users of the Evia app; (2) determine the characteristics of hot flashes and night sweats for users of the Evia app; (3) determine the self-reported sleep quality of users of the Evia app; (4) determine the self-reported mental health of users of the Evia app; and (5) determine the relationship between hot flash frequency and anxiety and depression for users of the Evia app.

Methods: This study analyzed data collected from participants who have downloaded the Evia app. Data were collected at 1 time point from a self-report questionnaire that assessed the demographic and clinical characteristics of users. The questionnaire was given to users when they downloaded the Evia app. Users of the Evia app fill out a questionnaire upon enrolling in the program and prior to beginning the intervention. This included 9764 users.

Results: Results showed that the mean age of users was 49.31 years. A total of 41.6% (1942/4665) of users reported experiencing 5 or more hot flashes per day, while 51.2% (1473/2877) of users reported having difficulty falling asleep each night and 47.7% (1253/2626) of users reported their sleep quality to be terrible. In addition, 38.4% (1104/2877) of users reported that they often feel anxious or depressed. There was a small, significant, and negative correlation between hot flash frequency and self-report frequency of anxiety and depression ($r = -0.09$).

Conclusions: This study showed that the average age of app users is in line with the median age of natural menopause. A large percentage of users reported experiencing 5 or more hot flashes per day, while 51.2% (1473/2877) of users reported having difficulty falling asleep each night and 47.7% (1253/2626) of users reported their sleep quality to be terrible. In addition, 38.4% (1104/2877) of users reported that they often feel anxious or depressed. There was a small, significant, and negative correlation between hot flash frequency and self-report frequency of anxiety and depression ($r = -0.09$).

Keywords: hypnotherapy; hot flashes; smartphone app; mHealth; mobile health; app; apps; applications; hypnosis; menopause; menopausal; gynecology; usage; women's health; user; users; demographics; demographic; characteristic; characteristics; mental health; alternative; complementary; mind body; hypnotism
Introduction

Hot flashes are common during menopause, which is marked by 12 months of amenorrhea or absence of the menstrual period. It is estimated that approximately 75% of women experience hot flashes during menopause [1]. During the transition into menopause, hormone levels fluctuate, and this leads to many physical changes and symptoms, with hot flashes being known as the “hallmark” symptom of menopause. Hot flashes are distressing for women who experience them, and they are associated with a lower quality of life [2] and psychological well-being [3]. Hot flashes are also associated with sleep disturbances [4-6]. Given the many consequences of hot flashes, it is important to find treatments to reduce them.

The most effective treatment for reducing hot flashes is hormone therapy; however, this has been shown to increase the risk of breast cancer and its use is contraindicated in many situations such as in breast cancer survivors [7]. Other treatments for hot flashes such as selective serotonin reuptake inhibitors, selective norepinephrine reuptake inhibitors, clonidine, gabapentin, cognitive behavioral therapy, and mindfulness-based stress reduction have not been found to be as effective as hormone therapy, and they often include side effects that make their long-term use undesirable [8-13]. Therefore, effective and safe nonpharmacological treatment options are needed to relieve hot flashes.

Hypnotherapy, the use of hypnosis for a medical disorder or concern, has been shown to be effective in reducing hot flashes [14,15]. In a clinical trial that examined the efficacy of hypnotherapy for reducing hot flashes among breast cancer survivors, it was found that hypnotherapy reduced hot flashes by approximately 68% from baseline to intervention end point and this was significantly greater than controls [14]. In another clinical trial that examined the efficacy of hypnotherapy for reducing hot flashes among postmenopausal women, hot flashes were reduced by approximately 80% from baseline to week 12 and these reductions were significantly greater than the structured attention control condition [15].

Although hypnotherapy has been shown to be effective in clinical trials, it is not routinely used in clinical practice. Very few clinicians are trained to administer hypnotherapy for hot flashes. One solution to close this implementation gap is to administer hypnotherapy for hot flashes via a smartphone app. Evia is a smartphone app that delivers hypnotherapy for hot flashes. It is based on a hypnotherapy protocol that has been rigorously tested in randomized clinical trials [14,15] and was developed in consultation with an expert in the field of hypnotherapy for hot flashes. Evia includes a 5-week hypnotherapy intervention with daily tasks that include listening to the audio-recorded hypnotherapy session, tracking hot flashes, and psychoeducational readings. Evia has made hypnotherapy more widely accessible for women who are experiencing hot flashes; however, the app has yet to undergo empirical testing. Additionally, research on user characteristics is lacking. This is the first study to report the characteristics of users of Evia. The information gathered from this study will allow for the Evia app to be tailored and optimized toward its users. This study will also tell us whether individuals who seek relief from hot flashes via a smartphone app are different from those who seek treatment in a clinical setting or a clinical trial, and will allow us to better understand who is downloading and engaging with the digital hypnotherapy intervention. Some previous research has shown that smartphone app users are generally younger [16], and this study will explore whether users of Evia are younger than women who seek treatment in a clinical setting or clinical trial. Additionally, the study will explore how app users may differ in the length of time they have experienced menopause symptoms or symptom severity. In addition, this study is a first step to conducting further research on the Evia app, including clinical trials to evaluate efficacy.

This study analyzed data collected from participants who have downloaded the Evia app. These data were collected by Mindset Health, the developers of the Evia app. Users of the Evia app fill out a questionnaire upon enrolling in the program and prior to beginning the intervention. These data will be analyzed to achieve the study’s aims. This study aims to (1) determine the average age, stage of menopause, and length of menopause for users of the Evia app; (2) determine the characteristics of hot flashes and night sweats for users of the Evia app; (3) determine the self-reported sleep quality of users of the Evia app; (4) determine the self-reported mental health of users of the Evia app; and (5) determine the relationship between hot flash frequency and anxiety and depression for users of the Evia app.

Methods

Overview

Data were collected at 1 time point from a self-report questionnaire that was given to users when they downloaded the Evia app between October 5, 2021, and July 8, 2022. Users filled out the self-report questionnaire directly on the Evia app. These data were collected prior to the user beginning the intervention and were analyzed retrospectively.

Participants

Participants included individuals who downloaded the Evia app between October 5, 2021, and July 08, 2022. Participants are users of the Evia app and were recruited via Facebook advertisements and the app. Any person who downloaded the Evia app during this time period was prompted to fill out a questionnaire with demographic and clinical information prior to beginning the intervention. This included 9764 participants.

Measures

Aim 1

Participants were asked to fill in their age. Participants were asked, “How would you classify your stage of menopause?” Response options included “perimenopause,” “menopause,” “postmenopause,” “I’m not menopausal,” and “I’m not sure.” To determine the amount of time that users have been experiencing menopause symptoms, participants were asked, “How long have you been experiencing menopause symptoms?” Response options included “0-6 months,” “6-12 months,” “1-2 years,” “2-3 years,” “3-5 years,” “5-10 years,” and “10+ years.”
**Aim 2**
Participants were asked, “How many hot flashes do you experience each day?” Response options ranged from “0” to “5+.” To assess the severity of hot flashes that users are experiencing, participants were asked, “How severe are your hot flashes?” Response options included “very mild”, “mild”, “moderate”, “severe”, and “very severe.” To assess the characteristics of users’ hot flashes, they were asked, “What do your hot flashes feel like?” Response options included “sudden feeling of warmth,” “perspiration,” “flushed appearance,” “rapid heartbeat,” “anxiety/an aura,” “dizziness/weakness,” “nausea,” and “other.” Users were able to select multiple responses for this item. To assess the number of night sweats that users experience, they were asked, “How many night sweats do you experience each night, on average?” Response options ranged from “0” to “5+.”

**Aim 3**
Users were asked to report sleep difficulty with 1 self-report item that reads, “Many women also struggle with sleep during menopause. Do you find it difficult to fall asleep each night?” Response options include “no,” “a little bit,” and “yes.” Users were also asked to rate their sleep quality with 1 self-report question that read, “How would you rate the quality of your sleep?” Response options included “terrible,” “fair,” “good,” and “excellent.”

**Aim 4**
Users were asked 1 self-report question that read, “Some women also struggle with mental health during menopause. How often do you feel anxious or depressed?” Response options included “never,” “sometimes,” “often,” and “constantly.”

**Aim 5**
Users were asked, “How many hot flashes do you experience each day?” Response options ranged from “0” to “5+.” Users were asked, “Some women also struggle with mental health during menopause. How often do you feel anxious or depressed?” Response options included “never,” “sometimes,” “often,” and “constantly.”

**Ethical Considerations**
This study was reviewed by the institutional review board at Baylor University and was considered to be exempt. The institutional review board determined the study to be secondary research for which consent is not required. The data analyzed were deidentified.

**Data Analyses**
Descriptive statistics (means, SDs, and frequencies) were calculated for the variables of interest. For aim 5, which was to determine the relationship between hot flash frequency and self-report frequency of anxiety and depression, a Pearson correlation analysis was used.

**Results**

**Aim 1: Age, Stage of Menopause, and Length of Menopause Symptoms**
The mean age of users was 49.31 (SD 6.69(1,973),(92,999)) years. Out of 9103 valid responses, 0.5% (n=49) of users reported being not menopausal, 31.2% (n=2837) of users were perimenopausal, 14.2% (n=1293) of users were menopausal, 13.1% (n=1188) of users were postmenopausal, 7.6% (n=688) of users were uncertain, and 33.5% (n=3048) of users reported unknown (see Table 1). Out of 3127 valid responses reporting length of experiencing menopause symptoms, 20.4% (n=639) of users reported 0-6 months, 20.9% (n=655) of users reported 6-12 months, 20% (n=624) of users reported 1-2 years, 13.5% (n=421) of users reported 2-3 years, 11.3% (n=354) of users reported 3-5 years, 9.3% (n=290) of users reported 5-10 years, and 4.6% (n=144) of users reported 10+ years (see Table 1).
Table 1. Stage of menopause and length of menopause symptoms for Evia users. This table lists the frequencies and percentages of Evia app users who are in each stage of menopause and their length of menopause symptoms. These data were collected from Evia app users who downloaded the Evia app. Data were collected via a questionnaire given on the Evia app to users prior to beginning the intervention and were analyzed retrospectively. These data were collected between October 5, 2021, and July 8, 2022. The Evia app includes a hypnotherapy intervention to help reduce hot flashes.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage of menopause (n=9103)</td>
<td></td>
</tr>
<tr>
<td>Not menopausal</td>
<td>49 (0.5)</td>
</tr>
<tr>
<td>Perimenopausal</td>
<td>2837 (31.2)</td>
</tr>
<tr>
<td>Menopausal</td>
<td>1293 (14.2)</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>1188 (13.1)</td>
</tr>
<tr>
<td>Uncertain</td>
<td>688 (7.6)</td>
</tr>
<tr>
<td>Unknown</td>
<td>3048 (33.5)</td>
</tr>
<tr>
<td>Length of menopause symptoms (n=3127)</td>
<td></td>
</tr>
<tr>
<td>0-6 months</td>
<td>639 (20.4)</td>
</tr>
<tr>
<td>6-12 months</td>
<td>655 (20.9)</td>
</tr>
<tr>
<td>1-2 years</td>
<td>624 (20)</td>
</tr>
<tr>
<td>2-3 years</td>
<td>421 (13.5)</td>
</tr>
<tr>
<td>3-5 years</td>
<td>354 (11.3)</td>
</tr>
<tr>
<td>5-10 years</td>
<td>290 (9.3)</td>
</tr>
<tr>
<td>10+ years</td>
<td>144 (4.6)</td>
</tr>
</tbody>
</table>

Aim 2: Hot Flash Characteristics

Out of 4665 valid responses, 13.5% (n=630) of users reported experiencing 0 hot flashes per day, 9.4% (n=437) of users reported experiencing 1 hot flash per day, 13% (n=607) of users reported experiencing 2 hot flashes per day, 12.8% (n=595) of users reported experiencing 3 hot flashes per day, 9.7% (n=454) of users reported experiencing 4 hot flashes per day, and 41.6% (n=1942) of users reported experiencing 5+ hot flashes per day (see Table 2). Out of 4183 valid responses, 1.9% (n=80) of users reported that their hot flashes were very mild, 30.4% (n=1273) of users reported that their hot flashes were mild, 40.9% (n=1710) of users reported that their hot flashes were moderate, 20.3% (n=851) of users reported that their hot flashes were severe, and 6.4% (n=269) of users reported that their hot flashes were very severe (see Table 2). Out of 4180 valid responses, 83% (n=3469) of users reported that their hot flashes felt like a sudden feeling of warmth, 59.2% (n=2474) of users reported perspiration, 47.5% (n=1987) of users reported a flushed appearance, 34.5% (n=1442) of users reported rapid heartbeat, 32.7% (n=1367) of users reported anxiety or an aura, 24.5% (n=1023) of users reported dizziness or weakness, and 16.4% (n=686) of users reported nausea (see Table 2). Out of 321 valid responses, 13.4% (n=43) of users reported experiencing 0 night sweats, 25.5% (n=82) of users reported experiencing 1 night sweat, 23.1% (n=74) of users reported experiencing 2 night sweats, 21.8% (n=70) of users reported experiencing 3 night sweats, and 16.2% (n=52) of users reported experiencing 4 night sweats (see Table 2).
Table 2. Number, severity, and characteristics of hot flashes for Evia users. This table lists the frequencies and percentages for Evia app users’ number of hot flashes, severity of hot flashes, hot flash characteristics, and number of night sweats. These data were collected from Evia app users who downloaded the Evia app. Data were collected via a questionnaire given on the Evia app to users prior to beginning the intervention and were analyzed retrospectively. These data were collected between October 5, 2021, and July 8, 2022. The Evia app includes a hypnotherapy intervention to help reduce hot flashes.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of hot flashes (n=4665)</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>630 (13.5)</td>
</tr>
<tr>
<td>1</td>
<td>437 (9.4)</td>
</tr>
<tr>
<td>2</td>
<td>607 (13)</td>
</tr>
<tr>
<td>3</td>
<td>595 (12.8)</td>
</tr>
<tr>
<td>4</td>
<td>454 (9.7)</td>
</tr>
<tr>
<td>5+</td>
<td>1942 (41.6)</td>
</tr>
<tr>
<td><strong>Severity of hot flashes (n=4183)</strong></td>
<td></td>
</tr>
<tr>
<td>Very mild</td>
<td>80 (1.9)</td>
</tr>
<tr>
<td>Mild</td>
<td>1273 (30.4)</td>
</tr>
<tr>
<td>Moderate</td>
<td>1710 (40.9)</td>
</tr>
<tr>
<td>Severe</td>
<td>851 (20.3)</td>
</tr>
<tr>
<td>Very severe</td>
<td>269 (6.4)</td>
</tr>
<tr>
<td><strong>Hot flashes feel like (n=4180)</strong></td>
<td></td>
</tr>
<tr>
<td>Sudden feeling of warmth</td>
<td>3469 (83)</td>
</tr>
<tr>
<td>Perspiration</td>
<td>2474 (59.2)</td>
</tr>
<tr>
<td>Flushed appearance</td>
<td>1987 (47.5)</td>
</tr>
<tr>
<td>Rapid heartbeat</td>
<td>1442 (34.5)</td>
</tr>
<tr>
<td>Anxiety or an aura</td>
<td>1367 (32.7)</td>
</tr>
<tr>
<td>Dizziness or weakness</td>
<td>1023 (24.5)</td>
</tr>
<tr>
<td>Nausea</td>
<td>686 (16.4)</td>
</tr>
<tr>
<td>Other</td>
<td>307 (7.3)</td>
</tr>
<tr>
<td><strong>Number of night sweats (n=321)</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>43 (13.4)</td>
</tr>
<tr>
<td>1</td>
<td>82 (25.5)</td>
</tr>
<tr>
<td>2</td>
<td>74 (23.1)</td>
</tr>
<tr>
<td>3</td>
<td>70 (21.8)</td>
</tr>
<tr>
<td>4</td>
<td>52 (16.2)</td>
</tr>
</tbody>
</table>

**Aim 3: Sleep**

Out of 2877 valid responses, results demonstrated that 18% (n=519) of users reported no to having difficulty falling asleep each night, 30.8% (n=885) of users reported a little bit of difficulty falling asleep each night, and 51.2% (n=1473) of users reported yes to having difficulty falling asleep each night. Out of 2626 valid responses, results also demonstrated that 47.7% (n=1253) of users reported their sleep quality to be terrible, 44% (n=1155) of users reported their sleep quality to be fair, 7.8% (n=206) of users reported their sleep quality to be good, and 0.5% (n=12) of users reported their sleep quality to be excellent.

**Aim 4: Mental Health**

Out of 2877 valid responses, 5.9% (n=169) of users reported that they never feel anxious or depressed, 38.4% (n=1104) of users reported that they often feel anxious or depressed, and 16.2% (n=465) of users reported that they constantly feel anxious or depressed.

**Aim 5: Relationship Between Hot Flash Frequency and Anxiety and Depression**

There was a small, significant, and negative correlation between hot flash frequency and self-report frequency of anxiety and depression ($r(2875) = -0.09, P<.001$).
**Discussion**

**Principal Findings**

Results demonstrated that the average age of app users (49.31 years) is in line with the median age of the beginning of perimenopause (47.5 years) and the median age of natural menopause (51.3) [17]. Although previous studies have shown that, overall, app users are generally younger [16], users of Evia are the age we would expect based on the age of menopause onset and symptoms emerging. Additionally, a large percentage of users (2837/9103, 31.2%) reported being in perimenopause, and this is oftentimes when menopause symptoms, including hot flashes, begin. An unexpected finding in this study was that a large percentage of users did not know their stage of menopause (3048/9103, 33.5%). This shows that users of Evia may not be aware of the various stages of menopause and what the menopause transition entails. This suggests that it may be beneficial for the Evia app to include some psychoeducational components about the menopause transition and stages of menopause. These psychoeducational components may help users of Evia better understand their menopause transition and experiences. Many users reported experiencing menopause symptoms for 0-6 months (639/3127, 20.4%), 6-12 months (655/3127, 20.9%), or 1-2 years (624/3127, 20%) with users less commonly reporting experiencing menopause symptoms for 5-10 years (290/3127, 9.3%) or 10+ years (144/3127, 4.6%). These results demonstrate that users are starting to use Evia toward the beginning of menopause symptoms. It is possible that those who have been experiencing symptoms for longer have learned how to cope with or treat their symptoms.

The largest percentage of users reported experiencing 5 or more hot flashes per day and reported that their hot flashes were moderate intensity. This finding is in line with previous studies that assessed hot flash frequency at baseline [15]. In a clinical trial by Elkins et al [15], participants experienced approximately 10 hot flashes per day on average at baseline. This study showed that, on average, Evia app users experience 5 or more hot flashes per day, which is in line with participants in previous clinical trials.

In line with previous studies regarding hot flashes and sleep, a majority of users (1473/2877, 51.2%) reported difficulty falling asleep each night and reported their sleep quality to be terrible (358/2877, 12.5%) or poor (316/2877, 11.0%). There were also limitations regarding how variables in the study were measured. For example, sleep quality and frequency of anxiety and depression were measured. For example, sleep quality and frequency of anxiety and depression were measured using single items. There are limitations to using single-item scales, including that they may not accurately or fully capture the construct. In addition, the items may have been worded in a suggestive manner. For example, “Many women also struggle with sleep during menopause. Do you find it difficult to fall asleep each night?” By starting the item with, “Many women also struggle with sleep during menopause...” it may be suggested for the individual to respond in a certain way. By identifying these limitations of measurement, this study will help to inform the optimization of data collection and measurement with the Evia app.

The results of this study will inform the optimization of the hypnotherapy intervention delivered via the Evia app. The results of this study will also inform the optimization of data collection via the Evia app. Once the Evia app is optimized, users will be better able to report their symptoms and experiences. This study also provides valuable information for future research.

**Limitations**

This study has several limitations that should be addressed. First, several characteristics of users remain unknown and were not addressed by this study. For example, information about geographic location and race or ethnicity was not reported. The study was not able to describe several important demographic factors of the sample, including race or ethnicity and education level.

There were also limitations regarding how variables in the study were measured. For example, sleep quality and frequency of anxiety and depression symptoms were measured using single items. There are limitations to using single-item scales, including that they may not accurately or fully capture the construct. In addition, the items may have been worded in a suggestive manner. For example, “Many women also struggle with sleep during menopause. Do you find it difficult to fall asleep each night?” By starting the item with, “Many women also struggle with sleep during menopause...” it may be suggested for the individual to respond in a certain way. By identifying these limitations of measurement, this study will help to inform the optimization of data collection and measurement with the Evia app.

**Future Directions**

The results of this study will inform the optimization of the hypnotherapy intervention delivered via the Evia app. The results of this study will also inform the optimization of data collection via the Evia app. Once the Evia app is optimized,
further research is warranted. Future studies should first examine the feasibility of the Evia app in a randomized controlled trial design. Then, future studies should examine the efficacy of the Evia app in comparison to a control group or even a face-to-face delivery.

Conclusions
This was the first study to report on the characteristics of users of the Evia app, which delivers hypnotherapy for hot flashes. Results showed that the average age of app users is in line with the average age of menopause onset. Results also showed that the largest percentage of users reported experiencing 5 or more hot flashes per day and reported that their hot flashes were moderate intensity. In line with previous research regarding hot flashes and sleep, a majority of users reported difficulty falling asleep each night and reported their sleep quality to be terrible or fair. In addition, a majority of users reported that they sometimes or often feel anxious or depressed. The results of this study will help to inform the optimization and tailoring of the hypnotherapy intervention delivered via the Evia app.

Acknowledgments
The authors would like to acknowledge Mindset Health and its employees for their collaboration in this project, as well as Clair Hall and Hanna Burrows for assistance with data collection.

Data Availability
The data set analyzed during this study is not publicly available due to confidentiality but is available from the corresponding author upon reasonable request.

Conflicts of Interest
GRE is a consultant to Mindset Health but was not involved in data collection. GRE discloses that he receives a consultant salary from Mindset Health and holds stock options within the company. However, this salary and equity were not dependent on the results of this research.

References


©Morgan Snyder, Gary R Elkins. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Adapting the Number of Questions Based on Detected Psychological Distress for Cognitive Behavioral Therapy With an Embodied Conversational Agent: Comparative Study

Kazuhiro Shidara¹, MEng; Hiroki Tanaka¹, PhD; Hiroyoshi Adachi², MD, PhD; Daisuke Kanayama², MD, PhD; Takashi Kudo², MD, PhD; Satoshi Nakamura¹, PhD

¹Nara Institute of Science and Technology, Ikoma, Japan
²Health and Counseling Center, Osaka University, Toyonaka, Japan

Corresponding Author:
Kazuhiro Shidara, MEng
Nara Institute of Science and Technology
8916-5, Takayama-cho
Ikoma, 630-0192
Japan
Phone: 81 80 4687 8116
Email: shidara.kazuhiro.sc5@is.naist.jp

Abstract

Background: The high prevalence of mental illness is a critical social problem. The limited availability of mental health services is a major factor that exacerbates this problem. One solution is to deliver cognitive behavioral therapy (CBT) using an embodied conversational agent (ECA). ECAs make it possible to provide health care without location or time constraints. One of the techniques used in CBT is Socratic questioning, which guides users to correct negative thoughts. The effectiveness of this approach depends on a therapist’s skill to adapt to the user’s mood or distress level. However, current ECAs do not possess this skill. Therefore, it is essential to implement this adaptation ability to the ECAs.

Objective: This study aims to develop and evaluate a method that automatically adapts the number of Socratic questions based on the level of detected psychological distress during a CBT session with an ECA. We hypothesize that this adaptive approach to selecting the number of questions will lower psychological distress, reduce negative emotional states, and produce more substantial cognitive changes compared with a random number of questions.

Methods: In this study, which envisions health care support in daily life, we recruited participants aged from 18 to 65 years for an experiment that involved 2 different conditions: an ECA that adapts a number of questions based on psychological distress detection or an ECA that only asked a random number of questions. The participants were assigned to 1 of the 2 conditions, experienced a single CBT session with an ECA, and completed questionnaires before and after the session.

Results: The participants completed the experiment. There were slight differences in sex, age, and preexperimental psychological distress levels between the 2 conditions. The adapted number of questions condition showed significantly lower psychological distress than the random number of questions condition after the session. We also found a significant difference in the cognitive change when the number of questions was adapted based on the detected distress level, compared with when the number of questions was fewer than what was appropriate for the level of distress detected.

Conclusions: The results show that an ECA adapting the number of Socratic questions based on detected distress levels increases the effectiveness of CBT. Participants who received an adaptive number of questions experienced greater reductions in distress than those who received a random number of questions. In addition, the participants showed a greater amount of cognitive change when the number of questions matched the detected distress level. This suggests that adapting the question quantity based on distress level detection can improve the results of CBT delivered by an ECA. These results illustrate the advantages of ECAs, paving the way for mental health care that is more tailored and effective.

(JMIR Form Res 2024;8:e50056) doi:10.2196/50056
KEYWORDS
cognitive behavioral therapy; psychological distress detection; embodied conversational agents; automatic thoughts; long short-term memory; multitask learning

Introduction

Background

Cognitive behavioral therapy (CBT) is an established and effective therapeutic approach for treating a wide range of mental illnesses, including depression and anxiety disorder [1,2]. This approach, rooted in the cognitive model of emotional responses, posits that our thoughts, feelings, and behaviors are interconnected. Its central principle asserts that negative automatic thoughts must be corrected because they deleteriously affect emotions and behaviors [3]. CBT has been widely adopted in clinical practice and in preventive health care for the general public [2,4]. Despite its efficacy, broad dissemination faces challenges [5,6]. The World Health Organization [5] asserts that factors such as a shortage of trained therapists, geographical and financial limitations, and societal stigma toward mental health contribute to a substantial treatment gap.

Researchers have attempted to provide CBT through embodied conversational agents (ECAs) to address social issues [7-9]. ECAs are computer-programmed interfaces that simulate human-like conversations with users by using natural language processing techniques. Examples include messaging app–like chatbots [10,11], robots [12,13], and ECAs displayed as computer graphics on a screen [14]. Text-based agents or chatbots are especially widespread and commercially available as smartphone apps. Inkster et al [11] and Fitzpatrick et al [10] are 2 examples of such agents whose ability has been extensively validated for targeting and helping individuals who are experiencing mild to moderate symptoms of depression and anxiety. These agents provide therapy for daily mental health care rather than intensive clinical treatment. Studies have shown their effectiveness in reducing symptoms of depression and anxiety, increasing mental well-being, and fostering user resilience [15].

ECAs, or chatbots, increase accessibility, affordability, and anonymity, potentially reaching a broader population that might benefit from CBT. However, the effectiveness of these digital therapies depends on several factors, including therapeutic techniques, empathy, user interface, and personalization. Asay and Lambert [16] categorized the factors affecting therapy effectiveness into four major groups: (1) extratherapeutic changes, which include clients’ personality and environmental factors that support recovery, independent of therapy participation; (2) common factors such as empathy and therapeutic alliance, which are found in various therapy approaches; (3) expectancy, which involves improvement resulting from clients’ anticipation of assistance and their belief in the therapy’s rationale and effectiveness; and (4) techniques, which are specific factors unique to particular therapies and adjusted to address specific issues. Among these 4 groups, the techniques are crucial for enhancing the quality of therapy. One critical CBT therapeutic technique strategically uses Socratic questioning [17]. On the basis of the pedagogical style of the ancient Greek philosopher Socrates, this method is a form of guided discovery that encourages clients to critically examine and articulate their thought patterns to facilitate cognitive change and distress reduction [18,19]. Socratic questioning helps clients deeply contemplate their problems and perspectives. Questions such as “Why do you think that way?” “What evidence supports this perspective?,” and “Have you considered other perspectives?” can be adjusted to specific problems or situations. These flexible templates can be applied to various concerns and situations. The process involves carefully considering the content, timing, and sequence of questions adjusted to an individual’s cognitive processes and emotional state, emphasizing the quality of queries when engaging the client in Socratic questioning [2,4,20]. Therapists observe the content of clients’ utterances, changes in their voices, and facial expressions to understand changes in their mood [2,4].

Recent studies have started to investigate the application of Socratic questioning in automated CBT systems such as ECAs. Kimani et al [14] and Shidara et al [21] explored ways to implement Socratic questioning in CBT delivered by ECAs. The former work provided Socratic questions according to a dialogue context using a flowchart of selectable inputs and reduced the specific anxiety about giving a presentation. The latter study set up 2 dialogue scenarios, with and without Socratic questioning, and evaluated them in a comparative experiment. The group with the scenario containing Socratic questioning showed a larger reduction in negative moods. Both studies suggest that Socratic questioning by ECAs promotes cognitive changes and distress reduction.

However, there is a significant gap in current research. Existing studies have not explored how to automatically adapt Socratic questioning. Ideally, ECAs should adapt CBT strategies according to users’ psychological states, just like human therapists do. Specifically, there is limited research on this topic. We must focus on the adjustment of the number of questions or its types. Therefore, we set our goal to personalize CBT strategies based on an individual’s current psychological state.

This study addresses this gap by introducing an adaptive method to improve the effectiveness of CBT using ECAs. This method adjusts the number of questions during a therapy session based on the detected level of psychological distress. The main goal is to provide questions according to the user’s level of distress to improve the therapeutic interaction. To achieve this goal, we formulated the following research questions:

1. Research question 1: Can modifying the number of questions based on the user’s level of psychological distress lead to better therapeutic outcomes?

To address this question, our system uses a machine learning model with natural language processing techniques to estimate a user’s psychological state during a CBT session. Our system can automatically detect psychological distress, allowing it to adapt to the user’s mental state without using psychological...
scales such as the Kessler Psychological Distress Scale (K6) in each session. Furthermore, by integrating the distress detection model, our system can estimate the user’s distress in real time during the CBT session. This functionality is intended to improve the effectiveness of CBT with an ECA. In this sense, we present the hypothesis 1.

Hypothesis 1: In a condition with the adaptive number of questions, improvement in distress and emotion and cognitive change will be better than in a condition with a random number of questions.

2. **Research question 2**: Is there a clear benefit to modifying the number of questions to match the user’s specific psychological state rather than setting a fixed optimal number for all?

For this question, we emphasized the value of customization, recognizing that the most effective balance of questions may vary from user to user. Therefore, we aimed to clarify the therapeutic benefit of determining the distress reduction moments. Through this investigation, we aimed to identify the most effective strategy: changing the number of questions or setting a user-specific number of questions. Therefore, we propose hypothesis 2.

Hypothesis 2: The adapted question amount has better improvement than both the too few questions amount and the too many questions amount in terms of the users’ distress improvement, emotion improvement, and cognitive change.

To validate these hypotheses, we tested it experimentally by comparing 2 conditions: the adapted number of questions condition and the random number of questions condition. We assessed the reduction of psychological distress and negative emotional states and cognitive changes in both conditions. We also analyzed the influence of deviations in the number of questions from the model’s estimated value on the effectiveness of the ECAs.

The incorporation of a distress detection model into an ECA system is essential to our method. In the Psychological Distress Detection and Conversational Scenario section, we describe our psychological distress model and the design of the agent architecture.

**Psychological Distress Detection and Conversational Scenario**

**System Overview**

Figure 1 shows the overview of our system, including the abstract scenario. The detailed scenario can be found in Multimedia Appendix 1. This ECA uses an ECA toolkit called Greta (Institut des Systèmes Intelligents et de Robotique in Paris) [22]. Its animation’s appearance uses a version identified as more acceptable to Japanese people [23]. We used the ja-JP-Wavenet-B of Google’s Text-to-Speech service to generate the synthesized speech of a Japanese female voice. We used a laptop (i7, 32GB RAM, Dell Inc) for the ECA implementation and recorded the participants’ facial expressions and speech with its built-in camera. A microphone and a speaker were built into a headset (Sennheiser). Figure 2 shows a conversation between a user and our ECA.

---

Figure 1. A framework of an embodied conversational agent and a conversational scenario for cognitive behavioral therapy. MPEG-4: Moving Picture Experts Group Phase 4.
In this study, we developed a question scenario based on the principles of cognitive restructuring [2,4], a key component of CBT. The scenario was supervised and reviewed by a psychiatrist to ensure clinical accuracy and relevance. The CBT process involves 3 main steps: identifying automatic thoughts, correcting them, and examining new balanced thoughts.

The first step in the scenario involves identifying automatic thoughts by describing the situation, the mood, and the automatic thoughts themselves. Typically, the situation and mood are described first, followed by identifying the automatic thoughts that occur between the situation and the mood. When describing mood, participants also report the intensity of their moods using a score, which is a percentage ranging from 0 (not stressed at all) to 100 (the most stressed they have ever felt). In this step, the users select their strongest automatic thought. Although multiple automatic thoughts may arise from a single situation, focusing on the strongest thought allows for the largest reduction of mood score.

Once the automatic thoughts have been identified, participants correct their thoughts by comparing them with the facts of the situation. The CBT system poses questions designed to increase participants’ awareness of overlooked facts and provides distance from their automatic thoughts to view them more objectively. After correcting their automatic thoughts, the participants were asked to state their new balanced thoughts. This step involved considering the insights gained while correcting automatic thoughts and generating new balanced thoughts. At the end of this process, participants reported their mood score again to confirm a reduction of their negative mood scores after deriving new balanced thoughts.

**Adaptation of Number of Questions**

In our developed scenario, multiple questions were included for correcting automatic thoughts. This step ensured that the standardized scenario effectively improved participants’ mood by incorporating multiple questions. In actual CBT sessions, questions are adjusted according to an individual patient’s condition and level of distress.

Our ECA posed questions aimed at helping participants notice overlooked facts and gain distance from their automatic thoughts.
to view them more objectively. In the developed scenario, multiple Socratic questions were included for correcting automatic thoughts. This step ensured that the standardized scenario effectively improved participants’ mood by incorporating multiple questions. The number of questions posed for correcting automatic thoughts was adapted based on the detected level of the participant’s psychological distress. After posing questions that correct an automatic thought, the participant’s responses were used to identify their current psychological distress state. If distress is detected, the system continues to pose additional questions that prompt the participants to correct their thinking from various perspectives. This process gently nudges participants to reconsider their automatic thoughts, offering them the opportunity to alter their perspective and reduce their distress. The system was designed using a pool of 21 Socratic questions to prompt the correction of automatic thoughts. If a participant continues to exhibit high levels of psychological distress, the agent can pose up to 20 additional questions to assist them in working through their distress.

If the estimated value of the detection model is nondistressed, the agent proceeds to the next step, where the participant composes a balanced alternative thought. This step was based on the insights gained in the previous step, which is the correction of automatic thoughts. After this process, the participant again reported a negative mood score, which was expected to confirm a less negative mood.

**Pretraining of Detection Models to Address Japanese Conversation**

In our quest to dynamically adapt the number of questions asked during a CBT session, psychological distress in users must be accurately detected. In this context, machine learning models have been developed to estimate psychological distress [25-29], often using publicly available data sets, such as the Distress Analysis Interview Corpus/Wizard-of-Oz (DAIC-WOZ) database [30]. Unfortunately, this data set primarily contains English language data focused on assessing depressive tendencies rather than providing mental health care. Therefore, our challenge is to adapt these models to suit a Japanese language context, especially for the CBT domain.

To address this challenge, we first applied a deep learning architecture of the depression detection model proposed by Li et al [31]. Their model identifies depressive tendencies in interactions with ECAs by classifying presence or absence into 2 categories. It combines recurrent neural network and long short-term memory networks and uses multitask learning for depression detection. Multitask learning enables simultaneous training on multiple tasks, leveraging shared knowledge to improve generalization, use relationships between tasks, and reduce overfitting. In the study by Li et al [31], the model achieved a high $F_1$-score of 0.71.

Figure 3 [30,32] shows the training flow of the psychological distress detection model [30,32]. For the training process, we used a combination of pretraining and transfer learning. First, we translated the original training data in the study by Li et al [31] from English into Japanese using machine translation software and pretrained the model with reference to the settings in the study by Li et al [31]. We then constructed a Japanese CBT data set using crowdsourcing and conducted transfer learning on the detection model.
We retained the original implementation details of the study by Li et al [31]. We trained the model for a maximum of 100 epochs with early stopping based on the macro \( F_1 \) metric for depression classification. We used cross-entropy loss and a batch size of 1 for the DailyDialog and DAIC-WOZ. Tokenization was performed using the MeCab library [36], and we constructed word embeddings with a default dimension of 128. The turn-level encoder consisted of 1 hidden layer and 128 output neurons. We tuned the document-level recurrent neural network layers within a range of \([1, 2, 3]\) and the hidden size within a range of \([128, 256, 512]\). The model parameters were optimized using the Adam optimizer [37] with a learning rate of 0.001. Dropout rates were set to 0.1 for both the turn and document encoders.

Transfer Learning for Real-Time Distress Detection in CBT

After pretraining the model with the translated data, we implemented transfer learning by replacing the fully connected layer. This transfer learning process allowed us to leverage the original model architecture while shaping it to our specific distress detection task in CBT. In this transfer learning phase, we incorporated only 3 types of user utterances corresponding to the situation, mood, and automatic thought responses within the CBT process. This approach was adopted to enable real-time distress detection during CBT sessions with the ECA. When the detection model makes inferences, it inputs the 3 most recent utterances. In this way, the distress detection model enables real-time detection in a CBT session.

Our crowdsourced data set was initially collected from 100 crowd workers from the general public. However, upon visual inspection of the data, it became apparent that the responses from some participants were not adequately informative or valid for our purpose. Therefore, we excluded data from 6 participants as likely outliers or nonresponsive. Thus, the final data set for our analysis consisted of responses from 94 participants. In the data collection, we assessed the level of psychological distress in our crowdsourced data set with the Japanese version of K6 [38] (Japanese version [39]). This scale is a concise and reliable self-report tool that assesses the level of psychological distress experienced by individuals in the previous 30 days. This 6-item scale measures the frequency of symptoms related to anxiety and depression, making it a valuable instrument for detecting and evaluating mental health disorders in both clinical and research settings. The items are rated on a 5-point Likert scale, from 0 (none of the time) to 4 (all of the time), leading to a total score range of 0 to 24. Higher scores indicate greater psychological distress. In this study, following Sakurai et al [40], a K6 score of \( \geq 5 \) was considered distressed, and a score of \( <5 \) was considered nondistressed. We constructed a psychological distress detection model that classified these 2 classes. Of the 94 participants, 63 (67%) were categorized as distressed (K6 score of \( \geq 5 \)) and 31 (33%) as nondistressed (K6 score \( <5 \)). We used stratified sampling to automatically split the data into training, validation, and testing subsets. We used 55% (52/94) of the participants for training, 19% (18/94) of the participants for validation, and 26% (24/94) of the participants for testing.

Methods

Ethical Considerations

The research ethics committee of the Nara Institute of Science and Technology reviewed and approved this study (2019-1-24-3). We engaged a human resources company to advertise and recruit participants. All participants provided written informed consent before participating in the experiment. The participants were compensated with an honorarium, which was paid by the human resources company. Study data were anonymized.

Participants

This study was conducted in February and March 2023. Eligibility criteria for the experimental participants were as follows: (1) aged between 18 and 65 years, (2) having no hearing impairments, and (3) ability to speak Japanese. For the analysis of this research, 49 participants were allocated into 2 conditions to assess the impact of varying the number of Socratic questions based on detected levels of psychological distress during CBT sessions with an ECA.

In addition to these participants, a separate group (of 26 participants) was involved in a distinct experimental condition, designed for a different research focus. Although data for this separate group were collected concurrently with the main study, the data from this separate group are not relevant to the current analysis and have not been included in this paper. The objective of collecting the data from this separate group was to analyze whether an extreme number of questions were stressful and discouraged cognitive changes. It is expected that participants in this separate group might experience or perceive the system rather than personal emotional challenges such as depression or anxiety. Specifically, participants may be dissatisfied with how the system works or interacts rather than with their internal struggles. Although this system-related dissatisfaction is also important, it is beyond the scope of this study. Only the 2 conditions were used in this study to focus on the person’s nonadaptive thoughts and moods regarding their problems.

Table 1 shows the demographics and baseline characteristics of the participants who were recruited through an external participant recruitment service. We also included 1 participant with a history of a mental health issue in the adapted number of questions condition. This participant was currently in remission and not taking any medication. This participant was recruited through a company that provides employment transition services for people with a history of mental health issues.
Table 1. Participants’ demographics and baseline characteristics (N=49).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Adapted number of questions condition</th>
<th>Random number of questions condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of participants, n (%)</td>
<td>25 (100)</td>
<td>24 (100)</td>
</tr>
<tr>
<td>Without mental health issues, n (%)</td>
<td>24 (96)</td>
<td>24 (100)</td>
</tr>
<tr>
<td>With a history of a mental health issue, n (%)</td>
<td>1 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>12 (48)</td>
<td>13 (54)</td>
</tr>
<tr>
<td>Female</td>
<td>13 (52)</td>
<td>11 (46)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>41.72 (12.79)</td>
<td>38.88 (14.89)</td>
</tr>
<tr>
<td>K6(^a), mean (SD)</td>
<td>4.08 (4.31)</td>
<td>4.38 (3.10)</td>
</tr>
</tbody>
</table>

\(^a\)K6: Kessler Psychological Distress Scale.

To confirm whether there was any bias in the age of the participants and the preexperiment K6 scores between the 2 conditions, we conducted a 2-sided Welch t-test and calculated Cohen d values. As a result, for age, Cohen d was 0.21 and the P value was .48; for preexperiment K6 scores, Cohen d was −0.08 and the P value was .78. These results indicate that there was no significant bias in age and preexperiment K6 scores between the 2 conditions.

**Experimental Conditions**

Figure 4 illustrates the conversation scenarios for our 2 experimental conditions: the adapted number of questions and the random number of questions. In the adapted number of questions condition, the ECA used a distress detection model. This model continuously estimated the participant’s level of distress after each answer for a thought-correction question. If the model detects distress, it prompts the agent to ask a new question to correct the automatic thought. This process is repeated until the model detects no distress or the maximum number of questions is reached. In contrast, the condition of a random number of questions does not use the distress detection model. Instead, the system randomly determines the number of questions without considering the participant’s distress state. In both conditions, the number of questions asked to correct automatic thoughts varied from a minimum of 1 to a maximum of 21.

![Figure 4. Diagram of the experimental conditions. The adapted number of questions condition, which includes the psychological distress detection model, versus the random number of questions condition, which does not include the distress detection model.](image-url)

The condition assignment was not concealed from the experimental examiner. Participants were automatically assigned to conditions based on their age, sex, and K6 scores (a self-reported psychological distress scale). The day before the experiment, the participants responded to these 3 items using a questionnaire administered through Google Forms. On the basis of the questionnaire results, the conditions were divided such that no imbalance existed between the conditions in terms of K6, age, and sex. The participants conducted the experiment according to their predetermined condition assignments. They were unaware of their condition assignment and were not informed of the differences between the conditions.
Before the experiment, an experimenter explained this to the participants who signed the consent forms. They then completed the K6 questionnaire and read a publicly available leaflet explaining CBT [41]. The experiment consisted of the same questions as those in the ECA dialogue scenario created in our previous study [21]. The experiment was completed within 1 hour, including its briefing and preparation.

**Measures**

**Psychological Distress**

K6 was used as the rating scale to assess the psychological distress of our participants. Before the experiment, we obtained the K6 results for the previous 30 days. After the experiment, the participants were asked if any of the K6 items had changed after the CBT. The postexperimental measures differed from the original protocol, and we asked about the changes solely for analyzing our experiment. We calculated the change in psychological distress using equation 1.

\[
\text{Change in psychological distress} = \text{K6 score (pre)} - \text{K6 score (post)} \quad (1)
\]

Before the session, we measured the Quick Inventory of Depressive Symptomatology [42] (Japanese version [43]), which assesses depressive tendencies.

**Cognitive Change**

The Cognitive Change-Immediate scale [44] is a 5-item self-report measure that assesses the extent of the cognitive changes experienced by users during a single CBT session within therapy sessions. Items were rated on a scale from 0 (not at all) to 6 (completely), yielding a total score ranging from 0 to 30. Higher scores indicate a greater degree of cognitive change experienced by the participants.

**Mood Change**

A mood score is a numerical representation of the intensity of self-reported negative emotions experienced by the participants. The scores ranged from 0 to 100, with higher scores indicating more severe distress. Participants verbally provided their mood scores at 2 time points during scenarios involving CBT sessions with the ECA. The initial mood score was used for prethought correction, and the subsequent score was used for postthought correction. Specifically, participants were asked, “On a scale of 0 to 100, where 0 is no problem at all and 100 is a huge problem, how intense is that feeling?” These questions are commonly used to assess relief from negative moods [2,4]. Participants provided their responses verbally during the CBT. The participants’ moods were characterized using such language as anxiety, depression, sadness, feelings of inferiority, and fatigue. In this study, these negative emotional states were collectively assessed under an umbrella term, negative mood. In this study, these negative emotional states were collectively assessed under an umbrella term, negative mood.

This calculation is a measure of the degree of change in a participant’s negative mood, where a larger positive value indicates a decrease in negative mood.

**State–Trait Anxiety**

The State–Trait Anxiety Inventory [46] (Japanese version [47]) is a validated, self-report questionnaire that assesses both state anxiety (temporary and situational anxiety) and trait anxiety (general and stable anxiety) in adults. The State–Trait Anxiety Inventory consists of two 20-item scales: the State–Trait Anxiety Inventory-State (STAI-S) and the State–Trait Anxiety Inventory-Trait (STAI-T). The STAI-S measures the respondent’s current anxiety levels and feelings of apprehension, tension, and nervousness. It assesses the extent to which an individual is experiencing anxiety in response to a specific situation or event. Items are scored from 1 (not at all) to 4 (very much), where higher scores indicate more severe state anxiety. The STAI-T evaluates a respondent’s general tendency to experience anxiety as a stable personality characteristic. It focuses on the feelings of anxiety that are not associated with a particular event or situation; instead, it reflects an individual’s overall tendency to become anxious. Items are scored from 1 (almost never) to 4 (almost always), where higher scores represent greater trait anxiety levels. Both STAI-S and STAI-T scores can range from 20 to 80, and higher scores denote greater anxiety. We calculated the change in STAI-S using equation 3.

\[
\text{Change in STAI-S} = \text{STAI-S (pre)} - \text{STAI-S (post)} \quad (3)
\]

This calculation is a measure of the degree of change in a participant’s state anxiety, where a larger positive value indicates a decrease in state anxiety.

**Statistical Analysis**

We conducted an investigation to address hypothesis 1. This hypothesis states that an adaptive selection of the number of questions based on detected psychological distress would have superior effects in improving mental states. This approach was compared with the one in which the number of questions was randomized. For this investigation, we focused on several metrics, including K6 scores, mood scores, STAI-S, and cognitive change. Both within-condition and between-condition comparisons were conducted within these metrics. Owing to the inability to confirm the normal distribution of these measures in both the adapted number of questions condition and the random number of questions condition, we chose to use nonparametric tests for our analyses.

For within-condition comparisons, we assessed the differences in scores before and after the CBT session for each measurement in both conditions. The Wilcoxon signed-rank test was used to analyze these within-condition differences, illustrating the effect of the session on the adapted and random number of questions conditions individually.

For between-condition comparisons, we compared the pre- and postsession scores for each measurement between the adapted and random number of questions conditions. The Mann-Whitney U test was used for this analysis. In addition, we conducted a comprehensive analysis across 4 key parameters: change in psychological distress, cognitive change, mood change, and...
change in the STAI-S. This analysis involved multiple 1-sided Mann-Whitney \( U \) tests under the assumption that the adapted number of questions condition would show higher effects on each measure than the random number of questions condition.

We conducted a subgroup analysis to examine hypothesis 2, which asserts that adapting the number of questions according to the participant’s psychological distress is more effective. We categorized the participants into 3 subgroups based on the difference between the number of questions asked and the distress detection model’s estimation: the same group (participants who were asked the same number of questions as determined by the distress detection model), the fewer group (fewer questions were asked than determined by the distress detection model), and the more group (more questions were asked than determined by the distress detection model). In this analysis, we combined participants from both the adapted number of questions condition and the random number of questions condition. We did this because we expected too few participants in the same group within the random number of questions condition. We compared the same group with the fewer and more groups, expecting superior outcomes for the same group. This comparison used a 1-sided Mann-Whitney \( U \) test to compare the cognitive change, mood change, change in K6 scores, and change in the STAI-S across the subgroups.

In addition, we conducted further analysis to validate hypothesis 2. In this step, we examined the differences in various measurements across specific question counts. We focused on groups divided by question amount, such as 10, in the condition of a random number of questions. These were then compared with the participants in the adapted number of questions condition. Owing to the limited sample size in each category, we restricted our examination to qualitative analysis. In this comparative analysis, we did not conduct statistical testing. Furthermore, we reported the number of nondistressed and distressed values automatically detected for each individual question. These analyses aimed to demonstrate the effectiveness of a flexible approach to the number of questions presented.

Although we collected usability questionnaires and subjective evaluation feedback from the participants, this information was not included in the primary analysis of this study.

## Results

### Model Construction Results

Pretraining our detection models to address Japanese conversation yielded the following results: for the English data set (original condition and reproductive experiment), the accuracy was 0.72, precision was 0.67, recall was 0.58, and \( F_1 \)-score was 0.62. When the data set was translated by an automatic translation model into Japanese, the models achieved an accuracy of 0.64, precision of 0.59, recall of 0.60, and an \( F_1 \)-score of 0.59, demonstrating that we maintained equivalent performance levels despite the change in language.

We further applied transfer learning to the pretrained model with Japanese conversation for real-time distress detection in CBT. The results of implementing this transfer learning are as follows: In the condition where only pretraining was applied, the model achieved an average accuracy of 0.55, precision of 0.47, recall of 0.48, and an \( F_1 \)-score of 0.48 across 5 evaluation trials. When both pretraining and transfer learning were implemented, the average scores improved, with the model achieving an accuracy of 0.70, precision of 0.60, recall of 0.63, and an \( F_1 \)-score of 0.61. The model that demonstrated the highest performance during these trials, exhibiting an accuracy of 0.75, recall of 0.69, precision of 0.72, and an \( F_1 \)-score of 0.70, was selected for integration into the conversational scenario.

### Experimental Results

Table 2 presents the pre- and postsession scores for both the adapted and random number of questions conditions across various metrics.
Table 2. Pre- and postsession measures for the adapted and random number of questions conditions.

<table>
<thead>
<tr>
<th></th>
<th>Adapted number of questions condition (n=25)</th>
<th>Random number of questions condition (n=24)</th>
<th>Between-condition comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cliffs delta</td>
<td>P value</td>
<td></td>
</tr>
<tr>
<td>K6a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presession measures, mean (SD)</td>
<td>4.08 (4.31)</td>
<td>4.38 (3.10)</td>
<td>-0.07</td>
</tr>
<tr>
<td>Postsession measures, mean (SD)</td>
<td>3.04 (4.19)</td>
<td>4.30 (3.17)</td>
<td>-0.17</td>
</tr>
<tr>
<td>Within-condition comparison, effect size r (P value)</td>
<td>-0.49 (.01)</td>
<td>-0.09 (.65)</td>
<td>_b</td>
</tr>
<tr>
<td>Mood score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presession measures, mean (SD)</td>
<td>57.44 (22.30)</td>
<td>65.63 (20.76)</td>
<td>-0.10</td>
</tr>
<tr>
<td>Postsession measures, mean (SD)</td>
<td>46.12 (24.53)</td>
<td>57.92 (17.38)</td>
<td>-0.11</td>
</tr>
<tr>
<td>Within-condition comparison, effect size r (P value)</td>
<td>-0.66 (.001)</td>
<td>-0.44 (.03)</td>
<td>_b</td>
</tr>
<tr>
<td>STAI-Sd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presession measures, mean (SD)</td>
<td>41.40 (12.43)</td>
<td>40.75 (9.53)</td>
<td>-0.01</td>
</tr>
<tr>
<td>Postsession measures, mean (SD)</td>
<td>38.40 (10.46)</td>
<td>38.21 (10.27)</td>
<td>0.02</td>
</tr>
<tr>
<td>Within-condition comparison, effect size r (P value)</td>
<td>-0.29 (.14)</td>
<td>-0.39 (.05)</td>
<td>_b</td>
</tr>
<tr>
<td>QIDSd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presession measures, mean (SD)</td>
<td>5.00 (4.62)</td>
<td>4.83 (3.67)</td>
<td>-0.03</td>
</tr>
<tr>
<td>STAI-Td</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presession measures, mean (SD)</td>
<td>44.10 (12.5)</td>
<td>42.1 (10.8)</td>
<td>0.04</td>
</tr>
<tr>
<td>Time spent (second), mean (SD)</td>
<td>494.96 (161.10)</td>
<td>661.40 (289.20)</td>
<td>-0.10</td>
</tr>
<tr>
<td>Number of questions, mean (SD)</td>
<td>4.76 (4.48)</td>
<td>8.83 (6.52)</td>
<td>-0.17</td>
</tr>
</tbody>
</table>

aK6: Kessler Psychological Distress Scale.

bNot available.
cSTAI-S: State–Trait Anxiety Inventory-State.
dQIDS: Quick Inventory of Depressive Symptomatology.
eSTAI-T: State–Trait Anxiety Inventory-Trait.

The average time spent on the dialogues in the adapted number of questions condition was 494.96 (SD 161.10) seconds, whereas the random number of questions condition spent an average of 661.40 (SD 289.20) seconds. However, it should be noted that owing to system issues, we were unable to record the time spent by one participant in each condition. As a result, the reported mean and SD for time spent exclude the data from these 2 participants. All other measurements were successfully recorded for these participants.

The average number of questions asked to correct an automatic thought was 4.76 (SD 4.48) for the adapted number of questions condition and 8.83 (SD 6.52) for the random number of questions condition.

We evaluated the distress detection model’s performance in the adapted number of questions condition. Specifically, we evaluated its accuracy in correctly identifying the nondistressed states, indicating when to stop asking Socratic questions. The model’s performance was measured by comparing its output (0 denoting nondistressed) with the participants’ actual postexperiment psychological distress states. The model correctly identified nondistressed state in 72% (18/25) of the instances.

In the presession phase, the K6 scores did not significantly differ between the 2 conditions. However, a significant difference was observed in the postsession phase; the adapted number of questions condition recorded a notably lower K6 score, indicating a substantial reduction in distress compared with the random number of questions condition. On other measures, the pre- and postsession scores were not significantly different between the 2 conditions.

A closer look at the within-condition comparisons revealed that the adapted number of questions condition experienced a decrease in scores across several metrics following the session, albeit to varying degrees. The adapted number of questions condition showed more marked improvements, especially in the K6 and mood scores, with significant reductions after the session, as evidenced by Cliff delta values of −0.49 (P=.01) and −0.66 (P=.001), respectively. In contrast, although the random
number of questions condition also showed a reduction in the mood score, characterized by a medium effect size (Cliff delta=-0.44; \( P=0.03 \)), the change in the K6 score was negligible between the pre- and postsession periods.

Table 3 presents the comparative results of changes in each measurement. We conducted a 1-sided Mann-Whitney \( U \) test to compare the cognitive change, mood change, change in the K6 scores, and change in the STAI-S between the adapted number of questions condition and the random number of questions condition. The results indicated that there was no significant difference in the change in psychological distress, mood change, and change in STAI-S between the 2 conditions, with \( P \) values of .11, .16, and .54, respectively. However, for cognitive change, a relatively large effect size was observed with a Cliff delta of 0.12, and the \( P \) value was .07, showing a trend toward significance, suggesting that the adapted number of questions condition exhibited a tendency for greater cognitive change compared with the random number of questions condition.

Table 3. Comparative results of changes in each measurement using nonparametric analysis of 1-sided Mann-Whitney \( U \) test for differences between conditions.

<table>
<thead>
<tr>
<th></th>
<th>Adapted number of questions condition, mean (SD)</th>
<th>Random number of questions condition, mean (SD)</th>
<th>Cliff delta</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in psychological distress</td>
<td>1.04 (1.95)</td>
<td>0.08 (1.28)</td>
<td>0.09</td>
<td>.11</td>
</tr>
<tr>
<td>Cognitive change</td>
<td>16.3 (6.38)</td>
<td>13.7 (5.59)</td>
<td>0.12</td>
<td>.07</td>
</tr>
<tr>
<td>Mood change</td>
<td>0.22 (0.31)</td>
<td>−0.05 (0.88)</td>
<td>0.08</td>
<td>.16</td>
</tr>
<tr>
<td>Change in STAI-S(^a)</td>
<td>3.00 (8.10)</td>
<td>2.54 (5.99)</td>
<td>−0.009</td>
<td>.54</td>
</tr>
</tbody>
</table>

\( a \)STAI-S: State–Trait Anxiety Inventory-State.

To investigate hypothesis 2, we conducted a subgroup analysis as described in the Statistical Analysis subsection. We categorized the 49 participants from the adapted and random number of questions conditions based on the deviation between the differences in the questions asked and the estimated value of the model. Of the 21 participants, 7 (14%) received fewer questions than estimated and 14 (29%) received more questions than estimated, and for 28 (57%) participants, the number of questions matched the estimated value, denoted as the same group.

Table 4 shows the mean and SD for each subgroup. For cognitive changes, comparisons between the same and fewer groups showed a small effect size (Cliff delta=0.24; \( P=0.03 \)), indicating a significant difference. In contrast, comparisons between the same and more groups showed a negligible effect size (Cliff delta=0.04; \( P=.65 \)), indicating no significant difference. For the change in psychological distress, the comparisons between the same and fewer groups showed a negligible effect size (Cliff delta=0.03; \( P=.40 \)), indicating no significant difference. Similarly, comparisons between the same and more groups showed a negligible effect size (Cliff delta=0.04; \( P=.34 \)), indicating no significant difference. For mood changes, the comparisons between the same and fewer groups showed a negligible effect size (Cliff delta=0.08; \( P=.26 \)), indicating no significant difference. Similarly, the comparisons between the same and more groups showed a negligible effect size (Cliff delta=0.01; \( P=.44 \)), indicating no significant difference. Finally, for changes in the STAI-S, comparisons between the same and fewer groups showed a negligible effect size (Cliff delta=−0.03; \( P=.46 \)), indicating no significant difference. The comparisons between the same and more groups also showed a negligible effect size (Cliff delta=−0.07; \( P=.77 \)), indicating no significant difference.

Table 4. Mean and SD in the same, more, and fewer groups.

<table>
<thead>
<tr>
<th></th>
<th>Same, mean (SD)</th>
<th>Fewer, mean (SD)</th>
<th>More, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in psychological distress</td>
<td>0.86 (1.93)</td>
<td>−0.14 (1.78)</td>
<td>0.36 (1.01)</td>
</tr>
<tr>
<td>Cognitive change</td>
<td>15.57 (6.51)</td>
<td>10.71 (4.79)</td>
<td>16.07 (5.14)</td>
</tr>
<tr>
<td>Mood change</td>
<td>0.20 (0.30)</td>
<td>−0.45 (1.57)</td>
<td>0.13 (0.32)</td>
</tr>
<tr>
<td>Change in STAI-S(^a)</td>
<td>2.53 (7.81)</td>
<td>1.86 (6.04)</td>
<td>3.71 (6.32)</td>
</tr>
</tbody>
</table>

\( a \)STAI-S: State–Trait Anxiety Inventory-State.

Table 5 illustrates the mean changes in various measurements such as distress, cognitive change, and mood change across different numbers of questions. As indicated in Table 4, the mean changes for the adapted number of questions condition were 1.04 for change in psychological distress, 16.3 for cognitive change, 0.22 for mood change, and 3.00 for change in STAI-S. The bar graphs of these data can be found in Multimedia Appendix 2. For the random number of questions condition, it was observed that in most instances, the mean change was lower compared with the overall average of the adapted number of questions condition. However, a notable exception was observed at a question count close to the average question count for the adapted number of questions condition (mean 4.76, SD 4.48), specifically at 5 questions. At this count, the random number of questions condition exhibited...
slightly higher scores in both distress change and cognitive change compared with the adapted number of questions condition. Furthermore, a trend was observed in cognitive change, where an increase in the number of questions correlated with a larger change, indicating a potential area for further exploration and validation in subsequent studies. Despite these observations, it was consistently found that the adapted number of questions condition manifested superior effectiveness across various numbers of questions, indicating robustness in its application regardless of the question count.

Table 5. Mean change in each questionnaire across different numbers of questions posed to participants in the random number of questions condition. Instances where the mean is “not available” indicate that the number of participants (n) is 0. Instances where the standard SD is “not available” are due to the number of participants (n) being either 0 or 1.

<table>
<thead>
<tr>
<th>Number of questions</th>
<th>Change in distress, mean (SD)</th>
<th>Mood change, mean (SD)</th>
<th>Cognitive change, mean (SD)</th>
<th>Change in STAI-S&lt;sup&gt;a&lt;/sup&gt;, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.67 (0.58)</td>
<td>0.21 (0.08)</td>
<td>12.67 (2.08)</td>
<td>1.33 (4.51)</td>
</tr>
<tr>
<td>2</td>
<td>-1.00 (1.41)</td>
<td>-0.03 (0.04)</td>
<td>4.50 (0.71)</td>
<td>4.00 (5.66)</td>
</tr>
<tr>
<td>3</td>
<td>0.50 (0.71)</td>
<td>0.08 (0.12)</td>
<td>11.00 (9.90)</td>
<td>4.00 (0.00)</td>
</tr>
<tr>
<td>4</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>12.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>&lt;b&gt;-5.00 (&lt;b&gt;—&lt;/b&gt;)&lt;/b&gt;</td>
</tr>
<tr>
<td>5</td>
<td>1.50 (2.12)</td>
<td>0.10 (0.14)</td>
<td>18.00 (2.83)</td>
<td>1.50 (9.19)</td>
</tr>
<tr>
<td>6</td>
<td>0.50 (0.71)</td>
<td>0.13 (0.19)</td>
<td>7.00 (8.49)</td>
<td>5.00 (8.49)</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>0.17 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>13.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>-4.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>-4.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>11.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>-9.00 (&lt;b&gt;—&lt;/b&gt;)</td>
</tr>
<tr>
<td>10</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>11</td>
<td>1.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>0.83 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>19.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>5.00 (&lt;b&gt;—&lt;/b&gt;)</td>
</tr>
<tr>
<td>12</td>
<td>0.50 (0.71)</td>
<td>0.20 (0.00)</td>
<td>19.50 (3.54)</td>
<td>3.50 (3.54)</td>
</tr>
<tr>
<td>13</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>14</td>
<td>-1.00 (0.00)</td>
<td>-0.33 (0.47)</td>
<td>16.50 (0.71)</td>
<td>-0.50 (6.36)</td>
</tr>
<tr>
<td>15</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>16</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>17</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>15.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>-1.00 (&lt;b&gt;—&lt;/b&gt;)</td>
</tr>
<tr>
<td>18</td>
<td>0.50 (0.71)</td>
<td>0.21 (0.06)</td>
<td>15.50 (4.95)</td>
<td>12.50 (9.19)</td>
</tr>
<tr>
<td>19</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>17.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>4.00 (&lt;b&gt;—&lt;/b&gt;)</td>
</tr>
<tr>
<td>20</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>21</td>
<td>0.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>0.43 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>20.00 (&lt;b&gt;—&lt;/b&gt;)</td>
<td>2.00 (&lt;b&gt;—&lt;/b&gt;)</td>
</tr>
</tbody>
</table>

<sup>a</sup>STAI-S: State–Trait Anxiety Inventory–State.

<sup>b</sup>Not available.

Table 6 displays the number of detected values classified as nondistressed and distressed for each question aimed at correcting automatic thoughts. The bar graphs of these data can be found in Multimedia Appendix 3. In the adaptive question amount condition, questioning concluded with the detection of a participant’s first nondistressed value. Conversely, the process continued regardless of the values detected in the random question amount condition. Consequently, the total number of participants decreased as the number of questions increased. A significant proportion of participants were identified with nondistressed values, particularly within the initial 4 questions. However, as the number of questions reached a certain amount, the frequency of nondistressed value detection decreased, with none identified from questions 16 to 21. Meanwhile, distressed value detection persisted, with at least 1 individual exhibiting distressed values through the 21st question.
Table 6. Number of nondistressed and distressed values detected for each question aimed at correcting automatic thoughts, including only those participants who were presented with the same or a greater number of questions.

<table>
<thead>
<tr>
<th>Number of questions</th>
<th>Number of destressed values</th>
<th>Number of nondestressed values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

The primary aim of this study was to evaluate a method that adapts the number of questions in a CBT session based on the user’s level of distress, aiming to enhance the session’s effectiveness.

Hypothesis 1 posited that an adaptive approach to the number of questions selection, based on detected psychological distress, would lead to superior cognitive change and reduce psychological distress as well as negative emotional states compared with the random number of questions condition. The results partially supported this hypothesis. We observed a reduction in psychological distress from before the session to after the session in the adapted number of questions condition, whereas no such reduction was evident in the random number of questions condition. Furthermore, the level of distress after the session was significantly lower in the adapted number of questions condition than in the random number of questions condition. In contrast, no significant difference was observed in the changes in each measurement between the 2 conditions. These findings overall indicate that our approach contributed to optimizing the health care process and its outcomes.

A plausible reason for the significant improvement observed only in distress is that the deep learning model incorporated in this study was designed to detect the presence or absence of psychological distress. This suggests that although the CBT system’s adjustment of the number of questions was adjusted for distress, it was not necessarily adjusted for other measures. In cognitive models, cognition, including automatic thoughts, and reactions such as distress are considered distinct entities [2]. Relevant studies [19,44,45] explore the ways in which cognitive shifts can influence stress levels. Moving forward, it might be necessary to analyze the content of CBT and propose methods that enhance both cognitive and distress changes.

Hypothesis 2 proposed that better improvement would occur when the number of questions corresponded with the model’s estimated value. The data partially supported this, indicating more significant cognitive changes when the number of questions was aligned with the model’s estimate compared with when fewer questions were asked. This insight suggests the importance of precision in the number of questions selected to enhance the effectiveness of CBT delivered by ECAs. It was also implied that a large number of questions does not necessarily work effectively. Table 6 (Multimedia Appendix 3) shows that none of the participants were detected as nondistressed from questions 16 to 21. This result suggests that further questioning might not lead to significant improvements when there is no improvement in distress after a certain number of questions. Therefore, it would be prudent to set an upper limit on the number of questions that the system can ask.
We personalized the therapeutic process by ensuring a timely adjustment of the number of questions when the distress alleviation was projected to be insufficient. Our experimental data suggest that this approach significantly bolsters the effectiveness of ECAs in integrating thought-correction techniques into the therapy process. The data also revealed that maintaining a balance in the number of questions was critical for improving CBT’s effectiveness. Our findings highlight the potential of dynamic and personalized strategies in enhancing mental health care.

**Comparison With Prior Work**

The field of mental health care has significantly evolved owing to advancements in artificial intelligence. Among these advancements, using ECAs to deliver mental health care services is particularly promising. Influential studies in this domain, including the work of Fitzpatrick et al [10], DeVault et al [48], and Fulmer et al [49], investigate the effectiveness of these text-based conversational agents or ECAs in mental health care. Furthermore, research by Inkster et al [11], Ghandeharioun et al [50], and Murali et al [51] has revealed the importance of these agents’ ability to convey empathy.

These empathetic elements are crucial for creating a therapeutic alliance and supportive mental health care. Among these advancements, using ECAs to deliver mental health care services is particularly promising. Influential studies in this domain, including the work of Fitzpatrick et al [10], DeVault et al [48], and Fulmer et al [49], investigate the effectiveness of these text-based conversational agents or ECAs in mental health care. Furthermore, research by Inkster et al [11], Ghandeharioun et al [50], and Murali et al [51] has revealed the importance of these agents’ ability to convey empathy.

In addition, research endeavors by Kimani et al [14] and Shidara et al [21] initiated the exploration of the application of Socratic questioning in CBT delivered by ECAs. These studies have shown promising results in fostering cognitive changes and reducing distress. However, they have not explored the dynamic adjustment of Socratic questioning based on users’ psychological states. This study aims to address this gap.

The integration of such adaptive strategies in mental health care, as demonstrated by our study, underscores the substantial contribution of our research in advancing the field. Our study introduces a novel component that modulates the number of questions according to the degree of the user’s psychological distress. The results of our study indicate that adapting to a user’s psychological distress by modulating the number of questions significantly reduced psychological distress compared with a random number of questions. This finding is in alignment with psychiatric insights suggesting that the modification of automatic thoughts through questioning may not yield sufficient effectiveness if such thoughts remain superficial [2].

**Limitations**

When interpreting the results of this study, several limitations must be addressed. First, the cross-sectional design used in our research did not capture the long-term effects of ECAs on psychological distress. A long-term experiment is important to gain a deeper understanding of the lasting effects of ECAs and their potential role in improving mental health outcomes. Another limitation of our study is the small number of participants who exhibited high depressive tendencies. To enhance the generalizability of our findings, future research should involve a more diverse sample of participants, covering a broad range of psychological distress levels. This sample should include individuals with varying degrees of depression and other mental health concerns, ensuring that the results are more widely applicable to different populations that are experiencing various psychological issues. In addition, further research is needed on the effective selection of the number of Socratic questions to maximize the effectiveness of CBT. Techniques such as the WOZ method, in which a therapist operates an ECA, might shed light on this aspect [52].

**Conclusions**

Our study provides evidence that adjusting the selection of questions based on an individual’s distress levels can significantly enhance CBT effectiveness. This adjusted approach allows for a more personalized health care, which can improve the therapeutic outcomes for individuals who are struggling with mental health issues, including anxiety and depression.

Our research highlights the importance of timely and appropriate reactions when an individual’s distress levels improve during therapy. By carefully monitoring and responding to these changes in distress, the ECAs can better support users’ progress. Overall, our study highlights the value of personalized and adaptive approaches in CBT, paving the way for more effective and responsive mental health care.

**Acknowledgments**

This work was funded by the Japan Science and Technology Agency Core Research for Evolutional Science and Technology (grant JPMJCR19M5).

**Data Availability**

The data sets generated during and analyzed during this study are not publicly available because of privacy and personal information protection concerns, but they are available from the corresponding author on reasonable request.

**Conflicts of Interest**

None declared.
Multimedia Appendix 1
Conversational scenario of an embodied conversational agent.

[DOCX File, 16 KB - formative_v8i1e50056_app1.docx]

Multimedia Appendix 2
Mean change across different numbers of questions posed to participants in the random number of questions condition. The blue bars represent the variation in mean change for measurements: (A) change in distress, (B) mood change, (C) cognitive change, and (D) change in State–Trait Anxiety Inventory-State (STAI-S). The green dashed line serves as a reference for comparison, indicating the average change in the adapted number of questions condition.

[PDF File, 248 KB - formative_v8i1e50056_app2.png]

Multimedia Appendix 3
Number of detected values for each question aimed at correcting automatic thoughts: (A) nondistressed and (B) distressed values, including only those participants who were presented with the same or a greater number of questions.

[PDF File, 68 KB - formative_v8i1e50056_app3.png]

References


41. Ono Y. Leaflet: using cognitive behavior therapy to improve mental skills. Woman Wave Co. URL: https://www.cbtjp.net/downloads/skillup/ [accessed 2023-06-03]


Abbreviations

CBT: cognitive behavioral therapy
ECA: embodied conversational agent
DAIC-WOZ: Distress Analysis Interview Corpus/Wizard-of-Oz
K6: Kessler Psychological Distress Scale
STAI-S: State–Trait Anxiety Inventory-State
STAI-T: State–Trait Anxiety Inventory-Trait
A Closed-Loop Digital Health Tool to Improve Depression Care in Multiple Sclerosis: Iterative Design and Cross-Sectional Pilot Randomized Controlled Trial and its Impact on Depression Care

Kyra Henderson¹, BA; Jennifer Reihm¹, MPA; Kanishka Koshal³, MSPH; Jaelene Wijangco¹, BS; Narender Sara¹, MBA, MS; Nicolette Miller¹, BS; Marianne Doyle¹, LCSW; Alicia Mallory¹, BA, BSN, RN; Judith Sheridan¹, PhD; Chu-Yueh Guo¹, MD; Lauren Oommen¹, DO; Katherine P Rankin¹, PhD; Stephan Sanders¹, PhD; Anthony Feinstein², MPhil, PhD; Christina Mangurian³, MAS, MD; Riley Bove¹, MMSc, MD

¹Department of Neurology, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States
²Department of Psychiatry, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, Canada
³Department of Psychiatry, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Riley Bove, MMSc, MD
Department of Neurology
Weill Institute for Neurosciences
University of California, San Francisco
1651 4th Street
San Francisco, CA, 94158
United States
Phone: 1 4155027209
Email: riley.bove@ucsf.edu

Abstract

Background: People living with multiple sclerosis (MS) face a higher likelihood of being diagnosed with a depressive disorder than the general population. Although many low-cost screening tools and evidence-based interventions exist, depression in people living with MS is underreported, underascertained by clinicians, and undertreated.

Objective: This study aims to design a closed-loop tool to improve depression care for these patients. It would support regular depression screening, tie into the point of care, and support shared decision-making and comprehensive follow-up. After an initial development phase, this study involved a proof-of-concept pilot randomized controlled trial (RCT) validation phase and a detailed human-centered design (HCD) phase.

Methods: During the initial development phase, the technological infrastructure of a clinician-facing point-of-care clinical dashboard for MS management (BRIDGE) was leveraged to incorporate features that would support depression screening and comprehensive care (Care Technology to Ascertain, Treat, and Engage the Community to Heal Depression in people living with MS [MS CATCH]). This linked a patient survey, in-basket messages, and a clinician dashboard. During the pilot RCT phase, a convenience sample of 50 adults with MS was recruited from a single MS center with 9-item Patient Health Questionnaire scores of 5-19 (mild to moderately severe depression). During the routine MS visit, their clinicians were either asked or not to use MS CATCH to review their scores and care outcomes were collected. During the HCD phase, the MS CATCH components were iteratively modified based on feedback from stakeholders: people living with MS, MS clinicians, and interprofessional experts.

Results: MS CATCH links 3 features designed to support mood reporting and ascertainment, comprehensive evidence-based management, and clinician and patient self-management behaviors likely to lead to sustained depression relief. In the pilot RCT (n=50 visits), visits in which the clinician was randomized to use MS CATCH had more notes documenting a discussion of depressive symptoms than those in which MS CATCH was not used (75% vs 34.6%; $\chi^2 = 8.2; P = .004$). During the HCD phase, 45 people living with MS, clinicians, and other experts participated in the design and refinement. The final testing round included 20 people living with MS and 10 clinicians including 5 not affiliated with our health system. Most scoring targets for likeability and usability, including perceived ease of use and perceived effectiveness, were met. Net Promoter Scale was 50 for patients and 40 for clinicians.
Conclusions: Created with extensive stakeholder feedback, MS CATCH is a closed-loop system aimed to increase communication about depression between people living with MS and their clinicians, and ultimately improve depression care. The pilot findings showed evidence of enhanced communication. Stakeholders also advised on trial design features of a full year long Department of Defense–funded feasibility and efficacy trial, which is now underway.

Trial Registration: ClinicalTrials.gov NCT05865405; http://tinyurl.com/4zkvr9x

(JMIR Form Res 2024;8:e52809) doi:10.2196/52809

KEYWORDS
depression; quality of life; bring your own device; mHealth; closed-loop; clinical trial; multiple sclerosis

Introduction

Background

Approximately 50% of people with multiple sclerosis (MS) have a depressive disorder, and people living with MS are 2-3 times more likely to be diagnosed with a depressive disorder than the general population [1]. However, depression in MS remains underreported, underevaluated [2], and undertreated [3,4], despite the prevalence of depression in people living with MS, low-cost tools available to screen for depression, and evidence-proven, society-recommended pharmacologic and nonpharmacologic treatment modalities [5,6]. In clinical practice, common barriers to these clinical goals include time constraints during clinical visits (Nelson, M, unpublished data, August 2020), stigma and discomfort discussing psychiatric symptoms [7], insufficient antidepressant medication dose, duration or mechanism of action [8], inadequate attention to other MS symptoms (fatigue, cognitive impairment, and urinary retention) that could interfere with mood and worsen symptom burden [2,9], and patient difficulty following treatments owing to difficulties with access, insurance, or finding specialists close to home [10], as well as the many competing demands on their time. Solutions to address these gaps in reporting, screening, and treatment are needed.

A closed-loop intervention could close these gaps in care and represent a pragmatic approach to address the suboptimal treatment of depression in real-world MS settings. Closed-loop interventions in health care refer to systems that minimize gaps in communication between patients and clinicians; these have been implemented effectively for a number of medical conditions [11-13]. For the purpose of improving depression care in MS, such an intervention would support relevant clinical information flow from patients to clinicians at the point of care and back to patients to support patient-centered care. Such a system would not focus on a specific “one size fits all” treatment or intervention modality (eg, social intervention) or on a specific proprietary mood app, but could support developing a complex plan individualized for each patient’s symptoms, goals, and capability. To accomplish this, the tool should efficiently deliver patient-reported mood symptoms to clinicians in line with the “5 rights” [14] (right information, to the right person, in the right format, through the right channel, and at the right time in the workflow). Further, the tool must promote the behaviors (eg, reporting, screening, treatment recommendations, and following through with timely refills or referral scheduling) that are likely to lead to mood improvements. Finally, this tool should seek to streamline workflows and streamline care within an existing care team (neurologist, nurse, and interprofessional staff) and system of care (informatics, care delivery, and payment structure). Once developed, even after extensive multidisciplinary stakeholder input, a digital health solution must be thoroughly socialized within a health system to increase adoption.

Some prior studies support the feasibility of monitoring patient mood longitudinally in people living with MS as well as the possible effectiveness of a closed-loop approach. There is good concordance between patient reports and clinical depression, supporting the use of patient-reported tools for depression monitoring [15]. The 3-month pilot CoachMS randomized controlled trial (RCT; NCT03335618; n=21; people living with MS; [16]) showed that it was feasible, and acceptable, to monitor patients' bothersome symptoms (mood, ambulation, and bladder) and to act on them clinically in near real time (coaching patients to address these) [16]. This intervention showed some preliminary efficacy in supporting behavioral change [17,18] likely to address depression, including in one case, recognition of and urgent hospitalization for suicidality. A proposed closed-loop intervention could further deliver data on patient function directly to the clinician at the point of care to support effective response.

Objectives

This study details a 3-phase endeavor designed to develop a comprehensive, closed-loop system for monitoring and thoroughly treating depression to be tested in real-world clinical settings: MS CATCH (Care Technology to Ascertain, Treat, and Engage the Community to Heal Depression in people living with MS). In phase I, a prototype was developed. In phase II, it was piloted in a clinical setting to obtain preliminary cross-sectional efficacy data on a fundamental premise, namely that visualization of patient mood during the clinical encounter would improve clinical attention to mood during that visit. In phase III, having ensured that the tool was satisfactory to patients and clinicians and that its use did indeed improve attention to depression in the pilot study, the tool was then refined using an extensive process of human-centered design (HCD) [19,20].

Methods

Phase 1: Initial Technical Build

Overview

Members of the primary research team led earlier engagement efforts that informed MS CATCH development, namely,
creating the BRIDGE dashboard concept [21], testing the acceptability of patient-reported outcomes (PROs) in a point-of-care MS clinical dashboard (MS NeuroShare; [19]), and developing a closed-loop reporting system that collects and synthesizes data at the point of care (MS Falls Insight Track; MS-FIT); [22]). Each of these design processes informed the current technical build.

**BRIDGE**

BRIDGE [21] is a technologically scalable, institutionally approved, workflow friendly, cross-disease, modular precision medicine platform [23]. BRIDGE launches from within a patient’s encounter from the Epic electronic health record (EHR) at the University of California, San Francisco (UCSF) using industry standard integration (Figure 1; Table S1 in Multimedia Appendix 1 [24]), delivering a seamless experience for clinicians [21]. BRIDGE was designed using an extensive process of HCD [19,20]. The tool accesses and visualizes data that are currently available in various disparate sources (EHR, patient diaries, and publicly accessible websites), with the substantial innovation that they are presented here in one comprehensive and streamlined format. A disease-specific version of BRIDGE is live within multiple clinics at the UCSF, with the MS clinic BRIDGE used as a point-of-care dashboard for the current project.

**Figure 1.** Components of the MS CATCH (Care Technology to Ascertain, Treat, and Engage the Community to Heal Depression in people living with multiple sclerosis) tool, including the (A) patient-facing survey and log, (B) the clinician-facing dashboard, and (C) the closed-loop system, with in-basket alerts for the clinician.

**MS NeuroShare**

This closed-loop prototype designed for a collaborative health system integrated a patient-facing PRO app with an EHR-based clinician-facing dashboard [25]. Patient participants cited the perceived value of thinking about and recording information before their appointments, noting how it impacted discussions with their clinicians, adding that coviewing information with the physician put physicians and patients on the same page and promoted a conversation of equals. Clinicians perceived patients’ prospectively collected mood scores as more sensitive, accurate, and comprehensive than a patient’s recall during the visit, and felt that these could improve value and promote clinician engagement [25].

**MS-FIT Tool**

A closed-loop system to prospectively record, report, and prevent falls in people living with MS was designed using HCD in a process analogous to that intended for MS CATCH [22]. MS-FIT includes low-burden regular falls ascertainment that patients can access with one click from any computer, tablet, or device, triggering a clinician’s inbox message for new or serious falls. Clinicians can then access and launch, from the medical record, a comprehensive version of BRIDGE refined
MS CATCH Prototype

The prototype for MS CATCH used in this study was refined for the specific purposes of depression monitoring and treatment by the study team using the BRIDGE and MS-FIT technical scaffolding, and informed by the MS NeuroShare and MS-FIT design processes and experiences. The MS CATCH prototype used in this study is shown in Figure S1 in Multimedia Appendix 1.

Phase 2: Proof-of-Concept Pilot Randomized Controlled Trial

Research Setting and Participants

Design

This was a cross-sectional, randomized controlled pilot study.

Setting

The primary clinical setting was the UCSF Center for MS and Neuroinflammation, which specializes in providing care to >5000 adults with MS annually, and has an extensive track record of pivotal trials for MS [26,27], remote monitoring [28,29], and treatment [16].

Participants

MS clinicians were invited to participate in the study during the center’s monthly research meeting, and a convenience cohort of 6 clinicians was enrolled using a signed electronic consent form. Then, from the participating neurologists’ practices, adult patients with an MS diagnosis scheduled for upcoming in-person and video neurology visits were contacted before the clinic visit by the study coordinator via email and phone call. Interested patients were scheduled for an enrollment visit whether in-person, or via telephone, or Zoom (Zoom Video Communications, Inc; an institutionally approved, Health Insurance Portability and Accountability Act–secure, televideo platform). During this visit, they provided detailed informed consent via DocuSign (DocuSign, Inc). Once patients signed the informed consent form, they were asked to complete a series of PROs, including the 9-item Patient Health Questionnaire (PHQ-9) scale, via a secure REDCap (Research Electronic Data Capture; Vanderbilt University) link. Patients with scores of ≥5 (mild depression: 5-9, moderate: 10-14, moderately severe: 15-19, and severe: 20-27) met the study criteria and were seen in the clinic. For participants with PHQ-9 scores of ≥19, even if their clinician was not asked to use BRIDGE, the clinician was informed of their score near the conclusion of the clinic visit to ensure safe treatment of the patient. At the conclusion of the study visit, both clinician and patient feedback regarding the tool were solicited. During the proof of concept pilot, MS CATCH was used once during the visit in real time. For the full RCT, the patient-facing mood survey will be completed monthly, as described in the results for phase III.

Trial Outcomes

Usability: Participant Feedback on Tool

At the end of the study visit, patients were asked to provide feedback on the tool. There were 2 main prompted questions: “Tell us (with illustrative examples) of what worked well and what didn’t, with regards to BRIDGE, during your appointment.” This question was asked to both patients and clinicians, whereas “Please explain in detail, why it was or wasn’t useful for you to report your mood symptoms before the visit” was just addressed to the patients. Then, 2 questions informed from the System Usability Scale [30,31] and the health IT usability evaluation model [32] were completed, assessing likeability and perceived usefulness of the tool.

Efficacy: Effect of the Tool on Attention to Depression

At the end of the study, each clinical encounter was reviewed by a team member with an eye to whether mood was mentioned during the clinical visit. After the review was completed by one team member, another team member blinded to the treatment assignment audited the report. Only the second team member was blinded to treatment assignments. Any disagreements were resolved by a third party.

Ethical Considerations

This study was approved by the UCSF institutional review board (#18-26148).

Statistical Analysis

The participant demographics and usability outcomes were analyzed using descriptive statistics. The main efficacy outcome, percentage visits where depression was mentioned in the clinic notes, was compared for visits with and without the use of MS CATCH using chi-square analysis [33].

Phase 3: MS CATCH Tool Refinement (Stakeholder Engagement and Intended User Input Over 6 Months)

Study Design

HCD is a process that holds at its center the needs of the intended users. During the MS CATCH refinement phase, the overall goal was to fine-tune and optimize the intervention by partnering with patient and clinician users and multidisciplinary stakeholders in a sequence of iterative feedback sessions to inform and validate design decisions (Figure 2). To our knowledge, features assembled into the MS CATCH prototype

were reminded to launch BRIDGE via text or email before the patient appointment when applicable, and neither clinician nor patient participant were blinded to the intervention. Patients continued to be recruited until a target (n=50) had met the inclusion criteria and were seen in the clinic. For participants with PHQ-9 scores of ≥50, even if their clinician was not asked to use BRIDGE, the clinician was informed of their score near the conclusion of the clinic visit to ensure safe treatment of the patient. At the conclusion of the study visit, both clinician and patient feedback regarding the tool were solicited. During the proof of concept pilot, MS CATCH was used once during the visit in real time. For the full RCT, the patient-facing mood survey will be completed monthly, as described in the results for phase III.
had never before been available at the point of care in a clinically actionable format for depression treatment in MS. It was important to verify that the workflow and displays promoted behavioral change likely to improve depression care and to simplify the display into the features most likely to drive adoption of the tool and support depression care. Here, the “capability, opportunity, and motivation model for behavioral change” (COM-B) [18] was used. COM-B is a hub in the center of Michie’s behavioral change wheel [18] and can help inform a patient’s intention to engage in a planned behavior, which is considered the best predictor of that behavior [34,35]. The COM-B approach, and individual intervention functions, have been successfully applied to facilitate behavioral change in people living with MS [17].

**Figure 2.** The sequence of iterative feedback and human-centered design (HCD) development leading to tool optimization over 9 months, including activities, participants, and outcomes. COM-B: capability, opportunity, and motivation model for behavioral change; Health IT USES: Health IT Usability Evaluation Scale; MS: multiple sclerosis; MS CATCH: Care Technology to Ascertain, Treat, and Engage the Community to Heal Depression in people living with multiple sclerosis UCSF: University of California, San Francisco.

<table>
<thead>
<tr>
<th>Group</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholder Advisory Group</td>
<td>Meeting 1 at study kickoff: Refine project goals, inform patient and clinician engagement protocols, and advise on potential clinical, technological, or ethical challenges.</td>
</tr>
<tr>
<td></td>
<td>Meeting 2 when tool is ready for clinical trial: Review the tool optimization and feedback scores (see below) and confirm that the intervention is ready for launch in the trial.</td>
</tr>
<tr>
<td></td>
<td>Meeting 3 following trial start: Review enrollment and early qualitative feedback on clinical trial activated and develop a roadmap for the transition plan.</td>
</tr>
<tr>
<td>Discovery Interviews</td>
<td><strong>Process:</strong> Members of the primary research team with expertise in engagement and HCD conducted 1:1 interviews with patients and clinicians separately, via the UCSD* Zoom platform. The sessions were recorded to ensure that all feedback was captured.</td>
</tr>
<tr>
<td></td>
<td><strong>Goals:</strong> To identify new features and map all features and data types to COM-B principles to facilitate behavioral outcomes (reporting, evaluating, treating) [15] likely to support depression reduction.</td>
</tr>
<tr>
<td></td>
<td><strong>Outcomes:</strong> Feedback was used to generate a list of features to be modified (e.g., enhanced, clarified, removed, and simplified) by the developer. Clarifying feedback was also elicited on perceived ease of use and usefulness to ensure that only the components most likely to be effective would be used.</td>
</tr>
<tr>
<td>Think aloud Session</td>
<td><strong>Process:</strong> Members of the primary research team conducted 1:1 sessions where the tool was launched and tested with patients and clinicians. Patients were presented with specific scenarios for entering data into the platform (e.g., using the tool after a difficult life event vs for monitoring during periods of emotional stability) and the way they engaged with the tool was observed. Clinicians were exposed to interactive live demonstrations.</td>
</tr>
<tr>
<td></td>
<td><strong>Goals:</strong> Validate modifications made following discovery interviews and further refine and clarify each feature.</td>
</tr>
<tr>
<td></td>
<td><strong>Outcomes:</strong> Both patients and clinicians were asked to prioritize the individual data types by scoring their “perceived usefulness,” so that only the key drivers of behavioral changes were retained. If any specific data scored low (i.e., not perceived as useful), then they were either removed or modified in the visualization.</td>
</tr>
<tr>
<td>Final Tool Scoring</td>
<td><strong>Process:</strong> The final version of MS CATCH was scored by 20 diverse patients with MS and 10 MS clinicians using established scales such as the SUS, Health IT USES, and NPS.</td>
</tr>
<tr>
<td></td>
<td><strong>Goals:</strong> Assess clinician perception of usability with the System Usability Scale (SUS), Assess patient and clinician perceptions of usefulness and ease of use with Health IT USES-based questions, and perceptions of likeability with a single-item question and the Net Promoter Score (NPS). [26, 27, 28]</td>
</tr>
<tr>
<td></td>
<td><strong>Outcomes:</strong> Any domains scoring below target as defined in Table 3 would trigger another round of iterative edits, with input from the Stakeholder Advisory Group.</td>
</tr>
</tbody>
</table>

To maximize the likelihood that the tool would be adopted and effective, it was evaluated using the health IT usability evaluation model [32]. This model integrates multiple usability theories including the Technology Acceptance Model and evaluates both subjective and objective outcomes. Although refining the critical data and visualization elements, as well as technological and clinical workflow aspects, the 4 key variables proposed by Mathews et al [36] for digital health tool validation were examined to determine whether the tool (1) reflects HCD principles and (2) is likely to engage patients. These factors include usability, effectiveness, learnability or ease of use, and likeability. Usability was defined using the System Usability Scale, learnability and ease of use were measured with a subset of the Health Information Technology Usability Evaluation Scale–based questions, and likeability was assessed with a single Likert scale question: “Do you like the tool?” and the net promoter score (NPS). Effectiveness was determined based on the pilot study visit notes outcomes.

**Sequential Methods With Iterative Technological Modifications**

Over a 9-month period between September and June, sequential activities were conducted as outlined in Figure 2.
**Stakeholder Advisory Group**

This group was convened and led by members of the primary research team. The primary research team comprised investigators with a track record of collaboration as well as expertise in both scientific and technological aspects of this project, including MS clinical trials (RB), digital tools to evaluate and treat cognition and mood in MS [16] (RB and AF), psychiatric care for underserved populations and innovations in treatment of psychiatric conditions in diverse settings (CM) [37-44], implementation science (CM), statistics (Ann Lazar), patient engagement (RB and JR) [20,21,25], HCD of digital tools (JR, RB, NS, and NM) [20,21,25], and launching an institutionally and technologically sophisticated, scalable, cross-disease platform from the EHR at UCSF (RB and BRIDGE team). The stakeholder advisory group included the research team listed above, as well as a patient champion JS, patient advocacy group leader Linda Glassel, National Multiple Sclerosis Society, Northern California Chapter President, MS nurse expert AM, Registered Nurse, social worker MD licensed independent clinical social worker, and MS neurologist CYG.

**Participants**

MS clinicians and other experts (degrees including Doctor of Medicine, Nurse Practitioner, Registered Nurse, and Masters in Social Work) were identified within the UCSF MS clinic and the investigators’ broader professional network. Adults with MS were recruited from the UCSF MS center by the study team (convenience sampling). All participants provided informed consent to test the tool and provide preliminary data. Furthermore, 31% (10/32) of the patients participated in the phase II pilot study.

**Ethical Considerations**

Activities during this phase (HCD phase) were approved by the UCSF institutional review board (#22-36620).

**Results**

**Phase 1: Initial Technical Build**

The MS CATCH prototype used for the current phase II and phase III study phases, informed by prior HCD processes (BRIDGE, MS NeuroShare, and MS-FIT), is shown in Figure S1 in Multimedia Appendix 1. A patient-facing mood survey was selected to be administered via a secure REDCap link that could be accessed on any device with Wi-Fi or cellular data capabilities. The validated, self-administered PHQ-9 scale was then connected to the BRIDGE platform so that it could be visualized at the point of care and interpreted in BRIDGE in addition to the other elements of the patient’s MS history, as well as a list of resources including psychotherapy and psychiatry, available in the patient’s home area (California only). The PHQ-9 survey also triggered a clinician’s inbox message for worsening mood. The mechanism by which MS CATCH is postulated to work is by facilitating communication flows that are likely to support intended users’ COM-B behaviors (patient: reporting and follow through and clinician: evaluating and recommending) likely to lead to symptom improvement. With regard to usability, the intended users, that is, both clinicians and patients, can each access it with one click.

**Phase 2: Proof-of-Concept Pilot Randomized Controlled Trial**

**Participants**

Of the 200 patients contacted, 106 (53%) agreed to participate in this study. Of these, 50 (47.2%) met the inclusion criteria of a PHQ-9 score >4 and were cared for by 6 participating physicians. Distribution of PHQ-9 scores was as follows: 52% (26/50) mild depression (5-9); 32% (16/50) moderate depression (10-14); 10% (5/50) moderately severe depression (15-19); and 6% (3/50) severe depression (20-27). Patients were then randomized to visits where the BRIDGE dashboard was used (24/50, 48%) or not used (26/50, 52%). All participants were contacted before their clinical visit, which was conducted between May and December, 2021. This visit was where the MS CATCH intervention was used. The cohort included 40 (80%) females and 10 (20%) males, aged 25-75 (mean 50.46, SD 13.0) years, including individuals with all MS subtypes (32 relapsing remitting MS, 9 primary progressive MS, 1 progressive relapsing MS, 6 secondary progressive MS, 1 unspecified, and 1 MS likely).

**Tool Usability**

The mean likeability score was 4.3/5 for patients (93% of the patient participants agreed or strongly agreed that they liked the tool) and 4.4/5 for clinicians (100% of the clinicians agreed or strongly agreed that they liked the tool).

For perceived usefulness [30,31], among clinicians, 100% of them reported that the tool increased their attention to the patient’s mood, particularly for patients with mild to moderate symptoms that might otherwise have been missed. Clinicians provided some specific feedback on how to further customize the dashboard, suggesting that it would be better to have the categorical components of the PHQ-9 rather than just the raw score (3/6, 50%).

Among patients, 71% (20/28) reported feeling that it was useful to report their mood-related symptoms before the visit. In qualitative feedback, many patients reported that filling out the mood survey ahead of time gave them the opportunity to reflect on their own feelings: “It set the tone, upfront, about how I was feeling.” “It’s helpful for me to think about assessing my mood in a structured way,” “was useful because it made me stop and reflect,” and “it opened the door to explain how...new MS symptoms are affecting my mood.”

**Tool Effectiveness**

For the patients with a PHQ-9 score of ≥5, visit notes documented a discussion of depressive symptoms in 75% of the visits where MS CATCH was used versus 35% of the visits where it was not used (N=50; χ² = 8.2; P = .004; Table 1). One patient was referred to a psychiatrist for suicidal ideation.
### Phase 3: MS CATCH Tool Refinement

#### Study Participants

Overall, 45 individuals provided feedback on the tool during successive phases of development. The results are summarized in Table 2.

<table>
<thead>
<tr>
<th>Phase (cohort, round #)</th>
<th>Sample size</th>
<th>Age range(^a) (years)</th>
<th>Sex (females), n (%)</th>
<th>Clinical context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery interviews: Patients</td>
<td>5</td>
<td>30-57.6</td>
<td>3 (60)</td>
<td>MS(^b)</td>
</tr>
<tr>
<td>Discovery interviews: Clinicians</td>
<td>5</td>
<td>N/A(^c)</td>
<td>4 (80)</td>
<td>MS (MD(^d), NP(^e), and MSW(^f))</td>
</tr>
<tr>
<td>Think-alouds: patients</td>
<td>7</td>
<td>30.4-66</td>
<td>6 (86)</td>
<td>MS</td>
</tr>
<tr>
<td>Think-alouds combined with final tool testing: Clinicians</td>
<td>Think-alouds: 5; tool testing: 10</td>
<td>N/A</td>
<td>6 (80)</td>
<td>MS</td>
</tr>
<tr>
<td>Final tool testing: Patients</td>
<td>20</td>
<td>29-63</td>
<td>15 (88)</td>
<td>MS</td>
</tr>
</tbody>
</table>

\(^a\)If applicable.

\(^b\)MS: multiple sclerosis.

\(^c\)N/A: not applicable.

\(^d\)MD: Doctor of Medicine.

\(^e\)NP: Nurse Practitioner.

\(^f\)MSW: Masters in Social Work.

#### Discovery and Think-Aloud Sessions: Selected Findings

Detailed findings and scores from each round of engagement, as well as iterative development steps, are presented in Multimedia Appendix 1 and summarized here. During the discovery interviews (N=5 patients, N=5 clinicians), each tool component was found to be useful and usable. A number of features and data types were identified that mapped to the COM-B principles of behavior change, and these resulted in design changes and interventions promoted to address potential boosters or blockers (Table S2 in Multimedia Appendix 1). For example, patients universally felt that the PHQ-9 should be delivered monthly if not more often, despite concerns by the study team that it would be too burdensome. The patient-facing survey was edited to support free-text entry and addition of language at end of survey informing patients who scored >15 to seek care urgently, as their care team may not see the score for 36 hours, and instructing patients to seek immediate care if experiencing a mental health crisis. During the think-aloud sessions (N=7 patients, N=5 clinicians), some patients articulated concern for how answers might be interpreted by their clinicians, for example, concern about sending up unnecessary “red flags” with their survey answers as symptoms addressed on the survey can overlap with MS symptoms. Relatedly, the free text box proved to be a popular design feature as it presents an opportunity to offer context for potentially concerning or MS symptom-related survey answers. Additional features were noted to be of value; for example, neighborhood walkability was specifically identified by patients as useful context for their clinicians to have within the clinical decision support view. Patients also universally expressed interest in having dashboard views available in a summary display for reference following appointments. All surveyed participants preferred an electronic summary format distributed through the patient portal (MyChart) as opposed to an alternative format that was not integrated with their EHR portal. Clinician feedback led to a reduction in the visual “wordiness” of the original clinical decision support widget. One decision that could not be accommodated was the ability to include patient insurance information, which substantially influences prescribing choices and referral decisions, as there is unfortunately no “ground truth” source for this information.

#### Final Changes Made to Tool Before Scoring

The patient survey was further developed to include an in-survey pop-up directing patients who screened positive for suicidal intent to seek immediate medical care (with contact information), a catalog of mental health resources at the end,
and a companion patient-facing mood tracker displaying the patient’s longitudinal mood survey results. On the dashboard side, the clinical decision support screen was further refined to be less “visually overwhelming.”

**Qualitative Feedback on the Closed-Loop System**
Clinicians’ qualitative feedback on the potential impact of MS CATCH on the care delivery experience and quality of conversations with patients and caregivers ranged from appreciation for the intervention’s holistic approach to concern that limited visit times will constrain the dashboards’ seemingly limitless potential.

- “Patients will feel empowered to make decisions using the information.”
- “The value comes from physicians being able to see the patient’s trajectory, and from patients being able to have a longitudinal view.”
- “It will allow me to come into visits with more of a plan.”
- “While [BRIDGE] may not be useful for every problem, it can be valuable for high stakes symptoms (depression, falls).”
- “Feel like it gives me more to offer my patients.”
- “Reinforces what we should do.”
- “Helps to have an objective measure of symptom worsening. Is more reliable than past notes.”

**Description of the Final MS CATCH Prototype**
The final prototype is displayed in Figure 1. The patient-facing survey is easily accessible via email. Along with receiving the short monthly survey via email, the patient’s response history and resources relevant to depression self-management can be viewed at the end of the survey (Figure S2 in Multimedia Appendix 1). PHQ-9 scores of 15 or higher or any suicidal intent trigger an alert to the patient’s MS clinician’s EHR in-basket. This is a critical feature allowing for timely follow-up within the usual care workflow, including an urgent referral to psychiatry. The clinician-facing dashboard (Figure 1) has a comprehensive mood evaluation and treatment dashboard that launches from the EHR. This dashboard can be copied and pasted into the clinical notes as well as the patient’s after visit summary to be accessed at the end of the visit and between visits. The tool uses a closed loop system, diagramed in Figure 1, which integrates the patient survey, in-basket messaging, and clinical dashboard tools.

**Final Tool Scoring**
A convenience sample of adults with MS (n=20) and clinicians (n=10) participated in the final tool testing. Of the participating clinicians, 5 (50%) practiced at external (non–UCSF-affiliated) institutions, and 5 (50%) were affiliated with UCSF. The target and achieved results are presented in Table 3.

Patient assessment ratings exceeded the goals in approximately all categories. Despite a patient NPS of 50, which is considered one point below excellent, slightly fewer (75%) than the goal of 80% agreed or strongly agreed that the tool was likeable (mean 4.25, SD 1.12). Of those (n=5) who scored the tool ≤4 with regard to likeability, approximately all (n=4, 80%) were neutral (score=3). Reasons varied from lack of personal applicability, “Mood symptoms are due to MS, not depression,” and “I have a lot of anxiety, but not depression”; to survey design, “I don’t like the multiple-choice options”; and to formatting, “Font is small.” Concrete follow-on actions taken by the development team in response to this feedback were as follows:

- Increasing the default survey font size, in addition to giving patients the option to adjust the font independently.
- Adding a note above the free text box indicating it can be used to add comments or clarify answers.
- Coaching physicians on how to communicate with patients about the survey, including how it can help detect anxiety and mood symptoms that may or may not be MS-related.

Similarly, clinician ratings exceeded the goals in approximately all categories. A clinician NPS of 40 indicated strong favorability in terms of recommending to peers. Perceived usefulness was slightly under goal at 95%. Although all but 1 clinician agreed or strongly agreed that the tool was useful (mean 4.35, SD 0.75) and 1 rated the tool ≤4 citing concerns about limited bandwidth and lack of financial incentives or reimbursable mechanisms for clinicians to address mood between clinical encounters. This feedback echoes concerns noted during discovery interviews about visit time limitations and the need for complementary workflows to ensure that patient mood issues can be monitored in a timely fashion, with the assistance of multidisciplinary staff (registered nurse and masters in social work) and elevated to physicians as needed.
Psychiatric and nonpsychiatric care in the EHR is often low, and better integration is associated with improved psychiatric care [51]. In this study, we showed that viewing patients’ self-reported mood at the point of care improved attention to depression in the clinical notes. This of course does not inform follow through on this intention. To that end, our 3-month pilot CoachMS RCT (NCT03335618; 21 people living with MS) showed that monitoring and responding to patients in real time [16] was feasible, acceptable, and showed preliminary effects on behavioral change [17,18] suggesting that sustained monitoring and feedback can promote behavioral change [16].

After development, implementation into clinical workflows represents the final goal of health-related tools. From prior experience in developing platforms for the clinical encounter, it was clear that socialization of a tool is a critical step. For patients, knowing that information about their mood will tie back to their care team and translate into improved attention to their care is an important component of tool use, as is the ability to self-monitor using the survey results. Furthermore, direct actionable steps were implemented based on patient feedback, such as providing a list of resources at the end of the survey to reduce the activation energy required to access resources during a depressive episode. Similarly, clinicians are more likely to use and recommend a tool if they know that patients will have completed their surveys [25] and are more likely to want access to patient-generated data if it is clinically actionable. Integrating comprehensive management resources will support individualization of the treatment plan rather than “one size fits all” treatment interventions.

Many treatment trials for depression have to date focused on “one size fits all” approaches testing a specific intervention or combination of interventions. Novel approaches tested within

---

### Table 3. Final tool scoring by patients (N=20) and clinicians (N=10): Likeability and Usability. Prespecified targets and achieved results are presented; bolded results are at or above target.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Patients (N=20)</th>
<th>Clinicians (N=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Likeability</strong></td>
<td>Target Result</td>
<td>Target Result</td>
</tr>
<tr>
<td>Single question (“Do you like this tool?”) graded on a Likert scale(^a); metric: score (&gt;4) (agree or strongly agree)</td>
<td>80% 75%</td>
<td>80% 100%</td>
</tr>
<tr>
<td><strong>Likeability</strong></td>
<td>Good: (&gt;0); favorable: (&gt;20): excellent: (&gt;50)</td>
<td>Good: (&gt;0); favorable: (&gt;20): excellent: (&gt;50)</td>
</tr>
<tr>
<td>Net promoter score (NPS)(^b); “how likely are you to recommend this tool to another patient with MS/clinician?” [46]</td>
<td>78 (SD 11.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td>SUS(^c); a rapid, valid, scalable industry standard, reliable with small sample sizes</td>
<td>—</td>
</tr>
<tr>
<td><strong>Perceived usefulness</strong></td>
<td>Health ITUES(^d)-based questions for perceived usefulness [30,31]; metric: score (\geq 4) (agree or strongly agree)</td>
<td>80% 88%</td>
</tr>
<tr>
<td><strong>Perceived ease of use</strong></td>
<td>Health ITUES-based questions for perceived ease of use [30,31]; metric: score (\geq 4) (agree or strongly agree)</td>
<td>80% 97%</td>
</tr>
</tbody>
</table>

\(^a\)Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

\(^b\)NPS responses (0-10 scale) were calculated by subtracting the percentage of detractors (those who scored 0-6) from the percentage of promoters (those who scored 9 or 10).

\(^c\)SUS: System Usability Scale.

\(^d\)Not available.

\(^e\)Health ITUES: Health IT Usability Evaluation Scale.

---

### Discussion

Depression is a serious comorbidity in MS that is associated with an increased risk of disability progression [47], yet it remains severely undertreated; by one estimate, two-thirds of individuals meeting the clinical criteria for major depressive disorder receive no antidepressant medications [48]. Depression is often missed altogether by neurologists, even when it is severe and accompanied by suicidal ideation [49,50]. We iteratively and extensively piloted and developed a closed-loop intervention designed to improve reporting, screening, and treatment of depression in a cohort of clinical patients with MS. The preliminary data presented here support our key premise underlying this closed-loop approach [11] that delivering reliable information about a patient’s mood in an interpretable, visual format to the clinician would increase the likelihood that depression was detected (including, in one case, suicidality) and addressed in the clinical encounter. Further, our robust process of HCD and tool evaluation allowed us to maximize the likelihood that this tool would be adopted by ensuring that it met the stringent goals of being likeable, easy to use, and perceived as being effective.

We are not aware of other closed-loop tools developed to support depression care in MS that are integrated with the EHR. Distinct from other mood apps on the market, the patient self-report app directly connects to the clinician’s in-basket so that the information can be efficiently streamlined. In full clinical use, the patient-facing survey would be distributed once per month, and clinicians will not be reminded to launch BRIDGE. Studies have generally shown that integration between psychiatric and nonpsychiatric care in the EHR is often low, and comprehensive management resources will support individualization of the treatment plan rather than “one size fits all” treatment interventions.

Many treatment trials for depression have to date focused on “one size fits all” approaches testing a specific intervention or combination of interventions. Novel approaches tested within...
MS include sequential blocks of medications [52] as well as combinations of behavioral and pharmacological interventions [53]. Extending beyond the usual screening and care components (such as asking about moods retrospectively), MS CATCH will comprehensively visualize the other functions contributing to depression, including other conditions (eg, sleep) and unmet social needs (eg, substance use counseling, neighborhood unsafe for walking exercise). These will support customized action prompts to help patients self-manage, and clinicians will tailor treatments, anticipate challenges, and make social and behavioral health referrals.

Some limitations were evident in the design and compromises had to be made. For example, the PHQ-9, which is widely adopted as a screening tool, may not be particularly specific. For example, in response to the “constant bother” question, one participant responded “MS is a constant bother.” However, this tool was selected because of its brevity, extensive validation across multiple conditions including MS, and because it comprised the 9 criteria that are identified in the diagnosis of depressive disorder per the DSM-IV [45]. Another limitation is the lack of information regarding insurance acceptance by mental health professionals. Unfortunately, this limits actionability in terms of referrals to mental health professionals in the United States. For the resources map, we chose to focus on mental health professionals near the patient who were considered to have expertise for individuals with MS, but we also included a link to a website that includes insurance information to allow cross-checking. A third challenge is scalability. Key technological and clinical features were intentionally selected to ensure that the tool’s modular infrastructure could be scaled to other symptoms, conditions, and clinical settings. Technological factors designed to support scalability include (1) the quality and content of static visualizations that can be disseminated broadly regardless of a clinic’s technology and (2) optimizing the industry technological standards used for the build so that the code can be shared by clinicians in other health settings (eg, other MS centers using Epic EHR). However, the integration of the build into other health systems ultimately depends on governance and motivation internal to that system.

MS CATCH is a comprehensive, low-burden, closed-loop platform designed to reduce depression severity and prevalence in people living with MS by supporting real-time communication and alerts, shared decision-making, and action prompts for comprehensive, personalized interventions. The research findings summarized here suggest that MS CATCH scores well on likeability and usefulness scales and that it improves attention to mood in individual clinic visits. In a complementary pilot study, we previously showed that remote monitoring of mood can lead to timely intervention [16]. Altogether, these findings support the results of a planned randomized single-center clinical trial evaluating its effect on behaviors that support depression reporting, ascertainment, and care. A similar model can be used to improve other clinical symptoms in MS and for managing other chronic medical conditions with high prevalence of depression.

Acknowledgments
This trial is funded by a Department of Defense US Army Medical Research Acquisition Activity Clinical Trials Award (project #W81XWH2210894; principal investigator: RB).

Authors’ Contributions
KH drafted the manuscript. KK and JW collected preliminary data and revised the manuscript. JR was involved in trial design, conducted interview and interpreted results, and revised the manuscript. AF, CM, MD, AM, CYG, and LO contributed expertise from their respective fields through stakeholder feedback sessions and other avenues and revised the manuscript. NS and NM were involved in tool iterative development and revised the manuscript. RB was involved in trial design, obtained study funding, and drafted the manuscript. KPR and SS revised the manuscript.

Conflicts of Interest
The authors report no disclosures relevant to this clinical trial. KH, JR, KK, JW, NS, NM, MD, AM, JS, LO, and KPR had no conflicts of interest to disclose. CYG reports personal advisory board or consulting fees from Genentech and Horizon. AF received grant support from Multiple Sclerosis (MS) Society of Canada, Canadian Institutes of Health Research, book royalties from Johns Hopkins University Press, and speaker’s honoraria from Novartis. CM is supported by several grants from the National Institutes of Health (NIH; National Institute of Mental Health, National Institute on Minority Health and Health Disparities, National Institute of Allergy and Infectious Diseases, and National Institute on Drug Abuse), Health Resources and Services Administration, Doris Duke Charitable Foundation, California Health Care Foundation, Genentech, Gilead, and United Health Group. RB received research support from NIH, Department of Defense, National Multiple Sclerosis Harry Weaver Award, Biogen, Novartis, and Hoffman-La Roche. RB has received personal advisory board or consulting fees from EMD Serono, Horizon, Janssen, and TG Therapeutics. SS receives research funding from BioMarin Pharmaceutical.

Multimedia Appendix 1
Supplementary tables and figures.
[PDF File (Adobe PDF File), 649 KB - formative_v8i1e52809_app1.pdf ]
References


45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]


Abbreviations

COM-B: capability, opportunity, and motivation model for behavioral change
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
EHR: electronic health record
HCD: human-centered design
MS CATCH: Care Technology to Ascertain, Treat, and Engage the Community to Heal Depression in people living with multiple sclerosis
MS: multiple sclerosis
NPS: net promoter score
PHQ-9: 9-item Patient Health Questionnaire
PRO: patient-reported outcome
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
UCSF: University of California, San Francisco

Edited by A Mavragani; submitted 18.09.23; peer-reviewed by J Simard; comments to author 17.10.23; revised version received 27.10.23; accepted 24.11.23; published 15.03.24.

Please cite as:
A Closed-Loop Digital Health Tool to Improve Depression Care in Multiple Sclerosis: Iterative Design and Cross-Sectional Pilot Randomized Controlled Trial and its Impact on Depression Care
JMIR Form Res 2024;8:e52809
URL: https://formative.jmir.org/2024/1/e52809
doi:10.2196/52809
PMID:38488827
Assessing a GPS-Based 6-Minute Walk Test for People With Persistent Pain: Validation Study

Joshua Simmich1,2, PhD; Nicole Emma Andrews1,2,3,4, PhD; Andrew Claus3,5, PhD; Megan Murdoch3,6, BPhty(Hons); Trevor Glen Russell1,2, PhD

1RECOVER Injury Research Centre, Faculty of Health and Behavioural Sciences, The University of Queensland, Brisbane, Australia
2STARS Education and Research Alliance, Surgical Treatment and Rehabilitation Service, The University of Queensland and Metro North Health, Brisbane, Australia
3The Tess Cramond Pain and Research Centre, Royal Brisbane and Women's Hospital, Metro North Hospital and Health Service, Brisbane, Australia
4Occupational Therapy Department, Royal Brisbane and Women's Hospital, Metro North Hospital and Health Service, Brisbane, Australia
5School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia
6Physiotherapy Department, Royal Brisbane and Women's Hospital, Metro North Hospital and Health Service, Brisbane, Australia

Abstract

Background: The 6-minute walk test (6MWT) is a commonly used method to assess the exercise capacity of people with many health conditions, including persistent pain. However, it is conventionally performed with in-person supervision in a hospital or clinic, therefore requiring staff resources. It may also be difficult when in-person supervision is unavailable, such as during the COVID-19 pandemic, or when the person is geographically remote. A potential solution to these issues could be to use GPS to measure walking distance.

Objective: The primary aim of this study was to assess the validity of a GPS-based smartphone app to measure walking distance as an alternative to the conventional 6MWT in a population with persistent pain. The secondary aim of this study was to estimate the difference between the pain evoked by the 2 test methods.

Methods: People with persistent pain (N=36) were recruited to complete a conventional 6MWT on a 30-m shuttle track and a 6MWT assessed by a smartphone app using GPS, performed on outdoor walking circuits. Tests were performed in random order, separated by a 15-minute rest. The 95% limits of agreement were calculated using the Bland-Altman method, with a specified maximum allowable difference of 100 m. Pain was assessed using an 11-point numerical rating scale before and after each walk test.

Results: The mean 6-minute walk distance measured by the GPS-based smartphone app was 13.2 (SD 46; 95% CI −2.7 to 29.1) m higher than that assessed in the conventional manner. The 95% limits of agreement were 103.9 (95% CI −107.7 to −71.1) m and −77.6 (95% CI −107.7 to −61) m, which exceeded the maximum allowable difference. Pain increased in the conventional walk test by 1.1 (SD 1.0) points, whereas pain increased in the app test by 0.8 (SD 1.4) points.

Conclusions: In individuals with persistent pain, the 2 methods of assessing the 6MWT may not be interchangeable due to limited validity. Potential reasons for the differences between the 2 methods might be attributed to the variation in track layout (shuttle track vs continuous circuit); poor GPS accuracy; deviations from the 30-m shuttle track; human variability in walking speed; and the potential impact of a first test on the second test due to fatigue, pain provocation, or a learning effect. Future research is needed to improve the accuracy of the GPS-based approach. Despite its limitations, the GPS-based 6MWT may still...
have value as a tool for remote monitoring that could allow individuals with persistent pain to self-administer frequent assessments of their functional capacity in their home environment.

*JMIR Form Res 2024;8:e46820*  doi: 10.2196/46820

**KEYWORDS**

GPS; mobile apps; exercise test; pain; chronic pain; mobile phone

**Introduction**

**Background**

Persistent pain affects 20% of the Australian population, has fueled the opioid epidemic, and cost the Australian government Aus $73.2 billion (US $54.7 billion) in 2018, the latest year for which figures are available [1,2]. People with persistent pain report that their pain negatively affects their ability to exercise and their physical functioning [3].

The 6-minute walk test (6MWT) is a standard method for measuring submaximal exercise capacity, which measures how far a person can walk over 6 minutes (6-minute walk distance; 6MWD). Conventionally, the 6MWT involves the participant walking multiple laps of a straight path that is 30 m in length [4]. Track length can influence the results of the 6MWT [5-8], likely due to a difference in the number of 180-degree turns, each necessitating acceleration and deceleration. Standard instructions are given and standard encouragement prompts are provided at preset intervals, as it is known that changes in instructions or encouragement can affect the performance during this test [9,10].

The 6MWT has been used in a range of conditions including chronic heart failure [11], peripheral artery disease [12], and chronic obstructive pulmonary disease [13]. It has also been used for older adults more generally [14]. Within the context of pain conditions, the 6MWT has been used as an outcome measure in both observational studies [15] and clinical trials [16,17]. The average 6MWD for people with persistent pain upon entry into pain programs has been reported to be 389 (SD 94) m [18] or 427 (SD 127) m [19]. In comparison, the reference values of the 6MWD for healthy men and women aged between 50 and 60 years are approximately 578 m and 534 m, respectively [20].

The minimum clinically important difference (MCID) for the 6MWD has been estimated for various conditions: 25 to 80 m in people with chronic obstructive pulmonary disease [21-23]; 22 to 42 m in people with lung cancer [24]; 33 m in people with pulmonary arterial hypertension [25]; 36 m in people with chronic heart failure [26]; 25 m in people with coronary artery disease [27]; and 18 m in older adults at risk of falls [28]. However, estimates in the population with persistent pain are less well defined, with different studies providing a range of estimates from a distance between 15 m and 30 m used in a clinical trial [16] to a distance between 156 m and 167 m calculated in fibromyalgia (a persistent pain disorder) [29]. Benaim et al [19] used an anchor-based approach to estimate MCID values of 60 m for chronic musculoskeletal pain of the spine and 75 m for chronic musculoskeletal pain of the lower limb. On the other hand, a recent study was unable to calculate an MCID in people with persistent pain undergoing a pain rehabilitation program due to a lack of correlation between the 6MWD and patient-reported outcome measures [18]. Therefore, further studies are needed to determine a more precise and consistent estimate of MCID values for the 6MWD in the population with persistent pain.

The conventional 6MWT is typically performed in a clinic or hospital. However, this is a significant limitation if patients cannot attend the facility due to distance or lack of access to transport. In addition, it can be difficult for some facilities to locate a straight section of corridor that is at least 30 m in length and largely free of obstacles and interruptions from other users of the corridor. The 6MWT can be performed outdoors, as an outdoor 6MWT appears to be comparable to an indoor 6MWT when the standard 30-m track is used [30], although outdoor testing can be limited by weather at the time of the test. The 6MWT can also be conducted at a participant’s residence, but this requires the clinician to travel to the patient, which increases the requirement for staff time and resources. In addition, locating a suitable 30-m track inside a house or in a backyard can be difficult. Holland et al [31] attempted to perform the 6MWT using the longest length of track that was practicable in the home environment but achieved an average track length of just 13 (SD 7) m for indoor tests and 20 (SD 10) m for outdoor tests. As such, the 6MWD estimates obtained from 6MWTs conducted inside or outside a home underestimated the 6MWD by an average of 30 (SD 69) m compared to the standard in-clinic tests using a 30-m track [31]. The variability in track length also contributed to wide limits of agreement (LoAs) between a conventional 6MWT conducted at a hospital and a 6MWT conducted at home (95% LoA 167 m-102 m). Finally, any method of performing the 6MWT that requires a clinician to assess the participant in person may be precluded if in-person contact needs to be limited, such as occurred worldwide during the COVID-19 pandemic. This has highlighted the need for alternative methods of performing the 6MWT.

Performing the 6MWT remotely would eliminate the problems associated with travel or in-person contact. Several studies have reported the use of accelerometers in smartphones or consumer-level wearable activity monitors to perform 6MWTs by detecting the number of steps or number of turns [32-38]. For instance, in a study conducted among healthy adults, the mean difference between accelerometer-based measurement and simultaneous clinician observation of a conventional 6MWT was 0, although the SD was 42 m [32]. However, this has not been replicated in clinical populations, and the app is not publicly available. Other accelerometer studies required their participants to use a holster or harness to affix their smartphone to a specific location on their body [38] or to wear a wrist-worn activity tracking device [33], both of which are potential barriers...
to use or uptake in a home environment. Furthermore, the 6MWT assessed via an app would ideally be undertaken on a 30-m-long track to best replicate the clinical setting, but it is unlikely that people with persistent pain would have access to such a length of track or the equipment to precisely measure its length in the home environment.

A potential solution to these issues could be to use GPS to measure the walking distance [39]. Modern smartphones include GPS receivers, and GPS-based walk tests can be conducted during an outdoor walk over level terrain (given this walk is free from the obstructions of tall buildings and does not involve many sharp turns). Although changing the 6MWT from the conventional shuttle to a continuous walk introduces some bias due to the removal of acceleration and deceleration when turning around, this might be less than the bias that would occur from using very short track lengths. For instance, in a population with respiratory problems, when compared to a 6MWT performed on a 30-m shuttle track, a 6MWT performed on a continuous circular track of 40-m circumference overestimates the 6MWD by 3% [40], but decreasing a shuttle track to 10 m underestimates the 6MWD by approximately 9% to 10% [8]. Furthermore, it may be possible to use an algorithm to compensate for the overestimation associated with continuous walking if it is consistent, but this is unlikely for the variable error that is associated with using tracks of varying and imprecisely measured lengths in a home environment.

Salvi et al [36] recently developed an app, called Timed Walk, which could use either an indoor algorithm using smartphone sensors or an outdoor algorithm using GPS. To the best of our knowledge, this is the only purpose-built app for conducting a 6MWT that is available on both Android and Apple iOS smartphones and the only app that uses GPS. This app has been tested in a population with pulmonary arterial hypertension in 2 studies that estimated the bias (the average difference between the app’s measurement and a reference) and the variability of the differences (given as SD of the differences). In the first study, Salvi et al [36] used a distance wheel as the reference measurement and found a larger bias but smaller variability when using the indoor algorithm (mean −2.0, SD 7.8 m) than when using the outdoor algorithm (mean −0.80, SD 18.6 m). However, in their later pilot trial, the indoor algorithm differed by 14.6 (SD 75) m when compared to the simultaneous performance of a conventional 6MWT in a clinic. This difference and SD is greater than that of the outdoor algorithm (mean 2.5, SD 47 m) compared to a clinic-based test performed within a 7-day period (before or after) [41]. The authors believe that much of this variation can be attributed to the app being used incorrectly, such as using the outdoor test on a tightly curved path. However, these studies compared the outdoor GPS mode of the app to a distance wheel [36] or to a conventional 6MWT performed within a 7-day period (before or after) [41]. Thus, interday variation in submaximal exercise capacity could also have contributed to the variation between these measurements. A comparison between the GPS-based test function of this app and a conventional 6MWT performed on the same day has not been attempted. In addition, this app has not been validated in the population with persistent pain.

**Aims and Hypothesis**

The primary aim of this study was to evaluate the concurrent validity of the GPS-based 6MWT using the Timed Walk smartphone app in an outdoor setting and conducted on the same day as the conventional 6MWT using a 30-m straight path. The secondary aim of this study was to estimate the difference between the pain evoked by the 2 test methods.

It was hypothesized that the agreement between the results of the 2 methods will be within the nominated maximum allowable difference and that the 2 methods will demonstrate an excellent level of correlation.

**Methods**

This was an observational study comparing 2 methods of estimating the 6MWD: a conventional 6MWT with a clinician and a 6MWT using a smartphone app with GPS.

**Participants**

The sample comprised a combination of participants recruited from 2 sources, to ensure a broad sample of people with persistent pain. The first recruitment method involved outpatients from a persistent pain clinic located in a large, tertiary public hospital in Australia. Potential participants were invited to participate in the study by their treating clinician. The second method of recruitment involved advertising the study via social media (Facebook and Twitter).

Participants were screened via an in-person or telephone conversation with a member of the research team. Individuals were eligible for inclusion if they were aged >18 years, had persistent pain (>9 months in duration), claimed that they were able to walk at least 100 m on flat ground, and owned a compatible smartphone on which they were willing to install the Timed Walk app. Individuals were excluded from participating if they had comorbidities impacting their ability to walk on flat ground (and did not have medical clearance to participate), were unable to speak English, or did not possess a compatible smartphone.

**Data Collection**

Participants performed 2 6MWTs: one in the conventional manner with a clinician and another with the Timed Walk app (version 0.2.0 or 0.3.0, depending on the date of the test).

The app test was conducted outdoors to use the GPS signal. Participants were instructed to walk either in a circuit around the paths of a local park or around the perimeter of a hockey playing field for the app test. Care was taken to select a mostly level path and to avoid sharp turns. As walking alongside a researcher walking at a comfortable distance behind each participant to provide supervision. The Timed Walk app automatically announced the standard encouragement prompts at appropriate times.

The 6MWT assessed in the conventional manner was completed in accordance with existing guidelines [4], using a straight track of 30 m in length, with the ends of the track marked with cones and the starting line marked with a clearly visible line on the
ground. A researcher (JS) conducted all the tests and provided the standard encouragement prompts.

For practical reasons and to minimize differences between the testing methods, the conventional 6MWT was also conducted outdoors at the same location as the GPS test. Outdoor testing has been shown to be no different to indoor testing in individuals with pulmonary disease [30]. Outdoor testing was only performed when weather conditions were mild (temperature of 10 °C-30 °C, no rain, and wind speed <20 km/h).

To account for potential fatigue or learning effects, the sequence in which the participants performed each form of the 6MWT was randomized on the day of testing. Urn design randomization was used [43]. Participants were randomized using a simulated urn initially containing 2 balls, with 1 ball representing each test method. After drawing a ball to allocate each participant to their starting test method, the ball was returned to the urn and an additional ball added for the opposite test method. To further minimize the effect of fatigue, participants were rested in a shaded location for 15 minutes between tests. Participants were instructed to wear comfortable clothing and appropriate footwear, to be well hydrated (and bring a water bottle), and to not perform any vigorous exercise 2 hours before and after the test.

Neither the participants nor the researcher performing the conventional 6MWT were blinded to the study hypothesis or to the methods of assessment being received.

**Outcome Measures**

The primary outcome of this study was the 95% LoAs between conventionally measured and GPS-measured 6MWD.

Secondary outcomes included the following: (1) the intraclass correlation coefficient (ICC) between the conventionally measured and GPS-based 6MWD and (2) the difference between the change in pain for the 2 types of 6MWTs. Pain intensity data were collected at four time points: (1) immediately before the first 6MWT, (2) immediately after the first 6MWT, (3) immediately before the second 6MWT, and (4) immediately after the second 6MWT. Pain was assessed on an 11-point numerical rating scale (NRS), verbally delivered to the participant at the abovementioned time points. This scale ranged from 0 (indicating no pain at all) to 10 (indicating pain as bad as it could be or worst pain).

Additional demographic data collected included participant age, gender, employment status, primary diagnosis (including pain source or body region), and years since diagnosis of the persistent pain condition. These data were collected using a paper-based questionnaire on the day of testing.

**Statistical Analysis**

The Bland-Altman method of analysis was used to estimate the 95% LoAs between the conventionally measured and GPS-measured 6MWD. The exact parametric 95% CIs for the upper and lower limits of the 95% LoAs, considered together as pairs, were calculated using the methods proposed by Carkeet [44]. The maximum allowable difference between measures was set at 100 m, approximately in between the 60-m to 75-m site undergoing occupational rehabilitation [19] and 156-m to 167-m MCID in people with fibromyalgia [29].

Outliers were excluded from the analysis if the difference between the 2 test methods was >3 SDs from the mean difference.

Assumptions of normality and homoscedasticity were assessed through visual inspection of quantile-quantile plots and residual plots. The ICC was calculated using a 2-way random-effects model for both absolute agreement and consistency, with values >0.8 considered to be good and those >0.9 considered to be excellent. The difference between the mean change in pain due to each test was presented with paired-sample CIs.

Several exploratory analyses that were not initially planned were conducted after the results for the main outcome measures were inspected. The exploratory analysis of the variance between recruitment groups was conducted using an F test. Welch 2-sample CIs were used for the exploratory comparisons between the 2 smartphone operating systems (Android and iOS) and the 2 sequences in which the tests were conducted (in-person test conducted first and app test conducted first). Paired-sample CIs were used to present the results of an exploratory comparison between pain at the start of the first test and pain at the start of the second test.

**Sample Size and Power**

The sample size required for the Bland-Altman analysis was calculated using the method outlined by Shieh [46]. A difference of 2.5 m was used for this calculation, based on a study comparing an unsupervised test using the GPS-based Timed Walk app performed at home and a clinic-based test conducted within a 7-day period (before or after) [41]. However, with the assumption that the tests performed on the same day and under supervision would result in slightly less variability, sample size calculation assumed an SD of 40 m rather than the previously reported value of 47 m. The maximum allowable difference between measures was set at 100 m as previously stated. Additional assumptions include the following: 90% power, α of .05, and a null central portion of 0.95 (ie, 95% of measurements lie between the 100-m maximum allowable difference boundaries). This calculation resulted in a requirement for 34 participants, which was increased to 38 participants to account for potential dropouts.

**Ethical Considerations**

All participants were provided with a participant information sheet that explained the study, and they provided their written consent to participate. Participants were compensated with a voucher worth Aus $20 (US $14-15) for their time and travel expenses. All data were deidentified. The study was approved by the Royal Brisbane and Women’s Hospital Human Research Ethics Committee (reference number HREC/2021/QRBW/75331) and ratified by the University of...
Queensland Human Research Ethics Committee (reference number 2021/HE001471).

Results

Sample Description

A total of 38 participants (n=17, 45% from the community and n=21, 55% from the outpatient pain clinic) were eligible and consented to participate. Of the 38 participants, 1 (3%) withdrew due to illness, 1 (3%) was excluded for no longer meeting the inclusion criteria, and 1 (3%) participant, whose GPS-based 6MWD was 88% more than their conventional 6MWD, was considered as an outlier and excluded from the analysis. The first participant was tested on October 27, 2021, and the final participant was tested on August 27, 2022 (reflecting issues with recruitment during the COVID-19 pandemic). A full description of the final sample of participants (N=35) is presented in Table 1.

Table 1. Description of individual samples and the combined sample.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Community (n=16)</th>
<th>Outpatient pain clinic (n=19)</th>
<th>Total (N=35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>50 (16)</td>
<td>49 (12)</td>
<td>49 (13)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>14 (88)</td>
<td>15 (79)</td>
<td>29 (83)</td>
</tr>
<tr>
<td>Men</td>
<td>2 (13)</td>
<td>4 (21)</td>
<td>6 (17)</td>
</tr>
<tr>
<td>Type or location of pain, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low back pain</td>
<td>8 (50)</td>
<td>9 (47)</td>
<td>17 (49)</td>
</tr>
<tr>
<td>Whole body pain or fibromyalgia</td>
<td>2 (13)</td>
<td>8 (42)</td>
<td>10 (29)</td>
</tr>
<tr>
<td>Lower limb pain</td>
<td>4 (25)</td>
<td>1 (5)</td>
<td>5 (14)</td>
</tr>
<tr>
<td>Neck or upper limb pain</td>
<td>2 (13)</td>
<td>1 (5)</td>
<td>3 (9)</td>
</tr>
<tr>
<td>Years since pain onset, mean (SD)</td>
<td>13 (15)</td>
<td>13 (11)</td>
<td>13 (13)</td>
</tr>
<tr>
<td>Baseline pain score (0-10 on NRS(^a)), mean (SD)</td>
<td>2.9 (2.2)</td>
<td>5.1 (2.5)</td>
<td>4.1 (2.6)</td>
</tr>
<tr>
<td>Smartphone operating system, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Android</td>
<td>9 (56)</td>
<td>15 (79)</td>
<td>24 (69)</td>
</tr>
<tr>
<td>Apple iOS</td>
<td>7 (44)</td>
<td>4 (21)</td>
<td>11 (31)</td>
</tr>
</tbody>
</table>

\(^a\)NRS: numerical rating scale.

After randomization, of the 35 participants, 17 (49%) participants performed the GPS-based 6MWT first and the conventional 6MWT second and the remaining 18 (51%) participants performed the tests in the reverse order. The results for the 2 types of 6MWTs, according to the order of the tests, are presented in Table 2.

Table 2. Values for the 6-minute walk distance (6MWD) according to method of measurement and order of tests.

<table>
<thead>
<tr>
<th></th>
<th>First 6MWD, mean (SD)</th>
<th>Second 6MWD, mean (SD)</th>
<th>Average 6MWD, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS-based 6MWD</td>
<td>426.5 (132.7)</td>
<td>466 (148.7)</td>
<td>446.8 (140.5)</td>
</tr>
<tr>
<td>Conventional 6MWD</td>
<td>453 (134.8)</td>
<td>413.1 (125.3)</td>
<td>433.6 (129.9)</td>
</tr>
<tr>
<td>Average 6MWD, mean (SD)</td>
<td>440.1 (132.5)</td>
<td>440.3 (138.4)</td>
<td>N/A(^a)</td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.

Concurrent Validity

The mean 6MWD measured by the GPS-based smartphone app was 13.2 m (SD 46 m; 95% CI −2.7 to 29.1) higher than the 6MWD assessed in the conventional manner. The 95% LoAs were 103.9 (95% CI 87.4-134.1) m and −77.6 (95% CI −107.7 to −61.0) m, which are not entirely within the maximum allowable difference of 100 m. A Bland-Altman plot is presented as Figure 1. As the distribution of the differences was leptokurtotic (an excess kurtosis of 2.2), an alternative, nonparametric Bland-Altman approach was applied, which yielded a median difference of 25.1 m and 95% LoAs (at the 2.5% and 97.5% quantiles) of −86.9 m and 76.4 m.
The ICC for agreement (2-way random-effects model) between the 2 6MWTs was 0.94 (95% CI 0.88-0.97) and the ICC for consistency (2-way random-effects model) was 0.94 (95% CI 0.89-0.97), both of which indicate good to excellent agreement.

The average of the two 6MWD measurements differed based on recruitment source, with those recruited from the community having a higher 6MWD (527 m, 95% CI 489-564) compared to those recruited from the outpatient clinic (368 m, 95% CI 304-431). The mean difference between the two 6MWD measurements also differed based on recruitment source, with outpatient measurements only differing by –4 m (95% CI –30 to 23) compared to a much higher difference of 33 m (95% CI 22-44) in the community sample. The apparent trend of overestimation by the GPS-based 6MWT relative to the conventional 6MWT with increased walking distance seems to be attributable to participants from the community displaying greater values than outpatients for both the average 6MWD and the difference between GPS-based and conventional 6MWD measurements (Multimedia Appendix 1). An exploratory comparison suggests that the difference between the two 6MWTs in the outpatient sample had a variance that was 6.5 (95% CI 2.3-17.4) times higher than the variance in the community sample (3009 m$^2$ vs 461 m$^2$).

In addition, an exploratory comparison found that the mean difference between the two 6MWT methods (GPS 6MWD minus conventional 6MWD) was 34.9 m (95% CI –5.29 to 75.2) greater when Android smartphones were used compared to that when Apple smartphones were used. The difference in 6MWDs when the GPS-based 6MWT was performed on Android smartphones was 34.2 (SD 61.6) m, whereas the difference in 6MWDs when the app test was performed on Apple smartphones was –10.8 (SD 57) m.
Another exploratory comparison suggests that the order in which the 2 tests were conducted has no clear effect on the difference between the tests. The mean difference between the two 6MWT methods (GPS 6MWD minus conventional 6MWD) when the GPS-based 6MWT was conducted first was 13.4 (SD 45.5) m, and the difference between the methods when the conventional 6MWT was performed first was 13.0 (SD 48.4) m. Therefore, when the app 6MWT was performed first, the mean difference between the tests was increased by 0.3 m (95% CI –31.9 to 32.6) versus when the conventional 6MWT was performed first.

**Change in Pain**

Pain increased in the conventional walk test by 1.1 (SD 1) points, whereas pain increased in the app test by 0.8 (SD 1.4) points. The conventional walk test produced a change in pain score that was 0.26 (95% CI –0.20 to 0.73) points higher than that produced by the app test.

An exploratory comparison suggests a statistically significant difference between pain at the start of the first test compared to pain at the start of the second test (t_{34}=2.969; P=.005). The pain rating before the first test averaged 4.1 (SD 2.5) points, and the pain rating before the second test (after the rest period) averaged 4.7 (SD 2.6) points. On average, pain at the beginning of the second test was 0.57 (95% CI 0.18-0.96) points higher than pain at the beginning of the first test. However, the median difference in pain was 0 (range –1 to 4) points. Approximately half (17/35, 49%) of the participants reported a pain level at the start of the second test that was identical to the level reported at the start of the first test. Figure 2 illustrates the distribution of individual differences between pain ratings at the start of each test.

**Discussion**

**Principal Findings**

This study investigated the validity of a GPS-based 6MWT smartphone app (Timed Walk) compared to that of the conventional method of conducting a 6MWT in people with persistent pain. To the best of our knowledge, this is also the first study to compare the Timed Walk app to a conventional 6MWT conducted on the same day in any population.

Participants walked an average of 433.6 (SD 129.9) m during the conventional 6MWT, which is typical of people with persistent pain [15,19,47]. The mean pain score at baseline for
participants of this study was 4.1 (SD 2.6), which is moderately lower than the averages of 6.0 and 6.4 noted during enrollment into a pain program in the literature [15,47]. This is likely due to our inclusion of participants from the community, who would be expected to have lower pain scores than those enrolled in a rehabilitation program. However, the baseline pain scores of the outpatients included in this study was 5.2 (SD 2.5), which is more consistent with previous estimates. Therefore, we can still be confident that the study sample was representative of people with persistent pain.

The results of the Bland-Altman analysis indicated uncertainty about the agreement between the 2 methods of measuring 6MWD. The 95% LoAs were 103.9 (95% CI 87.4-134.1) m and −77.6 (95% CI −107.7 to −61.0) m. These CIs are consistent with LoAs that are fully outside or within our a priori 100-m maximum allowable difference. Therefore, there is uncertainty about whether 95% of the differences will lie within this 100-m threshold. These LoAs are also larger than the estimated MCID in the 6MWT for chronic musculoskeletal pain, which is estimated to be 60 m for chronic musculoskeletal pain of the spine and 75 m for chronic musculoskeletal pain of the lower limb [19], but are within the much larger MCID estimate range of 156 m to 167 m for fibromyalgia [29]. Therefore, agreement between methods might not be sufficient for clinicians to use these two 6MWT methods interchangeably and still be able to reliably detect clinically important differences in the 6MWD for people with chronic pain of the limbs and spine. It is possible that the level of agreement is sufficient to detect clinically significant differences in people with fibromyalgia; however, this study was not performed specifically among people with fibromyalgia.

There are several potential reasons for differences between the two 6MWD measurements. First, the tests differed in how they were conducted, with one using a 30-m shuttle track and the other using various circuits around a public park. The GPS-based 6MWT involved fewer and more gradual turns, without the requirement to decelerate and accelerate with short turns on a shuttle track. Previous studies have demonstrated that continuous (circular or rectangular) tracks result in 6MWD estimates that are 3% to 10% higher than those obtained from a straight shuttle track [7,40,48]. Therefore, this is a likely explanation for the observed 6MWD derived from the Timed Walk app in this study being 13.2 m (95% CI −2.7 to 29.1) higher than that obtained from the conventional 6MWT—approximately a 3% increase. However, the CIs for the mean difference in the 6MWD between tests are wide enough to include both the possibility of no difference between the tests and a difference as large as approximately 30 m or approximately 7% of the total 6MWD. The 2 tests also differed in how encouragement phrases were delivered—by a researcher or through audio from the smartphone app—which could also impact walking distance. In addition, although the 30-m track was consistent for all conventional 6MWTs, the exact walking route used for the GPS-based 6MWT for each participant varied in terms of potentially influential factors such as the number of turns, radius of turns, and number of laps performed. It is possible that differences in the walking routes used could also partially explain the variability between the 2 tests.

Second, the tests differed in how they were measured, which could have introduced errors that could be observed as either systemic bias or as variability in the results. Inaccuracy of measurement in the conventional 6MWT might occur due to inaccurate measurement of the length of the track (or partial final lap), the participant not being able to stop instantly at 6 minutes, and the participant deviating from the track during walking or making wider turns around the markers. Although none of these issues are present for the GPS method of measurement, other sources of inaccuracy likely exist. In their initial 30 tests with 8 different smartphones, Salvi et al [36] compared the GPS-based algorithm of the Timed Walk app to a simultaneous measurement of distance with a trundle wheel, thereby eliminating variability between separate walking tests. Although the mean difference between those 2 measurements was minimal at 0.8 m, the SD was 18.6 m (or approximately 4.2% of the mean 6MWD of 438 m). As the Bland-Altman method calculates 95% LoAs as 1.96 SDs from the mean, approximately 36-m wide 95% LoAs (or 8.4% of the 6MWD) might be expected based on measurement error alone. This measurement error could be related to a poor-quality GPS signal (due to tree coverage or cloud cover), inaccurate GPS receivers in smartphones, and issues with the app’s algorithm. GPS accuracy is typically assessed in a static (nonmoving) condition or while driving [49,50], both of which differ significantly from a walking test. The accuracy of GPS measurements in the context of a dynamic, real-world activity such as walking is influenced by various factors, including signal processing and filtering algorithms specific to each app. Some previous studies suggest that the GPS receivers have different levels of accuracy across different models of smartphones [51]. The difference we observed between Android and Apple smartphones suggests some role for the hardware or operating system, which, as participants’ devices were effectively random, could introduce an additional source of variability.

Conducting 2 separate walking tests introduces an additional source of variability due to normal human variability in walking speed, in addition to the measurement error discussed previously. The conventional 6MWT, when performed 1 week apart, has a minimum detectable change (at 95% confidence) of 86 m in people with persistent pain [18] and from 50 m to 80 m in adults aged >60 years [52]. The minimum detectable change at 95% confidence is equivalent to the difference between the 95% LoA and the mean difference [53,54]. This suggests that the conventional method of assessing a 6MWT is only barely able to reliably detect a change of the same magnitude as our nominated 100-m maximum allowable difference. The test-retest reliability of the Timed Walk app has only been reported once in the literature. Salvi et al [41] have reported data about the test-retest reliability of the Timed Walk app in pulmonary hypertension, based on 89 pairs of outdoor tests performed 7 days apart in 10 patients. That study also reported a mean difference of 1.8 m (0.7% of the 6MWD) and an SD of 37 m (or 10.1% of the 6MWD) between pairs of GPS-based 6MWTs, giving a 95% LoA of approximately 74 m on either side of the mean. Therefore, this evidence suggests that it can be difficult to reliably detect a change of 100 m between two 6MWTs, even when both 6MWTs are conducted

https://formative.jmir.org/2024/1/e46820  JMIR Form Res 2024 | vol. 8 | e46820 | p.1341 (page number not for citation purposes)
using the same methods (either the conventional method or the GPS-based method).

Finally, there could be some effect of the first test on the results of the second test. For instance, performance might have decreased during the second test due to the provocation of fatigue or pain. Although a 15-minute rest period is sufficient for people with cardiac disease to recover from fatigue induced by walking [55], this may not be the case for recovery from pain. We found that some (15/35, 43%) of participants reported higher pain scores at the start of their second test than at the start of their first test. This residual pain could have altered their performance during the second test, as pain can increase the energy cost of walking [56,57]. On the other hand, a learning or practice effect whereby 6MWD increases for a second 6MWT conducted on the same day has been demonstrated in several studies in healthy adults [58,59] and people with respiratory disease [60] and chronic heart failure [61]. This learning effect may be due to decreased anxiety (more willingness to walk faster if they see that it does not provoke as much pain or fatigue) or improvements in technique (optimum stride length or pacing) during subsequent tests [4]. In addition, a learning effect may still occur regardless of the rest period between tests conducted on the same day. As the order in which the 2 tests were conducted was randomized in this study, any effect of test order would manifest as variance in the difference between the 2 methods. However, when the difference between 6MWD measurements was compared between the subgroup randomized to perform the app test first versus the subgroup performing the app test second, there was no clear evidence of a difference. As the 2 walk tests investigated in this study were very dissimilar, a learning effect may have been less likely to occur. Alternatively, a learning effect could have been countered by an approximately equal and opposite effect of pain or fatigue. Interestingly, the greater variability between the 2 tests seen in the outpatient subsample, who also had higher pain scores and more widespread pain than the community subsample, may also suggest that those who are experiencing greater pain have more variable walking speeds in general. This could be related to spatiotemporal gait mechanics, which are known to be affected by pain conditions [62,63].

Validity, as assessed by the ICC for agreement (2-way random-effects model) between the 2 6MWTs, was good to excellent at 0.94 (95% CI 0.88-0.97), suggesting that approximately 12% of the variation between the 2 measurements is due to bias or error. ICC for consistency was also good to excellent at 0.94 (95% CI 0.89-0.97). The near equality of these ICC values indicates that the validity is mostly being affected by random error rather than systemic error (bias). Despite the variability introduced by the previously mentioned differences between the GPS-based and conventional 6MWTs, the ICC for agreement in this study is similar to previously reported ICC values (agreement) of 0.91 and 0.92 for the test-retest reliability (with 1 day or 7 days between tests) of the conventional 6MWT in fibromyalgia [64,65]. Similarly, it is comparable to the ICC of 0.91 reported by Salvi et al [41] for the test-retest reliability of the Timed Walk app in pulmonary hypertension, based on 89 pairs of outdoor tests performed 7 days apart in 10 patients.

Pain

Secondarily, the study also compared the 2 methods to examine the differences in the degree of pain evoked by the test. Pain increased by approximately 1 point on NRS in both tests. Therefore, there is no indication of any additional advantage or disadvantage to either method with respect to pain provocation. However, this sample comprised people with a variety of different pain conditions. It is possible that a population with only hip or knee pain would have experienced a difference between the 2 tests, perhaps due to the increased frequency of turning in the conventional 6MWT. Moreover, pain was only assessed immediately after the walking tests, and both tests were completed 15 minutes apart; therefore, potential differences in pain experienced in the hours or days after the tests are beyond the scope of this study.

Strengths and Limitations

This study compared Timed Walk against a conventional 6MWT on the same day. People with persistent pain can have variability in day-to-day pain [66,67], which may result in variations in walking speed. Conducting the tests 15 minutes apart removed this source of variability. However, the pain caused by the first 6MWT was clearly still a factor during the second 6MWT for some individuals. It is unclear whether extending the rest period to 30 minutes or an hour would have been sufficient for this pain to resolve completely. Moreover, only pain intensity was assessed and not the quality of the pain. Recording descriptors of the quality of pain (such as sharp, dull, throbbing, or aching) or assessing pain-related fear of movement might have provided deeper insights into why some participants with significant pain increases did not exhibit changes in walking distance but others did.

The study used the Timed Walk app on participant’s own smartphones. Similarly, this study used a wide variety of heterogeneous walking tracks on grass or paved surfaces at a local park, with this heterogeneity precluding any analysis of the effect of walking track on the results. Although the observed agreement between methods may have been higher if a single smartphone and a set walking track were used consistently for all GPS-based walking tests, the results of this study have more practical relevance to people using their own smartphone in real-world environments. Similarly, as GPS is reported to have issues in measuring distances around sharp corners, a linear track for GPS may have been a more accurate measure of distance walked but would have been practically more difficult for participants to replicate at home and thus would be of less practical relevance.

This study calculated the 95% LoAs and their CIs under the assumption that the data were normally distributed. However, due to noticeable kurtosis (excess kurtosis of 2.2), the calculated 95% LoAs may be a slight overestimate. Although a nonparametric approach found that the 2.5% and 97.5% quantiles of the sample fell just within the nominated maximum allowable difference, it did not provide CIs to represent the uncertainty in the 95% LoAs. To obtain these CIs, the data could have been transformed to better approximate a normal distribution, or a bootstrap approach could have been used. However, regardless of the approach used, it is likely that the
95% LoAs or the associated CIs would be outside the maximum allowable difference.

The maximum allowable difference nominated in this study was selected in the absence of clearly established values for clinically relevant differences in the 6MWT in persistent pain conditions. A value of 100 m was selected as a compromise between the existing MCID estimates for the 6MWT in people with chronic pain of the back and lower limbs (60 m and 75 m, respectively) [19] and in those with fibromyalgia (which range from 156 to 167 m) [29]. The lack of knowledge about clinically acceptable differences for the 6MWT significantly limits the interpretation of the Bland-Altman analysis.

Implications for Future Studies and Clinical Practice

Future studies are required to develop more accurate methods for remotely performing the 6MWT in populations with chronic disease including persistent pain. This may involve accelerometer-based methods, GPS-based methods, or a combination of both. Future updates to the Timed Walk app will require reassessment of concurrent validity. In addition, future studies could further elucidate the sources of variability between the GPS-based 6MWT and conventional 6MWT. This could entail performing a GPS-based 6MWT with the additional use of a trundle wheel as a third measurement of the 6MWD, simultaneous measurements derived from multiple smartphones from various manufacturers, and a systematic comparison of specific walking tracks. It may also be wise for future research studies to ensure that participants unfamiliar with the 6MWT are given a practice test to minimize the potential learning effect. Finally, more research is needed to establish a more precise estimate of the MCID in persistent pain to allow for future studies to better define the maximum allowable difference between 6MWT methods.

As both tests in this study were performed under supervision, future investigations should consider the effect of unsupervised use of the GPS-based 6MWT app. Tests that are self-administered may be even more variable if care is not taken to avoid sharp turns, tall buildings, uneven terrain, or inconsistencies in weather conditions. Future studies are also required to investigate whether the test-retest reliability of the GPS-based 6MWT, when used unsupervised in people with persistent pain, is sufficient for use as a tool for remote monitoring. It is feasible that the imprecision in the estimates from the GPS-based app may be less of an issue if testing is conducted more frequently, as the average of a series of tests may provide a more stable estimate of the participant’s functional status over time.

Although conventional and GPS-based 6MWT methods do not demonstrate sufficient agreement to be used interchangeably in people with persistent pain, the ability to perform the test remotely using the GPS-based app may still have benefits in a clinical context. The ability for patients to perform the GPS-based assessment at home without the need for a clinician’s presence may increase patient autonomy and reduce the burden of frequent clinic visits. In addition, for individuals who are unable to regularly attend in-clinic evaluations, obtaining an approximate measurement of functional capacity more frequently via a GPS-based 6MWT may still be clinically useful, even if the estimate is imprecise, as it provides some information, which is better than a complete lack of data. GPS-based measurements may also offer more ecologically relevant assessments of functional capacity than in-clinic testing, better reflecting a patient’s natural walking abilities and providing an opportunity for training walking capacity in the patient’s everyday environment. Overall, although the in-clinic standard 6MWT performed by experienced personnel remains as the gold standard, the GPS-based app may still be considered as a complementary tool.

Conclusions

This study demonstrated that the concurrent validity between the GPS-based 6MWT using the Timed Walk app and the conventional 6MWT may not be sufficient for the 2 methods to be used interchangeably in people with persistent pain while still being able to detect clinically significant differences in the 6MWD. Despite limited validity, the GPS-based 6MWT may still have clinical application as a complementary tool to the conventional 6MWT performed in the clinic, especially for remote monitoring. In addition, the GPS-based 6MWT makes it possible to conduct more frequent assessments of functional capacity, which are self-measured without the presence of a clinician and conducted in a more ecologically relevant environment. Future studies are needed to improve the accuracy of the GPS-based 6MWT for remote monitoring.

Acknowledgments

The authors would like to thank the participants for their participation in this study. In addition, the authors would like to thank D Salvi for answering their questions regarding previous investigations of the Timed Walk mobile app. Finally, the authors acknowledge the use of ChatGPT (version 3.5; OpenAI; February 2023) as a writing assistant, specifically for rephrasing and summarizing content.

Data Availability

The data sets generated during and analyzed in this study are not publicly available due to lack of ethical clearance to disclose data to third parties.

Conflicts of Interest

None declared.
Multimedia Appendix 1
Bland-Altman plots depicting the difference between GPS-based and conventional 6-minute walk distances against the average of the 2 measurements, with linear regression trend lines estimated for either the whole sample or for each recruitment group.

[PDF File (Adobe PDF File), 220 KB - formative_v81e46820_app1.pdf]

Multimedia Appendix 2
Scatter plots depicting the difference in walking distance between the 2 pairs of tests (GPS-based test vs conventional test and type of walk test conducted first vs type of walk test conducted second) against the difference in pain levels at the start of each walk test.

[PDF File (Adobe PDF File), 208 KB - formative_v81e46820_app2.pdf]

References
18. Murdoch D, Window P, Morton C, O'Donohue R, Ballard E, Claus A. People at a persistent pain service can walk it, but some struggle to talk about it: reliability, detectable difference and clinically important difference of the six-minute walk test. Musculoskeletal Care 2023 Mar;21(1):221-231. [doi: 10.1002/msc.1687] [Medline: 36065494]


52. Arkesteyn M, Low D. Responsiveness of functional assessments to monitor change in balance, walking speed and strength of older adults: a systematic review of the minimal detectable change. medRxiv. Preprint posted online June 06, 2022 2024 [FREE Full text] [doi: 10.1101/2022.06.06.22276029]


Abbreviations

6MWD: 6-minute walk distance
6MWT: 6-minute walk test
ICC: intraclass correlation coefficient
LoA: limit of agreement
MCID: minimum clinically important difference
NRS: numerical rating scale

©Joshua Simmich, Nicole Emma Andrews, Andrew Claus, Megan Murdoch, Trevor Glen Russell. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Using Principles of Digital Development for a Smartphone App to Support Data Collection in Patients With Acute Myocardial Infarction and Physical Activity Intolerance: Case Study

Diana Isabel Cáceres Rivera1*, MSc, PhD; Luz Mileyde Jaimes Rojas1*, MSc; Lyda Z Rojas2*, MSc, PhD; Diana Canon Gomez2*, MD; David Andrés Castro Ruiz1*, MSc; Luis Alberto López Romero3*, MSc

1Facultad de Enfermería, Universidad Cooperativa de Colombia, Bucaramanga, Colombia
2Centro de Investigaciones, Fundación Cardiovascular de Colombia, Floridablanca, Colombia
3Departamento de Pediatría, de Obstetricia y Ginecología y de Medicina Preventiva y Salud Pública, Universidad Autónoma de Barcelona, Barcelona, Spain
*these authors contributed equally

Corresponding Author:
Diana Isabel Cáceres Rivera, MSc, PhD
Facultad de Enfermería
Universidad Cooperativa de Colombia
Torre 2 Apto 203
Bucaramanga, 64000
Colombia
Phone: 57 3014006658
Email: dianacaceresrivera@gmail.com

Abstract

Background: Advances in health have highlighted the need to implement technologies as a fundamental part of the diagnosis, treatment, and recovery of patients at risk of or with health alterations. For this purpose, digital platforms have demonstrated their applicability in the identification of care needs. Nursing is a fundamental component in the care of patients with cardiovascular disorders and plays a crucial role in diagnosing human responses to these health conditions. Consequently, the validation of nursing diagnoses through ongoing research processes has become a necessity that can significantly impact both patients and health care professionals.

Objective: We aimed to describe the process of developing a mobile app to validate the nursing diagnosis “intolerance to physical activity” in patients with acute myocardial infarction.

Methods: We describe the development and pilot-testing of a mobile system to support data collection for validating the nursing diagnosis of activity intolerance. This was a descriptive study conducted with 11 adults (aged ≥18 years) who attended a health institution for highly complex needs with a suspected diagnosis of coronary syndrome between August and September 2019 in Floridablanca, Colombia. An app for the clinical validation of activity intolerance (North American Nursing Diagnosis Association [NANDA] code 00092) in patients with acute coronary syndrome was developed in two steps: (1) operationalization of the nursing diagnosis and (2) the app development process, which included an evaluation of the initial requirements, development and digitization of the forms, and a pilot test. The agreement level between the 2 evaluating nurses was evaluated with the κ index.

Results: We developed a form that included sociodemographic data, hospital admission data, medical history, current pharmacological treatment, and thrombolysis in myocardial infarction risk score (TIMI-RS) and GRACE (Global Registry of Acute Coronary Events) scores. To identify the defining characteristics, we included official guidelines, physiological measurements, and scales such as the Piper fatigue scale and Borg scale. Participants in the pilot test (n=11) had an average age of 63.2 (SD 4.0) years and were 82% (9/11) men; 18% (2/11) had incomplete primary schooling. The agreement between the evaluators was approximately 80% for most of the defining characteristics. The most prevalent characteristics were exercise discomfort (10/11, 91%), weakness (7/11, 64%), dyspnea (3/11, 27%), abnormal heart rate in response to exercise (2/10, 20%), electrocardiogram abnormalities (1/10, 9%), and abnormal blood pressure in response to activity (1/10, 10%).

Conclusions: We developed a mobile app for validating the diagnosis of “activity intolerance.” Its use will guarantee not only optimal data collection, minimizing errors to perform validation, but will also allow the identification of individual care needs.
Introduction

In recent decades, the ability to produce, collect, and communicate data around the world has increased exponentially with access to technologies such as smartphones. These technologies have improved data storage as well as its handling and analysis [1]. In the field of health, electronic record systems facilitate data collection that can be used for various purposes, allowing data retrieval that promotes the improvement of research processes such as identification and recruitment of patients for clinical projects [2,3].

In addition to obtaining individual data from each patient, the collection of large amounts of data can be useful to obtain information that more effectively supports the exploration of diseases, treatment, and rehabilitation. This creates the need to develop research platforms that optimize the capacity to conduct informative and innovative research and enable scientific approaches where objective data can be obtained with a minimum of errors and expended resources [4].

As part of the health staff providing care to cardiovascular patients, nurses can be the first to identify individual needs. To aid this, tools are available such as the NANDA (North American Nursing Diagnosis Association) taxonomy, which identifies the response of a person, family, or community to real health problems and potential vital processes. However, these diagnoses and their respective defining characteristics must be validated according to the context where they will be assessed, which constitutes a challenge in research into the use, implementation, and dissemination of technologies of information [5-7]. For this purpose, the use of digital platforms has demonstrated its applicability from the early stages of research, such as the assessment of care needs [8,9].

Mobile apps in health, education, and work in Colombia are applications of medical informatics apps; mobile app; mobile applications; nursing diagnosis; nursing research; research data; software; validation

Methods

Overview

This was a descriptive study conducted with 11 adults (aged ≥18 years) with a suspected diagnosis of coronary syndrome who attended a health institution for highly complex needs between August and September 2019 in Floridablanca, Colombia. An app for clinical validation of the “activity intolerance” diagnosis (NANDA code 00092) in patients with acute coronary syndrome was developed in three steps, outlined in the following sections.

Step 1: Operationalization of the Nursing Diagnosis

The first step consisted in the operationalization of the defining characteristics of the nursing diagnosis [11] of activity intolerance (NANDA code 00092), defined by NANDA-I [6] as “the lack of sufficient physiological or psychological energy to tolerate or complete the required or desired daily activities.” This diagnosis is categorized as “Domain 4: Activity / Rest, Class 4: Cardiovascular / pulmonary responses Need: Move and Pattern Activity-exercise.” It is also related to an imbalance between oxygen supply and demand, a sedentary lifestyle, immobility, and bed rest; it has defined characteristics [12].

Through an extensive search of the literature, we selected scales or instruments to standardize the measurement of each defining characteristic of this nursing diagnosis [11]. An interdisciplinary group that included 2 nurses, an epidemiologist, and a cardiologist verified the face validity of the operationalization.

Step 2: App Development Process

Initial Requirements Evaluation

Health professionals, along with a systems engineer, carried out the structural design of the data collection forms or case report forms. The digitization process was carried out using CommCare [7], which is an open source, cloud-based platform that helps researchers develop data capture tools using mobile devices. An open source tool was also used to create an Android-based mobile app for a low-income setting. Mobile apps can be used as a tool to track beneficiaries through a service lifecycle and can also streamline data collection [13]. Our app used the HTTPS protocol, which made it cryptographically secure. Access to data was password protected. The CommCare [7] platform was selected because it has been widely used for health projects all over the world and because of its ease of use and compatibility with older versions of Android. CommCare is a platform that works on Android mobile phones from version 2.3, but the platform recommends reviewing the documentation for these older versions because they may have limitations in terms of functionality and compatibility with the latest features developed by CommCare, so it is recommended to have at least Android version 4.0.3 or later, a storage space of at least 100MB.
a minimum of 1GB of RAM, and a processor with at least 2 cores for a better user experience.

Finally, we did not use any programming language because we used a platform that prevents us from reaching that level. We worked directly with CommCare, which allowed us to create data collection applications without touching or programming source code (Figure 1).

CommCare requires the use of a password to access the app and the data stored on the platform. This helps to ensure that only authorized users can access information. The platform uses the secure HTTPS communications protocol, uses role-based access, and is in compliance with data security regulations and standards such as the European Union’s General Data Protection Regulation (GDPR). This ensures that the platform follows good practices in terms of privacy and personal data protection.

Figure 1. Commcare platform design.

Development and Digitization of the Forms
The principles of the Scrum methodology for agile software development were applied. This is a regularly applied process that includes a set of best practices to work collaboratively in teams and obtain the best possible outcome of projects. It is characterized by a strategy of incremental development, boosting the quality of the result by getting to know people in self-organized teams and matching the different phases of development, rather than doing one after the other in a sequential or cascading cycle [14]. Through this methodology, an app was developed to gather data. This phase included the following six steps: (1) specifying the forms to be digitized, which contained the questions or variables to be obtained in the field; (2) dividing the various sections of the form into smaller subforms, depending on the size of the questionnaire or the time of application; (3) defining the variables as the simple question-and-answer type or as more complex ones containing calculations, depending on others, or having a different logical flow; (4) building the form on the CommCare platform; (5) generating app versions (eg, test versions); and (6) testing the app with health professionals who simulated data from possible patients and followed the flow of questions within the app to check if the different flows worked correctly; if errors or possible improvements were found during the process, the entire procedure was repeated from step 4.

Ethical Considerations
The Ethics Committee of Universidad Cooperativa de Colombia thoroughly reviewed and approved the research (report 003; April 16, 2018), as did the Fundación Cardiovascular de Colombia ethics committee (report 450; May 22, 2018). The study was carried out in strict adherence to the established protocol, regulatory requirements, Good Clinical Practice, the Declaration of Helsinki, and the clinical investigation guidelines of Universidad Cooperativa de Colombia. All participants provided their informed consent by signing a form. Participation in this study was entirely voluntary, and no financial compensation or reimbursements were offered to the participants.

The information obtained has been securely stored in the archives of the Universidad Cooperativa de Colombia to safeguard the privacy of individuals. Each patient was assigned a code to ensure that their names or identification did not appear in the database. Access to the collected data was restricted to the researchers, and the data will be used exclusively for the study’s intended purposes. Personal information is being protected in compliance with Colombian Law 1581 of 2012, which pertains to the right of “habeas data.”

Results
After repeatedly performing the entire process and correctly digitizing all the forms proposed in advance, the last version of the app (the production version) was generated. The result of the development process was an app that allowed obtaining information using the forms shown in Table 1.

Table 2 shows the scales and instruments used for the operationalization of the defining characteristics of nursing. The resulting app allowed the simultaneous collection, data entry, and follow-up of patients in different stages of investigation. Two previously trained nurses conducted a pilot test with the first 11 patients included in the research. Taking

https://formative.jmir.org/2024/1/e33868
into account the inclusion and exclusion criteria, a cardiologist selected potential patients. Subsequently, the patient received an explanation of the study; if they agreed to participate, they provided informed written consent. The information was filled out on tablet-type mobile device. Once the data were collected, a process of sending or synchronizing the data with the database in the cloud was carried out, for which it was necessary to have an internet connection (Wi-Fi network).

### Table 1. General information included in the app.

<table>
<thead>
<tr>
<th>Information collected</th>
<th>Forms used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociodemographic data</td>
<td>Personal data of the patient</td>
</tr>
<tr>
<td>Registered patient forms</td>
<td>Information about the patient</td>
</tr>
<tr>
<td>Hospital admission data</td>
<td>Referral or admission information</td>
</tr>
<tr>
<td>Background</td>
<td>Medical, surgical, family, and toxicology history</td>
</tr>
<tr>
<td>Current drug treatment</td>
<td>Angiotensin-converting enzyme inhibitors, statins, β-blockers, angiotensin II receptor blockers, aldosterone antagonists, acetylsalicylic acid, diuretics, thiazide diuretics, digitalis, antiplatelet agents, anticoagulants, vasodilators, antiarrhythmics, analgesics, inotropics</td>
</tr>
<tr>
<td>Intensity of angina</td>
<td>Angina intensity level</td>
</tr>
<tr>
<td>Diagnostic means</td>
<td>Electrocardiogram, electrographic changes, electrocardiogram findings, arteriography findings, cardiac enzymes, and other paraclinical tools</td>
</tr>
<tr>
<td>TIMI-RS⁰</td>
<td>TIMI-RS (S-elevation and non-ST elevation; acute myocardial infarction/unstable angina)</td>
</tr>
<tr>
<td>GRACEᵇ score</td>
<td>GRACE score</td>
</tr>
<tr>
<td>Defining characteristics of the diagnosis</td>
<td>Electrocardiogram changes, generalized weakness, exertional discomfort, dyspnea on exertion, fatigue, abnormal heart rate in response to activity, and abnormal blood pressure in response to activity</td>
</tr>
<tr>
<td>Follow-up</td>
<td>Presence of fatigue, pain, dyspnea on exertion, and physical activity</td>
</tr>
</tbody>
</table>

⁰TIMI-RS: thrombolysis in myocardial infarction risk score.
ᵇGRACE: Global Registry of Acute Coronary Events.

### Table 2. Scales and validated instruments included in the app for diagnosing intolerance to activity.

<table>
<thead>
<tr>
<th>Defining characteristics</th>
<th>Scale or instrument used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrocardiogram changes (arrhythmias, conduction abnormality, ischemia)</td>
<td>Suggested changes by Clinical Practice Guideline of the Ministry of Social Protection of Colombia [15]</td>
</tr>
<tr>
<td>General weakness</td>
<td>Handgrip measurement (muscle strength in kg) and bioimpedance measurement</td>
</tr>
<tr>
<td>Exertional discomfort</td>
<td>Brief Disease Perception Questionnaire [16]</td>
</tr>
<tr>
<td>Exertional dyspnea</td>
<td>Grade of dyspnea according to the New York Heart Association [17]</td>
</tr>
<tr>
<td>Fatigue</td>
<td>Piper Fatigue Scale [18]; Ruffier test (aerobic capacity) [19]; Borg scale (perceived physical exertion) [20]</td>
</tr>
<tr>
<td>Abnormal heart rate in response to activity</td>
<td>The suggested values in the Clinical Practice Guide of the Ministry of Social Protection of Colombia [14]</td>
</tr>
<tr>
<td>Abnormal blood pressure in response to activity</td>
<td>2018 European Society of Cardiology/European Society of Hypertension Clinical Practice Guidelines for the Management of Arterial Hypertension [21]</td>
</tr>
</tbody>
</table>

Later, the systems engineer reviewed the database obtained through the web platform CommCare, which allows downloading information as a flat file or in spreadsheet format. In this way, the research team verified the correct operation of the app and its use in the field, obtaining positive results that allowed the continuity of the investigation with more patients. Figures 2-4 show 3 screenshots of the app.

The pilot test yielded descriptive data (n=11). The participants had an average age of 63.2 (SD 4.0) years, 82% (9/11) were men, and 18% (2/11) had incomplete primary schooling. We found that 64% (7/11) had a history of hypertension and 73% (8/11) had ever smoked. The defining characteristics present in this group of patients were exercise discomfort in 91% (10/11), electrocardiogram abnormalities in 9% (1/10), abnormal heart rate in response to exercise in 20% (2/10), dyspnea in 27% (3/11), weakness in 64% (7/11) and abnormal blood pressure in response to activity in 10% (9/10) (Table 3). The k agreement index ranged from 73% to 100%.
**Figure 2.** Screenshot of app start screen.

![App Start Screen](image)

**Figure 3.** Screenshot showing the list of forms that can be used with patients in the field.

![List of Forms](image)

**Figure 4.** Screenshot showing question on the personal data of the patient.

![Question on Personal Data](image)
Table 3. Presence of defining characteristics of the nursing diagnosis “activity intolerance” in patients (n=11) with acute coronary syndrome according to 2 independent evaluators.

<table>
<thead>
<tr>
<th>Defining characteristics</th>
<th>Participants evaluated by nurse 1, n (%)</th>
<th>Participants evaluated by nurse 2, n (%)</th>
<th>Agreement between the evaluators, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Electrocardiogram changes (arrhythmias, conduction abnormality, ischemia)</td>
<td>1 (9)</td>
<td>10 (91)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>General weakness</td>
<td>7 (64)</td>
<td>4 (36)</td>
<td>8 (73)</td>
</tr>
<tr>
<td>Exertional discomfort</td>
<td>10 (91)</td>
<td>1 (9)</td>
<td>9 (82)</td>
</tr>
<tr>
<td>Exertional dyspnea</td>
<td>3 (27)</td>
<td>8 (73)</td>
<td>3 (27)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0 (0)</td>
<td>11 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Abnormal heart rate in response to activity</td>
<td>2 (20)</td>
<td>8 (80)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Abnormal systolic blood pressure in response to activity</td>
<td>1 (10)</td>
<td>9 (90)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Abnormal diastolic blood pressure in response to activity</td>
<td>0.0 (0)</td>
<td>10 (100)</td>
<td>1 (10)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

We describe the development process of a mobile app for collecting health research data in an easy, agile, and practical way. This strategy may be used for the complete collection of samples in the process of clinical validation of the nursing diagnosis “activity intolerance.” In addition, a good rate of agreement was found among the evaluators thanks to the standardization used in the app.

In recent years there has been an increase in the use of computer technologies to replace paper records by means of mobile apps, web forms, and specialized software; likewise, it has become evident that these are key tools to improve quality in health care [22]. However, it is still a challenge to continue implementing new strategies, achieve their efficient use by health professionals, and make their implementation easier and more accessible.

This process enabled us to validate the app’s use for identifying prevalent nursing diagnoses, such as activity intolerance, in patients with acute myocardial infarction. Among the 9 defining characteristics we evaluated, there was an agreement of over 80% among the evaluators for 5 of them. This, in turn, helped us identify the most prevalent characteristics, namely dyspnea on exertion and heart rate alteration in response to activity. It is also noteworthy that none of the evaluators identified fatigue in any of the users.

Mobile Apps

We evaluated this strategy for identifying nursing diagnoses that require an objective definition of their characteristics and clinical judgment [23]. The precise operationalization of the defining characteristics through a predefined registry structure, as seen in this mobile app, enhances the precision of nursing diagnoses [1]. In this sense, it enables the evaluation of these characteristics, which can improve documentation for nursing staff, thereby aiding in the inference and evaluation of diagnoses [2]. Therefore, this app aims not only to enhance the quality and safety of care processes but also to promote the adoption of standardized nursing language, addressing the limitations in its use.

Another possible use of this app is in education, where it would potentially help to strengthen the precision of documentation in nursing diagnoses [3]. This strategy is adapted to current conditions, in which the use of virtual methods and mobile technologies has been shown to be a new basic input for the teaching process, making it necessary for professionals and trainers to make an adequate use of this type of strategy.

A relationship where nurse and patient can contribute to improving administrative processes that benefit others has been described in settings such as outpatient care [5]. This is expected to contribute to research scenarios that promote improved caregiving. Apps can assist in the assessment and generation of nursing diagnoses in hospital practice [24], and they have been used in research studies such as clinical trials for the self-management of angina [25].

Limitations

This work was limited to a specific nursing diagnosis. Future work should include other prevalent diagnoses in patients with cardiac disease. An evaluation of usability among end users could help improve our strategy, and more data is also needed to better specify the large-scale feasibility and cost of this strategy with other nursing diagnoses.

Other aspects to improve in the design of future research are to include scales and instruments used in health care to measure different variables. These sources of information should be updated according to the context, clinical conditions, and even environmental conditions. An additional challenge is the integration of these types of apps to existing health systems. A recent review with the objective to provide an overview of studies that have collected patient data using an app-based approach indicated that using mobile technologies could help to overcome challenges linked with data collection in epidemiological research. However, further feasibility studies need to be conducted to test the applicability and acceptance of these mobile apps for epidemiological research in various subpopulations [26].
Conclusions
We developed a mobile app for use in the validation process of the nursing diagnosis activity intolerance. This app enabled the evaluation of defining characteristics, which can enhance documentation for nursing staff, facilitate more effective inference and evaluation of diagnoses, and reduce errors in information recording. One significant potential of this app lies in its impact on education, as it aids in improving the precision of nursing diagnosis documentation and, as a result, enhances the quality of care planning.

Data Availability
The data sets generated during and/or analyzed during this study are available in the Mendeley repository [27].

Conflicts of Interest
None declared.

References
7. Commcare. URL: http://www.commcarehq.org/ [accessed 2020-02-09]


22. Lunney M. Critical need to address accuracy of nurses’ diagnoses. Online J Issues Nurs 2008 Jan 31;13(1) [FREE Full text] [doi: 10.3912/oijn.vol13no01ppt06]


Abbreviations

GDPR: General Data Protection Regulation
NANDA: North American Nursing Diagnosis Association

Edited by A Mavragani; submitted 10.02.23; peer-reviewed by T Behera, J King; comments to author 07.07.23; revised version received 18.10.23; accepted 02.01.24; published 18.03.24.

Please cite as:
Cáceres Rivera DI, Rojas LMJ, Rojas LZ, Gomez DC, Castro Ruiz DA, López Romero LA
Using Principles of Digital Development for a Smartphone App to Support Data Collection in Patients With Acute Myocardial Infarction and Physical Activity Intolerance: Case Study
JMIR Form Res 2024;8:e33868
URL: https://formative.jmir.org/2024/1/e33868
doi:10.2196/33868
PMID:38498019

©Diana Isabel Cáceres Rivera, Luz Mileyde Jaimes Rojas, Lyda Z Rojas, Diana Canon Gomez, David Andrés Castro Ruiz, Luis Alberto López Romero. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Preliminary Efficacy of a Cognitive Behavioral Therapy–Based Smartphone App for Smoking Cessation in China: Randomized Controlled Pilot Trial

Shanshan Chen¹, MD; Jinsong Tang¹, PhD; Congyang Wu², MBA; Ge Zhang², BM; Jing Zhang², PhD; Yanhui Liao¹, PhD

¹Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China
²Johnson & Johnson Pharmaceutical Company, Shanghai, China

Corresponding Author:
Yanhui Liao, PhD
Sir Run Run Shaw Hospital
Zhejiang University School of Medicine
3 East Qingchun Road
Hangzhou, 310016
China
Phone: 86 18814898844
Email: lioyanhui@zju.edu.cn

Abstract

Background: The overall prevalence of cigarette smokers in China is very high, and China’s total cigarette consumption makes up more than 40% of the world’s consumption. In view of the lack of smoking cessation services and social support in China and the effectiveness of mobile phone apps for quitting smoking in other countries, we carried out a smartphone app–based smoking cessation trial in China.

Objective: This study aimed to evaluate the efficacy of a cognitive behavioral therapy (CBT)–based smoking cessation smartphone app among smokers seeking treatment in China.

Methods: We conducted a randomized controlled, web-based pilot clinical trial in China between February 23 and June 27, 2021. Eligible participants were randomly assigned to the smoking cessation app intervention group or the control group in a ratio of 1:1. The intervention group received the CBT smoking cessation intervention using a smartphone app, and the control group received a “thank you” message. The intervention was 4 weeks long, and the patients were followed up for 4 weeks. The primary outcome was self-reported continuous smoking abstinence at week 4 after the quit date. The secondary outcomes included self-reported 7-day point prevalence of smoking abstinence; reduction of the number of cigarettes smoked per day at weeks 1, 2, 3, and 4; and program acceptability.

Results: A total of 973 people were recruited to quit smoking, of whom 262 completed basic information, 56 were excluded, and 206 were randomized and included in the final analysis. There were 189 (91.7%) men and 17 (8.3%) women, with an average age of 34.46 (SD 7.53) years and an average daily smoking rate of 15.93 (SD 7.10) cigarettes/day. We found 30 (29.7%) of the 101 participants in the intervention group and 7 (6.7%) of the 105 participants in the control group reported continuous smoking cessation after the quit date at week 4 (odds ratio 5.92, 95% CI 3.78-9.26; \( P<.001 \)). The 7-day point prevalence abstinence rate of the intervention group varied from 42.6% (43/101) to 46.5% (47/101) after 1, 2, 3, and 4 weeks, while the control group varied from 18.1% (19/105) to 26.7% (28/105). Compared to the control group, continued smokers consumed 1.5-3.0 fewer cigarettes per day in the intervention group. The overall program got positive user feedback with a high satisfaction rate (66/87, 76%) and an average Mobile Application Rating Scale user version score of 3.46.

Conclusions: Our pilot study provided preliminary evidence that the CBT-based smoking cessation smartphone app led to improved smoking quit rates versus control in Chinese smokers. The study demonstrated the CBT-based smartphone app may be an effective and feasible digital treatment model to help smokers quit, which may improve smoking cessation service quality and accessibility in China.

Trial Registration: ClinicalTrials.gov NCT04421170; https://clinicaltrials.gov/study/NCT04421170
International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-041985
Introduction

Tobacco smoking remains one of the leading causes of preventable death [1]. More than 8 million smokers worldwide die each year from smoking-related causes, of whom about 7 million die of diseases caused by smoking and about 1.2 million die from diseases caused by secondhand smoke exposure [2]. More than 1 million people in China lose their lives due to smoking every year. If effective action were not taken to significantly reduce the smoking rate, the number of deaths due to smoking would increase to 2 million per year by 2030 and 3 million per year by 2050 [3]. Since 1978, cigarette production in China has grown from 20% to over 40% of the world’s cigarette products [4]. Tobacco has caused a huge economic burden to society and individuals, including the high cost of medical care, more diseases, and premature death [5]. According to survey results on adult smoking prevalence in China in 2018, the smoking rate of people aged 15 years or older in China was 26.6% [6], which was still higher than in many other countries.

One of the key reasons for China’s low smoking cessation rate is limited smoking cessation services. The 2018 China Adult Tobacco Survey demonstrated that 90.1% of smokers who tried to quit in the past 12 months did not use any form of cessation aid [6]. A report in 2019 showed that 366 hospitals and primary health care institutions in China set up smoking cessation clinics, of which only 43% provided smoking cessation medications [7]. Few smokers were willing to go to the smoking cessation clinic for help [5], and most self-help cessation attempts failed [8,9]. Most smokers tend to relapse in the first few weeks after trying to quit smoking [10]. The success rate of quitting smoking after 1 year was 3%-5% for those without support, 7%-16% when smokers received behavioral intervention, and as high as 24% when smokers received drug treatment and behavioral help [11]. At present, it is necessary and meaningful to find accessible, effective, and scalable smoking cessation intervention methods to improve smoking cessation services in China.

Nowadays, mobile phone–based smoking cessation support provides a new channel for those who cannot get access to or lack the willingness to use face-to-face support [12]. About 45% of mobile phone subscriptions worldwide were related to smartphones, and this number continues to grow as 75% of new mobile phone sales were smartphones [13]. Around the world, mobile phones are becoming more and more useful in health information and health care service delivery [14]. Digital medical service has the advantages of economy, easy access [12], and easy promotion, which provides an opportunity for developing cost-effective smoking cessation digital interventions [15]. We previously carried out a study on smoking cessation intervention based on SMS text messaging (Happy Quit) in China, which supported the effectiveness and feasibility of digital medical services [16]. Smartphone apps for health and health care are increasing rapidly, but there are few in the smoking cessation area [17]. Smartphone apps can enable diverse functions, including audio and video materials, and can provide additional resources through the network [18]. Besides, smartphone apps could be more powerful than SMS text messaging programs for digital intervention because they have the potential to increase user engagement through diversified user interfaces and user experiences [19]. A study has found that under the same intervention content, the effect of a smartphone-based smoking cessation intervention app was stronger than that of a non–mobile device–based web page intervention. Mobile devices have the potential to make it easier for smokers to get smoking cessation support [20]. However, most of the currently available smartphone smoking cessation apps on the market have the problem of low compliance with standard clinical practice guidelines [21]. In 2018, a study of smoking cessation apps for the UK mobile phone app market found that most smoking cessation apps had low theoretical adherence, and the overall quality of smoking cessation apps was still unsatisfactory [22]. A review showed that there was insufficient evidence of the smartphone app’s effectiveness on smoking cessation support; thus, more randomized controlled trials (RCTs) were recommended to validate the smartphone-based digital smoking cessation interventions [12]. Specifically, as far as we know, there has been no mobile smoking cessation app with sufficient clinical evidence in China.

As an indispensable part of psychotherapy, cognitive behavioral therapy (CBT) plays an important role in the field of psychological and behavioral intervention [23]. One of the roles of CBT involves challenging behaviors that may trigger or sustain difficulties, such as smoking. CBT has been proven to help reduce cravings and promote smoking cessation by changing participants’ thoughts and behaviors [24]. The CBT quitting method consists of several parts, generally including preparation before quitting smoking, starting quitting smoking, and maintaining or preventing the recurrence treatment stage [25]. CBT may be a good choice for people who want to quit smoking, especially for those who want to quit smoking through nondrug methods [25,26]. Our previous study on “Happy Quit” SMS text message smoking cessation in China found that a CBT-based smoking cessation intervention can effectively improve the 24-week continuous smoking cessation rate [16,27]. As far as we know, there is little research to explore the effect of CBT-based mobile apps on smoking cessation intervention. Therefore, we are committed to designing a CBT-based mobile app for Chinese smokers who want to quit smoking, so as to find more effective ways to quit smoking.

Considering the above factors, in order to provide more feasible and effective smoking cessation services, we developed a scientific mobile smoking cessation app based on clinical practice guidelines. Smoking is an unhealthy behavior that can be altered, and CBT can help solve a wide range of smoking cessation problems. Various forms of interventions were developed in the app based on CBT to help smokers learn new...
skills, resist smoking cravings [28], and better deal with emotional disorders [29].

This Mandarin mobile smoking cessation app, based on CBT, integrates smoking cessation support and social skills training and finally achieves the goal of cognitive behavior change. A previous 1-arm study on the feasibility and acceptability of this CBT-based smartphone app was carried out for Chinese smokers who wanted to quit and showed that the smoking cessation app may become a new digital therapy model and have the potential to provide support for smoking cessation services in China [30]. We hypothesized in this study that the CBT-based smoking cessation app is feasible and acceptable and can significantly increase the quit rate.

**Methods**

**Objectives**

In the current trial, the objectives were to evaluate the feasibility and acceptability of this Chinese CBT-based app in a direct-to-participant clinical design and preliminarily evaluate the efficacy of the CBT-based app for smoking cessation in China. Given that CBT [23] is the current standard in behavioral intervention for smoking cessation, we tested the hypotheses that participation in this intervention will lead to significant improvement in the self-reported continuous smoking cessation rate at 4 weeks; self-reported 7-day point prevalence smoking abstinence and reduction of the number of cigarettes smoked per day at weeks 1, 2, 3, and 4; and that the program would be acceptable to participants.

**Study Design**

This was a randomized controlled, direct-to-participant clinical trial conducted in China. Researchers conducted preliminary conditional screenings for each participant. In the baseline assessment, all eligible participants were required to fill out a baseline questionnaire that included demographic information, motivation to quit smoking and willingness.

Using a randomization method by electronic data capture system, participants were randomly assigned to a smoking cessation app intervention group or to a control group in a 1:1 ratio after the completion of the screening, consent, and baseline questionnaires. No changes have been made to the trial design since its commencement. Participants in the intervention group received the CBT-based smoking cessation app, while those in the control group were encouraged to quit but were not provided with the CBT-based smoking cessation app. A control group was used since this was an exploratory study designed to assess both the feasibility, acceptability, and initial efficacy of the app intervention. Participants were required to complete the program acceptability assessment, including the App Satisfaction Assessment Scale and the Mobile Application Rating Scale, user version (uMARS; including engagement, functionality, aesthetics, information, and the average score of the uMARS total score), 4 weeks after the quit date. Each item is scored from 1 to 5, with a maximum total score of 5.0. The higher the score, the better the satisfaction [31]. Follow-up visits were conducted at weeks 1, 2, 3, and 4 after the participants started smoking cessation. The study design is depicted in Figure 1.
Figure 1. Flowchart of the study.

Participants
The eligibility criteria for participants are shown in Textbox 1.

Textbox 1. Eligibility criteria for participants.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cigarette smokers (smoked more than 100 cigarettes in their lifetime and currently smoke 5 or more cigarettes a day)</td>
</tr>
<tr>
<td>Aged 25 years or older</td>
</tr>
<tr>
<td>Able to read and write in Chinese</td>
</tr>
<tr>
<td>Owning a smartphone (operating system: iOS or Android)</td>
</tr>
<tr>
<td>Having experience in using apps</td>
</tr>
<tr>
<td>Expressing an interest in quitting smoking within the next month</td>
</tr>
<tr>
<td>Willing to provide informed consent to participate in the study</td>
</tr>
<tr>
<td>Able to follow up for at least 1 week</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsmokers or only use electronic cigarettes</td>
</tr>
<tr>
<td>Smokers without previous “serious” attempts to quit smoking (Motivation to Stop Scale score &lt;7)</td>
</tr>
<tr>
<td>Currently experiencing severe mental illness</td>
</tr>
<tr>
<td>Had already started their quit attempt or used any smoking cessation treatment at the time of registration</td>
</tr>
<tr>
<td>Unable to use smartphones and apps</td>
</tr>
<tr>
<td>Did not have sufficient command of Chinese to participate in the trial</td>
</tr>
</tbody>
</table>
Procedures

Sample Size
A sample size of about 200 participants was calculated to give 80% power and a 2-sided 5% significance for detecting a beneficial difference in the self-reported continuous smoking cessation rates between the CBT app intervention and the control intervention. These assumptions were made based on our previous RCT of CBT-based SMS text messaging interventions for smoking cessation [16] and other previous RCTs of smartphone app interventions for smoking cessation [15], as well as the consideration of the better efficacy of the CBT-based app than the non–CBT-based app for smoking cessation [32].

Recruitment
From February 23 to June 27, 2021, a total of 206 participants were recruited through advertisements on social media, such as WeChat (Tencent), Weibo, the website, and the principal investigator–affiliated hospital WeChat official account. Potential participants were screened through a mobile phone or WeChat phone contact. Of the 973 screened, 711 did not meet the preliminary inclusion and exclusion criteria. Of the 262 baseline assessments completed, 56 were deemed ineligible to participate. Therefore, a total of 206 participants were randomized and included in the analysis according to the intention-to-treat (ITT) principle.

Stratified Block Randomization
Randomization was conducted by an electronic data capture system, a computerized system designed for the collection of clinical data in electronic format for RCTs and other clinical trials. Randomization was at the level of the individual participant with a 1:1 ratio by the method of minimization stratified by balancing for (1) with or without previous quit attempts and (2) 10 cigarettes or more per day. The study investigators were blinded to participants’ treatment allocation until all data were collected. The investigators who analyzed the data were also blinded to participants’ allocated groups until analysis completion.

Interventions

Overview
Following randomization, participants set a smoking cessation day. On that day, participants began receiving personalized messages to help quit smoking through either a smartphone app (intervention group) or by simply receiving SMS text messages (control group). The intervention group also logged into the smoking cessation app, started a smoking cessation journey, completed tasks, and sought personalized smoking cessation assistance under the automatic guidance of the app, while the control group received regular SMS text messages.

Control Group
Participants in the control group received information to thank them for their participation and to remind them to complete their smoking status at each point.

Intervention Group
Participants from the intervention group were invited to download the CBT-based app. The app integrates cognitive-behavioral principles and tailored behavior-change skills in Chinese. More descriptive details of this smartphone app were available elsewhere [33]. Participants who were randomly assigned to the intervention group were instructed on how to access and use the app, including a schedule of intervention activities, tasks during the pre- and postpreparation stages, and expected completion dates. Participants could begin their assigned tasks immediately after completing the pretest assessment. The system limited participants to 3 tasks each day during the preparation stage. To encourage completion, SMS text message reminders were sent each week after the quit date.

Both Groups
All participants earned CNY 20 (US $2.8) phone top-ups for the completion of pre- and posttest questionnaires. Participants from both groups received a reminder by SMS text message to complete questions about their smoking status during weeks 1, 2, 3, and 4 after the quit date by electronic Patient Reported Outcomes (ePRO) software.

Outcome Measures

Participant Demographics and Smoking Behaviors at Baseline
Participant demographic and smoking behaviors at baseline, including gender, age, height, weight, education, cigarettes per day, quit history, motivation, self-efficacy to quit, quitting smoking gradually or abruptly, smoking craving, and severity of nicotine dependence, were evaluated using the Fagerstrom Test for Nicotine Dependence (FTND).

4-Week Continuous Smoking Abstinence
Continuous smoking abstinence for 4 weeks is defined as smoking no more than 5 cigarettes in the past 4 weeks since quitting, measured through response to the following item: (1) “How many cigarettes have you smoked in the past 4 weeks?” Response choices were “0,” “1–5,” or “more than 5”; and (2) “If more than 5, recorded the number of smoked cigarettes per week.” Within 4 weeks, participants who smoked more than
5 cigarettes indicated a recurrence [34]. This was the primary outcome measure.

**7-Day Point Prevalence Smoking Abstinence**

Self-reported abstinence of at least 7 days before the assessment day was assessed 1, 2, 3, and 4 weeks after each participant’s quit date. It is defined as smoking no more than 5 cigarettes in the past 1, 2, 3, and 4 weeks since quitting.

**Reduction of the Number of Cigarettes Smoked Per Day**

Daily cigarette consumption among participants who were still smoking after quitting days (participants who smoked more than 5 cigarettes in total but less than 1 cigarette per day were considered to have smoked 1 cigarette per day).

**Program Acceptability**

A questionnaire of program acceptability was assessed at 4 weeks post quitting, including the App Satisfaction Assessment Scale and the uMARS. Questions are shown in the Results section.

**Safety**

The safety of this program was evaluated by the collection and analysis of spontaneous adverse events reported by participants. No serious adverse events were reported by participants in our trial.

**Intervention Effects**

The primary outcome was the self-reported continuous smoking abstinence rates at 4 weeks after the quit date. Secondary outcomes included self-reported 7-day point prevalence smoking abstinence at weeks 1, 2, 3, and 4, reduction of the number of cigarettes smoked per day from week 1 to week 4, program acceptability, and the association of outcomes with baseline data. In all analyses, participants who dropped out or were lost to follow-up were considered treatment failures and smokers. Participants using smoking cessation methods not allowed in this study were not included in the efficacy analysis.

**Statistical Analysis**

Statistical analyses were conducted with R software (R Foundation for Statistical Computing) and SPSS (version 23; IBM Corp). All baseline data comparisons, including participant demographics and smoking behaviors, were conducted using a chi-square test. A total of 206 participants were included in the analysis according to the ITT principle. In both groups, nominal and ordinal demographics were compared using chi-square tests, and continuous variables were compared using independent samples 2-tailed t tests. The number of average cigarettes smoked per day was compared using a 2-sample t test between groups. The mixed-effects model was used to test the self-reported smoking abstinence rates in intervention groups and control groups. Program acceptability (treatment satisfaction ratings) for participants with the app intervention was evaluated by tallying the proportion of users answering each item on the usability measure. If the smoking status was not available after quitting, the participant was considered to have smoked the same number of cigarettes per day as before quitting. The odds ratio was used as a measure of outcomes in the intervention group compared to the control group. Multiple imputations will be used in chained equations. A 2-sided P<.05 was used to determine statistical significance.

**Ethical Considerations**

The study protocol was approved by the ethics committee of Sir Run Run Shaw Hospital, an affiliate of Zhejiang University School of Medicine (protocol number 20200129-33), and was published [33]. The trial was performed in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants. The purpose, procedures and measurements, potential risks, and benefits of the trial were explained to each participant before recruitment. The investigators who analyzed the data were also blinded to participants’ allocated groups until analysis completion.

**Results**

**Follow-Up Rate**

Figure 1 shows the process of screening, grouping, and follow-up of participants in this study. A total of 206 participants were recruited from February 23 to June 27, 2021. Of the 973 screened, 711 did not meet the preliminary inclusion and exclusion criteria. Overall, 262 completed baseline assessments, but 56 were deemed ineligible to participate. Therefore, a total of 206 participants were included in the analysis according to the ITT principle.

**Participant Characteristics**

Demographics and smoking characteristics at baseline for all participants are presented in Table 1. There were no statistically significant differences in the baseline characteristics of the 2 groups of participants (P>.05). A total of 206 participants were enrolled in the trial. There were 189 (91.7%) men and 17 (8.3%) women, with an average age of 34.46 (SD 7.53) years and an average daily smoking rate of 15.93 (SD 7.10) cigarettes per day.
Table 1. Baseline characteristics of study groups.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Intervention (n=101)</th>
<th>Control (n=105)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>94 (93.1)</td>
<td>95 (90.5)</td>
</tr>
<tr>
<td>Woman</td>
<td>7 (6.9)</td>
<td>10 (9.5)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>34.62 (8.03)</td>
<td>34.30 (7.04)</td>
</tr>
<tr>
<td><strong>Age (years), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>58 (57.4)</td>
<td>61 (58.1)</td>
</tr>
<tr>
<td>&gt;34</td>
<td>43 (42.6)</td>
<td>44 (41.9)</td>
</tr>
<tr>
<td><strong>Education (years), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤12</td>
<td>14 (13.9)</td>
<td>12 (11.4)</td>
</tr>
<tr>
<td>&gt;12</td>
<td>87 (86.1)</td>
<td>93 (88.6)</td>
</tr>
<tr>
<td><strong>Number of cigarettes smoked per day, mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>39 (38.6)</td>
<td>38 (36.2)</td>
</tr>
<tr>
<td>11-20</td>
<td>49 (48.5)</td>
<td>55 (52.4)</td>
</tr>
<tr>
<td>21-30</td>
<td>10 (9.9)</td>
<td>11 (10.5)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>3 (3)</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Number of previous quit attempts, mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>35 (34.7)</td>
<td>36 (34.3)</td>
</tr>
<tr>
<td>1-5 times</td>
<td>62 (62.4)</td>
<td>64 (61)</td>
</tr>
<tr>
<td>≥6 times</td>
<td>4 (4)</td>
<td>5 (4.8)</td>
</tr>
<tr>
<td><strong>FTND(^a) score, mean (SD)</strong></td>
<td>3.62 (2.40)</td>
<td>3.62 (2.19)</td>
</tr>
<tr>
<td><strong>BMI, mean (SD)</strong></td>
<td>24.39 (2.96)</td>
<td>24.54 (3.91)</td>
</tr>
<tr>
<td><strong>Smoking (years), mean (SD)</strong></td>
<td>13.85 (7.88)</td>
<td>14.50 (6.65)</td>
</tr>
<tr>
<td><strong>Motivation to quit, mean (SD)</strong></td>
<td>9.04 (1.48)</td>
<td>8.94 (1.31)</td>
</tr>
<tr>
<td><strong>Self-efficacy to quit, mean (SD)</strong></td>
<td>8.03 (2.37)</td>
<td>8.13 (1.89)</td>
</tr>
<tr>
<td><strong>VAS(^b) score, mean (SD)</strong></td>
<td>6.77 (2.53)</td>
<td>6.81 (2.16)</td>
</tr>
</tbody>
</table>

\(^a\)FTND: Fagerstrom Test for Nicotine Dependence.  
\(^b\)VAS: Visual Analogue Scale (VAS measures the degree of craving for smoking; 0-10 points, where 0 represents no craving and 10 represents strong craving.

The Outcome of Abstinence Rates

**The Primary Outcome**

Compared with 7/105 (6.7%) participants in the control group, self-reported continuous smoking abstinence at week 4 was higher (30/101, 29.7% participants) in the intervention group (odds ratio 5.92, 95% CI 3.78-9.26; \(P<.001\); Table 2).
Table 2. Primary and secondary outcomes.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Intervention participants (n=101), n (%)</th>
<th>Control participants (n=105), n (%)</th>
<th>OR (95% CI)</th>
<th>P value&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported continuous abstinence at 4 weeks</td>
<td>30 (29.7)</td>
<td>7 (6.7)</td>
<td>5.92 (3.78-9.26)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Secondary outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported 7-day point prevalence of abstinence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 week</td>
<td>47 (46.5)</td>
<td>28 (26.7)</td>
<td>2.39 (1.78-3.22)</td>
<td>.003</td>
</tr>
<tr>
<td>2 weeks</td>
<td>43 (42.6)</td>
<td>19 (18.1)</td>
<td>3.36 (2.43-4.64)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3 weeks</td>
<td>43 (42.6)</td>
<td>21 (20)</td>
<td>2.97 (2.16-4.07)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4 weeks</td>
<td>45 (44.6)</td>
<td>20 (19)</td>
<td>3.42 (2.48-4.70)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>OR: odds ratio.
<sup>b</sup>Bonferroni corrected P values.

The Secondary Outcomes

Table 2 details the results of self-reported 7-day point prevalence of smoking abstinence at weeks 1, 2, 3, and 4. It was shown that the self-reported 7-day point prevalence of smoking abstinence at weeks 1, 2, 3, and 4 of the intervention group were 46.5% (47/101), 42.6% (43/101), 42.6% (43/101), and 44.6% (45/101), respectively, while 7-day point prevalence of smoking abstinence in the control group was 26.7% (28/105), 18.1% (19/105), 20% (21/105), and 19% (20/105), respectively. The differences were statistically significant at all time points. Compared to the control group, continued smokers consumed 1.5-3.0 fewer cigarettes per day in the intervention group (Table 3).

Table 3. Cigarettes consumed per day in nonabstinent participants (only those who self-reported smoking more than 5 cigarettes [reaching the standard of relapse] after the quit date were included).

<table>
<thead>
<tr>
<th>Assessment point</th>
<th>Intervention group</th>
<th>Control group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants, n</td>
<td>Mean (SD)</td>
<td>Participants, n</td>
</tr>
<tr>
<td>Baseline</td>
<td>101</td>
<td>15.96 (7.46)</td>
<td>105</td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 week</td>
<td>54</td>
<td>8.81 (5.56)</td>
<td>77</td>
</tr>
<tr>
<td>2 week</td>
<td>64</td>
<td>9.53 (5.69)</td>
<td>94</td>
</tr>
<tr>
<td>3 week</td>
<td>68</td>
<td>9.21 (5.58)</td>
<td>95</td>
</tr>
<tr>
<td>4 week</td>
<td>71</td>
<td>9.50 (5.58)</td>
<td>98</td>
</tr>
</tbody>
</table>

The Acceptability and Feasibility of the CBT-Based Smoking Cessation App Program

Participants were required to complete the program acceptability assessment, including the App Satisfaction Assessment Scale and uMARS, 4 weeks after their quit date. All participants (n=101) in the intervention group were contacted to evaluate the program based on their experience, and we finally obtained responses from 87 people, including 87 on the App Satisfaction Assessment Scale and 84 on the uMARS. The overall program satisfaction was 76% (66/87) and the percentage of dislikes was less than 10% (9/87). The uMARS included the average scores of 5 items scored 1-5, namely, engagement (2.96), functionality (3.73), aesthetics (3.56), information (3.6), and the average score of the uMARS total score (3.46). More details are available in Tables S1 in Multimedia Appendix 1 and Table 4.

Table 4. Descriptive analysis of Mobile Application Rating Scale user version subdomains total score (n=84).

<table>
<thead>
<tr>
<th>Subdomain</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement</td>
<td>2.96 (0.76)</td>
<td>3.00 (2.60-3.40)</td>
<td>1.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Functionality</td>
<td>3.73 (0.81)</td>
<td>3.75 (3.25-4.25)</td>
<td>1.50</td>
<td>5.00</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>3.56 (0.81)</td>
<td>3.67 (3.00-4.00)</td>
<td>1.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Information</td>
<td>3.60 (0.92)</td>
<td>3.75 (3.00-4.25)</td>
<td>1.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Average of total score&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.46 (0.71)</td>
<td>3.47 (3.06-3.92)</td>
<td>1.53</td>
<td>5.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>Average of total score = total score (engagement + functionality + aesthetics + information)/4.
**Discussion**

**Principal Findings**

In this study, our main results indicate that (1) the 4-week continuous smoking cessation rate of the intervention group participants (30/101, 29.7%) was statistically significantly higher than that of control group participants (7/105, 6.7%); (2) the self-reported 7-day point prevalence smoking abstinence rate in the intervention group at the weeks 1, 2, 3, and 4 was higher than that in the control group (43/101, 42.6% to 47/101, 46.5% vs 19/105, 18.1% to 28/105, 26.7%); (3) compared with the control group, those who continued to smoke in the intervention group smoked 1.5-3.0 fewer cigarettes a day; (4) the app had high overall satisfaction rating (66/87, 76%) and average uMARS scores (3.46). These findings suggest that the CBT-based smartphone app is effective for smoking cessation, and it could be an effective, feasible, and easy-to-access quitting app in China. It is hopeful for this app to be introduced to a large-scale Chinese population to intervene in smoking cessation.

The effectiveness of this CBT-based smartphone app intervention on short-term smoking cessation rates was encouraging. About half (47/101, 46.5%) of participants in the intervention group and about a quarter (28/105, 26.7%) in the control group reported 7-day abstinence in the first week after quitting, which was comparable to the preliminary results of other randomized controlled digital intervention studies on smoking cessation [35,36]. Previous research results for smartphone apps and SMS text messaging smoking cessation programs were similar to ours [16,37]. Participants in the intervention group reported point prevalence that was 3 to 4 times higher than that in the control group from 1 to 4 weeks, and self-reported continuous abstinence at 4 weeks in the intervention group was similar to that in previous smartphone smoking cessation app studies [37,38].

The results of this study show that the smoking cessation rate is high, and there have been some similar or inconsistent studies in the past, although there are some differences in smoking cessation standards. A systematic evaluation of smartphone apps for quitting smoking found that the results of 11 RCTs were not consistent. Among these 11 studies, the abstinence rate of the test group was significantly higher than that of the control group in 4 studies, the abstinence rate of the test group was not significantly higher than that of the control group in 5 studies, and there was no difference between the test and control groups in 2 studies [1]. The social environment in China is still quite tolerant of smokers and appears less supportive of quitting behavior, so potential quitters in China may be more resistant to personal intervention than smokers in other countries [16]. As far as we know, there are no mobile smoking cessation platforms developed under the theoretical guidance of professional medical institutions in the domestic software development market, and most of the smoking cessation apps in China are developed by individuals or companies that may not have knowledge of clinical smoking cessation practice. Those apps comply less with the recommendations from the China clinical smoking cessation guidelines (2015 edition), and their smoking cessation effect is limited [21]. Therefore, for most Chinese smokers, using apps to quit smoking is a new way to quit smoking that needs further improvement.

China has introduced some smoking control measures, but the choice of a mobile smoking cessation service is relatively limited. The earliest method of smoking cessation based on mobile phones was mainly SMS text messages [39]. Our team previously carried out an SMS text message smoking cessation project, and the effect of smoking cessation was relatively significant [16]. China’s smoking cessation services are inadequate. An authoritative national survey found that in 31 provinces, only 366 hospitals and primary care centers offered smoking cessation clinics [7]. Among them, smoking cessation clinics are mostly attached to tertiary or secondary hospitals, but population-level interventions rely heavily on primary care settings [40], where fewer clinics exist. In addition, the limited use of smoking cessation medications in China is also a serious problem [7]. Additionally, many smokers who want to quit smoking have low awareness of and use of smoking cessation support services [41].

It is noteworthy that the overall program satisfaction (66/87, 76%) and the average score of the uMARS total score (3.46) were high. About 80% (69/87) of the participants would happily recommend the app to others, which was consistent with the results of a randomized clinical trial of a smartphone app based on acceptance and commitment therapy for quitting smoking [37]. In this study, except for engagement, the software’s functionality, aesthetics, information, and the average score of the uMARS total score were all around 3.5, indicating that the participant’s overall satisfaction level was relatively high. Similar to our results, a study showed that participants made suggestions, including making features more ramified and integrating with some social media platforms to increase app and user interactivity, although some smartphone smoking cessation apps were widely accepted [41].

This pilot study provided valuable learning for future research. A review study showed that the certainty of evidence comparing smartphone apps to very low-intensity smoking cessation support was very low, and more RCTs were needed to test these interventions [12]. Also, the recruitment of participants in the intervention and control groups of RCTs related to smartphone apps for smoking cessation was not easily supported [41]. It is worth mentioning that we recruited eligible participants in this preliminary study and carried out RCTs in accordance with the norms, obtaining relatively accurate data and also providing a reference and basis for subsequent RCT studies. These preliminary data still provide sufficient and reliable information for our subsequent larger, robust trials.

**Limitations and Strengths**

There are some limitations to our study. First, this study was a pilot study with a small sample size, and the results may be biased. At present, our large-sample trial has begun, which will help to further confirm our preliminary research results. Second, due to the need to master the basic ability to use the app, the enrolled smokers were relatively younger and had a higher education level, so the smoking cessation effect of the smoking cessation APP is uncertain for older or less educated smokers.
and more caution needs to be exercised if the results are
generalized to older or less educated smokers. Finally, some
important subgroup analyses, such as smoking cessation rates
between men and women, tobacco users, and e-cigarette users,
were not considered in this study. Although we matched the 2
groups of quitters on factors such as gender and smoking status,
these differences could potentially have an impact on smoking
cessation outcomes. Despite these limitations, this study’s
strengths included its conduct in China, the direct-to-participant
design, and the use of an RCT.

Conclusions
This pilot study of a CBT-based smartphone smoking cessation
app provided preliminary evidence that the app led to improved
smoking cessation rates versus control in Chinese smokers. This
study demonstrated that the CBT-based smartphone app may
be an effective and feasible digital treatment model to quit
smoking, which may improve smoking cessation services in
China.

Acknowledgments
We sincerely thank all participants, their families, and all workers for their support. We are grateful to Johnson & Johnson
Pharmaceutical Company for providing critical comments and suggestions for the study design and development of the digital
intervention app. The source of funding had no impact on the data collection and analysis of the study or the writing and submission
of the manuscript.

The research is supported by Johnson & Johnson Pharmaceutical Company (K-20201478). The funder was collaboratively
involved in the study design, development of the data collection tool, and app software development, but the company had no
involvement in enrollment, providing an informed consent form, contacting participants, addressing safety issues, or monitoring
and collecting data.

Data Availability
All data in this study will be available from the corresponding author on reasonable request and upon completion of the data user
agreement.

Authors' Contributions
YL and JT developed and designed the study, and SC collected and analyzed the data. SC and YL took the lead in drafting the
manuscript protocol, with contributions by JT, who advised on the study design and coordinated study approval. YL and JT read
and proposed critical comments, as well as approved the manuscript for publication. All authors read and approved the final
manuscript.

Conflicts of Interest
YL received funding from Johnson & Johnson Pharmaceutical Company for the study. SC, JT, CW, GZ, IZ, and YL have no
potential conflicts of interest to declare.

Multimedia Appendix 1
Questions for assessing program satisfaction.
[DOCX File, 17 KB - formative_v8i1e48050_app1.docx ]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1408 KB - formative_v8i1e48050_app2.pdf ]

References
1. Chu KH, Matheny SJ, Escobar-Viera CG, Wessel C, Notier AE, Davis EM. Smartphone health apps for tobacco cessation:
2023-12-07]
URL: http://www.tobaccoatlas.org/ [accessed 2014-10-07]


Abbreviations

CBT: cognitive behavioral therapy
ePRO: electronic Patient Reported Outcomes
FTND: Fagerstrom Test for Nicotine Dependence
ITT: intention to treat
RCT: randomized controlled trial
uMARS: Mobile Application Rating Scale user version

Caitlin M Drover\(^1\), MPH; Adam S Elder\(^1\), PhD; Brandon L Guthrie\(^{1,2}\), PhD; Debra Revere\(^3\), MLIS; Nicole L Briggs\(^4\), MPH; Laura M West\(^4\), MS; Amanda Higgins\(^4\), MSHI; William B Lober\(^{2,5,6}\), MD; Bryant T Karras\(^4\), MD; Janet G Baseman\(^1\), PhD

\(^1\)Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, United States
\(^2\)Department of Global Health, School of Public Health, University of Washington, Seattle, WA, United States
\(^3\)Department of Health Systems and Population Health, School of Public Health, University of Washington, Seattle, WA, United States
\(^4\)Department of Health, Washington State, Tumwater, WA, United States
\(^5\)Department of Biobehavioral Nursing and Health Informatics, School of Nursing, University of Washington, Seattle, WA, United States
\(^6\)Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

Corresponding Author:
Adam S Elder, PhD
Department of Epidemiology
School of Public Health
University of Washington
Hans Rosling Center for Population Health
3980 15th Avenue NE
Seattle, WA, 98195-1616
United States
Phone: 1 5056159180
Email: eldera3@uw.edu

Abstract

Background: WA Notify was Washington State’s smartphone-based COVID-19 digital exposure notification (EN) tool, which was used to help limit the spread of COVID-19 between November 30, 2020, and May 11, 2023. Following the 2022 Washington State Public Health Association Annual Conference, attendees who had WA Notify activated began receiving ENs alerting them to a possible COVID-19 exposure during the conference. A survey was emailed to all conference attendees to measure WA Notify adoption, mechanisms through which attendees received ENs, and self-reported engagement in protective behaviors postexposure.

Objective: This study aimed to learn more about the experiences of WA Notify adopters and nonadopters who may have been exposed to COVID-19 at a large group gathering.

Methods: A web-based survey administered through REDCap (Research Electronic Data Capture; Vanderbilt University) was sent to all attendees of the Washington State Public Health Association conference. Self-reported demographic information and characteristics of respondents were summarized. Regression models were used to estimate relative risks to compare WA Notify adoption and testing behaviors between groups.

Results: Of the 464 total registered attendees who were sent the survey, 205 (44%) responses were received; 201 eligible attendees were included in this analysis. Of those, 149 (74%) respondents reported having WA Notify activated on their phones at the time of the conference. Among respondents with WA Notify activated, 54% (n=77) reported learning of their potential exposure from a WA Notify EN. Respondents who reported that they did not have WA Notify activated and learned of their potential exposure via the event-wide email from conference organizers were 39% less likely to test for COVID-19 compared to respondents with WA Notify activated who learned of their potential exposure from the email (relative risk 0.61, 95% CI 0.40-0.93; \(P=.02\)), and this gap was even larger when compared to respondents who learned of their exposure from a WA Notify EN. The most commonly cited reason for not having WA Notify activated was privacy concerns (n=17, 35%), followed by not wanting to receive ENs (n=6, 12%) and being unaware of WA Notify (n=5, 10%).

Conclusions: Digital EN systems are an important tool to directly and anonymously notify close contacts of potential exposures and provide guidance on the next steps in a timely manner. Given the privacy concerns, there is still a need for increasing
transparency surrounding EN technology to increase uptake by the public if this technology were to be used in the future to slow the spread of communicable diseases.

*JMIR Form Res 2024;8:e50716*  doi: [10.2196/50716](https://doi.org/10.2196/50716)

**KEYWORDS**
COVID-19; exposure notification; digital public health tool; survey analysis; conference; online survey; digital tool; public health; contact tracing

### Introduction

#### Background

Smartphone-based COVID-19 digital exposure notification (EN) tools were developed at the beginning of the COVID-19 pandemic to help limit the spread of the SARS-CoV-2 virus. These tools alert users of possible exposure to COVID-19 based on Bluetooth-based proximity to other users’ devices and other users informing the system of a positive COVID-19 diagnosis [1]. According to the Association of Public Health Laboratories, which helped support EN systems across the United States, 28 states have implemented a digital EN tool, including Washington State, which launched its EN system, WA Notify on November 30, 2020 [2,3]. Like most of these tools, WA Notify was built using the Google-Apple Exposure Notification (GAEN) Bluetooth system and its GAEN Express platform [1-4]. GAEN-based digital COVID-19 EN tools preserve the users’ privacy; adoption or activation of the tool is anonymous, users and their locations are not tracked, and data protections ensure ENs are distributed anonymously. As of April 11, 2023, WA Notify had been activated on smartphones over 3.92 million times. During the period in which WA Notify was operating, approximately 230,000 positive COVID-19 cases were reported to the system and an estimated 1.94 million ENs were shown to users [5].

The effectiveness of EN tools like WA Notify is presumed to include providing rapid notification of possible exposure to users [6] and using effective strategies to encourage individuals shown an EN to engage in behaviors to reduce further spread, such as testing and isolating [7,8]. While direct observation of WA Notify users’ protective behaviors after being shown an EN is not possible, we can ask users who were shown an EN about their anticipated and completed protective behaviors through optional, anonymous surveys. The results from these surveys suggest that ENs do influence behavior, and it has been estimated that digital EN tools such as WA Notify have averted thousands of new COVID-19 cases in the states in which they have been deployed and widely adopted [9-11].

In October 2022, the Washington State Public Health Association (WSPHA) held its Annual Conference. Several conference attendees tested positive for COVID-19 either during or following the conference. In the week following, WA Notify users who attended the conference began receiving ENs of possible exposure during the time period when they were attending the conference. This event presented a unique opportunity to capture WA Notify user interactions in the context of a large public gathering as well as the experiences of public health professionals—a population potentially more likely to both adopt digital EN tools and engage in protective behaviors—undertaking protective health behaviors in response to receiving an EN. We describe the results of a postconference web-based survey emailed to all attendees to measure WA Notify adoption, mechanisms through which attendees were alerted to a possible COVID-19 exposure (digital EN, personal communication, etc), and self-report of engagement in protective behaviors after learning of the exposure.

#### Objectives

We aimed to learn more about the experiences of WA Notify adopters and nonadopters who may have been exposed to COVID-19 at a large group gathering.

#### Methods

##### How WA Notify Works

WA Notify (and other similar EN systems) can be activated either through the settings on iPhones or downloaded as an app on Android phones [3,5]. When in close proximity to another activated phone, the devices will exchange random, anonymous cryptographic Bluetooth keys. If the owner of either device later tests positive for COVID-19, they may anonymously confirm their positive result using the tool. Depending on the length and proximity of the phones’ interaction, ENs will be shown on phones belonging to individuals who have exchanged Bluetooth keys with the infected individual. While ENs may be displayed on phones for exposures happening up to 10 days earlier, the average time between exposure and the EN alert was 4.6 days in October 2022. The examples of EN messages from WA Notify can be found in Multimedia Appendix 1.

##### Survey Instrument

A web-based survey was designed to capture WA Notify engagement (adoption and reasons for not having WA Notify activated), awareness that someone who attended the conference later tested positive for COVID-19, respondents’ COVID-19 status (presence of symptoms, testing, and result), vaccination status, and demographics (race or ethnicity, age, gender, and state residency). The survey was piloted internally to assess clarity, question order, and timing. The final survey was programmed and administered through REDCap (Research Electronic Data Capture) [12]. A copy of the survey can be found in Multimedia Appendix 2.

Survey invitations were distributed via email to a list of all WSPHA conference attendees on October 28, 2022, a total of 15 days after the conference, by the conference organizer. Reminders were sent on November 6, 2022, and the survey closed on November 10, 2022.
Inclusion Criteria
Only respondents who attended the WSPHA Annual Conference on October 11 through 13, 2022, were included. Reports of ENs distributed outside of October 11 through 23 were excluded since these ENs fall outside of the 10-day window for ENs plausibly resulting from interactions at the WSPHA conference.

Statistical Analysis
Survey responses were tallied and stratified by the primary outcome of interest, WA Notify activation status. Variables of interest included how respondents learned that someone who attended the conference later tested positive for COVID-19, the development of COVID-19 symptoms, engagement in protective behaviors after being shown an EN, and vaccination status. Univariate Poisson regression models were used to estimate relative risks (RRs) of having WA Notify activated and postconference testing behavior between groups.

Ethical Considerations
The University of Washington institutional review board determined that this project was a public health quality improvement or surveillance project and nonhuman subjects research. The data used in this study were obtained from a voluntary survey that did not provide any compensation for participation and did not collect any personally identifiable information.

Results
Respondent Demographics and Characteristics

Demographics
Of the 464 total registered attendees who were sent the survey, 205 (44%) responses were received. Four surveys were excluded because the respondents did not attend the conference in person. A total of 201 completed surveys were included in the analysis, of which 149 (74%) reported having WA Notify active on their phones during the conference (Table 1).

The majority of the sample identified as female (n=158, 79%), White (n=156, 78%), and non-Hispanic, any race (n=174, 87%). Sixty-four percent (n=128) resided in Western Washington and 92% (n=184) reported having completed the primary COVID-19 vaccination series and receiving at least 1 booster.
Table 1. Respondent characteristics by WA Notify activation status (N=201).

<table>
<thead>
<tr>
<th>WA Notify activated</th>
<th>Yes (n=149)</th>
<th>No (n=49)</th>
<th>Total4 (N=201)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23 (15.4)</td>
<td>9 (18.4)</td>
<td>33 (16.4)</td>
</tr>
<tr>
<td>Female</td>
<td>119 (79.9)</td>
<td>37 (75.5)</td>
<td>158 (78.6)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>3 (2)</td>
<td>1 (2)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>3 (2)</td>
<td>0 (0)</td>
<td>3 (1.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.7)</td>
<td>2 (4.1)</td>
<td>3 (1.5)</td>
</tr>
<tr>
<td><strong>Raceb, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>4 (2.7)</td>
<td>2 (4.1)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>Asian</td>
<td>18 (12.1)</td>
<td>1 (2)</td>
<td>19 (9.5)</td>
</tr>
<tr>
<td>Black</td>
<td>9 (6)</td>
<td>3 (6.1)</td>
<td>12 (6)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>White</td>
<td>113 (75.8)</td>
<td>41 (83.7)</td>
<td>156 (77.6)</td>
</tr>
<tr>
<td>Other</td>
<td>8 (5.4)</td>
<td>0 (0)</td>
<td>9 (4.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>4 (2.7)</td>
<td>4 (8.2)</td>
<td>8 (4)</td>
</tr>
<tr>
<td><strong>Hispanic ethnicityc, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>16 (10.7)</td>
<td>4 (8.2)</td>
<td>21 (10.4)</td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>130 (87.2)</td>
<td>42 (85.7)</td>
<td>174 (86.6)</td>
</tr>
<tr>
<td>Missing</td>
<td>3 (2)</td>
<td>3 (6.1)</td>
<td>6 (3)</td>
</tr>
<tr>
<td><strong>Region, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Washington</td>
<td>26 (17.4)</td>
<td>15 (30.6)</td>
<td>41 (20.4)</td>
</tr>
<tr>
<td>Western Washington</td>
<td>103 (69.1)</td>
<td>24 (49)</td>
<td>128 (63.7)</td>
</tr>
<tr>
<td>Missing</td>
<td>20 (13.4)</td>
<td>10 (20.4)</td>
<td>32 (15.9)</td>
</tr>
<tr>
<td><strong>Work from home status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All of the time</td>
<td>61 (40.9)</td>
<td>21 (42.9)</td>
<td>84 (41.8)</td>
</tr>
<tr>
<td>Most of the time</td>
<td>42 (28.2)</td>
<td>7 (14.3)</td>
<td>50 (24.9)</td>
</tr>
<tr>
<td>Some of the time</td>
<td>34 (22.8)</td>
<td>14 (28.6)</td>
<td>48 (23.9)</td>
</tr>
<tr>
<td>Not at all</td>
<td>12 (8.1)</td>
<td>7 (14.3)</td>
<td>19 (9.5)</td>
</tr>
<tr>
<td><strong>Vaccination status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unvaccinated</td>
<td>2 (1.3)</td>
<td>0 (0)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Primary series only</td>
<td>6 (4)</td>
<td>8 (16.3)</td>
<td>15 (7.5)</td>
</tr>
<tr>
<td>Primary series+booster</td>
<td>141 (94.6)</td>
<td>41 (83.7)</td>
<td>184 (91.5)</td>
</tr>
<tr>
<td><strong>Days since the last dose</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>124 (152)</td>
<td>154 (191)</td>
<td>134 (165)</td>
</tr>
</tbody>
</table>
**WA Notify Adoption**

Reported WA Notify adoption was highest among younger age groups, with 86% (n=55) of those 18-34 years old having WA Notify activated on their smartphones compared to 38% (n=3) of those 65 years and older (Table 2). We found that those aged 35-44 years were 22% less likely to have WA Notify activated than those aged 18-34 years (RR 0.78, 95% CI 0.63-0.97; P=.03).

Reported adoption also varied across racial groups. Respondents who identified as Asian were 29% more likely to have WA Notify activated compared to those who did not identify as Asian (RR 1.29, 95% CI 1.13-1.49; P<.001). Respondents who identified as “other” race were 35% more likely to have WA Notify activated compared to those who did not identify as “other” (RR 1.35, 95% CI 1.24-1.47; P<.001). We found no statistically significant association between WA Notify adoption and Hispanic ethnicity of any race.

Respondents living in Eastern Washington were 22% less likely to have WA Notify activated compared to those living in Western Washington (RR 0.78, 95% CI 0.61-1.00; P=.05).

---

<table>
<thead>
<tr>
<th>WA Notify activated</th>
<th>Yes (n=149)</th>
<th>No (n=49)</th>
<th>Total (N=201)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median (IQR)</td>
<td>42.0 (0-665)</td>
<td>48.5 (0-772)</td>
<td>43.5 (0-772)</td>
</tr>
<tr>
<td>Missing, n (%)</td>
<td>6 (4)</td>
<td>3 (6.1)</td>
<td>9 (4.5)</td>
</tr>
</tbody>
</table>

*a* Three respondents reported “I’m not sure” to whether they had WA Notify activated on their phones.

*b* Respondents could select more than 1 answer and percentages can add to over 100%.

*c* Hispanic origin is considered here as a grouping often considered by public health researchers and the United States Census Bureau.
Table 2. Respondent characteristics associated with having WA Notify activated (n=198).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>55/64 (86)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td>35-44</td>
<td>35/52 (67)</td>
<td>0.78 (0.63-0.97)</td>
<td>.03 *</td>
</tr>
<tr>
<td>45-54</td>
<td>30/39 (77)</td>
<td>0.90 (0.73-1.09)</td>
<td>.27</td>
</tr>
<tr>
<td>55-64</td>
<td>26/34 (77)</td>
<td>0.89 (0.72-1.10)</td>
<td>.28</td>
</tr>
<tr>
<td>65+</td>
<td>3/8 (38)</td>
<td>0.44 (0.18-1.07)</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23/32 (72)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td>Female</td>
<td>119/156 (76)</td>
<td>1.06 (0.84-1.34)</td>
<td>.62</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>3/4 (75)</td>
<td>1.04 (0.57-1.91)</td>
<td>.89</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>4/6 (67)</td>
<td>0.88 (0.50-1.56)</td>
<td>.67</td>
</tr>
<tr>
<td>Asian</td>
<td>18/19 (95)</td>
<td>1.29 (1.13-1.49)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Black</td>
<td>9/12 (75)</td>
<td>1.00 (0.71-1.40)</td>
<td>.98</td>
</tr>
<tr>
<td>White</td>
<td>113/154 (73)</td>
<td>0.90 (0.76-1.06)</td>
<td>.21</td>
</tr>
<tr>
<td>Other</td>
<td>8/8 (100)</td>
<td>1.35 (1.24-1.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Hispanic ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>130/172 (76)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td>Hispanic</td>
<td>16/20 (80)</td>
<td>1.06 (0.84-1.34)</td>
<td>.64</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Washington</td>
<td>103/127 (81)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td>Eastern Washington</td>
<td>26/41 (63)</td>
<td>0.78 (0.61-1.00)</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Work from home status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All of the time</td>
<td>61/82 (74)</td>
<td>1.18 (0.82-1.70)</td>
<td>.38</td>
</tr>
<tr>
<td>Most of the time</td>
<td>42/49 (86)</td>
<td>1.36 (0.95-1.95)</td>
<td>.10</td>
</tr>
<tr>
<td>Some of the time</td>
<td>34/48 (71)</td>
<td>1.12 (0.76-1.65)</td>
<td>.56</td>
</tr>
<tr>
<td>Not at all</td>
<td>12/19 (63)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td><strong>Vaccination status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unvaccinated</td>
<td>2/2 (100)</td>
<td>N/A c</td>
<td>N/A</td>
</tr>
<tr>
<td>Primary series only</td>
<td>6/14 (43)</td>
<td>0.55 (0.30-1.02)</td>
<td>.06</td>
</tr>
<tr>
<td>Primary series+booster</td>
<td>141/182 (78)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
</tbody>
</table>

*Italic formatting in this table indicates that the β coefficient from the corresponding regression is significantly different from 0 at the P=.05 level.

b Respondents could select more than 1 answer. The reference category for each comparison was those who did not select each respective response.

**COVID-19 Vaccination Status**
Only 2 respondents reported being unvaccinated (Table 2). Respondents who reported receiving only the primary vaccination series were 45% less likely to have WA Notify activated than those with at least 1 booster, although this association was not statistically significant (RR 0.55, 95% CI 0.30-1.02; P=.06).

**EN Receipt**
The WSPHA Annual Conference took place from October 11 through 13, 2022, in Wenatchee, WA. From October 13 through October 23, there were 42 conference attendees who reported receiving their first WA Notify EN. Figure 1 illustrates the dates of the first reported WA Notify EN among participants who remember the date of the EN and who had WA Notify activated during this period. We do not know when attendees tested positive or confirmed their positive tests using WA Notify.
However, due to the 10-day storage period for keys used to determine exposure, the plausible dates for an EN shown due to a potential exposure during the conference were between October 11 and 23. Three respondents reported ENs outside of this range (October 2, 10, and 31), and these ENs were not included in the analysis.

The only EN dates collected with this survey were for the first WA Notify EN shown to those who reported having WA Notify activated and who reported being aware that someone who attended the conference later tested positive for COVID-19. Dates for other ENs were unavailable, except for the WSPHA conference organizer email EN sent to all attendees on October 17.

**Figure 1.** The plausible dates for an exposure notification shown because of potential exposure during the conference were between October 11 and 23 due to the 10-day storage period for keys used to show exposure notifications. Data were only available for respondents with WA Notify activated who remembered the date of the first WA Notify exposure notification received. Three respondents reported exposure notifications outside of this range (October 2, 10, and 31), and these exposure notifications were not included in the analysis. WSPHA: Washington State Public Health Association.

**Symptoms, Testing, and WA Notify Adoption**

Following the conference, 17% (n=35) of respondents reported experiencing at least 1 symptom and 63% (n=126) took a test for COVID-19 (Table 3).

Among those who tested, 92.1% (n=116) reported taking an “at-home” rapid antigen test for COVID-19, while the others received a polymerase chain reaction test. Testing was most frequently prompted by receiving an EN from WA Notify (n=58, 29%) or email notification from WSPHA conference organizers (n=56, 28%), although many respondents also reported testing being a routine behavior when they travel (n=44, 22%). Respondents who reported having WA Notify activated were 74% more likely to test for COVID-19 following the conference compared to those who did not have WA Notify activated (RR 1.74, 95% CI 1.22-2.47; P=.002). Nine (7%) respondents reported testing positive following the conference.
Table 3. Respondent testing behaviors and postconference health outcomes by WA Notify activation status (N=201).

<table>
<thead>
<tr>
<th>WA Notify activated, n (%)</th>
<th>Yes (n=149)</th>
<th>No (n=49)</th>
<th>Total(^a) (N=201)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptoms postconference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>112 (75.2)</td>
<td>35 (71.4)</td>
<td>148 (73.6)</td>
</tr>
<tr>
<td>1</td>
<td>8 (5.4)</td>
<td>3 (6.1)</td>
<td>11 (5.5)</td>
</tr>
<tr>
<td>2+</td>
<td>19 (12.8)</td>
<td>4 (8.2)</td>
<td>24 (11.9)</td>
</tr>
<tr>
<td>Missing</td>
<td>10 (6.7)</td>
<td>7 (14.3)</td>
<td>18 (9)</td>
</tr>
<tr>
<td><strong>Tested postconference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>105 (70.5)</td>
<td>20 (40.8)</td>
<td>126 (62.7)</td>
</tr>
<tr>
<td>No</td>
<td>43 (28.9)</td>
<td>29 (59.2)</td>
<td>74 (36.8)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.7)</td>
<td>0 (0)</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td><strong>Test reason(^b,c)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WA Notify EN(^d)</td>
<td>56 (37.6)</td>
<td>2 (4.1)</td>
<td>58 (28.9)</td>
</tr>
<tr>
<td>Conference EN</td>
<td>44 (29.5)</td>
<td>12 (24.5)</td>
<td>56 (27.9)</td>
</tr>
<tr>
<td>Personal EN</td>
<td>14 (9.4)</td>
<td>1 (2)</td>
<td>15 (7.5)</td>
</tr>
<tr>
<td>Symptoms</td>
<td>13 (8.7)</td>
<td>3 (6.1)</td>
<td>16 (8)</td>
</tr>
<tr>
<td>Routine</td>
<td>38 (25.5)</td>
<td>6 (12.2)</td>
<td>44 (21.9)</td>
</tr>
<tr>
<td>Other</td>
<td>11 (7.4)</td>
<td>2 (4.1)</td>
<td>14 (7)</td>
</tr>
<tr>
<td><strong>Test method(^c)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home rapid test</td>
<td>96 (91.4)</td>
<td>19 (95)</td>
<td>116 (92.1)</td>
</tr>
<tr>
<td>Testing center</td>
<td>9 (8.6)</td>
<td>1 (5)</td>
<td>10 (7.9)</td>
</tr>
<tr>
<td><strong>Tested positive(^e)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7 (6.7)</td>
<td>1 (5)</td>
<td>9 (7.1)</td>
</tr>
<tr>
<td>No</td>
<td>98 (93.3)</td>
<td>19 (95)</td>
<td>117 (92.9)</td>
</tr>
</tbody>
</table>

\(^a\)Three respondents reported “I'm not sure” to whether they had WA Notify activated on their phones.

\(^b\)Respondents could select more than 1 answer and percentages can add to over 100%.

\(^c\)Among respondents who reported testing for COVID-19 following the conference (n=126).

\(^d\)EN: exposure notification.

Pathways for Receiving First EN

Most respondents (n=194, 97%) reported that they were aware that someone who attended the conference later tested positive for COVID-19 (Table 4). Respondents learned about the attendees who tested positive for COVID-19 through a variety of pathways. Among those who were aware, respondents most frequently reported learning of the COVID-19–positive attendees via an event-wide email notification from WSPHA conference organizers (n=93, 48%) or receiving a WA Notify EN (n=77, 40%). A small number of respondents learned about the COVID-19–positive attendees via personal communication from another attendee who tested positive for COVID-19 (n=7, 4%), communication from someone else who had been exposed (n=6, 3%), or an internal email notification from the Washington State Department of Health (n=4, 2%). Among respondents with WA Notify activated, 54% (n=77) reported learning of their exposure from WA Notify, while 33% (n=48) learned of their exposure from the conference organizer email. It is possible that some portion of the 33% (n=48) of respondents with WA Notify activated who received the email notification first were not considered exposed based on Bluetooth proximity or the risk score used in WA Notify. Alternatively, this group may have also been shown a WA Notify EN after the email notification, but data were not collected on additional notifications.
Table 4. Awareness of conference attendees who later tested positive for COVID-19 by WA Notify activation status (N=201).

<table>
<thead>
<tr>
<th>Aware of COVID-19–positive attendees</th>
<th>WA Notify activated, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=149)</td>
</tr>
<tr>
<td>Yes</td>
<td>144 (96.6)</td>
</tr>
<tr>
<td>No</td>
<td>5 (3.4)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\textsuperscript{b}Three respondents reported “I’m not sure” to whether they had WA Notify activated on their phones.

**How COVID-19–positive attendees first heard about**\textsuperscript{b}

<table>
<thead>
<tr>
<th></th>
<th>Yes (n=194)</th>
<th>No (n=149)</th>
<th>Total\textsuperscript{a} (N=201)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email from WSPHA\textsuperscript{c}</td>
<td>48 (33.3)</td>
<td>42 (91.3)</td>
<td>93 (48.2)</td>
</tr>
<tr>
<td>WA Notify EN\textsuperscript{d}</td>
<td>77 (39.9)</td>
<td>0 (0)</td>
<td>77 (39.9)</td>
</tr>
<tr>
<td>Personal communication from a COVID-19–positive attendee</td>
<td>7 (4.9)</td>
<td>0 (0)</td>
<td>7 (3.6)</td>
</tr>
<tr>
<td>Notified by a contact tracer</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Other\textsuperscript{e}</td>
<td>12 (8.3)</td>
<td>4 (8.7)</td>
<td>16 (8.3)</td>
</tr>
</tbody>
</table>

\textsuperscript{b}Among respondents who were aware that someone who attended the conference later tested positive for COVID-19 (n=194).

\textsuperscript{c}WSPHA: Washington State Public Health Association.

\textsuperscript{d}EN: exposure notification.

\textsuperscript{e}Reasons included communication from someone else who was exposed (n=6), Washington State Department of Health email notification (n=4), and cannot remember (n=2).

**Associations Among Pathway of First Received EN, Symptoms, Vaccination History, and Testing**

Respondents who reported that they did not have WA Notify activated and learned that someone who attended the conference later tested positive for COVID-19 via the email from conference organizers were 39% less likely to take a test for COVID-19 than respondents with WA Notify activated who learned of their potential exposure from the email (RR 0.61, 95% CI 0.40-0.93; \( P=.02; \) Table 5).

Respondents with WA Notify activated who learned of their potential exposure from a WA Notify EN were slightly more likely to test compared to those with WA Notify who received the email notification first, although this association was not significant (RR 1.14, 95% CI 0.90-1.45; \( P=.28 \)).

The number of symptoms experienced was associated with testing for COVID-19 following the conference. Respondents who reported experiencing 1 symptom were 62% more likely to test than those who experienced no symptoms (RR 1.62, 95% CI 1.28-2.05; \( P<.001 \)), and those who experienced 2 or more symptoms were 78% more likely to test than those who experienced no symptoms (RR 1.78, 95% CI 1.55-2.06; \( P<.001 \); Table 5). With the exception of 1 person who reported a sore throat, everyone who reported symptoms also reported testing for COVID-19.
Table 5. Associations among the method of receiving an EN\textsuperscript{a}, symptoms, vaccination history, and testing postconference (N=201).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Postconference testing, n/N (%)</th>
<th>RR\textsuperscript{b} (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How COVID-19–positive attendees first heard about\textsuperscript{c}</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WA Notify EN</td>
<td>58/77 (75)</td>
<td>1.14 (0.90-1.45)</td>
<td>.28</td>
</tr>
<tr>
<td>Email from WSPHA\textsuperscript{d} conference organizers—WA Notify activated</td>
<td>31/47 (66)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td>Email from WSPHA conference organizers—WA Notify not activated</td>
<td>17/42 (41)</td>
<td>0.61 (0.40-0.93)</td>
<td>.02\textsuperscript{e}</td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>83/148 (56)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td>1</td>
<td>10/11 (91)</td>
<td>1.62 (1.28-2.05)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2+</td>
<td>23/23 (100)</td>
<td>1.78 (1.55-2.06)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Vaccination status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unvaccinated</td>
<td>0/2 (0)</td>
<td>N/A\textsuperscript{f}</td>
<td>N/A</td>
</tr>
<tr>
<td>Primary series only</td>
<td>7/15 (47)</td>
<td>0.72 (0.41-1.25)</td>
<td>.24</td>
</tr>
<tr>
<td>Primary series+booster</td>
<td>119/183 (65)</td>
<td>1 (Reference)</td>
<td>Reference</td>
</tr>
<tr>
<td><strong>Days since last the vaccine dose</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-day difference</td>
<td>N/A</td>
<td>0.98 (0.96-1.01)</td>
<td>.13</td>
</tr>
</tbody>
</table>

\textsuperscript{a}EN: exposure notification.
\textsuperscript{b}RR: relative risk.
\textsuperscript{c}Among respondents who were aware that someone who attended the conference later tested positive for COVID-19 (n=194).
\textsuperscript{d}WSPHA: Washington State Public Health Association.
\textsuperscript{e}Italic formatting in this table indicates that the \( \beta \) coefficient from the corresponding regression is significantly different from 0 at the \( P=0.05 \) level.
\textsuperscript{f}N/A: not available.

Associations Between Pathway of First Received EN and Developing COVID-19

No statistically significant association was found between how respondents learned of the conference attendees who later tested positive for COVID-19 and whether the respondent developed COVID-19 using the Fisher exact test (\( P=0.21 \); Table 6).

Table 6. Associations between the method of receiving an EN\textsuperscript{a} and developing COVID-19 among those who were aware of COVID-19–positive attendees and tested following the conference (N=120).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values, n/N (%)</th>
<th>P value\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How COVID-19–positive attendees first heard about\textsuperscript{c}</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email from WSPHA\textsuperscript{d} conference organizers</td>
<td>6/49 (12)</td>
<td></td>
</tr>
<tr>
<td>WA Notify EN</td>
<td>2/58 (3)</td>
<td></td>
</tr>
<tr>
<td>Personal communication from a COVID-19–positive attendee</td>
<td>1/6 (17)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0/7 (0)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}EN: exposure notification.
\textsuperscript{b}Fisher exact \( P \) value.
\textsuperscript{c}There was an option for “Notified by a contact tracer,” but 0 respondents selected it.
\textsuperscript{d}WSPHA: Washington State Public Health Association.

Reasons for Not Activating WA Notify

Among those who reported not having WA Notify activated (N=49), there were a variety of reasons reported with some respondents (n=7) reporting multiple reasons. The most commonly cited reason was privacy concerns (n=17, 35%), followed by not wanting to receive ENs (n=6, 12%), and being unaware of WA Notify (n=5, 10%). Other reasons for not activating WA Notify included not understanding it (n=3, 6%), thinking it is cumbersome or bothersome (n=3, 6%), not getting around to it (n=3, 6%), having limited phone storage (n=3, 6%).
Discussion

Overview

This analysis allowed us to compare the characteristics of WA Notify adopters and nonadopters among those attending the WSPHA Annual Conference. It also explored testing behavior following a possible exposure and factors that may influence EN tool uptake. Our survey population consisted of public health professionals who, compared to the general public, are likely to be more aware of COVID-19 risks and prevention measures and are more inclined to use digital EN tools.

Principal Results

Our findings indicate that respondents who identified as WA Notify users were more likely to be younger compared to their counterparts, although this is unsurprising as younger age groups may be more comfortable with digital tools and technology in general. WA Notify users were also more likely to receive COVID-19 vaccinations relative to their counterparts, which would align with a higher likelihood of taking health precautions in general. Key findings included high levels of reported privacy concerns among nonadopters, that the majority (n=77, 54%) of WA Notify users reported first learning about their possible exposure from the tool, and that WA Notify users were more likely to test for COVID-19 compared to nonusers regardless of how they were first notified about their exposure.

Among nonusers, the most common reason for not having WA Notify activated was privacy concerns (n=17, 35%), despite stringent privacy protections, which make it nearly impossible for any government agency or corporation to obtain individual-level information [13,14]. We recognize that social desirability bias may have led nonusers to more frequently cite privacy concerns for not adopting WA Notify and assume that the general public, whose affiliation with public health is lower, likely has equal or greater privacy concerns than our sample. Still, this finding underscores the need for improved communications surrounding the privacy and data security of digital EN systems.

WA Notify users were significantly more likely to test for COVID-19 irrespective of how they first learned of their potential exposure. It is possible that testing was encouraged by the more personalized WA Notify EN or receiving multiple notifications or the combination of these 2 factors. Unfortunately, we cannot determine the extent to which WA Notify ENs encouraged testing. It is possible individuals with WA Notify activated were also more likely to test for COVID-19 regardless of whether they received an EN.

However, our findings indicate that WA Notify played an important role in distributing timely ENs to conference attendees following COVID-19 exposures at the WSPHA conference. Additionally, we believe a proximity-based digital EN system could be particularly useful in other large gatherings to alert close contacts of their exposure and reduce further transmission. The WSPHA conference was unique in that the attendees were public health professionals, and the sponsoring organization’s mission focused on health and public safety. Organizers in other event contexts may not be informed about infectious individuals attending the event, and if informed, organizers may be unable to or choose not to distribute their own EN to attendees.

Our results add to the evidence of the value of digital EN systems, but work remains to encourage adoption and engagement with these systems. The adoption of these systems is influenced by how easy they are to use (which is improved by the GAEN Express platform). Additionally, conducting pilot studies of the system, strong advertising campaigns, and engaging a wide array of stakeholders all likely contributed to WA Notify’s successes and can improve the adoption of similar systems in the future [3].

Limitations

There are several limitations to our findings. The survey population was a subset of individuals from the WSPHA conference and thus could have limited generalizability to the general population since the survey respondents are likely highly invested in public health and public health measures. These individuals could be more likely to use EN tools and take additional precautions against COVID-19 infection compared to the population of all WSPHA conference attendees or the general public.

There are also limitations to our survey design. The survey did not ask about prior COVID-19 infection, so natural immunity could not be accounted for. The question about vaccination status allowed an option for primary series booster; respondents who selected that response may have received either 1 booster or multiple boosters because the new bivalent booster had recently become available. We did not assess the respondents’ role within public health, such as if they worked at the state or local level or were a student. Knowing these characteristics may have provided more insight into the differences between these groups. When assessing the type and number of ENs received, data were only collected regarding the first EN method. The EN date information was only available for those who reported receiving a WA Notify EN; dates of other ENs were unknown if multiple were received.

Conclusions

Digital EN systems are important tools that have been used to notify close contacts of potential COVID-19 exposures. Our data suggest that having WA Notify activated encouraged individuals to engage in postexposure protective behaviors relative to those who did not have the tool activated. For digital EN technologies to sustain and expand their use in slowing the spread of communicable diseases moving forward, more work must be done to encourage uptake in groups less comfortable with the technology. Developing clear, consistent, and transparent messaging around the privacy and data security protocols built into these tools is needed to reduce these barriers to adoption. Additionally, understanding how to communicate evidence of public health value both to the general public and within the public health community is needed. Investment in future digital public health tools would benefit from an improved evidence base that includes systematic research on the impact of tailored messaging and the attributes and components of such
messaging for encouraging the adoption and sustained use of public health digital tools like WA Notify.

Acknowledgments
The work described in our manuscript was supported through an Interagency Agreement between the Washington State Department of Health and the University of Washington (HED 25742).

Authors’ Contributions
JGB, NLB, DR, LMW, AH, WBL, and BTK conceptualized, reviewed, and implemented the survey. CMD analyzed the data with support from BLG and ASE. CMD wrote the manuscript with support from BLG, ASE, DR, and JGB. All authors reviewed and approved the final version of the manuscript for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Examples of EN messages from WA Notify.
[DOCX File, 305 KB - formative_v8i1e50716_app1.docx]

Multimedia Appendix 2
REDCap survey sent to conference participants.
[PDF File (Adobe PDF File), 44 KB - formative_v8i1e50716_app2.pdf]

References


Abbreviations

EN: exposure notification
GAEN: Google-Apple Exposure Notification
REDCap: Research Electronic Data Capture
RR: relative risk
WSPHA: Washington State Public Health Association

©Caitlin M Drover, Adam S Elder, Brandon L Guthrie, Debra Revere, Nicole L Briggs, Laura M West, Amanda Higgins, William B Lober, Bryant T Karras, Janet G Baseman. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Developing a Text Messaging Intervention to Prevent Binge and Heavy Drinking in a Military Population: Mixed Methods Development Study

Chase A Aycock, PsyD; Indika Mallawaarachchi, MS; Xin-Qun Wang, MS; Daniel G Cassidy, PhD; Jordan M Ellis, PhD; Robert C Klesges, PhD; G Wayne Talcott, PhD; Kara Wiseman, MPH, PhD

1 United States Air Force, 37th Human Performance Squadron, Joint Base San Antonio-Lackland Air Force Base, San Antonio, TX, United States
2 Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States
3 United States Air Force, 37th Training Wing, Joint Base San Antonio-Lackland Air Force Base, San Antonio, TX, United States
4 United States Air Force, Wilford Hall Ambulatory Surgical Center, 59th Medical Wing, Joint Base San Antonio-Lackland Air Force Base, San Antonio, TX, United States

Corresponding Author:
Kara Wiseman, MPH, PhD
Department of Public Health Sciences
University of Virginia
P.O. Box 800717
Charlottesville, VA, 22908
United States
Phone: 1 4342348126
Email: kara.wiseman@virginia.edu

Abstract

Background: Alcohol misuse is the fourth leading cause of death in the United States and a significant problem in the US military. Brief alcohol interventions can reduce negative alcohol outcomes in civilian and military populations, but additional scalable interventions are needed to reduce binge and heavy drinking. SMS text messaging interventions could address this need, but to date, no programs exist for military populations.

Objective: We aimed to develop an SMS text messaging intervention to address binge and heavy drinking among Airmen in Technical Training in the US Air Force.

Methods: We implemented a 2-phase, mixed methods study to develop the SMS text messaging intervention. In phase 1, a total of 149 respondents provided feedback about the persuasiveness of 49 expert-developed messages, preferences regarding message frequency, timing and days to receive messages, and suggested messages, which were qualitatively coded. In phase 2, a total of 283 respondents provided feedback about the persuasiveness of 77 new messages, including those developed through the refinement of messages from phase 1, which were coded and assessed based on the Behavior Change Technique Taxonomy (BCTT). For both phases, mean persuasiveness scores (range 1-5) were calculated and compared according to age (aged <21 or ≥21 years) and gender. Top-ranking messages from phase 2 were considered for inclusion in the final message library.

Results: In phase 1, top-rated message themes were about warnings about adverse outcomes (eg, impaired judgment and financial costs), recommendations to reduce drinking, and invoking values and goals. Through qualitative coding of suggested messages, we identified themes related to warnings about adverse outcomes, recommendations, prioritizing long-term goals, team and belonging, and invoking values and goals. Respondents preferred to receive 1 to 3 messages per week (124/137, 90.5%) and to be sent messages on Friday, Saturday, and Sunday (65/142, 45.8%). In phase 2, a total of 283 respondents provided feedback about the persuasiveness of 77 new messages, including those developed through the refinement of messages from phase 1, which were coded and assessed based on the Behavior Change Technique Taxonomy (BCTT). For both phases, mean persuasiveness scores (range 1-5) were calculated and compared according to age (aged <21 or ≥21 years) and gender. Top-ranking messages from phase 2 were considered for inclusion in the final message library.

Conclusions: This study involved members from the target population throughout 2 formative stages of intervention development to design a BCTT-informed SMS text messaging intervention to reduce binge and heavy drinking, which is now being tested in
an efficacy trial. The results will determine the impact of the intervention on binge drinking and alcohol consumption in the US Air Force.

*(JMIR Form Res 2024;8:e55041) doi: 10.2196/55041*

**KEYWORDS**
text messaging; alcohol reduction; binge drinking; US; United States; US military; alcohol misuse; military; functioning; readiness; health; career; careers; text message; text messages; short message service; SMS; SMS intervention; drinking; Air Force; Airmen; mixed methods approach; message; messages; development study; qualitative coding; drinking alcohol; alcohol consumption; survey; descriptive statistics

**Introduction**

**Overview**

Alcohol misuse contributes to 140,000 deaths per year, making it the fourth leading cause of death in the United States [1]. Alcohol use also contributes significantly to the epidemic of *deaths of despair*, both as a primary cause of death and as a contributor to approximately 25% of opiate overdose deaths and suicides [2,3]. Deaths from alcohol have increased by 37% since 2005 [4]. Binge drinking and the associated health and social consequences are substantial public health concerns, with a high prevalence among young adults [5,6].

Alcohol misuse is a significant problem in the US military because of its impact on military functioning and readiness and its effect on the health and careers of service members. The Millennium Cohort Study, sampling >77,000 troops (active duty and reserve/guard), indicated that 18.5% of respondents reported a history of alcohol problems, and 7% to 8% reported current weekly heavy drinking (ie, >14 and >7 drinks per week for men and women, respectively) [5]. In 2018, overall, 34% of all Department of Defense (DoD) service members reported binge drinking at least once a month, defined as ≥5 drinks per sitting for men and ≥4 drinks per sitting for women, with significant differences based on gender (35.2% and 28.2% for men and women, respectively, P<.05) [7]. Specifically, in the Air Force, 24.1% reported binge drinking in the past month. Overall, young adults drink more than any other age group, and on average, young adults in the military drink 22% more than demographically similar civilians (33.1% vs 27.1%) [5,8].

Fortunately, brief alcohol interventions (BAIs) have been validated in college students and community-dwelling young adults who report elevated drinking levels and alcohol-related health or social problems. Overall, BAIs (typically 1-2 individual counseling sessions that include personalized feedback on drinking that are delivered in a motivational interviewing style) result in 30% to 50% reduction in binge drinking, with positive treatment effects reported up to 4 years after the intervention [9-11].

While individually delivered BAIs are effective, they are challenging to implement in large populations, such as the US military, due to logistics and cost. Therefore, scalable and cost-effective approaches for reducing the prevalence of binge and heavy drinking in this population are needed. One very promising approach to cost-effectively disseminate BAI content to many thousands of at-risk service members is through the use of automated SMS text messages sent at key times during training or when participants may be at the greatest risk for binge drinking. For example, Cadigan et al [12] evaluated specific personalized feedback during football tailgating, an event heavily associated with binge drinking. College students (N=130) who reported a history of tailgating and binge drinking were randomized to 1 of 2 SMS text message groups: an event-specific SMS text message or generic alcohol education messages. At the 1-month follow-up, those in the event-specific SMS text message group reported lower estimated peak blood alcohol concentration than those who received generic alcohol education messages.

In another study, 765 young adults who screened positive for hazardous drinking were randomized to SMS text message assessments and feedback group, SMS text message assessments only group, or a no SMS text message control group over a 12-week period [13,14]. Overall, there were decreases in the number of self-reported binge drinking days from baseline in the SMS text message assessments and feedback group, whereas there were increases in binge drinking days in the SMS text message assessments only and no SMS text message groups. Overall weekend bingeing was also lowest in the SMS text message assessments and feedback group.

Finally, Teeters et al [15] tested a mobile phone–based intervention with personalized SMS text messages to determine its impact on driving after drinking. Participants included 84 college students who endorsed driving after drinking ≥3 drinks at least twice in the past 3 months. Participants were randomized to either information about driving after drinking (eg, how alcohol affected the brain and nervous system) or mobile phone–based BAI related to driving after drinking that included personalized feedback and interactive SMS text messaging. The mobile phone–based interventions included personalized feedback delivered via secure SMS text message (eg, normative feedback about drinking level and frequency of drinking and driving). The feedback was delivered in the context of a 15- to 20-minute, SMS text-based “session” that included open-ended, motivational interviewing–style SMS text messages asking participants about their reaction to the feedback, the potential personal and professional costs of being arrested for driving under the influence of alcohol (eg, losing their college scholarship), and their plans for avoiding drinking and driving in the future. The study clinicians made tailored empathic reflections about the participants’ reactions and encouraged specific goal setting/planning. The primary outcome variable was self-reports of driving after drinking at a 3-month follow-up. Results indicated that students receiving the mobile phone–based BAI reported significantly greater reductions in the likelihood
of driving after drinking and the number of drinks consumed before driving than those in the information group.

While these findings suggest a potentially promising impact of SMS text message interventions on hazardous drinking outcomes, 2 recent meta-analyses in specific settings have shown potentially less robust efficacy of SMS text message interventions on heavy drinking when delivered as the only intervention [16] and when delivered as a preventive intervention [17]. For example, when delivered as the only intervention, Bendtsen et al [16] determined that while heavy episodic drinking and weekly alcohol consumption were low among participants who received the SMS text message intervention, the CI of these estimates both included the null, indicating a lack of a statistically significant association. In addition, the included studies were judged to be of low quality [16]. Therefore, more studies are needed to fully understand the potential impact of SMS text message interventions on binge and heavy drinking.

To date, the few studies that evaluated SMS text messages for binge drinking have been conducted in civilian settings. However, a large percentage of military personnel are also similar in age to college populations, making them potentially similarly receptive to mobile health interventions. Military personnel may also have confidentiality and health stigma concerns, which mobile health interventions are well suited to mitigate. Furthermore, military populations have expressed general interest in digital interventions [18,19]. As mentioned previously, the military has a particular problem with binge and heavy drinking. In all service branches, there is a protracted ban of alcohol in Basic Military Training and the first part of Technical (advanced) Training. In the Air Force, the ban is for approximately 12 weeks and tolerance to alcohol, even if the Airman (note that Airman is the term used by the Air Force for its service members, regardless of gender or position in the Air Force) used to drink alcohol before basic training, tolerance is low at the completion of training. Our research team has successfully implemented a group-based BAI and observed a significant reduction in the odds of an alcohol-related incident (ARI: getting in trouble while drinking) over the year of the intervention compared to the previous year (odds ratio 0.56, 95% CI 0.38-0.81; P=.002) [20]. Implementing an SMS text message intervention to complement this group-based BAI when Airmen begin to be allowed to leave base and access alcohol for the first time in approximately 3 months has enormous potential.

As the types of messages that civilians find persuasive may not be persuasive to military personnel, it is crucial to develop messages using feedback directly from Airmen in Technical Training to find messages that will be compelling and influential for this population. While individually tailoring the messages is often presumed to be beneficial in targeting addictive behaviors, there are also benefits to having universally acceptable messages. Nontailored messages require no information about the recipient, potentially diminishing the service members’ concerns regarding anonymity. In addition, nontailored messaging can increase feasibility and future dissemination if the intervention is determined to be effective.

**Objectives**

This paper describes a 2-phase mixed methods study designed to develop an SMS text message intervention to reduce binge and heavy drinking among Airmen in Technical Training. Specifically, the aims of this study were to (1) determine the persuasiveness of SMS text messages suggested by experts and Airmen designed to discourage binge and heavy drinking, (2) obtain data about the dose and timing of messages, and (3) create a final message library for efficacy testing.

**Methods**

**Phase 1**

**Study Population**

Participants included 149 Airmen in Technical Training located across 4 US Air Force bases who were attending, as part of standard training, a squadron-level group BAI session. Surveys were administered in each squadron at the end of the BAI session.

**Procedures**

**Message Development**

An initial set of 49 messages were created by content and population experts (eg, military, retired military, and nonmilitary psychologists and health behavior change experts). To reduce participant burden, the 49 messages were divided randomly into 3 sets, with 16 to 17 messages in each set.

**Message Evaluation**

All Airmen were asked to provide their age, gender, and past-month binge drinking frequency (“During the past 30 days, how many times did you have 4 (for women)/5 (for men) or more alcoholic drinks on one occasion?” with response options, “never,” “1 or 2 times,” “3 to 5 times,” “6 to 10 times,” and “11 or more times”). Overall, 3 questions assessed general beliefs about the importance of being successful in Technical Training, perceived impact of receiving a punishment related to alcohol while in the Air Force, and level of concern about having problem drinking, all of which used a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). The Airmen were then presented with their set of messages to be evaluated. All 3 sets of messages were included equally in each survey administration, so that all 49 items were assessed evenly within each squadron. For each message, the Airmen responded to 2 items (ie, “This statement makes me concerned about the impact of drinking while in Technical Training” and “This statement discourages me from wanting to drink alcohol excessively”) using a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree) [21]. Following message rating, the Airmen were asked, “What might be a good text message that you can think of that might discourage an Airmen from drinking too much during Technical Training?” and provided space for a write-in response. Finally, the Airmen were asked about message timing and frequency and for specific points during their training when information of this nature might be particularly useful to receive.
Data Analysis
Following data entry, all responses were summarized using descriptive statistics. The Airmen’s age was dichotomized to <21 years or ≥21 years to reflect US legal drinking limits. Airmen who reported any binge drinking in the past month were coded as binge drinkers. To explore the messages’ persuasiveness, means, medians, and SDs were calculated for the persuasiveness item. “This statement discourages me from wanting to drink alcohol excessively” for each message. The messages were ranked based on persuasiveness and compared according to age and gender using Wilcoxon rank sum tests to identify the top-ranked messages with no statistically significant differences in rating based on age or gender. The resulting top messages were qualitatively coded for themes by 3 study authors (CAA, DGC, and JME). The frequency and timing of message delivery was explored using frequencies and percentages. Finally, messages suggested by Airmen were qualitatively coded for themes.

Phase 2
Study Population
Participants included 283 Airmen in Technical Training located across 4 US Air Force bases who were attending, as part of standard training, a squadron-level group BAI session. Surveys were administered in each squadron at the end of the BAI session.

Procedures
Message Development
Messages suggested by Airmen from phase 1 were refined, often by reformatting them into Airmen quotes, and coded based on the Behavior Change Technique Taxonomy (BCTT). Highly rated messages from phase 1 were also refined and coded based on the BCTT. Informed by the study by Michie et al [22], additional expert-created messages were developed to ensure a pool of messages that addressed key behavior change techniques (BCTs) previously identified to successfully reduce alcohol consumption. The final pool of 77 messages covered the following BCTs: problem solving, goal setting (outcome), action planning, discrepancy between current behavior and goal, review outcome goals, self-monitoring of behavior, information about consequences, information about social and environmental consequences, anticipated regret, social comparison, behavior substitution, comparative imagining of future outcomes, future punishment, reducing negative emotions, restructuring the social environment, avoiding/reducing exposure to cues for the behavior, identification of self as a role model, valued self-identity, and verbal persuasion about capability. To reduce participant burden, the 77 messages were divided randomly into 5 sets, grouped according to BCTs, with 15 to 16 messages in each set.

Message Evaluation
All Airmen were asked to provide their age, gender, and past-month binge drinking frequency (“During the past 30 days, how many times did you have 4 (for women)/5 (for men) or more alcoholic drinks on one occasion?” with the following response options: “never,” “1 or 2 times,” “3 to 5 times,” “6 to 10 times,” and “11 or more times”). Overall, 8 questions were included to assess general beliefs about the importance of being successful in Technical Training, perceived impact of receiving a punishment related to alcohol while in the Air Force, level of concern about having problem drinking, perceived culture of drinking in the Air Force, and opinions about alcohol reduction strategies and not drinking while in Technical Training, all of which used a 5-point Likert-type response scale ranging from 1 (strongly disagree) to 5 (strongly agree). The Airmen were then presented with their set of messages to be evaluated. All 5 sets of messages were included equally in each survey administration, so that all 77 items were assessed evenly within each squadron. For each message, the Airmen responded to 4 items to assess persuasiveness (ie, “This statement makes me concerned about the impact of drinking while in Technical Training”; “This message makes me think about how to be successful in Technical Training”; “This statement discourages me from wanting to drink alcohol excessively”; and “This message would be useful if I wanted to reduce or limit my drinking”) using a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree) [21]. Following message rating, the Airmen were asked to select 3 messages from their set that they believed would be most likely and least likely to help Airmen make good choices about alcohol.

Data Analysis
Following data entry, all responses were summarized using descriptive statistics. The Airmen’s age was dichotomized to <21 years or ≥ 21 years to reflect US legal drinking limits. Airmen who reported any binge drinking in the past month were coded as binge drinkers. To determine message suitability for inclusion in the final message pool, means, medians, and SDs were calculated for each item regarding message persuasiveness. The messages’ persuasiveness was compared according to age and gender using Wilcoxon rank sum tests to identify any messages with statistically significant differences in rating based on age or gender. The messages in phase 2 were ranked based on persuasiveness and considered within each BCT and overall to develop a final message pool that could address multiple BCT targets. Any messages that were statistically different based on age or gender where 1 group had a mean score below a threshold of 3.5 were excluded from consideration.

Ethical Considerations
In both phases, participants were asked to complete an anonymous survey. Both surveys were approved by the Joint Base San Antonio institutional review board as quality improvement, and thus, a full review was not required. The voluntary, fully informed consent of participants was obtained as required by 32 Code of Federal Regulations (CFR) 219 and Department of Defense Instruction (DODI) 3216.02. The study staff emphasized that participation in the survey was optional and anonymous, and >99% of the population consented to participate. No information was obtained about the individuals who did not choose to participate.
Results

Phase 1
Among the 149 respondents, 113 were men, 76 were aged <21 years, and 131 reported no heavy drinking in the 30 days before starting military training (Table 1). Approximately all (140/148, 94.6%) reported agreement with the statement, “Being successful in Technical Training is extremely important to me.” Motivation to be successful in Technical Training was also high; 89.9% (133/148) agreed to the statement, “Receiving punishment for an alcohol related incident (ARI) in Technical Training could harm or end my career,” and 91.2% (135/148) agreed that “Whether I drink alcohol or not, it is very important to me to NOT have a problem related to my drinking” (data not shown).

Table 1. Descriptive characteristics of the study populations in phases 1 and 2.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Phase 1 (n=149), n (%)</th>
<th>Phase 2 (n=283), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>33 (22.6)</td>
<td>90 (32.3)</td>
</tr>
<tr>
<td>Man</td>
<td>113 (77.4)</td>
<td>189 (67.7)</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;21</td>
<td>76 (51.4)</td>
<td>125 (46.3)</td>
</tr>
<tr>
<td>≥21</td>
<td>72 (48.6)</td>
<td>145 (53.7)</td>
</tr>
<tr>
<td>Binge drinking in the past 30 days</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>17 (11.5)</td>
<td>73 (26.4)</td>
</tr>
<tr>
<td>None</td>
<td>131 (88.5)</td>
<td>204 (73.6)</td>
</tr>
</tbody>
</table>

The values do not add up to 100% owing to missing data.

The 10 expert-created messages with highest median persuasiveness scores and statistically similar responses based on age (<21 years and ≥21 years) and gender (man and woman) are shown in Table 2. Qualitative coding indicated that 9 of the 10 (90%) top-rated messages had a theme related to warning about adverse outcomes, with specific adverse events related to the following: impaired judgment (4/9, 44%), career or receiving an ARI (3/9, 22%), assault (1/9, 11%), and financial costs (1/9, 11%). Of the 10 top-rated messages, 5 (50%) also included recommendations, with specific recommendations related to the following: planning (3/5, 60%), moderation (1/5, 20%), and low-risk drinking (1/5, 20%). Of the 10 top-rated messages, 1 (10%) invoked values and goal reminders.
Table 2. Top-rated, expert-created messages among respondents, with theme coding performed by the study team in phase 1<sup>a,b,c</sup>.

<table>
<thead>
<tr>
<th>Themes and expert-created messages</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Warning about adverse outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Career</td>
<td></td>
</tr>
<tr>
<td>“You should develop a plan if you drink alcohol to avoid the risk of harming your Air Force career.”</td>
<td>4.21 (0.90)</td>
</tr>
<tr>
<td>“No one who gets an ARI&lt;sup&gt;d&lt;/sup&gt; ever planned to get one.”</td>
<td>3.92 (1.27)</td>
</tr>
<tr>
<td>“An ARI is potentially career ending in the Air Force.”</td>
<td>3.78 (1.34)</td>
</tr>
<tr>
<td>Impaired judgment</td>
<td></td>
</tr>
<tr>
<td>“Drinking can lower your inhibitions, leading to poor social judgement. Limit your drinking or don’t drink at all.”</td>
<td>4.13 (1.08)</td>
</tr>
<tr>
<td>“Drinking impairs your judgement. Plan to keep your drinking under control so you won’t make decisions that you later regret.”</td>
<td>4.13 (1.02)</td>
</tr>
<tr>
<td>“Set your limit ahead of time. Once you’ve had two drinks, you’re no longer the best judge of what should come next.”</td>
<td>4.02 (1.13)</td>
</tr>
<tr>
<td>“Drinking can lead to aggression (verbal and physical). If you are going to drink, avoid people and places that could lead to conflicts.”</td>
<td>3.96 (0.98)</td>
</tr>
<tr>
<td>Assault</td>
<td></td>
</tr>
<tr>
<td>“WARNING, almost all sexual assaults in Technical Training involved heavy drinking.”</td>
<td>4.06 (0.88)</td>
</tr>
<tr>
<td>Financial costs</td>
<td></td>
</tr>
<tr>
<td>“You’ve earned that paycheck, so make it last. Consider slowly sipping on a drink instead of gulping it. Non-alcoholic drinks cost much less than alcoholic ones. Water is free.”</td>
<td>4.00 (1.07)</td>
</tr>
<tr>
<td>Recommendation</td>
<td></td>
</tr>
<tr>
<td>Planning</td>
<td></td>
</tr>
<tr>
<td>“You should develop a plan if you drink alcohol to avoid the risk of harming your Air Force career.”</td>
<td>4.21 (0.90)</td>
</tr>
<tr>
<td>“You want to have a successful plan for your future in the Air Force and are willing to take the time to work at making responsible decisions when it comes to drinking alcohol.”</td>
<td>4.10 (0.95)</td>
</tr>
<tr>
<td>“Set your limit ahead of time. Once you’ve had two drinks, you’re no longer the best judge of what should come next.”</td>
<td>4.02 (1.13)</td>
</tr>
<tr>
<td>Moderation</td>
<td></td>
</tr>
<tr>
<td>“Drinking can lower your inhibitions, leading to poor social judgement. Limit your drinking or don’t drink at all.”</td>
<td>4.13 (1.08)</td>
</tr>
<tr>
<td>Low-risk drinking</td>
<td></td>
</tr>
<tr>
<td>“Drinking impairs your judgement. Plan to keep your drinking under control so you won’t make decisions that you later regret.”</td>
<td>4.13 (1.02)</td>
</tr>
<tr>
<td>Reminder of values and goals</td>
<td></td>
</tr>
<tr>
<td>“You want to have a successful plan for your future in the Air Force and are willing to take the time to work at making responsible decisions when it comes to drinking alcohol.”</td>
<td>4.10 (0.95)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Mean scores for the item, “This statement discourages me from wanting to drink alcohol excessively.”

<sup>b</sup>Response range 1-5 [21].

<sup>c</sup>Messages could have multiple theme codings and appear more than once in the table.

<sup>d</sup>ARI: alcohol-related incident.

Exemplar messages suggested by Airmen are shown in Table 3. Qualitative coding indicated that among the 13 exemplar messages, 11 (85%) had a theme or themes related to warning about adverse outcomes, with specific adverse events related to career (6/11, 55%) and impaired judgment (1/11, 9%) or nonspecified consequences (5/11, 45%). Of the 13 exemplar messages, 5 (39%) included recommendations, all with specific recommendations related to prioritizing long-term goals. In addition, of the 13 messages, 2 (15%) contained themes related to team and belonging and 2 (15%) contained themes related to values and goal reminders.
Table 3. Messages suggested by Airmen respondents to discourage peers from excessive drinking during Technical Training, with theme coding performed by the study team in phase 1.

<table>
<thead>
<tr>
<th>Messages suggested by Airmen</th>
<th>Theme coding</th>
</tr>
</thead>
</table>
| “Alcohol lasts a night; paperwork goes on your record. Don’t risk it.” | • Warning about adverse outcomes (career)  
• Recommendation (prioritize long-term goals) |
| “Choose carefully if you want to ruin your life just for one night of fun.” | • Warning about adverse outcomes (nonspecific)  
• Recommendation (prioritize long-term goals) |
| “Do not let actions taken while drunk ruin how far you’ve come and ruin the rest of your life. Think, ‘Is this what I would do if I were sober?’” | • Warning about adverse outcomes (nonspecific)  
• Warning about adverse outcomes (impaired judgment)  
• Recommendation (prioritize long-term goals) |
| “Don’t throw away your money and career this early on alcohol just to have some ‘fun’” | • Warning about adverse outcomes (career)  
• Recommendation (prioritize long-term goals) |
| “Enjoy yourself off base; however, just remember that your actions have consequences, which includes drinking.” | • Warning about adverse outcomes (nonspecific) |
| “First the man takes the drink, then the drink takes the man.” | • Warning about adverse outcomes (nonspecific) |
| “Live in the moment but make sure that moment doesn’t destroy your future.” | • Warning about adverse outcomes (nonspecific) |
| “If you drink, we won’t be on the same team anymore.” | • Warning about adverse outcomes (career)  
• Team and belonging |
| “Remember what the Air Force represents.” | • Reminder about values and goals |
| “Remember, having casual drinks with friends is fine, but know your limit and be responsible. Aim High Airman!” | • Warning about adverse outcomes (career)  
• Values and goals reminder |
| “We have made so many memories together and have many more to go on. Make good choices.” | • Team and belonging |
| “Why risk a lifelong career for a few drinks?” | • Warning about adverse outcomes (career)  
• Recommendation (prioritize long-term goals) |
| “Your career could be in that bottle.” | • Warning about adverse outcomes (career) |

Most Airmen preferred to be sent 1 to 3 messages per week (124/137, 90.5%), followed by 4 to 6 messages per week (10/137, 7.3%). The most frequently selected days to receive messages were Friday, Saturday, and Sunday in combination (65/142, 45.8%) and Thursday, Friday, and Saturday in combination (20/142, 14.1%). Most respondents preferred messages to be sent in the early evening (84/149, 56.4%) or late evening (80/149, 53.7%; data not shown).

Phase 2
Among the 283 respondents, 189 were men, 125 were aged <21 years, and 204 reported no heavy drinking in the 30 days before starting military training (Table 1). Approximately all (243/259, 93.8%) reported agreement with the statement, “Being successful in Technical Training is extremely important to me.” Motivation to be successful in Technical Training was also high, as 90.3% (233/258) agreed to the statement, “Receiving punishment for an alcohol related incident (ARI) in Technical Training could harm or end my career,” and 88.7% (228/257) agreed that “Whether I drink alcohol or not, it is very important to me to NOT have a problem related to my drinking” (data not shown).

The final messages selected for inclusion in the SMS text message intervention are shown in Table 4. The final message library includes 28 BCTT-informed messages across 13 different BCTs and 5 messages (not shown) directly referencing content originally introduced in the group BAI. The program is designed to be delivered over 6 weeks, sending 3 to 4 messages per week, and most messages are sent on Friday and Saturday in the late afternoon and early evening. Mean scores for the included messages ranged from 3.31 (SD 1.29) to 4.21 (SD 0.90). Of the top 5 highest-rated messages in the final message library, 4 were categorized into 2 BCTs: valued self-identity and information about health consequences. Specific recommendations for behavior differ for those who are and those who are not legally allowed to drink in the United States, as those with ARIs and aged <21 years face harsher disciplinary action. Therefore, while designed to be a universally delivered program, for Airmen aged 21 years or those who will be aged 21 years during the intervention period, safe drinking–related messages were prioritized, whereas for Airmen aged <21 years, abstinence-related messages were prioritized. The total number of messages received by each Airmen is the same. Therefore, the distribution of BCTs differs according to age, with most action planning messages (3/4, 75%) only being sent to Airmen aged <21 years.
aged ≥21 years. Messages for Airmen aged <21 years have a high percentage of messages addressing the consequences of alcohol use. The inclusion of targeted messages based on age also resulted in 2 new messages needing to be created that were informed by phase 2 results but not evaluated by Airmen. Of the final 28 messages, 8 (28.6%) were informed by messages suggested by Airmen from phase 1.
Table 4. The final message library of an SMS text message intervention to prevent binge and heavy drinking among Airmen in Technical Training, with coding based on Behavior Change Technique Taxonomy.

<table>
<thead>
<tr>
<th>Themes and messages</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem solving</strong></td>
<td></td>
</tr>
<tr>
<td>“Write down your money and career goals. How does alcohol fit into the picture? How would a DUI\textsuperscript{a} or an ARI\textsuperscript{b} impact those goals?”</td>
<td>4.07 (1.19)</td>
</tr>
<tr>
<td><strong>Goal setting (outcome)</strong></td>
<td></td>
</tr>
<tr>
<td>“Only you can choose how to spend your free time. What are options that are most in line with your long-term goals and plans?”</td>
<td>3.82 (1.22)</td>
</tr>
<tr>
<td><strong>Action planning</strong></td>
<td></td>
</tr>
<tr>
<td>“If you drink, eat a meal beforehand and consider having a glass of water between drinks.”\textsuperscript{c}</td>
<td>3.69 (1.21)</td>
</tr>
<tr>
<td>“Before going out, think about choosing a Wingman who will not drink at all and who will be a designated driver.”\textsuperscript{c}</td>
<td>3.62 (1.29)</td>
</tr>
<tr>
<td>“If you drink tonight, don’t drive. Call an Uber [link to Uber] or a Lyft [link to Lyft]”\textsuperscript{c}</td>
<td>3.57 (1.15)</td>
</tr>
<tr>
<td>“You can help prevent an ARI for you or a friend by helping them get home safely. Call them an Uber [link to Uber] or a Lyft [link to Lyft]”\textsuperscript{c}</td>
<td>3.58 (1.26)</td>
</tr>
<tr>
<td>“Your tolerance to alcohol will probably be very low after not drinking for several weeks. If you are going to drink, consider limiting your alcohol to three or less drinks.”\textsuperscript{c}</td>
<td></td>
</tr>
<tr>
<td><strong>Review outcome goals</strong></td>
<td></td>
</tr>
<tr>
<td>“Advice from a fellow Airman: ‘Think about what truly matters to you in the long run.’”\textsuperscript{f}</td>
<td>3.90 (1.19)</td>
</tr>
<tr>
<td><strong>Self-monitoring of behavior</strong></td>
<td></td>
</tr>
<tr>
<td>“If you choose to drink, partner with a Wingman to keep your drink count to less than 3 per outing and 1 per hour.”\textsuperscript{c}</td>
<td>3.43 (1.25)</td>
</tr>
<tr>
<td><strong>Information about consequences</strong></td>
<td></td>
</tr>
<tr>
<td>“Imagine failing Technical Training because of a drinking incident. Is it worth the risk? What’s your plan to make sure alcohol use doesn’t derail your career goals?”\textsuperscript{d,f}</td>
<td>4.16 (1.08)</td>
</tr>
<tr>
<td>“Alcohol lasts a night, but paperwork stays forever on your record. What can you do to make sure today’s decisions do not hurt your tomorrows?”\textsuperscript{d,f}</td>
<td>4.07 (1.08)</td>
</tr>
<tr>
<td>“Underage drinking or legal drinking that results in an adverse incident can both lead to an ARI. What would be the impact of an ARI on your long-term goals?”\textsuperscript{d}</td>
<td>3.96 (1.08)</td>
</tr>
<tr>
<td>“Advice from a fellow Airman: ‘Live in the moment but make sure that moment doesn’t hurt your future.’”\textsuperscript{f}</td>
<td>4.00 (0.99)</td>
</tr>
<tr>
<td>“Choosing to not drink can lead to better sleep and feeling refreshed and rested each morning. What would you like to get out of each day?”</td>
<td>3.85 (1.17)</td>
</tr>
<tr>
<td>“We are all in this together for the good times and the bad. Remember, if one of us gets in trouble, the whole squadron will have to suffer the consequences.”\textsuperscript{d,f}</td>
<td>3.87 (1.37)</td>
</tr>
<tr>
<td><strong>Social comparison</strong></td>
<td></td>
</tr>
<tr>
<td>“Many of your fellow Airmen choose not to drink at all during Tech Training to avoid getting side-tracked. What will you choose to do?”</td>
<td>3.71 (1.23)</td>
</tr>
<tr>
<td>“46% of Airmen choose not to drink in technical school. Consider calling a friend and seeing if they’d like to do something alcohol free this weekend.”</td>
<td>3.31 (1.29)</td>
</tr>
<tr>
<td><strong>Behavior substitution</strong></td>
<td></td>
</tr>
<tr>
<td>“One of the best ways to prevent hazardous drinking is to participate in fun alcohol-free group activities. What plans can you make this weekend to keep everyone entertained, safe, and out of trouble?”\textsuperscript{d}</td>
<td>3.83 (1.06)</td>
</tr>
<tr>
<td>“Interested in an alcohol-free weekend? Click here [Link to study webpage listing alcohol-free activities by base] to see the top things people do to have fun around here.”</td>
<td>3.81 (1.25)</td>
</tr>
<tr>
<td><strong>Comparative imagining of future outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>“Take a moment to reflect on what you hope to achieve in the next year, academically, professionally, or athletically. How could alcohol get in the way?”</td>
<td>4.19 (0.95)</td>
</tr>
</tbody>
</table>
Themes and messages

<table>
<thead>
<tr>
<th>Score, mean (SD)</th>
<th>Themes and messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.65 (1.16)</td>
<td>“Underage drinking is a primary reason Airmen get Article 15s and those can lead to a separation. What are other ways you can have fun?”*df</td>
</tr>
<tr>
<td>—</td>
<td>“Drinking can be a reason why Airmen get Article 15s and those can lead to a separation. What are other ways you can have fun?”c</td>
</tr>
</tbody>
</table>

**Reduce negative emotions**

<table>
<thead>
<tr>
<th>Score, mean (SD)</th>
<th>Themes and messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.98 (1.13)</td>
<td>“Alcohol is not the only way to unwind. Many Airmen enjoy doing things that do not include alcohol. Local alcohol-free activities include movies, coffee/desert shops, trampoline parks, playing video games or sports, or visiting a museum. What might you like to do that is alcohol-free?”</td>
</tr>
<tr>
<td>3.94 (1.02)</td>
<td>“There are healthier, safer, and cheaper ways to manage stress than drinking. Consider going to the gym or calling a friend.”</td>
</tr>
<tr>
<td>3.86 (0.97)</td>
<td>“Tech Training is challenging, and it’s helpful to have ways to relax and unwind. Have you ever tried tactical breathing? Find out more: [link to breathing exercise website].”</td>
</tr>
</tbody>
</table>

**Identification of self as a role model**

<table>
<thead>
<tr>
<th>Score, mean (SD)</th>
<th>Themes and messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.91 (1.21)</td>
<td>“Evenings out are a good time to be a responsible Wingman. Set the example and others will follow.”*df</td>
</tr>
</tbody>
</table>

**Valued self-identity**

<table>
<thead>
<tr>
<th>Score, mean (SD)</th>
<th>Themes and messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.21 (1.08)</td>
<td>“Remember why you joined the Air Force. [link to inspirational video]. Don’t let alcohol be what stops you from achieving these goals.”*df</td>
</tr>
<tr>
<td>4.00 (1.17)</td>
<td>“You are among the 1% who chose a profession of arms. Consider the role you would like alcohol to have within your professional identity.”</td>
</tr>
</tbody>
</table>

*DU: driving under the influence of alcohol.
ARI: alcohol-related incident.
Only sent to Airmen who are of US legal drinking age or who will be before the end of the intervention period.
Only sent to Airmen who are younger than the US legal drinking age.
Indicates a message that was not included in phase 2 testing.
Original source was a message suggested by Airmen from phase 1.

**Discussion**

**Principal Findings**

This 2-phase study developed an automated SMS text message intervention designed to reduce binge and heavy alcohol use among young adults entering the military. In phase 1, a sample of US Air Force Technical School trainees rated the messages created by experts and provided suggestions. In phase 2, the messages suggested by Airmen and those that were highly rated in phase 1 were refined and rated alongside additional expert-created messages. These phases culminated in the creation of a BCTT-informed SMS text message program for Airmen in Air Force Technical Training and highlight an approach for message testing and program development, which incorporates mixed methods feedback from the target population. This approach allowed us to have a final program that was directly informed by the target population and led to more than one-fourth (8/28, 28.6%) of the final messages being in the Airmen’s own words.

Top-rated, expert-created messages from phase 1 included recommendations for planning or drinking in moderation and warnings about adverse outcomes associated with drinking in excess (eg, career impact, risk for sexual assault, and impaired judgment associated with drinking). In particular, most (9/10, 90%) of the top-rated messages in phase 1 were loss-framed messages. Messages suggested by both experts and Airmen included warnings about adverse outcomes, especially pertaining to one’s career. Other themes of the messages suggested by Airmen were reminders to prioritize long-term goals and values over short-term “fun” and emphasis on belonging to the same collective team. In phase 2, an expanded set of items organized based on BCTs allowed for a more nuanced analysis of the potential message pool. Specifically, it also allowed for the inclusion of more gain-framed messages. As in phase 1, the highly rated messages in phase 2 were related to themes such as adverse outcomes and long-term values and goals. The 2 highest-rated messages in the final program were related to the topics of valued self-identity (“Remember why you joined the Air Force. [link to inspirational video]. Don’t let alcohol be what stops you from achieving these goals.”) and comparative imagining of future outcomes (“Take a moment to reflect on what you hope to achieve in the next year, academically, professionally, or athletically. How could alcohol get in the way?”), which both had loss-framed messages around alcohol getting in the way. Thematic consistency of top-rated messages across both phases of the study suggests that the messages ultimately included in the final message pool are likely to resonate with the target population. The real-world effectiveness of these messages are being empirically evaluated in an ongoing clinical trial.

The final intervention is targeted at Airmen of the US legal drinking age, while also being designed to be uniformly acceptable across genders. Therefore, message ratings were
considered overall and according to age and gender. This allowed for a program that can still be scaled up easily while providing relevant content for specific subsets of the population.

If tailoring based on other factors becomes necessary, for example, the Airman’s specific role in the Air Force, this study provides a framework through which messages can be developed and evaluated with specific tailoring needs in mind. It will also be important to explore the potential moderators of program success after completion of the currently underway clinical trial. The presence of a strong effect modifier would suggest a potential need to tailor based on specific Airmen characteristics and would help justify the collection of additional personal information at the time of program sign-up.

Phase 1 also identified Airmen’s preferences for receiving SMS text messages, which were incorporated into the final intervention. Specifically, the Airmen preferred to receive messages at a frequency of approximately 3 to 5 times per week, in the evenings and on weekends. This could be an ideal window to receive a timely prompt in the form of an SMS text message, given that Airmen are most at risk for binge drinking from Friday evening to Sunday evening when they become able to leave their base (approximately 10-12 weeks after starting Basic Military Training). Having information about message timing and frequency preferences allowed us to develop an SMS text message program that should fit within each Airman’s daily routine and maximizes the number of messages sent at times when the Airmen may be considering and planning off-base activities, where alcohol is more readily available.

Across both phases, respondents reported interest in being successful in Technical Training, believed that an ARI would be detrimental to their careers, and did not want to have a problem associated with their drinking. These results should be interpreted with some caution, as respondents had just completed the group BAI; therefore, the Airmen might have been having heightened feelings regarding the potential harms of alcohol use. It could also show the potential of the group BAI to be a teachable moment, which the SMS text message intervention will attempt to capitalize on to sustain self-reflection and opinions about alcohol use and the desire for success during Technical Training. The sentiment of success during Technical Training is also referenced in the SMS text message intervention to increase the potential salience of the messages.

In addition to being developed specifically for a noncivilian population, there are other differences between the SMS text message intervention described in this paper and the previously developed alcohol reduction SMS text message interventions [12,13]. For example, the 2 previous studies designed to address binge drinking included assessments of drinking behavior. Specifically, Cadigan et al [12] collected information about participant drinking behavior while tailgating at baseline to provide normative feedback to participants during the intervention, which referenced back to the information each participant provided. Suffoletto et al [13] asked participants to report anticipated drinking behavior and the highest number of drinks consumed over the weekend, and responses were used to tailor other intervention messages. Owing to confidentiality concerns within the military, actual alcohol use behavior is not able to be captured by Airmen within the SMS text message intervention described in this paper. Thus, direct normative feedback cannot be delivered to participants. However, there is a message included to allow for social comparison. Despite this difference, there are also similarities between the proposed intervention and previous studies, such as sending messages when participants are at the highest risk of drinking (ie, weekends) and not tailoring the messages based on other personal characteristics.

In addition to the programmatic differences described previously, to the best of our knowledge, the SMS text message intervention described in this paper is the first to use BCTs to guide intervention development within the context of alcohol reduction [22,23]. The use of BCTs is helpful for understanding the hypothesized mechanisms of action through which the SMS text message intervention may reduce binge and heavy alcohol use. In addition, it allows for the testing of active components in future evaluations. This SMS text message intervention includes content that covers 13 distinct BCTs, which is more than the average number seen in previous behavior change interventions to reduce alcohol use (mean 9.1 BCTs) [24], but less than another SMS text message intervention that was coded based on BCT after its development and evaluation (23 BCTs) [23]. However, there may be other BCTs that could prove to be valuable. A future study could evaluate the impact of adding content that addresses additional BCTs. Moreover, the SMS text message intervention could be separated into unique BCT components and compared directly [25] to better understand which BCTs are most important for addressing binge and heavy drinking in this population.

Limitations
A limitation of this study is that this pool of messages is based solely on the Airmen’s predictions about which messages will influence future drinking behavior. Consequently, the final message pool must be empirically tested to determine its impact on alcohol-related behavior. This study is the first to develop an SMS text message intervention to reduce binge and heavy drinking within a US military population, specifically, the US Air Force. However, it is possible that the results of this mixed methods approach are not generalizable to other military settings.

Conclusions
This study included the involvement of members from the target population throughout the formative stages of intervention development, to design an SMS text message intervention to reduce binge and heavy drinking in the US Air Force. Message content was anchored to evidence-based behavior change techniques according to the BCTT in conjunction with integration of feedback from Airmen regarding SMS text messaging language, timing, and frequency, thus increasing the odds of a robust intervention effect.
Acknowledgments
This study was conducted in collaboration with the Readiness and Performance Lab (RiPL). This study was supported by a grant from the National Institute of Alcohol Abuse and Alcoholism (R01AA029074) and was a collaborative endeavor between the US Air Force and University of Virginia.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Disclaimer
The views expressed are those of the authors and do not reflect the official views or policy of the Department of Defense or its components.

Conflicts of Interest
None declared.

References
4. Radley DC, Collins SR, Hayes SL. 2019 scorecard on state health system performance: deaths from suicide, alcohol, drugs on the rise; progress expanding health care coverage stalls; health costs are a growing burden. The Commonwealth Fund. 2019. URL: https://tinyurl.com/3tzd5wb3 [accessed 2023-11-10]


Abbreviations

ARI: alcohol-related incident
BAI: brief alcohol intervention
BCT: behavior change technique
BCTT: Behavior Change Technique Taxonomy
CFR: Code of Federal Regulations
DoD: Department of Defense
DoDI: Department of Defense Instruction

©Chase A Aycock, Indika Mallawaarachchi, Xin-Qun Wang, Daniel G Cassidy, Jordan M Ellis, Robert C Klesges, G Wayne Talcott, Kara Wiseman. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,
provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Association of Digital Engagement With Relaxation Tools and Stress Level Reduction: Retrospective Cohort Study

Inbar Breuer-Asher¹, MPH; Marilyn D Ritholz², PhD; David L Horwitz³, MD, PhD; Omar Manejwala¹, MD; Ephraim Behar¹, MSc; Yifat Fundoiano-Hershcovitz¹, PhD

¹DarioHealth, Caesarea, Israel
²Joslin Diabetes Center, Harvard Medical School, Boston, MA, United States
³DLH Biomedical Consulting, Las Vegas, NV, United States

Corresponding Author:
Yifat Fundoiano-Hershcovitz, PhD
DarioHealth
Ofek 8, 5 Tarshish Street
Caesarea, 3079821
Israel
Phone: 972 525296979
Email: yifat@dariohealth.com

Abstract

Background: Stress is an emotional response caused by external triggers and is a high-prevalence global problem affecting mental and physical health. Several different digital therapeutic solutions are effective for stress management. However, there is limited understanding of the association between relaxation components and stress levels when using a digital app.

Objective: This study investigated the contribution of relaxation tools to stress levels over time. We hypothesized that participation in breathing exercises and cognitive behavioral therapy–based video sessions would be associated with a reduction in stress levels. We also hypothesized a significant reduction specifically in participants’ perceived sense of burden and lack of productivity when engaged with breathing exercises and video sessions.

Methods: Stress levels were evaluated in a real-world data cohort using a behavioral health app for digital intervention and monitoring change. This retrospective real-world analysis of users on a mobile platform–based treatment followed users (N=490) who started with moderate and above levels of stress and completed at least 2 stress assessments. The levels of stress were tracked throughout the first 10 weeks. A piecewise mixed effects model was applied to model the trajectories of weekly stress mean scores in 2 time segments (1-6 weeks and 6-10 weeks). Next, a simple slope analysis was used for interpreting interactions probing the moderators: breathing exercises and video sessions. Piecewise mixed-effects models were also used to model the trajectories of specific perceived stress item rates in the stress questionnaire in the 2 segments (1-6 weeks and 6-10 weeks) and whether they are moderated by the relaxation engagements. Simple slope analysis was also used here for the interpretation of the interactions.

Results: Analysis revealed a significant decrease in stress symptoms (β=−0.25; 95% CI−0.32 to −0.17; P<.001) during the period of 1-6 weeks of app use that was maintained during the period of 6-10 weeks. Breathing exercises significantly moderated the reduction in stress symptoms during the period of 1-6 weeks (β=−0.07; 95% CI−0.13 to −0.01; P=.03), while engagement in digital video sessions did not moderate stress scores. Engagement in digital video sessions, as well as breathing exercises, significantly moderated the reduction in perceived sense of burden and lack of productivity during weeks 1-6 and remained stable during weeks 6-10 on both items.

Conclusions: This study sheds light on the association between stress level reduction and specific components of engagement in a digital health app, breathing exercises, and cognitive behavioral therapy–based video sessions. Our findings provide a basis for further investigation of current and moderating factors that contribute to the personalization of digital intervention. In addition, results may aid in developing a more comprehensive understanding of how digital intervention tools work for mental health and for whom they are most effective.

(JMIR Form Res 2024;8:e50506) doi:10.2196/50506)
KEYWORDS
mental health; perceived stress; stress reduction; digital health; video sessions; behavioral health; relaxation; breathing exercises; CBT; anxiety; cognitive behavioral therapy

Introduction
Stress is defined as an emotional or physical reaction typically caused by an external trigger [1,2] and is known to be a high-prevalence global problem, negatively affecting mental and physical health [3]. According to Gallup 2019 Global Emotions Report, the percentage of the American population who experienced stress (55%) is one of the highest in the world, as compared to the global average of 35% [4,5]. Furthermore, it is reported that distress levels have only increased during the COVID-19 pandemic due to health concerns, altered everyday life routines [6], and the unpredictability of its future implications [3]. A recent US stress survey conducted, reports an elevated average reported stress level from prepandemic levels [7]. Moreover, an alarming proportion of adults reported that stress has an impact on their day-to-day functioning, with more than a quarter (27%) saying that most days they are so stressed they cannot function [7].

Stress has a major impact on managing daily life, as everyone experiences it in varying forms every day [8]. According to the American Psychological Association’s 2014 Work and Well-Being Survey, 31% of employed adults indicated that they felt tense or stressed out during the workday [9]. Several studies investigating the understanding of how employee health impacts work performance have found stress can cause burnout, absenteeism, and reduced efficiency and performance [10], which can negatively impact the organization [11]. Yet, some resources can be implemented to help meet the pressures and demands faced at work [12].

There is evidence of improvements in health status due to workplace wellness programs for mental health [13,14]. In the past decade, the delivery of digital interventions has become increasingly popular; these interventions target common mental health illnesses [15] within the workplace due to their easy implementation [16]. Web-based interventions are shown to offer several advantages that may overcome some of the limitations of face-to-face approaches [17] as they are accessible through any internet-connected device. Among the working population, these interventions could especially benefit those who do not seek regular mental health treatment because of negative perceptions of mental health needs in the workplace [14].

It is reported that a substantial proportion of adults with common mental disorders fail to receive treatment [18]. The barriers thought to impede appropriate mental health care seeking have been well documented in the literature [18,19]. Many patients experience accessibility issues such as prolonged wait times, cost, clinic location, and transportation [20]. Others report that the stigma associated with mental illness diagnosis is the main reason that decreases their willingness to seek treatment [19].

Thus, digital interventions have the potential to overcome many of the barriers associated with seeking and accessing mental health care [18] and to provide increased convenience for patients allowing them access to care at any time.

Furthermore, digital health solutions aim to foster patient empowerment [21,22] by encouraging them to be proactive in the care process. These interventions allow people the right guidance and support for self-management, consequently addressing their own health-related goals [23]. Digital solutions also enable users to be better informed about their health, share experiences, assess and monitor specified health states, reach treatment, and improve communication between them and health care professionals [24].

In recent years, there has been an increase in the delivery of digital therapeutics targeting common mental health conditions [25,26]. Depression and anxiety, which have the highest prevalence among mental health conditions, are at the forefront of these targeted programs [26]. Previous research has demonstrated the efficacy of web-based interventions on the prevention and management of these conditions [27,28] and these conditions were even reported as efficacious as traditional face-to-face treatments [29].

Most treatment provided in behavioral health apps uses module-based sessions and the teaching of coping and management strategies based on cognitive behavioral therapy (CBT) principles, problem-solving therapy, and psychoeducation [20,26]. The digital application of CBT is showing promising results in the treatment of depression and anxiety disorders [30]. Additional research suggests that a successful approach to addressing digital health interventions is through a multicomponent design [13], using features such as tailored messages, reporting of thoughts, feelings, or behaviors [31]; relaxations features [5]; integrated therapist contact and other supplementary worksheets and engagements [20].

Notwithstanding that stress itself is not defined as a disorder, it can lead to major psychological and physical implications that could place people at risk for illness [17,32]. Given that stress represents a major threat to public health, effective and scalable solutions to accommodate the demand for stress-management interventions are needed [17]. The concept of “stress management” refers to the emotional, psychological, and behavioral methods used to develop skills to manage and reduce stress [33]. Several types of stress management interventions have been reported to be effective at reducing stress levels, such as CBT, time management, relaxation, and meditation exercises [34]. However, applying these self-help skills to achieve improvement requires motivation and self-discipline.

One of the easiest methods for managing stress-related symptoms is deep breathing exercises [3]; these exercises have been proposed as first-line and supplemental treatments for emotional disorders [35]. Several studies reported that the use of different breathing techniques showed significant improvement in mood states and perceived stress [36,37].

https://formative.jmir.org/2024/1/e50506

Jmir Form Res 2024 | vol. 8 | e50506 | p.1397

(page number not for citation purposes)
resulting in an effective improvement in the management of stress in daily life [36]. Slow and deep breathing can have a significant impact on reducing stress feelings and activating the relaxation response. This exercise is efficient as it reduces the ventilation in the dead space of the lungs and decreases the effect of stress and strain on the body by shifting the balance toward the parasympathetic system [38].

In recent years, the number of studies on digitally delivered stress management has been rising, yet the overall effect of specific formats of treatment delivery remains unclear [17]. Moreover, existing web-based stress management programs differ in various aspects such as type, content, and guidance, and this may influence their efficacy [17].

A recurring element described in the literature is video-based self-administered intervention programs, which are mainly rooted in a CBT framework [39,40]. Previous studies reported that video-based CBT has shown effectiveness for anxiety, sexual pain, and insomnia [39-41]. However, these video-based programs have not yet been integrated into the treatment of stress in a digital form. This method is described as a third waveform of CBT [40], which is characterized by concepts such as acceptance and mindfulness that focus on the person’s relationship to thought and emotions [42].

In their meta-analysis review of digital mindfulness-based interventions, Spijkerman et al [43] suggested that these interventions can be effective in reducing perceived stress. Mindfulness is a complete therapy that can minimize stress levels and improve psychological well-being because it includes various components of relaxation, such as yoga, deep breathing techniques, and focusing attention [44]. However, despite the abundance of digital mindfulness-based interventions and considering the structured variability [45], it is presently unknown whether different components of these interventions are effective.

In summary, although there has been an established body of research evaluating the efficacy of mobile app platforms for stress management, there has been less focus on the effectiveness of engagement with specific components and relaxation strategies for stress reduction. The purpose of this study is to evaluate the effect of a relaxation engagement digital behavioral health app on perceived stress. Our primary hypothesis is that in users with moderate-to-high stress levels, improvements in stress scores will be observed during the initial weekly period of the intervention and will be followed by maintenance for several weeks after. We hypothesized that engagement with certain digital components would moderate the reduction in stress levels. Specifically, we hypothesized that breathing exercises and video sessions would moderate stress levels’ reduction. We also anticipated a significant reduction in the perceived sense of burden and perceived level of productivity ranking as these measures were designed to reflect a change in perceived stress levels over time.

### Methods

#### Behavioral Health App

The Dario Health behavioral health app is essentially a modular tool delivering emotion-focused support designed to apply to a range of mental health conditions. Members generally have access to the Dario Health behavioral health platform as part of their employee or health plan benefits. A clinically based screening tool assesses each person’s needs and guides users to the most efficient support. The app provides several CBT programs including depression, anxiety, anger, stress, and substance use, and can be used as a self-guided intervention or with the help of a certified coach. The structure of each program includes conceptual videos, textual skills, breathing exercises, and monitoring progress tools. This study focuses on tracking cohorts of users with stress scores measured by their responses to the stress questionnaire.

The stress program begins with a psychology education module about stress, identifying symptoms, as well as how it presents itself physically, mentally, and emotionally. The program includes short CBT-based and empirically supported whiteboard educational videos, introducing skills such as understanding the impact of stress, coping strategies, identifying stress signs and symptoms, and stress-related behaviors. By increasing the awareness of the individuals’ thoughts and feelings and gradually explaining specific situations leading to stress, these modules provide a way to promote stress self-management. During these skills-based sessions, various techniques are introduced to assist with reducing stress such as relaxation exercises, breathing exercises, monitoring progress, and coaching. The content is delivered to engage the users and appealingly convey information. Relaxation exercises provide a digital experience including audio or video sessions of mindfulness, muscle relaxation techniques, and automatic thoughts and emotions regulation.

The breathing exercises are instructed as a short up to 5-minute audio guide; each one educates about a different technique, that is, progressive muscle relaxation, diaphragmatic breathing, or 4-7-8 breathing.

Coaching is introduced to users for support and assistance and people who choose to obtain coaching services via the app.

#### Measures

Stress levels were assessed over 10 weeks using a stress questionnaire. The questionnaire that objectifies the degree of stress severity is a 5-item self-reported, in-app delivered questionnaire. Each of the 5 items is scored from 0 (not at all) to 3 (nearly every day; Table 1). The questionnaire aims to assess an individual’s level of stress and the factors contributing to it. The questionnaire items reflect perceived stress such as a sense of burden, lack of productivity, and symptoms of stress such as lack of sleep, body tension, and irritability. As a severity measure, scores range from 0 (absent of stress symptoms) to 15 (severe stress levels). Stress scores of ≤5, 6-10, and >10 represent low, moderate, and high stress levels, respectively. Internal reliability testing of the questionnaire demonstrated an acceptable internal reliability with a Cronbach α of 0.73 [46].
Table 1. The components in the stress questionnaire grading stress severity.

<table>
<thead>
<tr>
<th>Item</th>
<th>Questionnaire</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Have you felt overwhelmed or stressed about important responsibilities?</td>
<td>0=“Not at all”&lt;br&gt;1=“Several days”&lt;br&gt;2=“More than half the days”&lt;br&gt;3=“Nearly every day”</td>
</tr>
<tr>
<td>2</td>
<td>What about feeling as though you were not as productive as you should’ve been?</td>
<td>0=“Not at all”&lt;br&gt;1=“Several days”&lt;br&gt;2=“More than half the days”&lt;br&gt;3=“Nearly every day”</td>
</tr>
<tr>
<td>3</td>
<td>Had trouble sleeping because your mind was racing or worrying?</td>
<td>0=“Not at all”&lt;br&gt;1=“Several days”&lt;br&gt;2=“More than half the days”&lt;br&gt;3=“Nearly every day”</td>
</tr>
<tr>
<td>4</td>
<td>Felt tension in your neck, back, or shoulders?</td>
<td>0=“Not at all”&lt;br&gt;1=“Several days”&lt;br&gt;2=“More than half the days”&lt;br&gt;3=“Nearly every day”</td>
</tr>
<tr>
<td>5</td>
<td>Felt noticeably more irritable, restless, or agitated than usual?</td>
<td>0=“Not at all”&lt;br&gt;1=“Several days”&lt;br&gt;2=“More than half the days”&lt;br&gt;3=“Nearly every day”</td>
</tr>
</tbody>
</table>

Independent variables were included in the analysis as potential moderating factors such as app sessions, videos, breathing exercises, and coaching interactions.

All data were transferred and stored in compliance with HIPAA (Health Insurance Portability and Accountability Act) requirements, using MongoDB or BigQuery database services. All data were anonymized before extraction for this study.

Users

A retrospective data evaluation study was performed on the Dario database.

The stress cohort consisted of a group of 490 users who used the Dario Behavioral Health platform between 2019 and 2022. The population included users who started at a moderate and above level of stress (stress questionnaire score >5) and those who completed at least two stress assessments. The sample included 388 (79%) women, 95 (19%) men, and 7 (2%) others. The age distribution was as follows: ≤35 (38%), 36-55 (49%), 56-65 (12%), and 66-76 (1%) years. The average baseline score of the stress questionnaire was 9.8 (SD 2.7; median 10, IQR 4). The weekly average time of app use observed in this cohort was 26.4 minutes (SD 39.2; median 14, IQR 21)

Ethical Considerations

All data used for the analysis were anonymized before extraction for this study. This study received an exemption from the institutional review board. Ethical & Independent Review Services [47], a professional review board, issued the institutional review board exemption for this study (21235 - 01#).

Statistical Analysis

A classical linear longitudinal model assumes a single-slope growth pattern for changes in an outcome variable across time. In contrast, piecewise - based mixed effects models allow flexibility in the modeling of variable change trajectories across time [48]. Previous studies consistently demonstrated a 2-stage dynamic of clinical outcomes associated with digital interventions: an initial improvement is followed by stabilization of the outcome [49,50].

Following a visualization of the distribution of the stress estimates over time, a piecewise cutoff point for the model slopes was chosen at 6 weeks of product use, assuming a change in the time-related stress severity level trajectory after 6 weeks considering previous research as well [51,52]. Such a behavior including 2 trajectories is in line with our previous research on digital follow-up measures in chronic conditions management [53,54].

Here, a piecewise mixed effects model was conducted to model the trajectory of the stress questionnaire mean score in 2 segments to allow the data to exhibit different linear trends over different periods. We tested whether digital engagement moderated 2 piecewise time trajectories in stress (1-6 weeks and 6-10 weeks). The model included a person-based random intercept and random slope for the time trajectory after the piecewise cutoff. Simple slope analysis was used for the interpretation of the interactions probing the moderators at 1 SD below (low engagement) and 1 SD above (high engagement) the average.

Next, we used mixed model analysis to model the trajectory of questions 1 and 2 rates (Table 1) in the stress questionnaire in
the 2 segments (1-6 weeks and 6-10 weeks). Moreover, we tested whether these scores were moderated by engagement with specific components in the app. The models included a random intercept and random slope of the time trajectory. Simple slope analysis was used for the interpretation of the interactions probing the moderators at 1 SD below (low engagement) and 1 SD above (high engagement) the average.

**Results**

**First Analysis: Stress Score Reduction Is Associated With Breathing Exercises**

In total, 90% (441/490) of the users improved or maintained their stress level over this study’s period, that is, remained at a moderate level or improved from moderate or high to a low level of stress.

Piecewise mixed model analysis revealed a significant decrease in stress levels ($\beta = -0.25; 95\% \text{ CI} -0.32$ to $-0.17; P < .001$) during the period of weeks 1-6. There were no significant time-related trends in stress during the period of weeks 6-10 ($\beta = 0.00; 95\% \text{ CI} -0.21$ to 0.22; $P = .98$). **Figure 1A** demonstrates time-related fluctuations in stress levels.

Breathing exercises significantly moderated the reduction in stress score during the period of weeks 1-6 ($\beta = -0.07; 95\% \text{ CI} -0.13$ to $-0.01; P = .03$). The weekly average number of breathing exercises completed is 1.55 (2.34 SD). Users with increased breathing exercise completion (+1 SD) demonstrated stronger reduction in stress score ($\beta = -0.45; 95\% \text{ CI} -0.64$ to $-0.25; P < .001$) compared to users who completed fewer breathing exercises ($-1 \text{ SD}$), that was not significant ($\beta = -0.13; 95\% \text{ CI} -0.32$ to 0.05; $P = .15$). Breathing exercise did not moderate the time trajectory of stress levels during the period of weeks 6-10 ($\beta = -0.02; 95\% \text{ CI} -0.23$ to 0.19; $P = .87$).

**Figure 1B** presents low and high levels of breathing exercise engagement, showing a significant reduction in stress levels only for users with high engagement, while the decrease for users with low engagement was not significant.

The engagement in digital video sessions did not significantly moderate stress score during both segments of time ($\beta = -0.01; 95\% \text{ CI} -0.03$ to 0.00; $P = .14$ and $\beta = -0.03; 95\% \text{ CI} -0.11$ to 0.04; $P = .37$ to week 1-6 and 6-10 accordingly).

**Second Analysis: Improvement in Perceived Stress Is Moderated by Digital Video Sessions**

**Overview**

The perceived stress level was reflected by items 1 and 2 in the questionnaire.

**Question 1: “Have you felt overwhelmed or stressed about important responsibilities?”**

Piecewise mixed model analysis revealed that breathing exercises significantly moderated the reduction in question 1 ranking during the period of weeks 1-6 ($\beta = -0.02; 95\% \text{ CI} -0.04$ to $-0.01; P = .009$). Breathing exercises did not moderate the time trajectory of question 1 ranking during weeks 6-10 ($\beta = -0.00; 95\% \text{ CI} -0.06$ to 0.06; $P = .91$). The weekly average number of breathing exercises completion is 1.55 (2.34 SD). Users with increased breathing exercise engagement (+1 SD) demonstrated a stronger reduction in perceived stress prompted by important responsibilities ($\beta = -0.13; 95\% \text{ CI} -0.19$ to $-0.08; P < .001$) compared to users who were less engaged in breathing exercises ($-1 \text{ SD}$), that was not significant ($\beta = -0.02; 95\% \text{ CI} -0.07$ to 0.03; $P = .40$).

Further piecewise mixed model analysis revealed that engagement in digital video sessions significantly moderated
the reduction in question 1 ranking during the period of weeks 1-6 (β=-0.01; 95% CI –0.01 to –0.00; P=0.02), yet no significant reduction was seen during the period of 6-10 weeks (β=0.00; 95% CI –0.02 to 0.02; P=0.91). The weekly average number of video session completion is 6.38 (8.59 SD). In addition, it was found that users who completed a higher number of video sessions (+1 SD) demonstrated a greater reduction in question 1 ranking (β=-0.13; 95% CI –0.19 to –0.07; P<0.001) compared to those who were less engaged (–1 SD) with video sessions (β=0.02; 95% CI –0.08 to 0.03; P=0.38).

**Question 2:** “What about feeling as though you were not as productive as you should’ve been?”

A piecewise mixed model has shown that the completion of breathing exercises has also significantly moderated the reduction of question 2 ranking in the stress questionnaire during the period of weeks 1-6 (β=-0.02; 95% CI –0.04 to –0.00; P=0.04); however, the period of 6-10 weeks was not significant (β=0.01; 95% CI –0.05 to 0.07; P=0.64). The weekly average number of breathing exercises completion is 1.55 (2.34 SD). Users with an increased number of completed breathing exercises (+1 SD) displayed a significant reduction in the question 2 ranking (β=-0.07; 95% CI –0.13 to –0.01; P=0.02), while in users who completed fewer breathing exercises (–1 SD), the reduction was not significant (β=0.02; 95% CI –0.04 to 0.07; P=0.55).

Here as well, an additional piecewise mixed model has demonstrated that engagement in video sessions significantly moderated the reduction in question 2 rank during weeks 1-6 (β=-0.01; 95% CI –0.01 to –0.00; P=0.01), yet no significant reduction was found in weeks 6-10 (β=0.00; 95% CI –0.02 to 0.02; P=0.84). The weekly average number of video session completion is 6.38 (8.59 SD). Furthermore, results from the simple slope analysis showed that users with increased engagement in video sessions (+1 SD) were significantly associated with improvement in question 2 rank (β=-0.08; 95% CI –0.14 to –0.02; P=0.01), compared to users with lower engagement (–1 SD) to video sessions (β=0.03; 95% CI –0.02 to 0.08; P=.27).

**Discussion**

**Principal Results**

The Dario Health behavioral health stress program delivers an optimized solution that combines videos, textual skills, breathing exercises, and interaction with a coach to improve the users’ emotional wellness. The goal of this study was to assess changes in stress outcomes over time among users with moderate and above levels of stress scores at baseline. This study tested the effect of digital therapeutic use on stress levels over time, using a specific engagement instrument for stress conditions. Piecewise mixed model analysis indicated a significant decrease in stress levels during the period of the first 6 weeks of product use and maintained during the period from 6 to 10 weeks with no significant time-related trends. Engagement in breathing exercises significantly moderated the reduction in stress levels during weeks 1-6. During the period of 6-10 weeks, the reduced level of stress was maintained, and there were no significant time-related trends. An additional piecewise mixed model revealed that engagement in video sessions significantly moderated the reduction of perceived stress as shown by a decreased sense of burden (question 1) and a decreased lack of productivity (question 2), during weeks 1-6. It is worth noting that stress can have various causes and effects, and everyone experiences stress differently. However, a sense of burden and lack of productivity are commonly associated with stress, and identifying and acknowledging them can help individuals take steps to manage their stress levels and improve their well-being.

This real-world analysis presents new evidence regarding the dynamic efficacy of digital therapeutic solutions for people with moderate to high stress levels. The results are consistent with previous research that observed improvements in mental health metrics following digital, app-based interventions [26,29,33]. Our findings also provide insight into the nonlinear nature of stress level reduction during the first 10 weeks of a digital behavioral health program by showing the direct association with breathing exercises during the improvement stage in the first 6 weeks and emphasizing the key role of breathing exercises on stress reduction. In the next 6-10 weeks, the reduced levels of stress are maintained indicating that experiencing the use of a behavioral health digital program is a sustainable approach that may create a positive environment leading to a meaningful change. A nonlinear effect of digital therapeutics has previously been demonstrated in studies of mental health recovery via digital therapeutic intervention [50,55]. Additionally, the nonlinear effect of chronic condition management has been previously demonstrated in studies of chronic condition management such as glucose levels in diabetes, blood pressure, and pain level changes over time [53,54,56].

The study of nonlinear change requires multiple assessments over time and examination of the individual trajectories of variables [55,57]. If the variables of interest are assessed frequently over the course of product use, processes can be examined to better understand what facilitates and inhibits change. In this study, we examined time course data moving beyond the question of whether stress level change occurs and toward an understanding of how change occurs. A piecewise mixed model analysis indicated a significant decrease in stress level during the period of the first 6 weeks of product use and maintained during the period from 6 to 10 weeks with no significant time-related trends. This real-world intervention may have captured how change in coping may occur. In other words, people initially improve by learning to cope more effectively with stress but that needs to be incorporated, so this change can occur over only a limited period before it can be truly integrated.

Based on current findings it appears that it takes 6 weeks to learn and then 6-10 weeks to integrate before becoming a changed behavior.

Although this real-world study does not relate to gender-specific experiences, it is notable that the gender ratio predominantly consists of women by 79% (388/490). Per previous studies, women are more likely to use e-mental health programs, designed features, and practices to reduce symptoms of stress or depression although they do not differ in terms of their general internet use [58]. The prevalence of major depressive episodes in women is about 2 times higher than that in men [59]. This could be explained by the existence of gender norms,
men are less likely than women to disclose mental health symptoms and often delay seeking help until symptoms become severe [60,61]. From the social role perspective, women often have a multitasking agenda hence, efficiency and convenience of treatment may be more regarded [62,63].

In terms of physiological aspects, men and women were not seen to differ in terms of breathing practice and integrative body-mind training for dealing with stress conditions [64].

To date, there have been very few large-scale studies examining the effectiveness of digital therapeutic interventions for emotional disorders delivered in real-world settings [65-67]. There is also limited quantitative literature on the associations between the program components and direct methods with clinical outcomes [68].

We hypothesized that breathing techniques may be an enhanced method to address and reduce the bodily and mental processes associated with stress. Breathing exercises are an easily taught and practiced direct method, and they can realize almost immediate benefits and therefore allow for greater behavioral change. Stress causes sympathetic activation; the sympathetic nervous system is 1 of the 2 branches of the autonomic nervous system [35]. The sympathetic nervous system plays a critical role in the body’s response to stress. When it is activated, it triggers the “fight or flight” response, which prepares the body for immediate action in response to a perceived threat or danger [69]. The sympathetic nervous system is activated by the hypothalamus in response to a threat, which triggers the release of hormones such as adrenaline and noradrenaline [70]. Therefore, various physiological responses are caused such as increased heart rate, blood pressure, and respiratory rate, as well as the release of glucose from the liver to provide energy for the body’s response [71]. Our findings demonstrated that engagement with breathing exercises significantly moderated the reduction in stress levels during the period of weeks 1-6. Slow and deep breathing is known to activate the parasympathetic nervous system.

Stress can have various causes and effects, and everyone experiences stress differently. However, a perceived sense of burden and a perceived lack of productivity are commonly associated with stress. Identifying and acknowledging these stress-related feelings are the first steps in taking proactive measures to address them. Therefore, we also hypothesized that the ranking of the two items in the questionnaire, that is, (1) “Have you felt overwhelmed or stressed about important responsibilities?” and (2) “What about feeling as though you were not as productive as you should’ve been?” would be associated with engagement with certain digital components. We assumed that an improvement in perceived stress is reflected by the ways those 2 items would be linked to product use when combining video sessions with digital tools. Our findings showed that both items’ rankings 1 and 2 were moderated by engagement in video sessions as well as in breathing exercises. Engagement with CBT-based video sessions has the potential to be effective and efficient in treating symptoms of mental health conditions in real-world settings [72-74]. Managing stress and handling challenges are obtained by learning how to cope with stress in day-to-day life and adjusting it to the level of stress necessary for optimal functioning. The video sessions explain the core concepts of problem-solving and decision-making to reduce burden and modulate perceptions of responsibilities.

One of the strategies that helped to manage stress and improve well-being is setting realistic goals and prioritizing tasks [75]. Offering content and recommendations about breaking down tasks into manageable steps and setting realistic goals that combat the style of perfectionistic thinking possibly helped individuals to reduce feelings of burden and increase productivity reflected in items 1 and 2 in the questionnaire.

It was consistently demonstrated in our findings that breathing exercises moderated the improvement of the general stress score and the ranking of items 1 and 2 as well. Breathing exercises have a significant impact on reducing perceived stress and promoting relaxation [64,76]. Deep breathing exercises, such as diaphragmatic breathing or belly breathing, engage the diaphragm muscle and stimulate the parasympathetic nervous system. The parasympathetic nervous system is the other branch of the autonomic nervous system, and it is responsible for promoting rest, relaxation, and digestion [35]. It has been demonstrated that during respiration, inspiration inhibits sympathetic nervous system activity [77]. Therefore, breathing techniques could be used as first-line and supplemental treatments for stress and emotional disorders [35,78].

Deep breathing exercises can help slow down the heart rate, bringing it closer to a calm and regular rhythm, contributing to a feeling of relaxation [79]. By consciously focusing on deep breaths and consciously relaxing the muscles, breathing exercises help release tension and promote a sense of physical relaxation [79]. During breathing exercise oxygen supply is increased supporting better delivery of oxygen to the body’s tissues, including the brain [80]. It can improve mental clarity, focus, and overall well-being. Lastly, engaging in breathing exercises requires focusing attention on the breath. This helps shift the focus away from stressors, worries, and racing thoughts, promoting a mindful state of being in the present moment. Mindfulness has been shown to reduce stress and enhance overall psychological well-being [81].

Recent reviews in the field of mental health programs highlight the need for research that goes beyond examining the overall effects of such programs. Instead, there is a call for more nuanced investigations that delve into the associations between specific components or building blocks of these programs and the clinical outcomes observed. This approach aims to comprehend the relative contributions of various activities within digital health management protocols. Researchers and digital health organizations can gain a deeper understanding of what works best in digital health management protocols and improve the effectiveness of mental health interventions [65,68].

Our findings indicate that breathing exercises and CBT-based video sessions help people lower their stress levels. These findings suggest that increasing digital engagement to specific activities in a lower engaged population may be an efficient way to optimize digital platforms supporting patients with moderate levels of stress. New models of care and digital tools have the potential to transform health care by increasing
accessibility, promoting patient-centered care, improving
timeliness, enhancing equity, and optimizing efficiency. Digital
tools can empower patients to take an active role in their health
care journey. This shift toward patient-centered care promotes
collaboration and empowers individuals to make informed
choices about their health [82,83].

We expect that the analytical approach applied in this study will
be beneficial for personalizing interventions and optimizing
incentivization planning. This information could be used to
further personalize outreach to encourage users to maintain their
personal critical level of mental health activities. More research
is required to further understand the current and potential
moderating factors of digital intervention tools for mental health
and how independent use of these tools facilitates the care of
patients with day-to-day stress levels.

Limitations
In this study, we used a single-arm design that analyzed
retrospective real-world data that lacked a comparison group
or control group, contained a self-selection bias, and lacked
randomization. While this design allowed us to explore the
association between digital relaxation tools and stress levels,
we acknowledge the limitation of not having a control group.
The lack of a control group restricts our ability to establish
causal relationships and control potential confounding variables.
Future research efforts could consider incorporating a control
group to further validate and extend the implications of our
findings. Additionally, we acknowledged the gender distribution
predominantly consisting of 79% (388/490) women that may
potentially influence the results and interpretation of the results.
However, women have higher rates of anxiety and depression
and use more digital treatment [58,59], therefore our results
may be specifically applicable to the real world. In other words,
the real-world population of women who have more anxiety
and depression may be better served with the use of a mobile
app. Although this study’s research topics are not related to
gender-specific experiences, future research should strive for a
more balanced gender representation. It is also possible that
people who chose to use digital therapeutic instruments were
those who were most motivated to make a change in their
conditions. These factors could limit the generalizability and
create a challenge for drawing causal conclusions. The time
scale was designed to reflect a weekly interval change over 10
weeks; however, the relationships of interest in this study could
be potentially investigated in different scales emphasizing daily
or monthly outcomes. Owing to the difficulty in tracking daily
changes in stress levels, most real-world studies focus on weekly
fluctuations.

Stress level was the key outcome assessed in this study by using
a stress questionnaire constructed by our internal experts. This
questionnaire demonstrated a Cronbach α value of 0.73, an
adequate reliability coefficient, as >0.70 is considered a measure
of internal consistency [84,85]. It was decided to use our stress
questionnaire although the Perceived Stress Scale-14, a 14-item
scale, is the most widely used for measuring perceived stress
[86] since it has fewer items and may be more useful in
increasing the answer rate among platform users.

Conclusions
This study provided evidence that an increase in mental health
management via digital engagement, specifically with relaxation
tools, can help reduce stress levels over time. Although stress
can be experienced in various forms, it is established that using
an app as an integrated part of the health self-management
process significantly improves clinical outcomes. It was
demonstrated that breathing exercises have a significant impact
on promoting relaxation. From a behavioral perspective,
engaging with CBT content helps recognize stress causes and
provides strategies to self-manage the feelings of burden,
ultimately improving well-being. These findings provide a basis
for encouraging users to take an active role and make informed
choices about their health. Future work is needed to investigate
current and other potential moderating factors and to develop
a more comprehensive understanding of the importance of the
app, how digital intervention tools work for mental health, and
for whom they are most effective. This knowledge will guide
the development and implementation of personalized and
evidence-based digital interventions, ultimately improving
mental health outcomes for individuals in a variety of contexts.

Conflicts of Interest
IB-A, OM, EB, and YF-H are employees of Dario Health. MDR and DLH are members of the Dario Health Scientific Advisory
Board. OM is a stock and option holder of Dario Health.

References


44.国资委 [Medline: 30100934]


The circulatory system and oxygen transport. Regulation of Tissue Oxygenation. URL: https://www.ncbi.nlm.nih.gov/books/NBK54112/ [accessed 2023-05-22]


Abbreviations

CBT: cognitive behavioral therapy

HIPAA: Health Insurance Portability and Accountability Act

©Inbar Breuer-Asher, Marilyn D Ritholz, David L Horwitz, Omar Manejwala, Ephraim Behar, Yifat Fundoiano-Hershcovitz. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
An Ecological Mobile Momentary Intervention to Support Dynamic Goal Pursuit: Feasibility and Acceptability Study

Ciarán O'Driscoll¹, DClinPsy, PhD; Aneesha Singh², PhD; Iya Chichua¹, BSc; Joachim Clodic¹, MSc; Anjali Desai¹, BSc; Dara Nikolova¹, BSc; Alex Jie Yap¹, BSc; Irene Zhou¹, BSc; Stephen Pilling¹, PhD

¹CORE Data Lab, Centre for Outcomes Research and Effectiveness, University College London, London, United Kingdom
²UCL Interaction Centre, University College London, London, United Kingdom

Corresponding Author:
Ciarán O'Driscoll, DClinPsy, PhD
CORE Data Lab, Centre for Outcomes Research and Effectiveness
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 207679 1897
Email: c.odriscoll@ucl.ac.uk

Abstract

Background: Individuals can experience difficulties pursuing their goals amid multiple competing priorities in their environment. Effective goal dynamics require flexible and generalizable pursuit skills. Supporting successful goal pursuit requires a perpetually adapting intervention responsive to internal states.

Objective: The purpose of this study was to (1) develop a flexible intervention that can adapt to an individual’s changing short to medium-term goals and be applied to their daily life and (2) examine the feasibility and acceptability of the just-in-time adaptive intervention for goal pursuit.

Methods: This study involved 3 iterations to test and systematically enhance all aspects of the intervention. During the pilot phase, 73 participants engaged in an ecological momentary assessment (EMA) over 1 month. After week 1, they attended an intervention training session and received just-in-time intervention prompts during the following 3 weeks. The training employed the Capability, Opportunity, Motivation, and Behavior (COM-B) framework for goal setting, along with mental contrasting with implementation intentions (MCII). Subsequent prompts, triggered by variability in goal pursuit, guided the participants to engage in MCII in relation to their current goal. We evaluated feasibility and acceptability, efficacy, and individual change processes by combining intensive (single-case experimental design) and extensive methods.

Results: The results suggest that the digital intervention was feasible and acceptable to participants. Compliance with the intervention was high (n=63, 86%). The participants endorsed high acceptability ratings relating to both the study procedures and the intervention. All participants (N=73, 100%) demonstrated significant improvements in goal pursuit with an average difference of 0.495 units in the outcome (P<.001). The results of the dynamic network modeling suggest that self-monitoring behavior (EMA) and implementing the MCII strategy may aid in goal reprioritization, where goal pursuit itself is a driver of further goal pursuit.

Conclusions: This pilot study demonstrated the feasibility and acceptability of a just-in-time adaptive intervention among a nonclinical adult sample. This intervention used self-monitoring of behavior, the COM-B framework, and MCII strategies to improve dynamic goal pursuit. It was delivered via an Ecological Momentary Intervention (EMI) procedure. Future research should consider the utility of this approach as an additional intervention element within psychological interventions to improve goal pursuit. Sustaining goal pursuit throughout interventions is central to their effectiveness and warrants further evaluation.

(JMIR Form Res 2024;8:e49857) doi:10.2196/49857

KEYWORDS

goal pursuit; ecological momentary intervention; ecological momentary assessment; mood; dynamics; network analysis; MCII; COM-B; support; pilot study; training; feasibility; acceptability; self-monitoring; implementation; psychological; effectiveness
Introduction

Background
We have all experienced difficulties in developing, pursuing, and achieving our goals. This can be influenced by our internal state (eg, mood and motivation), environmental factors (eg, competing goals or demands), and resources (eg, skills and opportunity). We need the ability to prioritize goal pursuit in a dynamic environment, determining how much effort to allocate and deciding when to shift our attention to other goals [1]. Managing this is a dynamic, within-person process that varies over time based on how much progress one has made toward their goals. These goals or tasks related to goals vary in how demanding they are, and our ability to pursue those goals will be influenced by capacity. Pursuing goals requires effort, both physical and mental, not just due to the difficulty of the task [2] but also for maintaining a mental representation of the goal [3]. Moreover, we need to make decisions around the allocation of resources to the task (eg, breaking down a task or abandoning it), affecting successful pursuit. Effort can be considered synonymous with motivation when measured objectively [4], with motivation influenced by expectancy or certainty and value attributed to the outcome. People exert more effort if the outcome is perceived to be more likely, important, and rewarding [5,6]. In addition, mood, particularly anhedonia, has been associated with behavioral reward processing deficits [4]. The application of theory to intervention suggests a need for a “perpetually adapting” intervention [7].

Successful goal pursuit involves numerous steps: option generation, cost-benefit decision leading to option selection, initiation, and pursuit [8]. Failure at any point can reduce the likelihood of pursuit, and there is a need to anticipate obstacles [9]. By considering obstacles, the individual is better able to anticipate and plan for challenges that may arise as they work toward their goal [10]. The individual must also be able to employ metacognitive strategies such as planning, self-monitoring, and flexibility to overcome challenges, and they may benefit from prompts and support to facilitate these strategies [11].

People’s intentions do not always translate into action: medium-to-large changes in intentions only lead to small-to-medium changes in behavior [12]. Most interventions focus on altering specific behaviors within specific contexts, and the results are not conclusive. Personalized feedback, goal setting, and self-monitoring appear promising, but they are not consistently effective across behaviors and contexts [13]. It is also unclear whether these skills generalize to other behaviors and contexts. Simple strategies or microinterventions can provide easy access and low-effort solutions to increase or maintain engagement in behavior change [14,15]. These strategies may be simple but difficult to sustain without practice.

Ecological Momentary Intervention (EMI) involves providing feedback or intervention to participants in real time based on the data collected from Ecological Momentary Assessment (EMA). The deployment of interventions via mobile devices (eg, smartphones) provides the opportunity to deliver intervention on scale as either a standalone or adjunct intervention. EMA alone can act as a form of self-monitoring, facilitating an awareness of thoughts, emotions, and behavior. Interventions can be personalized, based on momentary assessments, and delivered in anticipation of a change in the target behavior. In the context of adaptive ecological momentary interventions, push/pull strategies involve delivering interventions either proactively by the system (push) or in response to a user's request (pull), while just-in-time interventions are provided at opportune moments when they are likely to be most effective, based on a real-time assessment of the individual’s context and state [17]. This method can benefit the generalization of skill acquisition, where the just-in-time intervention prompts the individual to allocate increased resources toward skill acquisition.

We propose an intervention that aims to bolster skill acquisition (ie, effective goal pursuit) through a combination of evidence-based strategies and EMI implementation. These strategies include frequent self-monitoring, shown to improve goal attainment [13]; mental contrast with implementation intentions (MCI) [18], shown to produce a moderate effect on health behaviors [19]; and the Capability, Opportunity, Motivation, and Behavior (COM-B) framework [20] for goal setting. In addition, the intensive measurement of relevant goal-pursuit processes can be used to model the dynamics in daily life [21]. To develop an intervention for implementation within the clinical sample, we optimized the design following a research model for developing digital health interventions through iterations [22].

Study Aim
This study aimed to develop, evaluate, and implement a just-in-time adaptive intervention to improve goal pursuit. The intervention sought to provide training for a goal pursuit strategy that could easily be incorporated into participants’ daily lives in terms of time and effort. The goal was for this intervention to serve as an adjunct to behavioral interventions (whether psychological or health-related) in future research.

The aim was addressed by (1) developing a personalized approach to implementing a goal pursuit intervention in daily life; (2) identifying barriers and facilitators and monitoring the implementation process of the intervention through several iterations; and (3) piloting the intervention to evaluate its feasibility and acceptability, efficacy, and individual change processes by combining intensive single-case experimental design (SCED) and extensive methods. Figure 1 outlines the procedure for optimizing and piloting the ecological momentary intervention via 3 iterations.

https://formative.jmir.org/2024/1/e49857

JMIR Form Res 2024 | vol. 8 | e49857 | p.1409
(page number not for citation purposes)
Figure 1. Procedure for optimizing and piloting the ecological momentary intervention, outlining the steps undertaken within each iteration.

**Methods**

**Study Design**

This study involved 3 iterations to test and gradually improve all features of the intervention to ultimately inform a randomized control trial evaluation of the intervention. Two iterations focused on optimization, while the third iteration was an uncontrolled pilot study. After each iteration, feedback was used to improve the design and intervention to increase the feasibility and acceptability. Questionnaires and strategy training were delivered via Qualtrics and EMA/EMI using the m-path platform [23].

**Ethical Considerations**

Ethical approval was granted by the University College London (UCL) Research Ethics Department (21883/001). All participants were provided with details of the study, confirmed eligibility, and provided informed written consent. The study data were deidentified. All participants (including those who chose to withdraw partway through) were entered into a prize draw to win a £50 (US $64) voucher. They received £20 (US $25) and a personalized report outlining their results if they reached 70% completion of the EMA at the end of the study.

**Recruitment**

The participants were recruited through a convenience sampling approach from various social media platforms. Individuals were eligible if they (1) were over the age of 18 years; (2) able to read and understand English; (3) were based in the United Kingdom; and (4) had a personal smartphone with Android or iOS operating software. Because we were seeking to pilot in a “healthy” sample, our exclusion criteria were (1) current mental health difficulties, including anyone scoring above 15 on the 9-item Patient Health Questionnaire (PHQ-9) or >1 on the suicidal ideation item and 12 on the Generalized Anxiety Disorder 7 (GAD-7); and (2) anyone undergoing psychotherapy at the time the study was conducted. Figure 2 shows the study flowchart.

EMA studies require commitment from participants over a long period (up to 6 times a day for 28 days); therefore, dropout rates can be high. Appropriate incentives can encourage compliance. Therefore, we created a 2-fold incentive as outlined in the previous section. Participants were also incentivized through app functions such as earning badges and data visualization.
Developing the Intervention

Developing the intervention relied on adapting existing evidence-based strategies for goal pursuit. We aimed to develop an intervention that was adaptable to each participant’s goals and generalizable enough to accommodate shifts and changes in goal focus during their daily lives. A self-guided training session was developed and supplemented with abbreviated prompts that could be delivered throughout the intervention period (Multimedia Appendix 1).

Feasibility Assessment

Within each iteration, we collected information relating to the experience of participating in the study, specifically undertaking EMA and the intervention. This included Likert scale questions assessing the ease, helpfulness, and intrusiveness of the EMA; perceived effectiveness and utility of undertaking the intervention; and helpfulness of the strategy training guide. Participants could also provide qualitative feedback on their goal progress and experience. Feasibility was assessed through data relating flow of participant recruitment and retention throughout the intervention. Acceptability was assessed post intervention. This feedback took the form of a questionnaire assessing how feasible and acceptable the EMA schedule was for them; whether it was easy, helpful, or enjoyable; or whether it was intrusive and impeded optimal goal pursuit.

Assessment of Outcomes

Primary Outcome

Goal pursuit was used as the primary outcome to test the preliminary efficacy of our intervention, goal. This was measured using EMA, and participants were also asked about goal attainment post intervention.

Secondary Measures

To test whether the intervention influenced goal pursuit–related processes, we assessed participants’ pre-post changes on 5 goal pursuit–related measures, which are outlined in Textbox 1.
1. **Action orientation**: The Action Control Scale (ACS-90) [24] is divided into 3 subscales: Hesitation (8 items), the ability to initiate a task; Preoccupation (8 items), the ability to actively work on the task; and Volatility (6 items), the ability to stay action-oriented until completion.

2. **Defeatist Performance Beliefs**: The 15-item Defeatist Performance Beliefs (DPB) scale [25] measures overgeneralized negative thoughts in goal-striving.

3. **Prospective imagery**: The Prospective Imagery Test (PIT) [26] measures the ability to vividly imagine positive and negative future-orientated scenarios.

4. **Intertemporal choice**: The 27-item Money Choice Questionnaire (MCQ) [27] measures preferences between small immediate rewards and large delayed rewards. A general delay discounting parameter (“k”) is estimated, wherein greater k values represent steeper delay discounting.

5. **Difficulties in Emotion Regulation**: The 18-item Difficulties in Emotion Regulation Scale (DERS) [28] measures participants’ emotion regulation abilities.

### Ecological Momentary Assessment
The participants completed a series of questions numerous times a day over 28 days. The number changed across iterations. The questions aimed to (1) track mood and motivation; (2) assess goal characteristics such as reward, meaning, and importance; and (3) estimate the extent to which people are acting toward a goal, can visualize it, and feel self-efficacious.

In terms of goal pursuit, “I am acting toward a goal” was the primary outcome. Other questions measured mood, such as “How do you feel right now?” (smiley visual analog scale), anhedonia (“I’m enjoying what I am doing”), motivation (“I feel motivated”, expectation (“I feel hopeful”), energy (“I feel energized”), and questions related to goals, domain (“What I am doing is related to recreation/relaxation, education, relationships, work or health”), difficulty (“What I am doing is difficult”), importance (“What I am doing is important”), reward (“What I am doing is rewarding”), meaning (“What I am doing is meaningful”), implementation (“I know how I am going to reach this goal”), and representation (“I can picture myself achieving this goal”).

### Ecological Momentary Intervention
The study design was guided by the Risk of Bias in N-of-1 Trials (RoBiNT) scale to ensure the methodological quality of intervention studies employing single-case methodology (Multimedia Appendix 2). Details of each iteration and adaptation can be found in Multimedia Appendix 3.

### Pilot Component
Following the completion of baseline measures, the participants completed a 7-day EMA monitoring period, during which they completed 4 EMA surveys per day, occurring at semirandom intervals within an individualized 12-hour waking period (baseline). Following the completion of 7-day monitoring, the participants completed the self-guided COM-B/MCII training via the internet.

During the intervention period (the subsequent 21 days), all participants received a prompt at the start of each day, following which the goal pursuit variable alone was used to trigger prompts (falling 1 SD below their rolling average). If the individual indicated low goal pursuit, they were not asked subsequent questions relating to the goal (domain, difficulty, importance, reward, meaning, implementation, and representation), as this was considered aversive in previous iterations.

### Statistical Analyses

#### Acceptability / Feasibility
To evaluate the acceptability of study procedures, participants’ experiences were assessed in the postintervention questionnaire and analyzed descriptively.

#### EMA Analysis: Preprocessing Period
Participants with less than 48 time points were removed from the EMA analysis. In addition, participants were excluded when there was a baseline trend (ordinary least squares standardized beta coefficient >+/–0.3) on the basis that differences between phases can be difficult to interpret if improvement trends are observed in phase 1 (ie, due to natural improvement unrelated to the intervention).

#### SCED Analysis
The data obtained from the single-case experimental phase design used in this study have a hierarchical 2-level structure with observations (level 1) nested within individuals (level 2). We estimated design-comparable between-case standardized mean differences (BC-SMD) using restricted maximum likelihood methods [29]. We modeled baselines including both fixed and random effects for each level. The treatment phase was modeled with linear trends with both fixed and random effects at the level and slope. Assumptions were set around the likelihood methods [29]. We modeled baselines including both fixed and random effects for each level. The treatment phase was modeled with linear trends with both fixed and random effects at the level and slope. Assumptions were set around the session level error structure—an autoregressive model of order 1 (AR1) with variance differing by phase.

In addition, we estimated the differences in scores between the 2 phases using Ruscio A [30]. This metric reflects the probability that a randomly selected time point in phase 2 is larger than a randomly selected time point in phase 1 (calculated via Monte Carlo simulation: 10,000 runs) [31]. We also estimated the unstandardized difference in scores between the median values in the 2 phases.

#### Pre-Post Change
Differences between baseline and postintervention assessment measures were estimated using paired-sample t test.

#### Network Modeling
To explore the theoretical conceptualization of goal pursuit dynamics, we estimated a temporal network analysis. We
generated 2 models, the first with mood-related variables (mood, motivation, energy, hope, and interest) and goal pursuit (as these were captured at each time point), along with a separate model with goal-specific variables (difficulty, meaning, reward, importance, implementation, and representation) and goal pursuit. These were assessed separately, as the participants only rated the goal-specific variables if their goal pursuit was >1 SD below their rolling average.

We estimated multilevel vector autoregression networks via the mlVAR package on R software (R Foundation for Statistical Computing) [32], using the method lmer (sequential univariate multilevel estimation) with orthogonal estimation. We then visualized the autoregression networks with the qgraph package [33].

Both a temporal network (how the variable is predicted by all other variables after controlling for all other temporal effects) at the previous time point and a contemporaneous network (how variables are associated at the same time point, controlling for the influence of all other variables and temporal effects) were used within the results. The model assumes stationarity; as such, all items were detrended before including them in the network.

Results
Overview
As the purpose of the initial 2 iterations was to inform the optimization of the EMI, only data from iteration 3 (the pilot) are reported in this section.

Participant Characteristics
Sample characteristics are presented in Table 1. The total sample size was 73 participants (participants undertaking the EMI), with 65 (89%) completing postintervention measures. There were more female than male participants. While there was variation in ethnicity, there were no Black participants, and most of the sample were students. This was a healthy sample with low to no symptoms of depression (PHQ-9: mean 3.71, SD 3.73; range 0-15) and anxiety (GAD-7: mean 3.45, SD 3.65; range: 0-12).

| Table 1. Sociodemographic characteristics of the study sample (N=73). |
|--------------------------|-------------------|
| Characteristics          | Value             |
| Age (years), mean (SD)   | 25.05 (7.49)      |
| Female gender, n (%)     | 46 (63)           |
| Ethnicity, n (%)         |                   |
| Asian                    | 27 (37)           |
| Chinese                  | 15 (21)           |
| Mixed                    | 1 (1)             |
| Other                    | 4 (5)             |
| White                    | 26 (36)           |
| Employment status, n (%) |                   |
| Employed                 | 25 (34)           |
| Student                  | 46 (63)           |
| Unemployed/unable to work| 2 (3)             |

Acceptability Assessment
Post intervention, the participants completed a survey on the acceptability of the intervention (Figure 3, Multimedia Appendix 4). A total of 22 (35%) respondents stated they achieved the goal “a little better than expected,” 16 (25%) reported that they achieved the goal “much better than expected,” and only 2 (3%) reported not achieving the goal or experiencing a decline in their ability to reach a desired objective. When asked to assess the general ability to effectively pursue goals, 14 (22%) participants said it became much better and 33 (52%) reported a slight improvement. Moreover, 2 (3%) participants indicated that their overall competence in accomplishing goals had deteriorated. One participant’s goal was to get a promotion, and they did not get it; it wasn’t clear why the other participant thought their competence had deteriorated.

Open feedback on user experience was largely positive. In general, participants endorsed the simplicity and interactivity of the app, which had a visualization component to help track their responses. They suggested decreasing the number of notifications during the day and making a more varied list of EMAs to avoid respondent fatigue. The participants stated the MCII strategy was helpful, and engaging with the app helped them become more aware of their own goal-related behaviors:

...Helped to keep reminding myself what I needed to accomplish and keep it in the forefront of my mind. [Participant #15]

Whenever I see a reminder from the app, I seem to be persuaded to do something to change the current situation, even though I might have answered the questions with a negative emotion. [Participant #32]
Feasibility Assessment
To assess feasibility, we examined compliance with the EMA. Among all those who started the EMA (N=73), 10 (14%) were excluded from statistical analysis due to low compliance. Participants were sent 112 EMA surveys. On average, all 73 (100%) participants completed a mean of 85 (SD 33) or 76% of the EMAs across the month. While the strategy was presented at the first beep of each day, the prompt was also triggered by changes in mood on average 18 (SD 13) times per participant over a 21-day period.

Primary Outcome: Goal Pursuit SCED Analysis
We assessed changes in goal pursuit between the baseline and intervention phases for 73 participants (Multimedia Appendix 5). Throughout the experiment, the participants’ goal pursuit domains were recreation/relaxation (1024 observations, 29%), education (882 observations, 25%), relationships (559 observations, 16%), work (581 observations, 16%), and health (508 observations, 14%). There was a small effect size (BC-SMD 0.15, 95% CI 0.03-0.27). The intervention had an immediate significant effect, increasing the participants’ goal pursuit by 0.495 units (standard error: 0.152) (P<.001) but no significant additional improvement during the intervention period, at an intervention trend of 0.002 (P=.002). The probability of superiority (Ruscio A) was 0.59 (95% CI 0.54-0.63). There was a large amount of heterogeneity (I² =78.2%; H²=4.6), with 20 (27%) participants demonstrating CIs >0.5, 3 (4%) below 0.5 (suggesting poorer performance during the intervention phase), and the remaining 50 (69%) unclear (CIs crossing 0.5). The median difference between phases was 0.41 (SD 1.29).

Secondary Outcomes
A total of 65 (89%) participants completed the postintervention measures. No significant pre-post changes were noted for Defeatist Performance Beliefs (DPB) (t_{64}=0.36, P=.72), Prospective Imagery Test (PIT) (t_{64}=0.47, P=.64), DERS (t_{64}=0.47, P=.64), MCQ (t_{52}=1.36, P=.19), PHQ-9 (t_{61}=1.65, P=.10), and GAD-7 (t_{64} = –0.24, P=.81). GAD-7 and PHQ-9 were affected by floor effects. On the Action Control Scale (ACS) subscales, hesitation was significant (t_{61}=2.1121, P=.04) but not volatility (t_{63}= –0.76017, P=.45) or preoccupation (t_{61}=1.0326, P=.31). There was no significant change between baseline and intervention phases for the other EMA variables: mood (BC-SMD 0.03, 95% CI –0.07 to 0.13), motivation (BC-SMD 0.03, 95% CI –0.08 to 0.14), energy (BC-SMD –0.01, 95% CI –0.13 to 0.11), anhedonia (BC-SMD 0.01, 95% CI –0.12 to 0.10), and expectancy (BC-SMD 0.06, 95% CI –0.07 to 0.19).

Contemporaneous Networks
To elucidate dynamic processes during goal pursuit, we also estimated the network of associations between variables (Figure 4). The contemporaneous network visualized the partial correlations between variables at the same time point (controlling for the influence of all other variables and temporal effects). In the primary network, mood-related variables were associated as expected (mood, motivation, energy, and expectation) and anhedonia, motivation, and energy directly were associated with goal pursuit (explained variance for goal pursuit: $R^2= 0.24$). Meanwhile, in the secondary network, goal-specific variables were strongly associated, with representation, implementation, importance, and reward directly associated with goal pursuit (explained variance for goal pursuit: $R^2= 0.07$).
**Temporal networks**

The temporal network demonstrated how the variables predicted each other from 1 time point to the next. Within the primary temporal network, all variables demonstrated strong autocorrelations; only motivation predicted pursuit at the next time point, with motivation predicted by energy. There was a bidirectional relationship between expectation and anhedonia, with anhedonia negatively predicting mood. Expectation and mood both influenced each other at the next time point. Within the secondary network, there were strong autocorrelations for all variables except implementation and reward. Pursuit was predicted by implementation but had a stronger influence on implementation. Representation also predicted implementation at the next time point, as well as importance and meaning. Meanwhile, implementation negatively predicted meaning at the next time point. Reward negatively predicted difficulty, while difficulty only predicted itself.

**Discussion**

**Principal Findings**

This study aimed to develop, evaluate, and implement a just-in-time adaptive intervention to improve goal pursuit. Overall, the results suggest that the digital intervention was feasible and acceptable to participants. Our results show that participants endorsed high acceptability ratings relating to both the study procedures and the intervention. While there was a high level of attrition between baseline measures and those setting up the app, there was also a high level of retention and completion for those who did begin the EMI. There was a significant improvement in goal pursuit (between baseline and intervention) with most participants achieving their primary
goal and reporting that they would continue using the strategy, supporting its potential effectiveness in promoting positive behavior change. There was no improvement in pre-post measures measuring processes associated with goal pursuit.

This study focuses on the dynamics of goal pursuit including the reality that people will switch between multiple goals. There is little in the intervention literature that covers dynamic within-person processes that help individuals pursue their goals over time. Similar digital intervention studies have examined the use of employing within-person dynamic data to inform prompts. Korinek et al [34] used dynamical systems modeling within their adaptive intervention to increase walking behavior in participants who were overweight to set an “ambitious but doable” goal for themselves. Fallon et al [35] used a microrandomized control design to randomize participants to an intervention option (providing goal or social feedback relating to a physical activity goal) based on each participant’s specific state. Notably, their results suggested that the effectiveness of the intervention depended on the stage of their goal pursuit (how close they were to attainment). This study contributes to the literature by using the within-person variation on goal pursuit to prompt the intervention, leading to improved goal pursuit over time.

The only change in the associated goal pursuit measures was for hesitation—the ability to initiate intended actions. This construct appeared to align with implementation intentions, looking to improve the ability to translate specific intentions into behavior. This construct has been suggested to be particularly important for goal-striving across numerous domains [36]. Given the emphasis on mental representation within the strategy, it was surprising that there was no improvement in representation. This may be due to the sensitivity of the measures, where similar measures have not been associated with task performance [37]. The measures did not capture the vividness or intensity of the imagery, which would be important phenomena underpinning scene construct [38], and were expected to be targeted. While we assessed processes related to goal pursuit, we did not assess the mechanisms of change related to all strategies. Self-monitoring, for instance, is thought to improve mental health and well-being by increasing emotional self-awareness [39,40]. Future studies should endeavor to identify measures related to the mechanisms of change.

The network analysis can inform our understanding of dynamic goal pursuit highlighting a complex cyclical process involving interdependence, influence, and self-sustaining processes. There were no changes in the associated goal pursuit measures, either in terms of pre-post measures or EMA items. Contemporaneously, motivation and reward were directly associated with goal pursuit, while anhedonia was negatively associated with it. In terms of goal-specific constructs, representation, implementation, importance, and reward were directly associated with goal pursuit. Only motivation and implementation (knowing how the goal could be achieved) predicted goal pursuit at the next time point, with goal pursuit itself being the strongest predictor of goal pursuit. This would seem to suggest that the intervention directly targets goal pursuit rather than indirectly through an associated process (eg, motivation). Indeed, within the sample, motivational levels were high. Pursuit also predicted implementation, suggesting that goal pursuit is self-sustaining and boosts confidence in knowing how to achieve the goal. This is in line with the GOAL (goal-orientated action-linking) architecture, where motivation arises from the individual’s perception that their actions can impact the likelihood of achieving a goal [1]. Self-monitoring behavior and implementing the strategy may aid goal reprioritization, where the goal pursuit itself is a driver of further goal pursuit (within or between goals), as noted by the bidirectional relationship between pursuit and implementation.

**Strengths and Limitations**

The study had a few strengths and limitations. The development of the study procedure over multiple iterations facilitated adjustments including short momentary assessments, an efficient reporting process, low attrition, and high compliance. In addition to the app, the researchers provided consistent support through the intervention period, with reminder emails appearing particularly useful for improving compliance. For EMA technology to be successful in gathering accurate data and sustaining user interest over time, it is essential to engage users effectively. This is especially important when considering the translational application to mental health, where it can facilitate individuals taking a more active role in their recovery [41,42].

While there is an indication of an impact on goal pursuit, the single-case design reveals that while it might prompt change for some individuals, its effects remain unclear for the majority and entirely absent for others. In addition, the trend over the intervention did not indicate an incremental benefit; however, it may have been a sustaining pursuit, and without prompts, we may have seen a decline. Without a follow-up, it is also unclear whether the changes were maintained over time without prompts. Self-report requires self-awareness; indeed, we viewed the EMA as an active component enhancing awareness, but this could affect the measurement (either through meaning associated with items or change consequential to EMA), and this reactivity may interfere with causal claims [43].

This study aimed to personalize the just-in-time adaptive intervention. In the second iteration, we aimed to personalize the approach by identifying predictors of goal pursuit during the baseline period. However, this approach was hampered by poor compliance. In the pilot, we adopted a more conservative approach focusing solely on variation in goal pursuit; however, further studies may be able to improve the design by reinstating this approach. Finally, in relation to strategy training, the results from the initial iteration indicated that the web-based self-facilitated guide was considered optimal over video. There is evidence that the mode of learning may affect the effect size, where facilitator-led is stronger than self-facilitated [44]. Behavioral interventions that can be delivered via an app can address barriers that typically hamper engagement in intervention and may aid in study retention. Strategies for user engagement are a key aspect of EMI design. In this study, reducing participant burden was an important consideration; through iterations, we reduced the number of assessments and introduced branching of responses when not pursuing a goal. The number of assessments was still an issue for some participants, and this reduction comes at a cost to the availability.
and validity of data—for instance, nonrandom missing data (we had far fewer responses to model the secondary network). Passive monitoring helps reduce the burden but relies on proxy measures of goal pursuit and associated psychological processes, and it may not be as relevant for some goals as others [45].

Further research will need to consider piloting this EMI within a clinical sample before proceeding to a larger trial. The design has been optimized for easy integration into an individual’s daily life in terms of time and effort, with the intention that it could serve as an adjunct to behavioral interventions (psychological or health-related) in future research. Further considerations will need to be given to whether this should be tested as a standalone or an adjunct within an established intervention—for instance, facilitating behavioral activation for depression, or with cognitive behavioral therapy targeting the negative symptoms of psychosis. In this study, the participants were required to have enough motivation to pursue their goals. There is the question of whether this would be appropriate for those who lack motivation, as experienced in depression or psychosis. It is also unclear which aspects of the study design may need to vary, as it has been suggested that while compliance is related to fewer prompts in nonclinical samples, more frequent prompts led to higher compliance in studies with clinical samples [43]. Further research will need to carefully consider the design, intensity, and appropriateness of this EMI intervention for use within different clinical samples, while also accounting for differences in the mechanism of change between nonclinical and clinical samples.

**Conclusion**

This pilot just-in-time adaptive intervention used behavior self-monitoring, COM-B, and MCII strategies to improve dynamic goal pursuit. It was delivered via an EMI procedure and shown to be feasible and acceptable among a nonclinical adult sample. Given the potential feasibility, these results provide a foundation from which future research may implement a more rigorous methodology to assess efficacy within clinical populations that experience goal pursuit deficits. There was preliminary evidence of an effect on goal pursuit. However, this should be tested in a fully powered trial before drawing conclusions. Future research should consider the utility of this approach as an additional intervention element within psychological interventions to improve goal pursuit. Sustaining goal pursuit throughout interventions is central to their effectiveness and warrants further evaluation.

**Data Availability**

The data sets and scripts generated and analyzed during this study are available in the OSF repository [46].

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Details of the training session.

[DOCX File, 30 KB - formative_v8i1e49857_app1.docx ]

Multimedia Appendix 2

Methodological single-case experimental design (SCED) approach based on the Risk of Bias in N-of-1 Trials (RoBiNT) scale.

[DOCX File, 22 KB - formative_v8i1e49857_app2.docx ]

Multimedia Appendix 3

Details of each iteration and adaptions.

[DOCX File, 21 KB - formative_v8i1e49857_app3.docx ]

Multimedia Appendix 4

Acceptability questions and ratings.

[DOCX File, 20 KB - formative_v8i1e49857_app4.docx ]

Multimedia Appendix 5

Ecological Momentary Assessment (EMA) descriptives.

[DOCX File, 20 KB - formative_v8i1e49857_app5.docx ]

**References**


9. Hofmann W, Baumeister RF, F	e


46. OSF Repository. URL: https://osf.io/sa6h9/ [accessed 2024-03-06]

Abbreviations

ACS: Action Control Scale
AR1: autoregressive model of order 1
BC-SMD: between-case standardized mean difference
COM-B: Capability, Opportunity, Motivation, and Behavior framework
DBP: Defeatist Performance Beliefs
EMA: Ecological Momentary Assessment
EMI: Ecological Momentary Intervention
GAD-7: Generalized Anxiety Disorder 7
GOAL: goal-orientated action linking
MCII: mental contrasting with implementation intentions
PHQ-9: 9-item Patient Health Questionnaire
PIT: Prospective Imagery Test
RoBiNT: Risk of Bias in N-of-1 Trials
SCED: single-case experimental design
Effectiveness of a Web-based and Mobile Therapy Chatbot on Anxiety and Depressive Symptoms in Subclinical Young Adults: Randomized Controlled Trial

Stanisław Karkosz1*, MA; Robert Szymański1*, BA; Katarzyna Sanna2*, PhD; Jarosław Michalowski1*, PhD

1Laboratory of Affective Neuroscience in Poznan, SWPS University, Warsaw, Poland
2Center for Research on Personality Development in Poznan, SWPS University, Warsaw, Poland
* all authors contributed equally

Corresponding Author:
Stanisław Karkosz, MA
Laboratory of Affective Neuroscience in Poznan
SWPS University
Chodakowska 19/31
Warsaw, 03-815
Poland
Phone: 48 22 517 96 00
Email: skarkosz@swps.edu.pl

Abstract

Background: There has been an increased need to provide specialized help for people with depressive and anxiety symptoms, particularly teenagers and young adults. There is evidence from a 2-week intervention that chatbots (eg, Woebot) are effective in reducing depression and anxiety, an effect that was not detected in the control group that was provided self-help materials. Although chatbots are a promising solution, there is limited scientific evidence for the efficacy of agent-guided cognitive behavioral therapy (CBT) outside the English language, especially for highly inflected languages.

Objective: This study aimed to measure the efficacy of Fido, a therapy chatbot that uses the Polish language. It targets depressive and anxiety symptoms using CBT techniques. We hypothesized that participants using Fido would show a greater reduction in anxiety and depressive symptoms than the control group.

Methods: We conducted a 2-arm, open-label, randomized controlled trial with 81 participants with subclinical depression or anxiety who were recruited via social media. Participants were divided into experimental (interacted with a fully automated Fido chatbot) and control (received a self-help book) groups. Both intervention methods addressed topics such as general psychoeducation and cognitive distortion identification and modification via Socratic questioning. The chatbot also featured suicidal ideation identification and redirection to suicide hotlines. We used self-assessment scales to measure primary outcomes, including the levels of depression, anxiety, worry tendencies, satisfaction with life, and loneliness at baseline, after the 2-week intervention and at the 1-month follow-up. We also controlled for secondary outcomes, including engagement and frequency of use.

Results: There were no differences in anxiety and depressive symptoms between the groups at enrollment and baseline. After the intervention, depressive and anxiety symptoms were reduced in both groups (chatbot: n=36; control: n=38), which remained stable at the 1-month follow-up. Loneliness was not significantly different between the groups after the intervention, but an exploratory analysis showed a decline in loneliness among participants who used Fido more frequently. Both groups used their intervention technique with similar frequency; however, the control group spent more time (mean 117.57, SD 72.40 minutes) on the intervention than the Fido group (mean 79.44, SD 42.96 minutes).

Conclusions: We did not replicate the findings from previous (eg, Woebot) studies, as both arms yielded therapeutic effects. However, such results are in line with other research of Internet interventions. Nevertheless, Fido provided sufficient help to reduce anxiety and depressive symptoms and decreased perceived loneliness among high-frequency users, which is one of the first pieces of evidence of chatbot efficacy with agents that use a highly inflected language. Further research is needed to determine the long-term, real-world effectiveness of Fido and its efficacy in a clinical sample.

Trial Registration: ClinicalTrials.gov NCT05762939; https://clinicaltrials.gov/study/NCT05762939; Open Science Foundation Registry 2cqt3; https://osf.io/2cqt3
Introduction

Background

The interest in digital mental health apps has largely increased in recent years [1]. Their growing popularity results mainly from the pressure to use technology during the COVID-19 pandemic, which coincided with a fast deterioration in public mental health and the increasing quality of digital health technologies. Therapy applications have been proven to be helpful for clients not able to afford traditional therapy and for therapists seeking solutions to increase client engagement in therapy [2]. A recent meta-analysis [3] showed that using an internet-based intervention can be as effective as face-to-face therapy. Nevertheless, applications that are available on the market often lack appropriate scientific evidence of feasibility or efficacy [4].

One solution that seems to be especially promising is agent-guided cognitive behavioral therapy (AG-CBT [5]), in which interventions are provided by chatbots—applications backed by machine learning algorithms that mimic natural conversation while communicating with users via a chat interface [6]. The development of such applications has been approached in many different ways, but one bot that seems to be the most advanced so far is Woebot. Woebot is a self-help chatbot using CBT techniques such as psychoeducation, goal planning, and mood tracking to lower the levels of depression, anxiety, substance abuse, and, recently, postnatal depression [5,6].

In a randomized controlled trial (RCT), the use of Woebot for a period of 2 weeks has been proven to be more effective at reducing symptoms of anxiety and depression than the use of self-help materials prepared by the World Health Organization (WHO) [6]. Further studies have provided evidence that users can develop a bond with the Woebot on a similar level as the one that is built between the client and therapist during group CBT [5,7].

Previous experiences with English-speaking mental health care chatbots (eg. Woebot, Wysa, Youper [8-10]) have encouraged attempts to develop chatbots in other language versions, such as German [11], Chinese [12], Spanish [13], and Ukrainian [14], or even multilingual chatbots [15]. Currently, the development of such applications varies among high-income and low-income countries due to cross-cultural differences and specific obstacles [16]. In Poland, there is still a limited number of digital therapeutic solutions, although the need for them is growing. In the last few years, there has been a visible decline in mental health among Polish teenagers and young adults, which became especially severe during the COVID-19 pandemic [17].

To address that need, our team initiated the development of Fido—the first Polish therapy chatbot—that aims to provide mental health support to adolescents and young adults struggling with anxiety and depression. Our previous research on the interaction between humans and Fido showed that it is considerably user-friendly [18]. However, Fido still required an efficacy study, which is presented in this article along with the exploratory analysis of human–therapy chatbot interactions. We hope that this study will extend previous research in this field and enrich the discussion on agent-guided mental health treatment.

Objectives

A previous study on human-chatbot interactions using Fido provided satisfactory results, suggesting that Fido is pleasant to use; however, it hasn’t been optimized yet and required further development in the area of user experience (UX) [18]. After its UX optimization, Fido has been ready to use in an efficacy study. Therefore, we performed the first RCT aimed at testing the effects of using Fido to reduce subclinical depression and anxiety symptoms and compared them with the use of self-help materials from the book “Mind Over Mood” [19].

Based on previous clinical research of chatbots [6,8-10], we aimed to investigate the direct intervention effects and their stability. We hypothesized that, after a 2-week intervention and a 1-month follow-up, the following would occur:

1. The chatbot group reports lower depression, anxiety, and worry symptoms than those with the self-help book only.
2. The chatbot group has higher satisfaction with life than those using the self-help book only.
3. The positive affect is higher, and negative affect is lower than prior to intervention. Furthermore, this change is greater in the chatbot group than in the group with the self-help book.

Based on previous research on the chatbot-user interaction [8,18], we also hypothesized that:

1. Loneliness is lower in the chatbot group than in the group with the self-help book.
2. Participants form a bond with the chatbot, scoring at least 4 on the Working Alliance Inventory-Short Revised (WAI-SR) scale.
3. The users’ assessments of the chatbot's linguistic pragmatics correlate positively with the level of UX.

Methods

Trial Design

We used a 2 x 3 mixed factorial design with 2 intervention arms (Fido chatbot vs self-help book) and 3 time points (before the intervention [T1], immediately after the intervention [T2], and at a follow-up 1 month after the previous measurement [T3]).
The primary intervention lasted 2 weeks. After the intervention, the use of the technique was not obligatory (but it was not forbidden). For an overview of the procedure, see Figure 1.

**Figure 1.** Procedure flowchart with measures and time points (T0-T3).

---

**Participants**

Participants were recruited via Facebook and Instagram advertisements from May 2022 to June 2022 and screened using a Qualtrics online survey (measurement at T0). The estimated sample size was 80 participants, with an estimated dropout of around 15%. The sample size was based on previous studies of chatbot interventions [6,9], in which the estimated sample size (ie, N=70) allowed for detecting a moderate-large effect for depression. We increased the sample size to 80 due to the high dropout rate in previous studies (average of 7 participants per group) [6,9].

Eligibility criteria included (1) age between 18 years and 35 years; (2) not undergoing psychotherapy, coaching, nor psychopharmacological treatment; (3) no diagnosis of a neurological disorder; (4) declaring at least mild depressive or anxiety symptoms by achieving a total score of at least 16 points on the Center for Epidemiologic Studies Depression Scale Revised (CESD-R) [20,21] or at least 50 points on the Penn State Worry Questionnaire (PSWQ) [22,23]; and (5) being able to visit the study site in Poznań (Poland) to complete the follow-up measurements. Computer literacy and proficiency in Polish were implicitly presumed as all participants completed an online screening survey written in Polish.

After enrollment, the research team members performed simple randomization with a 1:1 ratio (via Python script). Because the intervention involved either using the chatbot or the book, participants were informed about their assignment. To reduce the effect of expectation bias, we intentionally masked our hypotheses about the superiority of chatbot-supported therapy. The research team members remained unblinded as well.

During the onboarding for the intervention, the participants were asked to use the assigned intervention technique as needed (ie, with no prespecified minimal time of use per day or week). The uptake of the intervention was monitored via regular commitment checks delivered by email, which could serve as a reminder to use the assigned intervention technique.

All participants received compensation for their involvement: ZŁ 90 (US $22.57) directly after the intervention and ZŁ 70 (US $17.56) at the follow-up.

---

**Ethics Approval and Informed Consent**

The study received approval from the Ethical Review Board at SWPS University of Social Sciences and Humanities (opinion no. 2022-158) and was registered during participant enrollment in the Open Science Framework (OSF) Registries [24].

We retrospectively registered the protocol and analysis plan under the clinical trial number NCT05762939. Participants were informed about each intervention, step-by-step procedure, their right to withdraw, and research data confidentiality. All participants expressed their consent via a checkbox on a screening survey. Should one of the interventions have proved more beneficial, we ensured the participants from the other arm could gain access to that intervention’s materials.

Due to the Messenger chatbot policies [25], the research team members were hypothetically able to access individual participant messages from the chatbot conversations. This limitation is due to the fact that, formally, Messenger chatbots are an interface to send messages to Facebook pages. Participants from the experimental condition were informed of this potential privacy breach via a consent form explicitly presented in the chatbot and Meta’s Data Policy [26].

**Interventions**

**Chatbot Intervention—Fido**

For the experimental group, we used a free, prerelease version of a therapy-supporting Polish-language chatbot, called Fido [27], integrated into Facebook Messenger. Participants were added as testers in Meta’s development website and accessed the chatbot via links embedded in individual emails. They were given no special training (apart from the initial email instructions) but were offered technical assistance in case of any problems.

The chatbot uses machine learning models for intelligent user intent detection and close-ended input methods (such as choosing 1 option from a list). It was developed using iterative co-development with focus groups consisting of therapists and potential users (for more information, see [18]). Moreover, external therapists provided quality assurance for all therapeutic methods used by the chatbot, while software engineers used standard testing procedures. During the trial, none of the features underwent any changes.
During the onboarding procedure, participants indicate their gender and receive information on user terms and data protection, as well as basic training in cognitive biases because understanding them is crucial for interaction with Fido. After onboarding, users can try different therapeutic techniques implemented in Fido using a tree-based structure.

The primary functionality of Fido is providing dialogue focused on intelligent recognition of cognitive biases and their subsequent modification using Socratic questioning. It also recognizes suicidal ideation and reacts to it by redirecting users to emergency hotlines.

To maximize Fido’s performance, one of its machine learning models implements the so-called ABC technique (known from CBT), which helps patients organize and differentiate between activating events, beliefs, and their emotional or behavioral consequences [28], as presented in Figure 2. Fido also provides psychoeducation about depression, anxiety, and emotions. Last, it embraces gratitude practice exercises [19].

**Figure 2.** An example of a Fido-patient conversation (translated from Polish to English): ABC technique and cognitive distortion recognition.

---

**Control Intervention—Materials From “Mind Over Mood”**

Chapters 1-6 and 12 from the Polish translation of “Mind Over Mood” [19,29] were used to familiarize control participants with similar content and tasks as those provided to the experimental group. The book contains psychoeducation content with self-help exercises and is a well-established therapist’s guide to depression.

**Outcomes**

**Primary Outcomes Measures**

**Overview**

Primary outcomes were assessed at the 3 time points: T1, T2, T3. All questionnaires were implemented in Qualtrics and administered mostly offline at the study site; when participants could not visit the lab, they were given a link to an online version of the survey. Participants completed validated Polish adaptations of the scales. For primary analyses, we used sum scores of every single primary outcome measured at 1 time point (not change scores).

**CESD-R**

Depression symptom severity was measured using the CESD-R [20,21]. It is a 20-item screening tool for major depressive disorder. The scale uses 9 symptom groups defined in the Diagnostic and Statistical Manual (DSM)-5 [30]: sadness, anhedonia, appetite, sleep, thinking/concentration, guilt, fatigue, agitation, and suicidal ideation.

**Patient Health Questionnaire-9**

Another measure used to assess participants’ depression was the Patient Health Questionnaire-9 (PHQ-9) [31,32] from the...
self-administered version of the Primary Care Evaluation of Mental Disorders (PRIME-MD) inventory [33]. This brief, 9-item scale is based on DSM-IV criteria for depression. It is used mainly for symptom severity monitoring in primary care.

**PSWQ**

We used the PSWQ [22,23] to measure worry tendency, which is the primary component of generalized anxiety disorder. The 16-item scale addresses worry excessiveness, generality, and uncontrollable dimensions.

**State-Trait Anxiety Inventory**

The State-Trait Anxiety Inventory (STAI) [34,35] is another questionnaire used to measure anxiety, both as a temporary state and as a relatively fixed trait of an individual. We used the 20-item trait scale of STAI.

**Positive and Negative Affect Schedule**

The Positive and Negative Affect Schedule (PANAS) [36,37] was used to measure the general affect during the last 2 weeks. We calculated 2 subscale sum scores: 1 for positive feelings (9 items) and 1 for negative affect (9 items). The last 2 items of the original scale were not administered due to human error in the online survey implementation.

**Satisfaction With Life Scale**

Primary measures also included a brief, 5-item Satisfaction With Life Scale (SWLS) [38,39], which was used to assess global life satisfaction in a cognitive-judgmental aspect. Together with positive and negative affect, life satisfaction is an important component of subjective well-being.

**Revised UCLA Loneliness Scale**

We used the 20-item Revised UCLA Loneliness Scale (R-UCLA) [40,41] to assess other specific aspects of well-being, namely subjective feelings of loneliness and social isolation.

**Secondary and Other Outcome Measures**

**Overview**

During the 2-week intervention, participants received 5 engagement check surveys, each containing a single question: “How much time have you spent on therapeutic chatbot/book use during the last 48 hours?” Answers were in 10-minute increments. We calculated the sum score of the engagement checks as the total amount of declared time spent on book or chatbot use in minutes. Each survey had to be completed in less than 24 hours.

After the intervention (T2), all participants also were asked an open-ended question: “How many times have you used the chatbot / read the book in the past two weeks?” If a participant declared a range of values (eg, “10-12 times”), we recoded their answer as the median value in this range.

Another secondary measure used in both arms at T2 was a 12-item test that assessed the participants’ acquired knowledge of psychoeducation topics covered both by Fido and the “Mind Over Mood” book (see Multimedia Appendix 1). The test was administered in a paper-and-pencil format at the study site.

Participants from the chatbot arm were also asked to complete several additional scales at T2, including the WAIS-R and UX.

**WAIS-SR**

The WAIS-SR [42] is a 12-item scale that measures therapeutic alliance in 3 key areas defined by Bordin [43], which are (1) agreement on the tasks and (2) goals of the therapy and (3) overall patient-therapist affective bond. As agent-guided therapy also requires some form of therapeutic chatbot–user alliance, we adapted this scale to be used in our study. All items were first translated to Polish using a forward and backward translation with reconciliation. Next, we modified phrases related to human-led therapy to increase their relevance for the chatbot therapy (see Multimedia Appendix 2 for the original WAIS-SR items and their adaptation for this study). Items from the tasks and goals subscales were not as relevant to this study (because Fido does not establish these elements of the intervention; they are determined ad hoc by the user), so alongside the total average scale score, we also made use of the average of the bond subscale.

**UX**

For the general assessment of UX, we included nonmandatory items from several scales used in other studies: Acceptability E-scale, Human-Agent Interaction Scale (HAIS), Language Pragmaticality Scale (LPS).

The Acceptability E-scale [44] is used to measure the overall acceptability and usability of health-related computer applications. The team prepared its Polish translation using a similar procedure to our adaptation of WAIS-SR (see Multimedia Appendix 3 for both the original and the translation). The original scale included 5 items. One of the items regarded the functionality of the program. We used this item 4 times, as we wanted to check the functionality of 4 different techniques (psychoeducation, cognitive bias recognition, suicidal thought recognition, and gratitude practice). This way, we ended up with 9 items and used their sum as a total score.

The HAIS [45] has answers that range from 1 to 7 (see Multimedia Appendix 4). The calculated sums from 6 subscales were used: Supportive Anthropomorphic Traits (3 items), Unsupportive Anthropomorphic Traits (4 items), Behavioral Traits (7 items), Uncanny Valley (6 items), Competence (6 items), Warmth (6 items).

We previously used the LPS [18] to assess the perceived ability of Fido to communicate in a pragmatically sound manner. The scale consists of 4 separately scored items with answers on a percentage scale (0%-100%): “What percentage of overall chatbot statements were adequate?” “What percentage of overall chatbot statements were neutral (neither adequate nor inadequate)?” “What percentage of overall chatbot statements were inadequate?” “In what percentage have you been feeling understood by a chatbot?”

The preregistered plan of the study also included chatbot metadata (frequency of use or time spent in the conversation) for each user after the intervention (T2). However, due to technical and ethical reasons, we could not link the questionnaire data with anonymous chatbot metadata. Therefore, we only used the metadata (frequency of use) at a group level.

https://formative.jmir.org/2024/1/e47960

Karkosz et al
Analytical Methods

For baseline, we conducted contingency analyses on the nominal and ordinal data and 2-tailed Welch $t$ tests on the continuous data, to measure whether groups were different in any way at the beginning of the study.

To assess the direct treatment effects on each primary outcome, we conducted repeated measures ANOVA with arm (experimental/control) as a between-group factor, as well as time point (T1/T2) as a within-group factor. We repeated the analyses to test treatment stability, including the scores obtained before versus 1 month after the intervention (T1/T3).

After the intervention (T2), we calculated descriptive statistics on human-agent interaction measures as well as the frequency of use (subjective measures and metadata), declared time spent on the intervention, and knowledge test scores. We also analyzed Pearson correlations between the human-agent interaction measures.

Results

Participants

Of the 245 people screened, 81 were admitted to the study. For an overview of participant flow, see the CONSORT (Consolidated Standards of Reporting Trials) diagram (Figure 3). Basic demographic characteristics as well as screening scores (CESD-R and PSWQ at T0) are presented in Table 1. Participants from the chatbot and control conditions did not differ in terms of age, sex, education, employment, university student status, or screening scores (at T0). Moreover, there was no major difference between the groups in any primary outcome measures at baseline ($t_{72}<1.57, P_s>.12$). Detailed $t$ test results are presented in Multimedia Appendix 5.

Figure 3. Flow of participants through each stage of the study.
Table 1. Participants’ demographic characteristics and screening scores (Center for Epidemiologic Studies Depression Scale Revised [CESD-R] and Penn State Worry Questionnaire [PSWQ]) as well as P values for the between-group tests before randomization (at T0).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Treatment group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chatbot (n=40)</td>
<td>Control (n=41)</td>
<td>P valuea</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>29 (73)</td>
<td>29 (71)</td>
<td>.36</td>
</tr>
<tr>
<td>Male</td>
<td>11 (28)</td>
<td>10 (24)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0</td>
<td>2 (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>26.60 (5.06)</td>
<td>24.76 (4.01)</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
<td></td>
<td>.75</td>
</tr>
<tr>
<td>Employed</td>
<td>26 (65)</td>
<td>27 (66)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>7 (2)</td>
<td>5 (12)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>7 (2)</td>
<td>9 (22)</td>
<td></td>
</tr>
<tr>
<td><strong>University student status, n (%)</strong></td>
<td></td>
<td></td>
<td>.65</td>
</tr>
<tr>
<td>Student</td>
<td>19 (48)</td>
<td>23 (56)</td>
<td></td>
</tr>
<tr>
<td>Not a student</td>
<td>21 (52)</td>
<td>18 (44)</td>
<td></td>
</tr>
<tr>
<td><strong>Highest level of education, n (%)</strong></td>
<td></td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>General secondary school</td>
<td>17 (43)</td>
<td>19 (46)</td>
<td></td>
</tr>
<tr>
<td>University degree (bachelor's or higher)</td>
<td>23 (58)</td>
<td>22 (54)</td>
<td></td>
</tr>
<tr>
<td><strong>CESD-R score at T0, mean (SD)</strong></td>
<td>27.58 (13.35)</td>
<td>31 (14.24)</td>
<td>.15</td>
</tr>
<tr>
<td><strong>PSWQ score at T0, mean (SD)</strong></td>
<td>59.78 (10)</td>
<td>59.27 (9.91)</td>
<td>.82</td>
</tr>
</tbody>
</table>

aCalculated using the Pearson χ² test for nominal variables and 2-tailed Welch t test for continuous variables.
bPercentages were rounded and may exceed 100.

Efficacy After 2 Weeks

We analyzed the efficacy of the intervention, comparing the data from baseline (T1) and after the 2-week intervention (T2). Only the main effects of time were significant at α=.05. For the between-subject and interaction effect results, see the tables in Multimedia Appendix 5.

For depressive symptoms, there were moderate effects of time. Scores for the CESD-R ($F_{1.66}=62.58, P<.001; \omega^2=0.08$) and PHQ-9 ($F_{1.66}=34.18, P<.001; \omega^2=0.06$) decreased in both groups.

In terms of anxiety and worry tendency, there was a small decrease in symptom severity in both groups, as measured with the PSWQ ($F_{1.66}=10.78, P=.002; \omega^2=0.01$) and STAI ($F_{1.66}=25.87, P<.001; \omega^2=0.03$).

For both study arms, we also detected small increases in satisfaction with life (SWLS: $F_{1.66}=13.59, P<.001; \omega^2=0.01$) and positive affect (PANAS-P: $F_{1.66}=16.54, P<.001; \omega^2=0.04$), while negative affect decreased (PANAS-N: $F_{1.66}=24.02, P<.001; \omega^2=0.05$). The decrease in feelings of loneliness was very small and not significant ($F_{1.66}=3.47, P=.07; \omega^2=0.00$; see Table 2). For a visual representation of treatment efficacy and stability, see Figure 4 [46].
Table 2. Primary outcomes at baseline (T1), after the intervention (T2), and at the follow-up (T3), as well as difference scores (T1 vs T2 and T1 vs T3).

<table>
<thead>
<tr>
<th>Outcome per arm</th>
<th>Score at T1, mean (SD)</th>
<th>Score at T2, mean (SD)</th>
<th>Score at T3, mean (SD)</th>
<th>T1 vs T2, mean difference (95% CI)</th>
<th>T1 vs T3, mean difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CESD-R&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>30.74 (13.94)</td>
<td>23.03 (14.30)</td>
<td>23.00 (14.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANAS&lt;sup&gt;b&lt;/sup&gt;-Positive</td>
<td></td>
<td></td>
<td></td>
<td>2.56 (1.30 to 3.81)</td>
<td>2.51 (1.21 to 3.82)</td>
</tr>
<tr>
<td>Chatbot</td>
<td>17.89 (6.51)</td>
<td>20.27 (6.29)</td>
<td>21.22 (5.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>17.71 (5.52)</td>
<td>20.66 (7.09)</td>
<td>20.13 (5.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANAS-Negative</td>
<td></td>
<td></td>
<td></td>
<td>–3.83 (–5.39 to –2.27)</td>
<td>–4.38 (–6.26 to –2.50)</td>
</tr>
<tr>
<td>Chatbot</td>
<td>31.19 (7.70)</td>
<td>27.00 (7.26)</td>
<td>26.81 (8.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>32.50 (7.86)</td>
<td>29.77 (8.76)</td>
<td>29.16 (8.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td>–2.56 (–3.44 to –1.69)</td>
<td>–2.82 (–3.91 to –1.72)</td>
</tr>
<tr>
<td>Chatbot</td>
<td>9.97 (4.98)</td>
<td>7.64 (4.53)</td>
<td>7.22 (4.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>11.87 (5.40)</td>
<td>9.11 (5.59)</td>
<td>9.50 (5.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSWQ&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td>–2.36 (–3.79 to –0.92)</td>
<td>–2.84 (–4.47 to –1.21)</td>
</tr>
<tr>
<td>Chatbot</td>
<td>9.97 (4.98)</td>
<td>7.64 (4.53)</td>
<td>7.22 (4.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>11.87 (5.40)</td>
<td>9.11 (5.59)</td>
<td>9.50 (5.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-UCLA&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td>–1.56 (–3.23 to 0.11)</td>
<td>–1.11 (–3.14 to 0.92)</td>
</tr>
<tr>
<td>Chatbot</td>
<td>45.44 (12.15)</td>
<td>43.06 (13.38)</td>
<td>41.67 (13.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>42.39 (13.25)</td>
<td>41.49 (13.42)</td>
<td>43.69 (12.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAI&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td>–3.31 (–4.61 to –2.01)</td>
<td>–3.40 (–4.99 to –1.80)</td>
</tr>
<tr>
<td>Chatbot</td>
<td>51.86 (8.75)</td>
<td>48.82 (9.25)</td>
<td>47.44 (9.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>52.42 (8.99)</td>
<td>48.54 (8.94)</td>
<td>49.53 (7.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWLS&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td>1.44 (0.66 to 2.22)</td>
<td>1.34 (0.34 to 2.34)</td>
</tr>
<tr>
<td>Chatbot</td>
<td>17.61 (6.67)</td>
<td>18.97 (7.03)</td>
<td>20.00 (7.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>19.50 (5.49)</td>
<td>20.91 (6.07)</td>
<td>19.91 (6.00)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>CESD-R: Center for Epidemiologic Studies Depression Scale Revised.
<sup>b</sup>PANAS: Positive and Negative Affect Scale.
<sup>c</sup>PHQ-9: Patient Health Questionnaire-9.
<sup>d</sup>PSWQ: Penn State Worry Questionnaire.
<sup>e</sup>R-UCLA: Revised UCLA Loneliness Scale.
<sup>f</sup>STAI: State-Trait Anxiety Inventory.
<sup>g</sup>SWLS: Satisfaction With Life Scale.
Figure 4. Mean (95% CI) scores at screening, at baseline (Pre), after the intervention (Post), and at the 1-month follow-up for the following primary outcomes: (A) Center for Epidemiologic Studies Depression Scale Revised (CESD-R), (B) Penn State Worry Questionnaire (PSWQ), (C) Satisfaction With Life Scale (SWLS), (D) State-Trait Anxiety Inventory (STAI), (E) Patient Health Questionnaire-9 (PHQ-9), (F) Revised UCLA Loneliness Scale (R-UCLA), (G) Positive and Negative Affect Scale (PANAS)-Positive, (H) PANAS-Negative. Confidence intervals were difference- and correlation-adjusted using the superb R library described by Cousineau et al [46].

Treatment Stability After 1 Month

The measurements from the follow-up (T3) were compared with those at baseline (T1) to determine if previously observed therapeutic effects were stable.

Depressive symptoms remained at a reduced level, as observed in both the CESD-R ($F_{1,57} = 32.10, P < .001; \omega^2 = 0.11$) and PHQ-9 ($F_{1,57} = 26.58, P < .001; \omega^2 = 0.07$) scores. A similar effect was found in anxiety (PSWQ: $F_{1,57} = 12.14, P < .001; \omega^2 = 0.01$) and
worry tendency (STAI: \( F_{1,57} = 18.12, P < .001; \omega^2 = 0.03 \)) scores. Stability in the therapeutic effect was also observed for satisfaction with life (SWLS: \( F_{1,57} = 7.22, P = .009, \omega^2 = 0.01 \)), positive affect (PANAS-P: \( F_{1,57} = 14.87, P < .001; \omega^2 = 0.04 \)), and negative affect (PANAS-N: \( F_{1,57} = 21.67, P < .001; \omega^2 = 0.06 \)). There was not sufficient evidence to conclude that loneliness (as measured using R-UCLA scores) changed from baseline to follow-up (\( F_{1,57} = 1.19, P = .28; \omega^2 = 0 \)).

As discussed previously, no group and interaction (of group and time) effects were detected for any primary outcomes measures (for detailed results, see Multimedia Appendix 5).

**Human-Agent Interaction Effects**

In terms of human-agent interaction characteristics, Fido achieved a mean 2.71 (SD 0.94) points on the WAI-SR scale, with an average score of 3.25 (SD 1.14) on the bond subscale. We detected a moderate correlation of language pragmaticality (item 1: % of messages rated as adequate) with the general UX level measured using the Acceptability E-scale (\( r = 0.60, 95\% \) CI 0.32 to 0.78; \( P < .001 \)).

Groups were similar in terms of the frequency of use they reported after the intervention (\( t_{50} = 1.44, P = .15; d = 0.36, 95\% \) CI -0.13 to 0.84) and scores on the knowledge test (\( t_{50} = 0.02, P = .99; d = 0.01, 95\% \) CI -0.51 to 0.52). However, the analysis of the time spent on the intervention via regular commitment checks suggested that the control group (mean 117.57, SD 72.40 minutes) actually spent more time reading the book than the chatbot group (mean 79.44, SD 42.96 minutes) spent on the interaction with Fido (\( t_{50} = 2.75, P = .008; d = 0.64, 95\% \) CI 0.17 to 1.11).

**Exploratory Analyses**

We extended our preregistered analyses after gaining insight into the data. The therapeutic alliance score (WAI-SR) correlated positively with the subjective sense of being, as understood from the LPS (\( r = 0.79, 95\% \) CI 0.60 to 0.89; \( P < .001 \)), overall acceptability (\( r = 0.77, 95\% \) CI 0.57 to 0.88; \( P < .001 \)), and scores for competence (\( r = 0.71, 95\% \) CI 0.48 to 0.85; \( P < .001 \)) and behavioral traits (\( r = 0.45, 95\% \) CI 0.12 to 0.69; \( P = .01 \)) from the HAIS.

We also detected a moderate, negative correlation between the WAI-SR score and age (\( r = -0.37, 95\% \) CI -0.63 to -0.03; \( P = .04 \)), which motivated us to split the experimental group into 2 subgroups by median (younger and older participants) and compare them using the Student t test. Younger participants formed a stronger human-chatbot bond (mean 3.59, SD 0.97) than older participants (mean 2.79, SD 1.23; \( t_{31} = 2.11, P = .04; d = 0.74, 95\% \) CI 0.02 to 1.45).

Moreover, in the chatbot group, loneliness (R-UCLA) scores after the intervention (at T2) were negatively correlated with the declared frequency of use (\( r = -0.48, 95\% \) CI -0.71 to -0.16; \( P = .006 \)). This finding motivated us to split the observations by median into subgroups of high- and low-frequency chatbot users. Repeated measures ANOVA of the R-UCLA scores with low-frequency and high-frequency users as the between-subjects effect and 2 time points (T1-T2) as the within-subjects effect yielded a statistically significant interaction of frequency and time point (\( F_{1,66} = 7.417, P = .01; \omega^2 = 0.01 \); see Figure 5). After 2 weeks, only high-frequency users reported lower levels of loneliness (\( t_{16} = 3.69, P_{\text{Bonf.}} = .005; d = 0.44, 95\% \) CI 0.07 to 0.81).

**Figure 5.** Revised UCLA Loneliness Scale (R-UCLA) scores for the high-frequency and low-frequency users at baseline and after the intervention. Confidence intervals were difference- and correlation-adjusted using the superb R library described by Cousineau et al [46].
Qualitative Feedback

The team also collected informal, qualitative data through individual in-depth interviews and an open-ended survey question after the intervention (T2). Volunteers (n=16) from the chatbot group answered questions about the perceived advantages and pitfalls of Fido, as well as other unintended effects that had not been assessed by the previously administered questionnaires. The most common response about the negative aspects of using Fido was that there were many failures in recognizing the user’s intent, which resulted in the chatbot telling the user that it did not know what they are talking about. Regarding the positive side effects, the most prominent was that using Fido encouraged the user to make a decision about starting traditional therapy.

Discussion

Principal Findings

This study is one of the few that presents an RCT of chatbot therapy efficacy in comparison with an active control, to the best knowledge of authors—the first one for the Polish language. Fido was effective in terms of reducing depression, anxiety, and worry symptoms, as well as for increasing satisfaction with life during the 2-week intervention. We also detected an increase in positive affect, along with a decrease in negative affect, after the intervention using Fido. All effects were stable for at least one month following the interventions. In contrast to some previous findings [6] about agent-guided therapy, we did not detect expected differences in the effectiveness between the intervention based on Fido and the one in which the book chapters were used.

Although there were no between-group effects in terms of symptoms or affect, or for obtained psychoeducational knowledge or the frequency of use, we did detect some other differences. Worth consideration is that the chatbot users spent less time on the intervention than the control group. Thus, although the effects of working with Fido were comparable to those achieved using self-help materials, they were achieved in a shorter period of time, which is an obvious benefit considering the concentration and motivation problems of patients in mental health crises. The differences in time spent on the 2 interventions may be related to differences in the form of the presented content: Even though both techniques are text-based, reading an ebook may be challenging, in contrast to absorbing shorter messages from the chatbot.

Contrary to our expectations, we did not detect a superior decrease in loneliness levels in the chatbot group when compared with the control group. However, we observed a drop in loneliness among participants who used the chatbot with higher frequency, an effect that did not occur either among participants who used Fido with low frequency or in the control group. Interestingly, although loneliness is linked to well-being, we did not detect a decrease in loneliness in the control group, but we did observe a therapeutic effect on all other primary outcome variables. In the face of this result, it is difficult to relate a decline in loneliness observed in high-frequency chatbot users to the decrease in depression or anxiety. A dissociation between treatment effects on loneliness versus anxiety and depression was observed in a study investigating the efficacy of internet-delivered CBT [47]. Some authors consider loneliness as a transdiagnostic phenomenon that can vary in intensity across different diagnoses [48]. It is possible that loneliness was differentially represented as a factor related to depressiveness and anxiety in our study sample, and we, therefore, did not observe a stable relationship between the effects observed for loneliness and those for the anxiety and depression levels. It is also possible that the duration of our study was simply too short to capture this relationship, as changes in levels of loneliness and depression seem to occur with different temporal dynamics [49]. Future studies should try to track the effects of therapy chatbots for longer time periods and replicate our findings regarding loneliness using nontherapeutic agents, as the interaction with such chatbots may also have a direct effect on loneliness with or without any effect on the mental health condition.

Younger participants formed a stronger relationship with Fido, bonding at an average score of 3.59 (on a 1-to-5 positively scored scale), which is higher than the bond reported for internet interventions but lower than Woebot or human-involved therapy (Woebot formed a bond at 3.8, traditional CBT formed a bond at 4, and group CBT formed a bond at 3.8 on the same scale [8]). Participants used the chatbot as frequently as the book in the 2-week intervention period, but the book readers declared spending more time using their psychoeducation materials than chatbot users.

Comparison With Prior Work and Limitations

We did not replicate the findings from the study on Woebot that demonstrated better treatment effects for the chatbot-based intervention than those for the self-help materials [6]. In addition, the alliance built with Woebot was higher than with Fido. This may suggest that, at this stage, Fido is less effective and user-friendly than Woebot.

However, we would argue that the lack of group differences in this study do not result from Fido’s insufficiency but is rather related to the fairly high effectiveness of the self-help materials that we used and the differences in our experimental procedure. First, the materials used in the study by Fitzpatrick et al [6] were mostly psychoeducation, and the materials used in the previous study required not only getting acquainted with psychoeducation but also performing a few cognitive exercises. Second, the materials used for the control group in the Woebot study were provided by the WHO, and those included in our study were derived from a CBT handbook that was empirically proven to be efficient [50]. Third, our study differed in the experimental protocol from the Woebot protocol; that is, in our study, engagement checks were sent to participants every 3 days, which may have served as a reminder to use the intervention. Although it provided valuable insight into participants’ behaviors, it may have also increased the treatment effects, making the 2 studies difficult to compare.

Even though we did not replicate the outcomes from the Woebot study [6], our data are in line with the results of other studies on digital therapeutics. Studies of mental health–supporting applications yield high efficacy mostly when compared with waiting list control samples but not with active control groups.
The lack of differences between the 2 groups investigated in this study prompts us to ask whether the interventions that we used would also be more efficacious when compared with a passive control group (ie, waiting control) and if they would match traditional CBT. To gain a deeper understanding of the patient-reported outcomes, further assessment methods could be used, such as clinical interviews [3] or even psychophysiological methods, that were previously used for human-agent interaction studies [45].

Another important factor when we consider the lack of replication of previous effects is the kind of intervention used in our control group, as different effect sizes are reported depending on the kind of intervention. In our study, the well-established handbook materials were used for the control condition, and the procedure ensured that participants used them regularly, which might have increased the effects of the control intervention [3]. It seems important for the entire area of intervention research to systematize the issue of the control condition in the future and to enable thoroughly selected active control conditions to a greater extent.

There are some other limitations. Our experimental procedure does not allow us to completely exclude the influence of the time factor. It may be that all participants improved only as a result of the passing time or regression toward the mean. The use of a waiting control group could enable control over this aspect of our study, providing information about differences in the outcome data without any intervention. However, if we compare the recruitment and baseline data, we can see that waiting for the beginning of treatment did not change the levels of depression and worry, which suggests that the overall affect and well-being were stable over time and changed only after the intervention.

Last, some elements of this study would be different in a routine application setting. The standard use of Fido or other self-help materials may not include frequent commitment checks or psychological assessments and certainly would not involve financial compensation. All of these factors could have influenced intervention uptake, effectiveness, and side effects. The generalizability of the results is also limited by the fact that subsequent free versions of Fido will use improved machine learning models (ie, achieving better precision and recall scores in intent recognition) while offering the same set of conversation functionalities. However, the level of human involvement and support should be very similar in the standard use of Fido.

**Future Research**

Prospective studies should try to replicate the effects presented here. One future direction would be to extend the timing of the intervention, as 2 weeks are relatively short. Thus, it would be beneficial to conduct at least a 4-week intervention and a 3- or 6-month follow-up to assess whether the effects are sustained after a longer period of time. Furthermore, future studies should consider using other control groups such as waiting lists or traditional CBT.

This article focused mainly on the therapeutic effects; however, during this experiment, we used several measures linked to human-chatbot interaction, which may be further explored. Future studies may provide insight not only into the quantitative measures linked to the use of chatbots but also into qualitative measures, as well as deep analysis of metadata, such as the distribution of cognitive biases and their relationship with the mental health condition. Those studies may extend our knowledge linked to digital therapeutics and provide a theoretical background for the development of further therapy applications.

**Data Availability**

The data sets generated during and/or analyzed during this study are available in the Open Science Framework repository [24].

**Acknowledgments**

This work was supported by a Ministry of Science and Higher Education grant (grant number 012/RID/2018/19). The authors would like to thank Norbert Szczepaniak and Iwona Gawrycka for their continuous contribution to the development of Fido, as well as Barbara Konat, who worked with us at the beginning of the project. It is worth noting that Fido was implemented by Emplocity and runs on natural language processing technology provided by this company.

**Conflicts of Interest**

All authors are actively involved in the development of Fido at a pre-release stage. Fido was commercialized in April 2023 in the form of a spin-off company, in which SK, KS, and JM have stock options.
Acceptability E-scale.  
[PDF File (Adobe PDF File), 111 KB - formative_v8i1e47960_app3.pdf]

Multimedia Appendix 4  
Human-Agent Interaction Scale.  
[PDF File (Adobe PDF File), 111 KB - formative_v8i1e47960_app4.pdf]

Multimedia Appendix 5  
Detailed Results.  
[PDF File (Adobe PDF File), 92 KB - formative_v8i1e47960_app5.pdf]

Multimedia Appendix 6  
CONSORT-EHEALTH checklist (V 1.6.1).  
[PDF File (Adobe PDF File), 388 KB - formative_v8i1e47960_app6.pdf]

References


27. Fido. URL: https://fido.ws/ [accessed 2023-04-02]


Abbreviations

AG-CBT: agent-guided cognitive behavioral therapy
CESD-R: Center for Epidemiologic Studies Depression Scale Revised
CONSORT: Consolidated Standards of Reporting Trials
DSM: Diagnostic and Statistical Manual
HAIS: Human-Agent Interaction Scale
LPS: Language Pragmaticality Scale
OSF: Open Science Framework
PANAS: Positive and Negative Affect Scale
PHQ-9: Patient Health Questionnaire-9
PRIME-MD: Primary Care Evaluation of Mental Disorders
PSWQ: Penn State Worry Questionnaire
RCT: randomized controlled trial
R-UCLA: Revised UCLA Loneliness Scale
STAI: State-Trait Anxiety Inventory
SWLS: Satisfaction With Life Scale
T0: recruitment
T1: baseline
T2: immediately after the 2-week intervention
T3: 1-month follow-up
UX: user experience
WAI-SR: Working Alliance Inventory-Short Revised
WHO: World Health Organization
Parents’ User Experience Accessing and Using a Web-Based Map of COVID-19 Recommendations for Health Decision-Making: Qualitative Descriptive Study

Samantha Cyrkot1, MSc; Lisa Hartling1,2, PhD; Shannon D Scott3, PhD; Sarah A Elliott1,2, PhD

1Alberta Research Centre for Health Evidence, Department of Pediatrics, University of Alberta, Edmonton, AB, Canada
2Cochrane Child Health, Department of Pediatrics, University of Alberta, Edmonton, AB, Canada
3Faculty of Nursing, University of Alberta, Edmonton, AB, Canada

Abstract

Background: The eCOVID19 Recommendations Map & Gateway to Contextualization (RecMap) website was developed to identify all COVID-19 guidelines, assess the credibility and trustworthiness of the guidelines, and make recommendations understandable to various stakeholder groups. To date, little has been done to understand and explore parents’ experiences when accessing and using the RecMap website for COVID-19 health decision-making.

Objective: To explore (1) where parents look for COVID-19 health information and why, (2) parents’ user experience when accessing and using the RecMap website to make health decisions, and (3) what knowledge mobilization activities are needed to increase parents’ awareness, use, and engagement with the RecMap website.

Methods: We conducted a qualitative descriptive study using semistructured interviews and a think-aloud activity with parents of children aged 18 years or younger living in Canada. Participants were asked to provide feedback on the RecMap website and to “think aloud” as they navigated the website to find relevant COVID-19 health recommendations. Demographic information was collected using a web-based questionnaire. A hybrid deductive and inductive thematic approach guided analysis and data synthesis.

Results: A total of 21 participants (13/21, 62% mothers) were interviewed and participated in a think-aloud activity. The data were categorized into four sections, representative of key elements that deductively and inductively emerged from the data: (1) parent information seeking behaviors and preferences for COVID-19, (2) RecMap website usability, (3) perceived usefulness of the RecMap website, and (4) knowledge mobilization strategies to increase awareness, use, and engagement of the RecMap website. Parents primarily used the internet to find COVID-19 information and focused on sources that they determined to be credible, trustworthy, simple, and engaging. As the pandemic evolved, participants’ information-seeking behaviors changed, specifically their topics of interest and search frequency. Most parents were not aware of the RecMap website before this study but found satisfaction with its concept and layout and expressed intentions to use and share it with others. Parents experienced some barriers to using the RecMap website and suggested key areas for improvement to facilitate its usability and perceived usefulness. Recommendations included a more user-friendly home page for lay audiences (separate public-facing user interface), improving the search and filter options, quicker navigation, clearer titles, more family-friendly graphics, and improving mobile-friendly access. Several strategies to disseminate the RecMap website were also expressed, including a mix of traditional and nontraditional methods (handouts and social media) in credible and high-traffic locations that parents frequent often.
Conclusions: Overall, parents liked the concept of the RecMap website but had some suggestions to improve its usability (language, navigation, and website interface). These findings can be used to improve the RecMap website for parents and offer insight for the development and dissemination of effective web-based health information tools and resources for the general public.

(JMIR Form Res 2024;8:e53593) doi: 10.2196/53593

KEYWORDS

awareness; COVID-19; credibility; credible; descriptive; guidelines; health evidence; information behavior; information needs; information seeking; information-seeking behaviour; interface; internet; interview; knowledge mobilization; parent; parenting; public health; qualitative; recommendation; recommender; SARS-CoV-2; think-aloud activity; think-aloud; trust; trustworthy; usability; user experience; web design; website

Introduction

During the COVID-19 pandemic, digital information sharing and exchange exploded as disease information emerged and messaging around various restrictions, such as lockdowns and physical distancing, came into effect. While this generated a wealth of web-based COVID-19 information, it presented challenges for the public, including parents and families, when navigating and recognizing trustworthy health recommendations [1,2].

In response, the eCOVID19 Recommendations Map & Gateway to Contextualization (RecMap) website was developed [3]. The overall aim of the RecMap is to identify and collate all COVID-19 guidelines, assess the credibility and trustworthiness of the recommendations, and make the recommendations understandable to various stakeholder groups. The RecMap was developed by Cochrane Canada in collaboration with the World Health Organization’s (WHO) Collaborating Center for Infectious Diseases, Research Methods and Recommendations at McMaster University, the GRADE (Grading of Recommendations Assessment, Development and Evaluation) centers, the Norwegian Institute of Public Health, the Guidelines International Network, the National Institute of Health and Care Excellence (NICE), WHO and the Pan American Health Organization (PAHO), and many other organizations [3,4]. The living catalogue of COVID-19 guidelines is freely accessible, in English and French [5].

On the RecMap website, the recommendations, which are typically written for health care professionals, are currently being adapted into plain language products. These products have been developed as easy-to-understand summaries of health recommendations [6,7]. Our group’s previous findings suggest that parents prefer and better understand plain language versions compared to standard language recommendations [8].

It is critical that the evidence-based information available to parents is effectively communicated, understandable, and engaging for lay audiences. This is necessary to guide uptake and create actionable impact [9-11]. However, previous anecdotal data have suggested that parents are unaware of the RecMap website when seeking COVID-19–related health information or recommendations. There is also little information about the RecMap website’s usability from the perspective of parents. Usability and perceived usefulness are key criteria to consider when developing effective web-based health information tools [12-15].

To address this gap, we aimed to explore (1) where parents look for COVID-19 health information and why, (2) parents’ user experiences (including barriers and facilitators) when accessing and using the RecMap website to make health decisions, and (3) what knowledge mobilization activities are needed to increase parents’ awareness, use, and engagement with the RecMap website. The results of this study can be used to inform the enhancement and knowledge mobilization of a web-based COVID-19 tool for the general public, as well as the future development of web-based resources on other health topics.

Methods

Ethical Considerations

Approval was received from the University of Alberta Health Research Ethics Board and the McMaster Research Ethics Board (Pro00126429, Pro15646), and all participants gave informed consent before any data collection. After study completion, participants were compensated with a CAD $50 (a currency exchange rate of CAD $1 = US $0.6518 is applicable) electronic gift card for their time.

Overview, Sampling, and Recruitment

We conducted a qualitative-descriptive study [16]. Participants were eligible and self-identified for study enrollment if they were a parent, legal guardian, or grandparent of a child aged 18 years or younger, were aged 18 years or older themselves, lived in Canada, could read and speak English, had access to email and the internet through a computer, tablet, or smartphone. The study was advertised nationwide on the web between February and April 2023 through our collaborators (eg, the RecMap Team, Pediatric Parent Consultancy Network, and Pediatric Parent Advisory Group), and networks (eg, Cochrane Canada, Translating Emergency Knowledge for Kids, and Children’s Healthcare Canada) through email, as well as on social media through Instagram, Twitter, and Facebook. Purposive sampling (based upon parenting role and ethnicity) was used to gather an in-depth understanding of Canadian parents’ experiences with the RecMap website [17]. Results are reported following the Consolidated Criteria for Reporting Qualitative Research checklist [18].
Sample Size
Sample size was shaped by data saturation, which was assessed concurrently throughout data collection and analysis to assess data comprehensiveness, variation, richness, and redundancy [19].

Study Components
Semistructured Interview and Think-Aloud Activity
Participants were invited to attend a web-based, one-on-one, semistructured interview using Zoom video conferencing software. The interview guide was field tested and adapted over 2 nonrecorded interviews using in-house research staff and a parent volunteer (Multimedia Appendix 1). Interviews were conducted by SC (a woman), who is a research coordinator with a Master of Science and has previous experience conducting qualitative interviews. Consenting participants had no previous relationship with the coordinator and were informed of the study objectives at the beginning of the interview. There were no other individuals present during each interview besides the participant and the research coordinator. Interviews were audio-recorded, deidentified, and transcribed verbatim using a third-party transcription service (SimplyTranscription). Field notes were made throughout the interviews, including during the think-aloud activity, in which the parent was observed navigating the website and shared their screen so the interviewer could ascertain the steps and clicks they used to complete the different tasks. The first part of the interview focused on understanding where parents look for COVID-19 information and for what purpose. Participants were then asked to visit the RecMap website on their electronic device of choice (eg, computer, tablet, or smartphone). The second part of the interview asked participants to provide feedback about the RecMap website and to “think aloud” as they were asked to navigate the website to find relevant health recommendations for children. The think-aloud activity is based in cognitive and psychological research, where participants talk aloud while performing a task to verbalize their thoughts that come to mind [20-22]. Think-aloud interview methods have formerly been used in combination to explore usability and perceived usefulness [12]. Participants were also asked to share possible knowledge mobilization strategies to increase parents’ awareness, use, and engagement with the RecMap website. No participants withdrew from the study, no repeat interviews were conducted, and transcripts were not returned to the participants for corrections or to provide feedback on the findings.

Demographic Questionnaire
Participants were asked to complete a short web-based demographic questionnaire (eg, parenting role, age, education, ethnicity, and child’s age) after finishing the interviews. Quantitative data were collected and managed using REDCap (Research Electronic Data Capture), hosted at the University of Alberta [23,24].

Data Analysis
Thematic Analysis
Thematic analysis was used to synthesize and identify common behaviors, processes, and preferences described in the semistructured interviews. Data management and analysis were facilitated using NVivo 14 Software (version 14.23.1; QSR International).

Data collection and analysis occurred iteratively, with data collection occurring until data redundancy was achieved. Interviews were coded by SC (the primary coder) and categorized to facilitate the development of themes. A hybrid deductive and inductive approach guided analysis in which data were categorized into 4 components. An established framework, the Technology Acceptance Model, was used to deductively organize the data, and additional codes were inductively gathered as they emerged from the data [12-15,25]. The analysis was guided by the following six main steps: (1) comparing the transcript with the recording and revising to ensure alignment (data cleaning); (2) reading transcripts and data familiarization; (3) generating initial codes (using a code manual and testing the reliability of codes through verification by a second reviewer [SAE]); (4) summarizing data and identifying initial themes guided by the model (deductive) and coding themes that extend beyond the model (inductive); (5) connecting the codes, identifying themes, and generating a “thematic map;” and (6) corroborating and legitimating coded themes. The coding system was refined throughout the iterative data collection and analysis stages, using a secondary coder (SAE) to code approximately 10% of the transcripts and compare them to maintain intrarater reliability [26]. Any discrepancies between the reviewers were discussed and resolved through consensus. The focused codes were further refined through collaboration between the 2 authors into themes and subthemes that identified common factors contributing to parental preference, usability, and perceived usefulness. All codes and transcripts were then re-examined to ensure the consistency and accuracy of the interpretation. Preliminary findings and interpretations were continuously reviewed and discussed among the research team. Other strategies for maintaining rigor were followed, such as detailed study logs and audit trails to ensure transparency [26].

Demographic Analysis
Demographic data were downloaded from REDCap into Microsoft Office Excel (2019; Microsoft Corp), and frequency distributions were used to describe the study sample.

Results

Sampling and Demographics
A total of 21 parents, mostly mothers (13/21, 62%) living in Canada, participated in interviews lasting between 30 and 60 minutes from February to April 2023. Participant demographics are presented in Table 1. Briefly, all parents reported a postsecondary education, with 57% (12/21) having a graduate or postgraduate degree. Parents identified most often as White (11/21, 52%) and South Asian (4/21, 19%). The RecMap website was accessed during the interviews by parents using a desktop or laptop computer (19/21, 90%), a tablet (1/21, 5%), and a smartphone (1/21, 5%).

<table>
<thead>
<tr>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size was shaped by data saturation, which was assessed concurrently throughout data collection and analysis to assess data comprehensiveness, variation, richness, and redundancy [19].</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semistructured Interview and Think-Aloud Activity</td>
</tr>
<tr>
<td>Participants were invited to attend a web-based, one-on-one, semistructured interview using Zoom video conferencing software. The interview guide was field tested and adapted over 2 nonrecorded interviews using in-house research staff and a parent volunteer (Multimedia Appendix 1). Interviews were conducted by SC (a woman), who is a research coordinator with a Master of Science and has previous experience conducting qualitative interviews. Consenting participants had no previous relationship with the coordinator and were informed of the study objectives at the beginning of the interview. There were no other individuals present during each interview besides the participant and the research coordinator. Interviews were audio-recorded, deidentified, and transcribed verbatim using a third-party transcription service (SimplyTranscription). Field notes were made throughout the interviews, including during the think-aloud activity, in which the parent was observed navigating the website and shared their screen so the interviewer could ascertain the steps and clicks they used to complete the different tasks. The first part of the interview focused on understanding where parents look for COVID-19 information and for what purpose. Participants were then asked to visit the RecMap website on their electronic device of choice (eg, computer, tablet, or smartphone). The second part of the interview asked participants to provide feedback about the RecMap website and to “think aloud” as they were asked to navigate the website to find relevant health recommendations for children. The think-aloud activity is based in cognitive and psychological research, where participants talk aloud while performing a task to verbalize their thoughts that come to mind [20-22]. Think-aloud interview methods have formerly been used in combination to explore usability and perceived usefulness [12]. Participants were also asked to share possible knowledge mobilization strategies to increase parents’ awareness, use, and engagement with the RecMap website. No participants withdrew from the study, no repeat interviews were conducted, and transcripts were not returned to the participants for corrections or to provide feedback on the findings.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demographic Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants were asked to complete a short web-based demographic questionnaire (eg, parenting role, age, education, ethnicity, and child’s age) after finishing the interviews. Quantitative data were collected and managed using REDCap (Research Electronic Data Capture), hosted at the University of Alberta [23,24].</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thematic Analysis</td>
</tr>
<tr>
<td>Thematic analysis was used to synthesize and identify common behaviors, processes, and preferences described in the semistructured interviews. Data management and analysis were facilitated using NVivo 14 Software (version 14.23.1; QSR International). Data collection and analysis occurred iteratively, with data collection occurring until data redundancy was achieved. Interviews were coded by SC (the primary coder) and categorized to facilitate the development of themes. A hybrid deductive and inductive approach guided analysis in which data were categorized into 4 components. An established framework, the Technology Acceptance Model, was used to deductively organize the data, and additional codes were inductively gathered as they emerged from the data [12-15,25]. The analysis was guided by the following six main steps: (1) comparing the transcript with the recording and revising to ensure alignment (data cleaning); (2) reading transcripts and data familiarization; (3) generating initial codes (using a code manual and testing the reliability of codes through verification by a second reviewer [SAE]); (4) summarizing data and identifying initial themes guided by the model (deductive) and coding themes that extend beyond the model (inductive); (5) connecting the codes, identifying themes, and generating a “thematic map;” and (6) corroborating and legitimating coded themes. The coding system was refined throughout the iterative data collection and analysis stages, using a secondary coder (SAE) to code approximately 10% of the transcripts and compare them to maintain intrarater reliability [26]. Any discrepancies between the reviewers were discussed and resolved through consensus. The focused codes were further refined through collaboration between the 2 authors into themes and subthemes that identified common factors contributing to parental preference, usability, and perceived usefulness. All codes and transcripts were then re-examined to ensure the consistency and accuracy of the interpretation. Preliminary findings and interpretations were continuously reviewed and discussed among the research team. Other strategies for maintaining rigor were followed, such as detailed study logs and audit trails to ensure transparency [26].</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demographic Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic data were downloaded from REDCap into Microsoft Office Excel (2019; Microsoft Corp), and frequency distributions were used to describe the study sample.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling and Demographics</td>
</tr>
<tr>
<td>A total of 21 parents, mostly mothers (13/21, 62%) living in Canada, participated in interviews lasting between 30 and 60 minutes from February to April 2023. Participant demographics are presented in Table 1. Briefly, all parents reported a postsecondary education, with 57% (12/21) having a graduate or postgraduate degree. Parents identified most often as White (11/21, 52%) and South Asian (4/21, 19%). The RecMap website was accessed during the interviews by parents using a desktop or laptop computer (19/21, 90%), a tablet (1/21, 5%), and a smartphone (1/21, 5%).</td>
</tr>
</tbody>
</table>
Table 1. Participant demographics (N=21). A currency exchange rate of CAD $1=US $0.6518 is applicable.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parenting role</strong></td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>13 (62)</td>
</tr>
<tr>
<td>Father</td>
<td>8 (38)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td>13 (62)</td>
</tr>
<tr>
<td>41-50</td>
<td>7 (33)</td>
</tr>
<tr>
<td>≥51</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Relationship status</strong></td>
<td></td>
</tr>
<tr>
<td>Partnered (married or common law)</td>
<td>20 (95)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>1 (5)</td>
</tr>
<tr>
<td>South Asian</td>
<td>4 (19)</td>
</tr>
<tr>
<td>East Asian</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Middle Eastern</td>
<td>1 (5)</td>
</tr>
<tr>
<td>White</td>
<td>11 (52)</td>
</tr>
<tr>
<td>Mixed Race</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Income (CAD)</strong></td>
<td></td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>3 (14)</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>1 (5)</td>
</tr>
<tr>
<td>100,000-149,999</td>
<td>8 (38)</td>
</tr>
<tr>
<td>≥150,000</td>
<td>8 (38)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Postsecondary certificate or diploma</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Postsecondary degree</td>
<td>8 (38)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>11 (52)</td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Number of children</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9 (43)</td>
</tr>
<tr>
<td>2</td>
<td>11 (52)</td>
</tr>
<tr>
<td>3</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Children’s age range (years)</strong></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>12 (35)</td>
</tr>
<tr>
<td>6-10</td>
<td>17 (50)</td>
</tr>
<tr>
<td>≥11</td>
<td>5 (15)</td>
</tr>
</tbody>
</table>

No participants identified as being a legal guardian, aged 30 years or younger, with an income of <CAD $50,000 (US $37,100.25), an education below postsecondary, having ≥4 children, or their child being aged 1 year or younger. All participants identified as being partnered, except for 1 participant who selected “prefer not to answer.”

Semistructured Interview and Think-Aloud Activity
The data were categorized into 4 sections, each representative of key elements that deductively and inductively emerged from the data. The first section addressed our first study objective of parent information-seeking behaviors and preferences for
COVID-19. The second and third sections addressed our second objective, where we analyzed user experience based on 2 dimensions, namely, RecMap website usability and perceived usefulness of the RecMap website. The fourth section addressed our final objective of knowledge mobilization strategies to increase awareness, use, and engagement with the RecMap website. Illustrative quotes to support themes and subthemes are presented in Multimedia Appendix 2.

Parent Information-Seeking Behaviors and Preferences for COVID-19
Relevant parent information-seeking behaviors and preferences about COVID-19 were grouped into 3 themes described in subsequent sections.

Parents Seek COVID-19 Information From a Variety of Sources
Parents reported seeking COVID-19 information from a variety of sources. The majority of parents used the internet (eg, web-based search engines, visiting health websites, and social media) as their primary source and reported being very comfortable using this modality to look for information. Secondary sources included the news, radio, and respected friends, family, and health care providers. Parents sought information applicable to their local jurisdiction (ie, province and community) to ensure familiarity with their current guidelines and mandates.

Parents Changed Their Information-Seeking Behaviors During the COVID-19 Pandemic
Parents reported a change in their information-seeking behaviors throughout the pandemic, particularly related to their topics of interest and information-seeking frequency. At the start of the pandemic, many parents reported seeking COVID-19 information on a daily or weekly basis, and their topics of interest included COVID-19 signs and symptoms, isolation requirements, and case numbers. At the time of conducting these interviews, parents reported seeking information less frequently (eg, on a monthly basis or less) and revealed looking for information on topics such as vaccines, masking, or travel guidelines when relevant to their situation and when needing to make an informed health decision for their family.

Parents Look for Specific Features and Elements When Seeking COVID-19 Information
The majority of parents described looking for COVID-19 information that was credible and trustworthy, which they characterized as being from a reputable source (eg, health authorities and the government). Parents also looked for information that was convenient to access (eg, mobile-friendly through a smartphone), relevant to their current environment (eg, email directly from their workplace or their child’s school), and aesthetically pleasing (eg, simple and engaging).

RecMap Website Usability
The data from the think-aloud activity were grouped into the following three themes related to RecMap website usability: (1) RecMap website purpose and target audience, which explored parents’ overall thoughts about the usefulness of the website; (2) RecMap website presentation and navigation, which explored their preferences for layout, formatting, language, and how parents used the website; and (3) RecMap website functionality, which highlighted parents’ experiences with website features and overall usability.

RecMap Website Purpose and Target Audience
The purpose of the RecMap website was well understood by most parents; however, the purpose of the RecMap website was not described anywhere on its home page. Parents wanted clearer messaging on the home page about what the website could offer families and the types of health recommendations available for children.

The target audience of the RecMap website was described as “for everyone” (ie, academics, health care providers, and the public), but again, parents felt that the home page lacked content directly addressing the public (or lay audiences), especially families. Some parents also questioned if the RecMap website was intended for a Canadian or global audience, as some content seemed specific to Canada (eg, McMaster University and Cochrane Canada logos on embedded videos and the website being only available in English and French), while other content felt geared toward a broader population (eg, recommendations presented based on European sources).

RecMap Website Presentation and Navigation
The overall layout and format of the RecMap website were appealing to parents and were described as professional and simple. However, some parents did not readily see who the website was developed by, particularly within the first few moments of navigating the home page, and this led them to initially question its credibility. The color scheme of the website was said to be aesthetically pleasing, but some parents felt that an extra pop of color with more family-friendly graphics could better engage a lay audience. Parents also noted the navigation buttons were not always intuitively located, and they found some headings to be repetitive, with the font size not consistently proportionate between headings, subheadings, and body text.

The recommendations in the format of a list compared to a map on the RecMap website were preferred by most parents, as they felt that it was more familiar to them and easier to understand and navigate (Figures S1 and S2 in Multimedia Appendix 3). However, many parents commented that the list was lengthy and wondered how it was organized (eg, alphabetically or by date), as that impacted their ability to find relevant information quickly.

Parents felt that some terminology used on the website (eg, adolescence and conditional) was too complex and that many lay audiences would struggle to understand the terms “RecMap” and “plain language,” which were frequently used. Parents noted that although the navigation button titled “plain language recommendations” was well-situated on the home page, they cautioned that parents who are unfamiliar with this term may not know how to navigate and seek out these recommendations, which are tailored to lay audiences. Parents cautioned that it may lead to others exploring standard language recommendations first when they could benefit from plain language. Parents suggested refraining from titling these
recommendations as “plain language” and generating an alternative title that is clearer and more suitable for lay audiences.

Parents thought the website was easy to navigate, especially the home page, but found that it required several clicks to reach a specific health recommendation. Parents also found that navigating the recommendations in map view was challenging as many were unfamiliar with this format and required a few moments to understand what the headings and numbers signified (Figure S2 in Multimedia Appendix 3).

Parents used a variety of navigation strategies when asked to find a recommendation specific to children. Most parents’ initial instinct was to use the search function. Others clicked on a preset option that was located on the home page (ie, either the map, list, or plain language recommendations button). Parents described their strategy rationale as being based on a combination of comfort levels (eg, using the search function), convenience (eg, quick), and taking into consideration the existing features and information available on the website. Parents who clicked on the “plain language recommendations” option from the start often felt less overwhelmed with information compared to those who initially browsed the standard language recommendations. Additionally, parents noticed that when using the search function on the home page, the results generated standard language recommendations, but parents felt that results for plain language recommendations should be prioritized for lay audiences.

RecMap Website Functionality

Parents found a variety of features to be useful and improve their user experience and overall website usability. The RecMap website’s search feature was emphasized by parents as appealing because it was well-located and gave the freedom to search for any topic of interest. However, challenges were experienced when the use of basic search terms led to unsuccessful findings (eg, irrelevant or no recommended results), and parents had to adjust their terms multiple times or try a new navigation strategy to obtain relevant results.

The filter option on the list of recommendations was frequently used by parents. This feature was reported as practical, but with some complex and irrelevant categories for a lay audience (eg, adolescent and grading approach). Based on the available categories, parents thought that age group (eg, infant, child, or youth), world region, source, and year of publication were most relevant and suggested adding age range. They also felt that these categories should be rearranged and listed at the top for easy access.

Parents were divided on whether they liked the feature of a new tab opening up (within their web browser) for each specific recommendation that they clicked on and explored. Some parents enjoyed this feature because they could easily navigate between tabs to different recommendations, but some did not notice the new tab, which made it challenging for them to return to the home page.

Parents noted a few undesirable features and some technical barriers as they explored the website. Many parents found that the mailing list pop-up (eg, subscribe to receive information about upcoming activities) was distracting and appeared too soon before they had a chance to explore the full website. Some parents also experienced extended loading times (eg, when trying to access the recommendations in map view), which impacted the overall usability and user experience of the website.

Perceived Usefulness of the RecMap Website

Regarding perceived usefulness, two themes emerged as follows: (1) intentions to use the RecMap website, which looked at parents’ thoughts about its usefulness and their desire to share and discuss it with others and seek supportive information; and (2) parents’ awareness and expectations of the RecMap website, which explored whether parents were familiar with and had previously accessed the website and if it met their needs and expectations.

Intentions to use the RecMap Website

After completing the think-aloud activity, most parents discussed their intentions to use and share the website with others (eg, friends and family). Parents who were hesitant to use or share it suggested that the home page needed to be more user-friendly (for a lay audience) with clearer titles, quicker navigation, and improved search and filter options to seek out relevant plain language recommendations in a timely manner.

Parents felt that if needed, they would discuss the RecMap website with their health care provider or seek out supplementary information (eg, other web-based sources) to enhance their understanding of what was available on the website in order to make an informed health decision for their child.

Awareness and Expectations of the RecMap Website

Very few parents reported being aware of or previously accessing the RecMap website before this study. Parents suggested that their unawareness may be due to limited promotion, a lack of health care referral, or that the RecMap website did not appear in their search engine results for COVID-19.

The overall concept of the RecMap website was liked by most parents, and they perceived this tool to be useful. Parents highlighted that it gives users an opportunity to search for COVID-19 information from a credible source, and some positively compared it to a large COVID-19 web-based search engine.

Some parents expressed that parts of the RecMap website were not what they expected and wanted to see more information tailored for parents and lay audiences. This included addressing COVID-19 topics related to possible child COVID-19 complications and symptom management, which parents felt were not clearly addressed. Additionally, most parents did not expect that when navigating the website, they would have the option to see recommendations in both standard and plain language. This had taken many parents a few moments to realize, particularly those who were unfamiliar with the term “plain language.” They also wanted more clarification about what the recommendation map was compared to the
recommendation list and if there was a difference between them beyond format.

**Knowledge Mobilization Strategies**

Knowledge mobilization strategies were grouped into two themes: (1) recommendations to tailor the RecMap website for parents, which highlighted suggestions to make the website more user-friendly and accessible for parents; and (2) approaches to dissemination, which explored where and how to share information about the RecMap website within the parent community to increase awareness, use, and engagement.

**Recommendations to Tailor the RecMap Website for Parents**

To better tailor the website for parents, it was often suggested to develop a separate user interface within the website to split the content into lay or public and health care professional or academic. This would facilitate relocating the standard language recommendations, which most parents find challenging to understand. This would also enable more tailored features, where some parents suggested a “frequently asked questions” section to dispel myths or an ability to insert anonymous data about your child (eg, age and vaccine status) into the RecMap website to generate relevant recommendations. A “bookmark” or “share” option to save recommendations was also suggested by parents, as they described having busy schedules and wanting quick and easy-to-find information.

Parents felt that the accessibility of the website was another key element to highlight. A mobile-friendly website and mobile apps were mentioned by many parents, who reported that they often use their smartphones as a convenient tool to look for information. Even though the RecMap is currently accessible through a smartphone, the map of recommendations is inaccessible. Parents also felt that being mindful of contrasting colors on the RecMap website was important for the accessibility of those with various visual abilities.

**Approaches to Dissemination**

Parents shared a variety of dissemination strategies to increase awareness of the RecMap website within the parent community. Parents felt that sharing information in convenient and high-traffic locations where parents frequent often (eg, social media, such as Facebook, Instagram, TikTok, and Twitter; health and immunization clinics; schools; parent groups; at extracurricular activities or events such as soccer games) would increase parents’ awareness and thus website use. Other strategies included collaborating with credible and trusted organizations and individuals (eg, health authorities, health care providers, and schools) to share information through their networks in person or through a web-based presence.

Parents suggested that approaches to increasing awareness of the RecMap website should not be limited to just one but rather should consist of multiple captivating methods to attract and engage parents with different preferences and means to access information. Methods suggested included QR codes, handouts, posters, social media posts, advertisements and sponsors (eg, on the radio, podcasts, in parenting magazines, school newsletters), media interviews (eg, on the news and radio), and word of mouth from trusted individuals (eg, health care providers).

**Discussion**

**Principal Findings**

The aim of this study was to explore parental COVID-19 information-seeking behaviors and preferences, along with their experience when accessing and using the RecMap website. Parents also shared several knowledge mobilization strategies to facilitate parents’ awareness, use, and engagement of the RecMap website.

**Information-Seeking Behaviors and Preferences**

This study identified that most parents report using the internet as a primary source to look for COVID-19 health information. This is not surprising given that recent reports document that approximately 70% of Canadians and Americans use the web to search for health information [27,28]. There is also now an array of COVID-19-specific information on the web that has become easily and conveniently accessible to the public [2]. Of note, parents reported a change in their information-seeking behaviors and preferences since the pandemic started, with parents looking for information less frequently as the pandemic evolved. This is important to consider when planning future knowledge mobilization strategies aimed at parents and the general public.

**Barriers to Using the RecMap Website**

When parents initially accessed the RecMap website, most struggled to identify who developed it; thus, many parents questioned its credibility. Web-based tools that are immediately recognizable (eg, well-branded and clear messaging on the home page) may help increase credibility and should be considered for future knowledge mobilization efforts among the general public.

Additionally, there was some content and language within the RecMap website that was deemed too advanced or complex by participants and thus a barrier, such as recommendations that were written in standard language. However, some participants expressed benefits to having access to standard language and suggested a public-facing user interface to separate laypersons and the public from health care professionals or academic content. Similar interface designs have been used on other health-related information websites across Canada, including Alberta Health Service and Ontario Health [30,31]. This format (lay-friendly material upfront with the ability to access standard language content) will also provide more opportunities to customize the RecMap website for this population and address additional suggestions mentioned by participants (eg, messaging that is focused on what the RecMap website can offer the general public).

Parents also experienced technical barriers with the loading time of the recommendations in the map view. A study on web usability testing in parents found that technological difficulties were a key barrier to successfully using a website and its features [32]. Load time is also an important web performance metric that can directly impact user engagement. Websites that take
longer than 3 seconds to load have reported that almost 55% of users leave the website [33]. In fact, some of our participants mentioned that they would potentially exit a website and find a new one if the loading time was excessively long. Other minor barriers noted in this study were shortcomings in the visual design (eg, color and graphics), unclear titles, and challenges with returning to the home page. Similar barriers have been reported in previous research that explored website user experience and usability testing [12,34], highlighting the need for end user input into the design at an early stage.

**Facilitators for Using the RecMap Website**

While parents liked the concept and felt the RecMap website could be useful to them, they described several facilitators to optimize use. The concept of the RecMap has previously been used and implemented by the World Health Organization to house living guidelines and recommendations on tuberculosis prevention and care [35]. At this time, usability testing has not been formally completed with parents or the general public. However, the RecMap’s digitized dissemination format has been compared for usability to a conventional tuberculosis website containing guideline recommendations, and the RecMap was found to be more accessible, improve understanding, and enable decision-making [36].

Other facilitators for parents using the RecMap website included a professional and simple presentation, the availability of an internal search engine, and a home page that was easy to navigate. These suggested elements are in line with standards that have been identified for the development of digital-based health information resources. In particular, when information is aesthetically pleasing, easy to find and navigate, this can contribute to increased user trust, readability, and facilitation of understanding [37]. Other research has found that simple aesthetics [29], clearly described headings and labels, information that is logically organized [38], limited technological barriers [32], and being usable in stressful situations [29] are equally preferred by parents. The use of tailored features to meet the needs and preferences of parents was also suggested to improve website usability and quality [32]. This was recommended by our participants in the form of a frequently asked questions section or the ability to bookmark recommendations for easy access.

**Knowledge Mobilization Strategies**

In 2018, it was reported that almost 90% of Canadian internet users aged 15 years and older have a smartphone and that almost 50% check their smartphone at least every 30 minutes [39]. Many parents in our study reported that they typically use their smartphones to look for and access health-related information, however, the initial version of the RecMap was not specifically designed to be smartphone compatible. It is clearly important that health-based websites are mobile-friendly and compatible on devices with smaller screens to optimize viewing and ease of navigation. Better accessibility may also support user experience, which can lead to better user engagement and increase usage.

Mobile apps are another avenue to explore for the RecMap, as there is a rise in parents using apps to support their parenting [40]. This modality is versatile and can encompass multimedia elements (eg, graphics and videos) to engage users and push notifications to alert users of new information or updates on the application. This is an interesting feature to consider given that the RecMap contains living recommendations that are frequently updated or changed. It is equally important for parents that mobile apps are of high quality, relevant, visually appealing, interactive, and seamless (ie, easy to navigate, do not freeze, or crash); otherwise, there is a risk that flawed apps will be deleted by end users within minutes after being downloaded [40].

It has been previously reported that dissemination strategies appropriate for parents include targeting credible sources such as community centers, public health institutions, schools, physicians, and government agencies [41]. This is comparable to our findings, where parents suggested targeting schools, health clinics, and collaborating with credible and trusted organizations and individuals to increase the RecMap website’s awareness. Furthermore, collaborating with reputable messengers (eg, physicians) can be an effective strategy. In fact, it is suggested that 70% of patients would like their physician to recommend web-based health information for them to review [42]. Similar to our findings, it has also been suggested that using a combination of dissemination methods can be most effective. This includes a multiplatform approach of digital (eg, social media and websites), traditional (eg, brochures and radio), and unique means (eg, community “parent nights”) [41].

**Strengths and Limitations**

This work has important implications for knowledge mobilization strategies to enhance and contextualize products to meet diverse end users’ needs as we move beyond the pandemic and for other public health issues, including regular seasonal outbreaks and future epidemic approaches.

A strength of this study is the use of multiple methods. The think-aloud activity allowed us to collect direct data on user experience and parent thoughts while using the RecMap website, and the interviews provided us with an opportunity to clarify initial thoughts and ask further questions. Nevertheless, the majority of participants self-identified as White and of higher socioeconomic status, which limited the generalizability of study findings to a wider audience and was not representative of Canada’s ethnically diverse population. However, we were able to recruit participants who identified as both mothers (13/21, 62%) and fathers (8/21, 38%), which enabled us to gather differing perspectives in terms of parenting roles. Future studies may benefit from developing a recruitment strategy that targets those with different parenting roles, visible minority groups, those of lower socioeconomic status, and those living in geographically diverse locations (eg, rural). In doing so, a better understanding of the diverse needs of parents across Canada could be examined, and strategies (digital and nondigital) to support knowledge mobilization around COVID-19 in various contexts could be developed. It is also worth noting that all participants lived in Canada and had access to a similar health care setting.

https://formative.jmir.org/2024/1/e53593

Cyrkot et al
Directions for Future Research

Future work related to the RecMap website should focus on using an integrated knowledge translation model where end users (e.g., parents) codevelop and implement a plan to increase the awareness, use, and engagement of the website among the general public and lay audiences. Our group will also work to improve the usability of the RecMap website and implement the changes based on our findings related to parents’ user experiences. This may also include further work on mobile-friendly access (e.g., smartphones and mobile apps). These steps will enhance the existing RecMap website and ensure that it is user-friendly and meets the needs of parents and the general public. Future projects similar to RecMap that seek to share health recommendations should involve various end users (e.g., parents, health care providers, and the general public) during the conceptual phase of web-based tool development to better tailor the final product toward their specific goals and needs.

As public health approaches to COVID-19 evolve and the focus shifts to post–COVID-19 conditions, researchers developing knowledge mobilization strategies can use these findings to inform decisions for implementing knowledge. While the RecMap was not developed specifically to meet parents’ information needs during the pandemic, it provides a model for how rapidly changing evidence can be synthesized and presented to support decision-making. However, rapidly changing contextual factors may affect the potential use, relevance, and positioning of the RecMap to support knowledge needs around COVID-19 in the future. The waning of attention to COVID-19 guidelines internationally and jurisdictional changes to COVID-19 management question the advantages of resources such as the RecMap post pandemic. Future research should explore sustainable models of knowledge mobilization to support parental information-seeking behaviors. The pandemic highlighted how families engage with web-based health information and how changes in knowledge needs and information-seeking behaviors need to be considered.

Conclusions

This study reports on the COVID-19 information-seeking behaviors and preferences of parents, the RecMap website user experience (including barriers and facilitators), and potential knowledge mobilization strategies. The concept of the RecMap website was appealing to parents and has the potential to become a user-friendly web-based tool for parents. However, understanding how, when, and why parents search for COVID-19 information, as well as what aspects of the user experience can be improved, has provided meaningful insight and recommendations to improve the usability and perceived usefulness of the RecMap website. Additionally, appropriately tailored knowledge mobilization strategies targeted at parents will increase their awareness, use, and engagement of the RecMap website. This data will contribute to the enhancement of the COVID-19 RecMap website for parents and can inform the development of effective web-based tools for the general public about other health topics.

Acknowledgments

The authors would like to thank members of the Pediatric Parent Advisory Group for supporting this work through the design, research materials, and pilot testing of the survey and interview guide. We also acknowledge the input of the RecMap Knowledge Mobilization Advisory Committee.

This study received funding from the Canadian Institutes of Health Research (CIHR; REC 183153), awarded to Dr Holger Schünnemann (McMaster University). CIHR had no involvement in the study design, collection, data analysis and interpretation, or report writing, and was not part of the decision to submit the paper for publication. LH is supported by a Canada Research Chair in Knowledge Synthesis and Translation. LH and SDS are distinguished researchers with the Stollery Children’s Hospital Foundation and the Women and Children’s Health Research Institute.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide questions and think-aloud activity processes.
[PDF File (Adobe PDF File), 201 KB - formative_v8i1e53593_app1.pdf]

Multimedia Appendix 2
Illustrative quotes to support the themes and subthemes.
[PDF File (Adobe PDF File), 183 KB - formative_v8i1e53593_app2.pdf]

Multimedia Appendix 3
The a) list and b) map view of the COVID-19 recommendations available on the RecMap website.
[PDF File (Adobe PDF File), 271 KB - formative_v8i1e53593_app3.pdf]

References


5. eCOVID RecMap. URL: https://covid19.recmap.org/ [accessed 2024-03-05]


27. Table 22-10-0137-01. Selected online activities by gender, age group and highest certificate, diploma or degree completed. Statistics Canada. 2021. URL: https://doi.org/10.25318/2210013701-eng [accessed 2024-02-27]


**Abbreviations**

GRADE: Grading of Recommendations Assessment, Development and Evaluation

NICE: National Institute of Health and Care Excellence

PAHO: Pan American Health Organization

RecMap: Recommendations Map & Gateway to Contextualization

REDCap: Research Electronic Data Capture

WHO: World Health Organization
Assessing Knowledge, Competence, and Performance Following Web-Based Education on Early Breast Cancer Management: Health Care Professional Questionnaire Study and Anonymized Patient Records Analysis

Michael Gnant, MD; Khatijah Lim Abdullah, PhD; Frances Boyle, MD; Chiun-Sheng Huang, MD; Katie Bickford, BSc; Sola Neunie, MSc; Alexander Noble, MSc; Anne Nunn, PhD; Caroline Sproat, PhD; Nadia Harbeck, MD; Carlos Barrios, MD

1Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria
2Department of Nursing, School of Medical and Life Sciences, Sunway University, Selangor, Malaysia
3Patricia Ritchie Centre for Cancer Care and Research, University of Sydney, Sydney, Australia
4Department of Surgery, National Taiwan University Hospital, Taipei City, Taiwan
5touch Independent Medical Education Limited, Stockport, United Kingdom
6Breast Center, Department of Obstetrics & Gynecology and Comprehensive Cancer Center Munich, University Hospital of Munich Ludwig Maximilian, Munich, Germany
7Latin American Cooperative Group and Oncoclínicas Group, Hospital São Lucas da Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

Corresponding Author:
Michael Gnant, MD
Comprehensive Cancer Center
Medical University of Vienna
Spitalgasse 23
BT86/E01
Vienna, 1090
Austria
Phone: 43 1 40160 35602
Email: mgnant@icloud.com

Abstract

Background: Web-based learning activities are key components of continuing medical education (CME) for health care professionals (HCPs). However, the published outcomes of web-based educational interventions for early breast cancer (EBC) are limited.

Objective: This study aims to objectively assess knowledge, competence, and performance among HCPs following participation in 2 EBC-focused CME activities and to identify the remaining educational gaps.

Methods: We developed 2 CME-accredited web-based educational activities addressing high-risk EBC, including integration of shared decision-making to optimize patient care (touchMDT) and stratification for early identification of high-risk patients and novel treatment strategies (touchPANEL DISCUSSION). Knowledge, competence, and performance were assessed before and after the activities against an expanded outcomes framework (levels 1-5) using self-reported questionnaires and an analysis of anonymized data extracted from patient records.

Results: Six months after the launch of the activity, 7047 and 8989 HCP participants engaged with touchMDT and touchPANEL DISCUSSION, respectively. The overall satisfaction was 82% (a total score of 20.6 out of 25) for the touchMDT and 88% (a total score of 21.9 out of 25) for the touchPANEL DISCUSSION. For the evaluation of knowledge and competence (50 respondents before the activity and 50 learners after the activity), there was a significant increase in the mean number of correctly answered questions from pre- to postactivity (touchMDT: median 4.0, IQR 3.0-5.0 to median 5.5, IQR 4.0-7.0; mean 4.00, SD 1.39 to mean 5.30, SD 1.56 and touchPANEL DISCUSSION: median 4.0, IQR 4.0-5.0 to median 6.0, IQR 5.0-7.0; mean 4.32, SD 1.30 to mean 5.88, SD 1.49; both \( P < .001 \)). A significant improvement in self-reported performance (50 respondents before the activity and 50 learners after the activity) was observed in a combined analysis of both activities (median 3.0, IQR 2.0-3.0 to median 4.0,
IQR 3.0-5.0; mean 2.82, SD 1.08 to mean 4.16, SD 1.45; P<.001). Patient record analysis (50 respondents before the activity and 50 learners after the activity) showed that the HCPs used a range of measures to determine EBC recurrence risk and revealed no significant differences in adjuvant therapies used before and after the activity (P=.97 and P>.99 for Ki-67 <20% and Ki-67 ≥20% tumors, respectively). The remaining educational gaps included strategies for implementing shared decision-making in clinical practice and the use of genetic and biomarker testing to guide treatment selection.

Conclusions: Brief, web-based CME activities on EBC were associated with an improvement in HCP knowledge, competence, and self-reported performance and can help identify unmet needs to inform the design of future CME activities.

(JMIR Form Res 2024;8:e50931) doi:10.2196/50931

KEYWORDS
continuing medical education; early breast cancer; performance; risk stratification; shared decision-making

Introduction

Clinical Decision-Making in High-Risk Early Breast Cancer
Clinical decision-making in cancer can be complex, requiring health care professionals (HCPs) to consider multiple clinical factors, patient preferences, and environmental factors [1-3]. A multidisciplinary team approach can improve the overall quality of decision-making [4-7], and there is evidence that shared decision-making (SDM) with patients can lead to better adherence to treatment and outcomes [1,8-9]. However, obstacles such as lack of resource availability and patient factors (eg, cultures and health literacy) have prevented the complete integration of SDM into cancer care [1,10-14]. HCPs need to understand these obstacles to deliver high-level care tailored to the individual.

The complexity of treating high-risk early breast cancer (EBC) has increased further with the emergence of complex new data on risk stratification and treatment. These include data on the prognostic and predictive value of biomarkers and the benefits of cyclin-dependent kinase 4 and 6 inhibitors, second-generation selective estrogen receptor antagonists and degraders, and protein kinase B (Akt) inhibitors [15-20]. Furthermore, the guidelines on the management of EBC have been updated recently to reflect recent data, which can be interpreted as influencing clinical practice [5,21-28]. Because of this rapid influx of complex information, HCPs who treat EBC are challenged to keep up with practice changes; consequently, it can take time for new and emerging data to be effectively integrated into practice, and patients with EBC may not receive optimal care [29-32].

Effectiveness of Web-Based HCP Education in Breast Cancer
Effective continuing medical education (CME) programs are crucial to ensure that HCPs implement up-to-date, evidence-based practices for their patients. Although CME for HCPs has traditionally been delivered using face-to-face formats, web-based educational activities now offer a viable alternative approach [33]. To date, several studies have demonstrated the effectiveness of web-based HCP education for breast cancer [34-36]. However, gaps remain in the provision of independent web-based medical education programs for high-risk EBC treatment and effective SDM. Although a small number of programs have been developed in EBC [37-39], these have typically been nonaccredited, therefore not providing HCPs with continuing professional development credits for attendance; pharmaceutical company–led, therefore with the potential to be perceived as promotional; and restricted or limited in terms of access; therefore only accessible to those with the means to do so. To address these gaps and complement the currently available programs, accredited, independent, expert-led education available on free-to-access platforms is required to support the knowledge, competence, and performance of HCPs.

Study Rationale, Objectives, and Aims
An expanded 7-level framework for assessing the outcomes of CME programs was developed by Moore et al [40] in 2009. This framework evaluates participation (level 1), satisfaction (level 2), knowledge (level 3), competence (level 4), performance (level 5), patient health status (level 6), and community health status (level 7). These levels are now widely used and included in many consensus documents and practice recommendations [41]. To date, published Moore’s level 5 [40] outcomes analyses in medical education have largely used subjective self-reported outcomes [42-44]. Although these can provide valuable insights, they can be open to bias, for example, recall bias or social desirability bias [45,46]. An analysis of anonymized patient records provides a complementary objective measure of performance that may enhance the understanding of physician behavior changes in practice.

In this study, we developed and implemented 2 faculty-led, CME-accredited web-based educational activities in high-risk EBC. The objectives of this study were to (1) measure changes in knowledge, competence, and self-reported performance after participation in the activities; (2) objectively evaluate changes in performance using anonymized patient records; and (3) identify the remaining educational gaps.

Methods

Study Design
This was a web-based study that analyzed the impact of 2 educational activities in high-risk EBC through CME questionnaires and deidentified patient records. The target audience comprised oncologists (including breast cancer surgeons), oncology nurse specialists, pathologists, and radiologists from Asia, Brazil, and Europe (excluding the United Kingdom). Satisfaction, knowledge, competence, and performance were assessed according to levels 2 to 5 of the
expanded framework for assessing the outcomes of CME programs by Moore et al [40] using self-reported questionnaires and anonymized data extracted from patient records. Questionnaire distribution and data collection were carried out by an independent third party (nuaxia Limited, United Kingdom).

Ethical Considerations

All consents and ethical confirmations for this study were obtained before the completion of both outcome questionnaires. The personal data of HCPs and patients were anonymized. General Data Protection Regulation [47] consents were obtained in the European Union and North America for each questionnaire; adherence to additional requirements in local markets was required to be confirmed in line with European Pharmaceutical Market Research Association professional guidance [48]. For respondents to the patient records questionnaire, consents and ethical confirmations were obtained as part of the questionnaire in accordance with the British Healthcare Business Intelligence Association professional guidelines [49] to ensure patient confidentiality. As part of their written consent to participate in the questionnaires, the HCPs were required to agree to the statement, “As with all research, your identity and personal data are strictly confidential and will not be revealed without your explicit further consent. You have the right to withdraw your consent to participate in market research at any time.” Complete patient records were not obtained by touch Independent Medical Education (touchIME) or any third party; the HCPs were instructed to extract the required anonymized information into a questionnaire. Ethics approval was not applicable for this study because this analysis was not classified as “health research” according to the UK Research and Innovation definition that guides whether research requires Research Ethics Committee review and Health Research Authority approval [50,51]. The study was defined as “market research” per rule 1.1 of the European Pharmaceutical Market Research Association code of conduct, which states that market research does not require a Clinical Research Ethics Committee or independent review board approval [48].

Educational Activities

Educational gaps and learning objectives were identified before the start of activity development in October 2021 by touchIME (a provider of IME for the global HCP community) through a review of the published literature. The expert faculty members further inputted into the educational gaps and learning objectives at the start of activity development. Expert faculty members with a background in breast cancer and SDM were identified by touchIME medical directors through searches of literature indexed in the PubMed database, congress websites, and web-based educational videos. Patient faculty were identified using social media searches across multiple channels. The faculty members (CB, FB, MG, NH, CSH and KLA) were blinded to the plans to assess participation and satisfaction with the education and to publish the outcomes. All the expert faculty (CB, FB, MG, NH, CSH and KLA) involved in the educational activities are authors of this manuscript.

A total of 2 web-based educational activities, touchMDT: “how can shared decision-making be successfully integrated to optimize care of patients with high-risk early breast cancer?” and touchPANEL DISCUSSION: “new horizons in high-risk HR+ HER2- EBC: Risk stratification for early identification and novel treatment strategies” were developed by touchIME in collaboration with the faculty members (CB, FB, MG, NH, CSH and KLA; complete details and learning objectives are included in Multimedia Appendix 1). The touchMDT activity comprised three 13- to 15-minute videos, providing 42 minutes of education in total, and the touchPANEL DISCUSSION activity comprised three 11- to 15-minute videos, providing 39 minutes of education in total. The activities were CME accredited by the Accreditation Council for Continuing Medical Education and the American Nurses Credentialing Center at the University of South Florida Health. Both activities were translated (video subtitles and downloadable slides) from English into Brazilian Portuguese, Chinese Mandarin, French, German, Italian, and Spanish by an independent third party (UnBabel, United States), and both activities were free to access for 1 year (the permissible lifetime of an accredited educational activity) on the touchONCOLOGY website from March 30, 2022, to March 30, 2023 (touchMDT) and from June 15, 2022, to June 15, 2023 (touchPANEL DISCUSSION). The 2 activities featured together on the host website, but participation in both activities was not mandatory.

Various communication channels were used to reach the target audience, including direct publicity emails to the touchONCOLOGY database within the first 12 weeks of the activity launch, with a further reminder at around 6 months; display banners on the touchONCOLOGY website; advertisements in the peer-reviewed journal touchREVIEWS in Oncology and Haematology; publicity via various relevant medical society partnerships; and social media partnerships on Facebook (Meta Platforms, Inc.), LinkedIn (Microsoft Corp.), and Twitter (rebranded as X; X Corp) targeted at the HCPs throughout the lifetime of the activity.

Assessment of Educational Outcomes

Levels 1 and 2 (Participation and Satisfaction)

Levels 1 (participation) and 2 (satisfaction) were assessed separately for each activity (Multimedia Appendix 2). Level 1 included the number of HCPs who engaged in the web-based educational activities and the average time they spent viewing the videos, and it was the only assessment not evaluated using a questionnaire.

Level 2 (satisfaction) was assessed by scoring 5 statements of satisfaction (using a Likert scale) in a postactivity questionnaire.

Levels 3 to 4 and Level 5 (Knowledge, Competence, and Performance)

Levels 3 to 4 and level 5 were assessed using questionnaires completed by relevant respondents (HCPs completing the questionnaire before the activity) and learners (HCPs completing the questionnaire following the education). No pretesting was performed. The target audience was predefined by specialty (eg, oncologists [including breast cancer surgeons], oncology nurse specialists, pathologists, and radiologists) and by region (Brazil, France, Germany, Italy, and Spain). To avoid any pre-exposure bias and to obtain a statistically representative
sample size, data were collected using an independent samples model for each activity. For levels 3 to 4, separate questionnaires were developed for each activity (each comprising 7 questions), whereas for level 5, a combined questionnaire was developed for both activities (3 questions each for touchMDT and touchPANEL DISCUSSION). All questions were developed by touchIME medical directors and writers and were approved for medical accuracy by the faculty. An overview of the topics and the complete questionnaires are provided in Multimedia Appendices 3 to 6.

For the touchPANEL DISCUSSION activity, changes in performance were also assessed through an evaluation of redacted patient records. The HCPs were sent a web-based questionnaire that captured data on (1) general patient demographics and history, (2) risk assessment and stratification to predict disease recurrence, and (3) treatment. Each HCP was requested to extract all relevant nonidentifiable information from a single anonymized patient record into the questionnaire.

Self-Reported Confidence and Intention to Change Practice

As part of the questionnaires, respondents and learners were asked, “How confident are you in treating patients with breast cancer?” (mutually exclusive responses: “not confident,” “a little confident,” “somewhat confident,” “moderately confident,” and “extremely confident”) and “As a result of your participation in this session, will you make a change in your practice?” (mutually exclusive responses: “yes,” “uncertain—more education needed,” “uncertain—practical limitations,” “no—more education needed,” and “no—practical limitations”).

Identification of Outstanding Educational Gaps

A total of 5 potential educational gaps were included in the levels 3 to 4 and level 5 questionnaires (Multimedia Appendix 7), and the learners were asked to rank them by importance. The results were analyzed using a single transferable vote system (Multimedia Appendix 8 [52]). In addition, questions that were answered incorrectly by ≥30% (15/50) of the learners in all postactivity questionnaires were identified as educational gaps. The 30% cutoff was determined by touchIME as an indicator of the requirement for further education based on analyses of data from previous outcomes.

Fielding of the Questionnaires

Preactivity levels 3 to 4 questionnaires were fielded 1 to 2 weeks before launch, whereas postactivity levels 2 to 4 questionnaires were fielded immediately after launch to a different set of HCPs who had participated in the education. The questionnaires were distributed to a database of 20,420 HCPs and then “closed” once a prespecified number had responded (n=50).

The level 5 questionnaire and first patient records questionnaire were fielded 1 to 2 weeks before launch—to a different set of HCPs than those who answered the levels 3 to 4 questionnaires—and closed after completion by a prespecified number of respondents (n=50). A total of 26 weeks after launch, the postactivity level 5 questionnaire and second patient records questionnaire were distributed to the same set of HCPs who completed the level 5 questionnaire and patient records questionnaire before launch. Because of drop out over time, additional HCPs who were matched to the dropouts were surveyed postactivity, if needed.

The HCP participants’ specialties were all validated by nuaxia Limited against third-party data with the assistance of artificial intelligence. This comprised any combination of data from professional associations, hospitals, professional publications, and prescribing and insurance databases.

The participants were required to answer “yes” to the screening question “Do you treat patients with early breast cancer?” to be eligible to participate in the questionnaire study. The HCPs eligible to participate received an honorarium in the form of a digital reward, which could be redeemed via a number of reward partners. In accordance with industry regulations and professional body guidance, the honorarium received was limited to the fair market value for an HCP of that seniority.

Statistical Analyses

Data were analyzed using SPSS Statistics software (version 28.0.1; IBM Corp). The sample size (50 respondents and 50 learners for each activity) was predetermined by the level of funding, and by applying a power analysis, we determined a resulting margin of error of 10% for both the touchMDT and touchPANEL DISCUSSION. Knowledge, competence, and self-reported outcomes were compared for the overall population using an independent samples 2-tailed t test. Data were subdivided by region and years of experience, and a 2-way ANOVA was used to assess the variation in results by these demographics. Individual questions were analyzed with a paired samples 2-tailed t test and a 1-way ANOVA, followed by a cluster analysis, to provide insights into the overall change in correct answers across the questions asked and the specific changes in all the answers given. Data on the number of questions answered correctly were measured as a continuous variable. At the individual question level, the answers were categorical. For performance, measured by the redacted patient records questionnaire, pre- and postactivity comparisons were made using paired and independent samples 2-tailed t tests, chi-square tests, and cluster analysis. These tests were used to assess patient history, use of risk assessment and stratification strategies, and patient treatment pathways.

Results

Assessment of Educational Outcomes

Levels 1 and 2 (Participation and Satisfaction)

By 6 months after launch, 7047 and 8989 participants had engaged with the touchMDT and touchPANEL DISCUSSION activities, respectively. The participants were predominantly oncologists, and the majority were based in Italy for the touchMDT and Brazil for the touchPANEL DISCUSSION activities (Multimedia Appendix 9). Overall satisfaction was 82% (a total score of 20.6 out of 25) for touchMDT and 88% (a total score of 21.9 out of 25) for touchPANEL DISCUSSION (Multimedia Appendix 10). Multichannel publicity reach and impact for the touchMDT and touchPANEL DISCUSSION activities are reported in Multimedia Appendix 11.
Levels 3 and 4 (Knowledge and Competence) A total of 50 respondents and 50 learners were questioned to assess levels 3 to 4 outcomes (Table 1).

Table 1. Demographics of participating health care professionals (HCPs) for the touchMDT and touchPANEL DISCUSSION activities.

<table>
<thead>
<tr>
<th>Demographics of HCPs</th>
<th>Levels 2 to 4 questionnaire</th>
<th>Level 5 questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>touchMDT</td>
<td>touchPANEL DISCUSSION</td>
</tr>
<tr>
<td></td>
<td>Respondents</td>
<td>Learners</td>
</tr>
<tr>
<td>Participants, N</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Specialty, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oncologist</td>
<td>40 (80)</td>
<td>42 (84)</td>
</tr>
<tr>
<td>Radiologist</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Pathologist</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Radiologist and pathologist</td>
<td>10 (20)</td>
<td>8 (16)</td>
</tr>
<tr>
<td>Years of practice, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1 to 5</td>
<td>6 (12)</td>
<td>5 (10)</td>
</tr>
<tr>
<td>&gt;5 to 20</td>
<td>29 (58)</td>
<td>25 (50)</td>
</tr>
<tr>
<td>&lt;10</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>10 to 20</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>0 to 20</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>&gt;20</td>
<td>15 (30)</td>
<td>20 (40)</td>
</tr>
<tr>
<td>Country, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>14 (28)</td>
<td>11 (22)</td>
</tr>
<tr>
<td>France</td>
<td>5 (10)</td>
<td>9 (18)</td>
</tr>
<tr>
<td>Germany</td>
<td>14 (28)</td>
<td>12 (24)</td>
</tr>
<tr>
<td>Italy</td>
<td>8 (16)</td>
<td>9 (18)</td>
</tr>
<tr>
<td>Spain</td>
<td>9 (18)</td>
<td>9 (18)</td>
</tr>
</tbody>
</table>

aData were collected 6 months after the launch of the touchMDT and touchPANEL DISCUSSION activities on September 29, 2022, and November 21, 2022, respectively. The questionnaire response categories differed between questionnaires and activities, meaning that some options were not applicable. The respondents and learners were the HCPs who completed the pre- and postactivity questionnaires, respectively.

bN/A: not applicable.

For the touchMDT and touchPANEL DISCUSSION activities, respectively, only 12% (6/50) and 18% (9/50) of the respondents answered at least 6 (86%) of the 7 questions correctly before the activity. This increased to 50% (25/50) and 70% (35/50) in the learners, respectively, after the activity (Figure 1). In Figure 1A, the heat maps show the proportion of respondents (N=50) and learners (n=50) who answered a specific number of questions correctly, as displayed by colors ranging from white (the lowest proportion of respondents and learners) to dark red (the highest proportion of respondents and learners). In Figure 1B, the box and whisker plots show the distribution of the number of questions correctly answered by all respondents and learners. In both plots, the horizontal red line within the box indicates the median, the “x” symbol represents the mean, the boxes indicate the IQR, and the vertical lines (whiskers) extend to the range of values, excluding outliers. The outliers are defined as values that fall outside a distance of 1.5 times the IQR from the upper and lower quartiles and are represented by empty circles. The respondents and learners were the HCPs who completed the pre- and postactivity questionnaires, respectively. For both activities, there was a statistically significant increase in the number of correctly answered questions from pre- to postactivity for all participants (touchMDT: median 4.0, IQR 3.0-5.0 to median 5.5, IQR 4.0-7.0; mean 4.00, SD 1.39 to mean 5.30, SD 1.39; P<.001 and touchPANEL DISCUSSION: median 4.0, IQR 4.0-5.0 to median 6.0, IQR 5.0-7.0; mean 4.32, SD 1.30 to mean 5.88, SD 1.49; P<.001; Figure 1). In subgroup analyses, significant increases in the number of correctly answered questions were observed irrespective of specialty or country for both activities and irrespective of years of experience for the touchPANEL DISCUSSION (Multimedia Appendices 12 and 13). Refer to Multimedia Appendix 14 for the responses to individual topics for the levels 3 to 4 questionnaires.
Figure 1. Summary of the number of correct responses for the levels 3 and 4 outcomes questionnaire before and after the launch of touchMDT and touchPANEL DISCUSSION. (A) Heat maps show the proportion of respondents (n=50) and learners (n=50) who answered a specific number of questions correctly, as displayed by colors ranging from white (the lowest proportion of respondents and learners) to dark red (the highest proportion of respondents and learners). (B) Box and whisker plots show the distribution of the number of questions correctly answered by all respondents and learners. In both plots, the horizontal red line within the box indicates the median, the “x” symbol represents the mean, the boxes indicate the IQR, and the vertical lines (whiskers) extend to the range of values, excluding outliers.

Level 5 (Performance)

A total of 50 respondents and 50 learners completed the level 5 questionnaire and the patient records questionnaire (Table 1). The dropout rate was 16% (8/50); therefore, 8 of the 50 learners were de novo HCPs who were matched to the dropouts.

Self-Reported

Preactivity, in total, 6% (3/50) of respondents answered at least 5 (83%) out of 6 questions with the best clinical option. This increased to 44% (22/50) of the learners after the activity (Figure 2). In Figure 2A, the heat maps show the proportion of respondents (n=50) and learners (n=50) who answered a specific number of questions correctly, as displayed by colors ranging from white (the lowest proportion of respondents and learners) to dark red (the highest proportion of respondents and learners).

In Figure 2B, the box and whisker plots show the distribution of the number of questions correctly answered by all respondents and learners. The horizontal red line within the box indicates the median, the “x” symbol represents the mean, the boxes indicate the IQR, and the vertical lines (whiskers) extend to the range of values, excluding outliers. The outliers are defined as values that fall outside a distance of 1.5 times the IQR from the upper and lower quartiles and are represented by empty circles. The respondents and learners were HCPs who completed the pre- and postactivity questionnaires, respectively. There was a statistically significant increase in the number of correctly answered questions from pre- to postactivity for all participants (median 3.0, IQR 2.0-3.0 to median 4.0, IQR 3.0-5.0; mean 2.82, SD 1.08 to mean 4.16, SD 1.45; P<.001). Refer to Multimedia Appendix 15 for the responses to individual topics for the level 5 questionnaires.

https://formative.jmir.org/2024/1/e50931

 JMIR Form Res 2024 | vol. 8 | e50931 | p.1454
In subgroup analyses of data by specialty, years of experience and country (Multimedia Appendix 16), increases in the number of questions with the correct clinical response were observed irrespective of specialty or years of experience, and in the subgroups of participants from Germany, Spain, and Italy.

**Patient Record Data**

The patient characteristics reported by the respondents and learners were broadly similar (Multimedia Appendix 17). The respondents and learners used a wide range of measures to determine the patient risk of recurrence (30 and 26 different combinations of measures, respectively; Multimedia Appendix 18). There was a significant difference between the number of respondents and learners selecting each measure ($\chi^2=508.474; P<.001$) and the number of measures selected ($\chi^2_{56}=203.125; P<.001$; Multimedia Appendix 19). The Ki-67 index use was >90% for both groups, and most of those who used the Ki-67 index reported patients who had tumors with a Ki-67 score of ≥20% (respondents: 30/48, 62.5% and learners: 30/47, 64%; Figure S1 in Multimedia Appendix 20). There was no significant difference between responders and learners regarding adjuvant therapy when analyzing patient record data by Ki-67 index, menopausal status, germline BRCA status, or the number of positive lymph nodes (Figure S2 in Multimedia Appendix 20).

**Identification of Outstanding Educational Gaps**

In ranking the unmet educational needs, touchMDT learners highlighted “strategies for implementing SDM in clinical practice” and “effective communication strategies in SDM to optimize patient outcomes” as key areas for future education, and touchPANEL DISCUSSION learners highlighted “using genetic and biomarker testing to guide therapy choice” and “individualization of treatment choice in high-risk hormone receptor–positive human epidermal growth factor receptor 2 negative (HR+ HER2–) EBC.” From the level 5 questionnaire, learners highlighted “using genetic and biomarker testing to guide therapy choice for patients with high-risk HR+ HER2- EBC” and “understanding the latest data on emerging treatments for high-risk HR+ HER2- EBC” as their most important unmet educational needs (Table S1 in Multimedia Appendix 20).

For the touchPANEL DISCUSSION, high levels of knowledge and competence were achieved for all 3 learning objectives. However, some unmet needs remained after participation in touchMDT, with 60% (30/50) and 36% (18/50) of the learners, respectively, being unable to demonstrate declarative knowledge (40/50; $P=.90$) for the level 5 questionnaire (Figure S3 in Multimedia Appendix 20).
of the current consensus on the use of patient decision aids to support SDM in breast cancer and being unable to demonstrate competence in supporting patients in understanding data and reaching a shared decision regarding neoadjuvant systemic therapy. In the level 5 questionnaire, ≥36% (18/50) of the learners were unable to provide the best clinical responses to the questions on reducing the patient burden (18/50, 36%), risk of recurrence (20/50, 40%), and chemotherapy treatment decisions (26/50, 52%).

Discussion

Principal Findings

This study evaluated the impact of 2 faculty-led, CME-accredited web-based learning activities on EBC. Participation in both activities resulted in high satisfaction scores and statistically significant improvements in self-reported knowledge, competence, and performance. Subgroup analyses indicated that the education was broadly beneficial, regardless of region, specialty, and experience. Patient record data analysis showed statistically significant differences in how the risk of recurrence was determined by responders and learners. The impact of the activities on therapeutic choice was limited, although this may be because of factors such as local guidelines, access to treatments, and individual patient characteristics. Approximately half of the learners (80/150, 53%) stated that they would change their practice in response to the educational material, further suggesting that the activities were valuable learning tools.

Participation in these web-based activities was considerably higher than could be achieved with traditional face-to-face programs and may also reflect the involvement of a multinational panel of recognized experts and the availability of translations. The latter is a major difference from traditional conference-based activities, as they are usually only available in 1 language, typically English.

Although many learners stated that they would change their practice, there remained a considerable number who were uncertain or who would not change their practice at the current time. The reasons for not changing practice were largely split between the requirement for more education and practical limitations. Although learners were not canvassed for specific details regarding practical limitations, we can speculate that they may relate to regional or national differences in institutional resources and structure or regulatory and reimbursement approvals of treatment. Indeed, documented barriers to change in clinical practice include institution-related issues such as flawed leadership and communication and insufficient organizational culture shift, and practical restrictions such as recently updated guidelines and access to emerging drugs [53]. In addition, a 2022 study on factors that influenced HCPs’ intent to put newly acquired learning into practice suggested that a lack of belief in one’s capabilities and the potential consequences of adopting new clinical behaviors can prevent HCPs from translating education into practice [54].

These data demonstrate the potential value of succinct web-based learning activities that are specifically designed to address knowledge gaps in complex and rapidly evolving medical fields. Although published data on outcomes following educational activities with similar methodologies remain limited, the results of this study are consistent with the results of a small number of previous studies in other disease areas [55-57]. Notably, the field of breast cancer is evolving rapidly, and it is essential for oncologists, particularly nonspecialists, to remain informed of the latest findings and recommendations [19-26]. Therefore, focused web-based learning activities may also be valuable for general oncologists to enable them to keep up-to-date with changes in specialist fields.

Strengths and Limitations

There were several strengths associated with this study:

1. The activities provided a holistic curriculum and unbiased education focused on clearly defined learning objectives.
2. The activities were translated into several languages and were available on a free-to-access website, which likely increased their global accessibility.
3. A broad target audience was successfully reached, highlighting the importance of the multichannel publicity strategy and the inclusion of faculty from often underrepresented geographies (Latin America and Asia-Pacific).
4. The multidisciplinary team approach and inclusion of a survivor of breast cancer ensured that the activities were directly applicable to the participants’ daily practice; the emotive element can also aid learning and retention [58].
5. The format used in the touchPANEL DISCUSSION allowed the latest evidence to be put into context for practicing physicians.
6. The development of the education in a short time frame allowed for emerging data to be communicated to the HCPs potentially ahead of publications and conventional educational formats.
7. A major strength was using an objective measure of performance to complement the more subjective self-reported data.

This study also has several limitations:

1. The conclusions that could be drawn from the main and subgroup analyses were speculative and limited because of the relatively small sample size.
2. All medical educational studies are affected by inevitable self-selection bias; that is, the HCPs who feel they lack knowledge on a specific topic are more likely to participate and, therefore, show a bigger impact of the education than might be the case for a wider audience.
3. This study only evaluated outcomes following 2 specific CME activities in EBC, and the conclusions may not be generalizable to other web-based learning formats.
4. An independent samples method was used for the levels 3 to 4 questionnaires and the self-reported level 5 questionnaire to avoid pre-exposure bias; however, this meant that the direct impact of education on each learner could not be determined. Therefore, we cannot rule out the possibility that other sources of education may have contributed to the reported outcomes.
Despite these limitations, this study offers the first comprehensive assessment of such activities using an accepted methodology.

**Identification of Needs for Further Education in High-Risk EBC**

Notably, only 30% (15/50) and 48% (24/50) of the touchMDT and touchPANEL DISCUSSION learners, respectively, answered all questions correctly in the levels 3 to 4 questionnaires. Similarly, only 20% (10/50) of the learners gave the best clinical response for all questions in the level 5 questionnaire. This may, in part, reflect a mixed audience but suggests that further education may be warranted to reinforce the content of these activities. To refine future educational activities, we identified several unmet needs during this study from the questions that were incorrectly answered after the activities. In addition, several self-reported educational gaps were identified, including implementing SDM in clinical practice, using genetic and biomarker testing to guide therapy choices, and understanding the latest data on emerging treatments.

**Future Directions**

Limitations of traditional CME approaches in affecting changes in physician performance have been previously reported [59]; however, the optimal format for continuing educational activities in both breast cancer and the wider medical field remains to be fully defined. This study and other similar studies show evidence that more engaging approaches have the potential to improve the effectiveness of CME [55-57,60,61]. Evaluation of the long-term impacts of CME on HCP performance, alongside assessment against Moore’s Levels 6 and 7 [40] to determine the effect of education on patient (level 6) and community (level 7) health will add further value to CME programs in the future. Activity formats that may be particularly relevant include experts sharing insights drawn from experience and their own clinical practice, conversational and panel discussion formats, and expert interviews supplemented by vignettes of consultation strategies. Furthermore, web-based delivery may offer scope for developing more individualized approaches. Web-based activities can also be made available as enduring materials, providing a resource designed for optimal HCP accessibility. Future studies will need to be designed to evaluate the comparative effectiveness of these different approaches to CME. In addition, in future studies, a larger sample size will increase the statistical power of the analyses.

**Conclusions**

Our study demonstrated that focused, free-to-access web-based CME activities on EBC were associated with improvements in HCP knowledge, competence, and self-reported performance. This study addressed the gaps in the provision of HCP-focused web-based IME programs that encompass recent advances in the high-risk EBC treatment landscape, the need for and approaches to effective SDM, and the gaps within the literature in terms of reporting outcomes from web-based CME activities. The identified unmet needs should be used to inform the design of future educational activities for this disease area. Further studies evaluating the long-term impact of education on HCP performance and on patient and community health will be valuable in defining the clinical impact of CME and the most effective channels for its delivery.

**Acknowledgments**

Medical writing support was provided by Steph Carter, and editorial support was provided by Cheryl Brant, both from touchIME, Cheshire, United Kingdom. The educational activities were jointly provided by the University of South Florida Health and touchIME. The authors would like to thank the touchIME audience outreach team (Joel Turner, Hannah Morton-Fishwick, and Louise Jarman) for the collection and analysis of the level 1 data. The authors would like to thank former employee of touchIME Sonia Mills for additional editorial support. Eli Lilly provided an independent medical education grant to fund the educational activity, medical writing, and editorial support but did not contribute directly to manuscript development, the decision to submit, or the submission process.

**Data Availability**

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

**Authors’ Contributions**

All authors contributed to the conceptualization, formal analysis of the data, and writing (review and editing) of the manuscript. KB, SN, A Noble, A Nunn and CS contributed to the investigation, methodology, project administration, visualization of the study, and writing (original draft) of the manuscript. SN contributed to the funding acquisition for and supervision of the manuscript. A Noble contributed to data curation.

**Conflicts of Interest**

CB reports receiving advisory board fees from AstraZeneca, Bayer, Boehringer Ingelheim, Eisai, GlaxoSmithKline, Lilly, MSD, Novartis, Pfizer, Genentech (Roche), Sanofi, and Zodiac Pharma; consultancy fees from AstraZeneca, Bayer, Boehringer Ingelheim, Eisai, GlaxoSmithKline, Lilly, MSD, Novartis, Pfizer, Genentech (Roche), Sanofi, and Zodiac Pharma; grants and research support from AbbVie, Abraxis Biosciences, AB Science, Amgen, Asana Biosciences, Astellas Pharma, AstraZeneca, BioMarin,
Boehringer Ingelheim, Bristol Myers Squibb, Daiichi Sankyo, Exelixis, GlaxoSmithKline, ImClone Systems, LEO Pharma, Lilly, Medivation, MSD, Merck KGaA, Merrimack, Millennium, Mylan, Novartis, Pfizer, PharmaMar, Polyphor, Genentech (Roche), Sanofi, Shanghai Henlius Biotech, and Taiho Pharmaceutical; and holding stock and shares (self-managed) in MEDSIR and Tummi. FB reports receiving advisory board fees from Eisai, Eli Lilly, Gilead, MSD, Novartis, Pfizer, and Roche and consultancy fees from Eli Lilly. MG reports receiving personal fees and travel support from AstraZeneca, Daiichi Sankyo, Eli Lilly, Menarini-Stemline, MSD, Novartis, Pierre Fabre, and Veracyte, and his immediate family member is employed by Sandoz. NH reports receiving consultancy fees from Gilead, Roche, Sandoz, and Seagen and speaker bureau fees from AstraZeneca, Daiichi Sankyo, Gilead, Lilly, MSD, Novartis, Pierre Fabre, Pfizer, Roche, and Seagen. CSH reports receiving advisory board fees from AstraZeneca, Lilly, Pfizer, Novartis, and Roche; grants and research support from AstraZeneca, Daiichi Sankyo, EirGenix, Lilly, MSD, Novartis, OBI Pharma, Pfizer, and Roche; and speaker bureau fees from AstraZeneca, Novartis, Pfizer, and Roche. KLA has no interests or relationships or affiliations to disclose. KB, SN, A Noble, A Nunn and CS are employees of touch Independent Medical Education Limited.

Multimedia Appendix 1
Activity development process.
[DOCX File , 25 KB - formative_v81e50931_app1.docx ]

Multimedia Appendix 2
Assessment of participation (level 1) and satisfaction (level 2).
[DOCX File , 22 KB - formative_v81e50931_app2.docx ]

Multimedia Appendix 3
Topics included in levels 3 and 4 and level 5 outcomes questionnaires.
[DOCX File , 13 KB - formative_v81e50931_app3.docx ]

Multimedia Appendix 4
Questions included in the levels 3 and 4 outcomes questionnaire for the touchMDT activity.
[DOCX File , 16 KB - formative_v81e50931_app4.docx ]

Multimedia Appendix 5
Questions included in the levels 3 and 4 outcomes questionnaire for the touchPANEL DISCUSSION activity.
[DOCX File , 16 KB - formative_v81e50931_app5.docx ]

Multimedia Appendix 6
Questions included in the level 5 outcomes questionnaire for the touchMDT and touchPANEL DISCUSSION activities.
[DOCX File , 16 KB - formative_v81e50931_app6.docx ]

Multimedia Appendix 7
Unmet educational need options included in levels 3 to 4 and level 5 questionnaires.
[DOCX File , 13 KB - formative_v81e50931_app7.docx ]

Multimedia Appendix 8
Single transferable vote system methodology.
[DOCX File , 12 KB - formative_v81e50931_app8.docx ]

Multimedia Appendix 9
Engagement results and demographics for the touchMDT and touchPANEL DISCUSSION activities.
[DOCX File , 14 KB - formative_v81e50931_app9.docx ]

Multimedia Appendix 10
Mean satisfaction scores of levels 2 to 4 questionnaire (out of a maximum of 5) completed by touchMDT and touchPANEL DISCUSSION learners.
[DOCX File , 13 KB - formative_v81e50931_app10.docx ]

Multimedia Appendix 11
Multichannel publicity reach and impact for the touchMDT and touchPANEL DISCUSSION activities.
Multimedia Appendix 12
Mean number of correct responses for the levels 3 and 4 outcomes questionnaire before and after the launch of touchMDT by (A) country, (B) years of experience, and (C) specialty subgroups of the respondents and learners.

Multimedia Appendix 13
Mean number of correct responses for the levels 3 and 4 outcomes questionnaire before and after the launch of touchPANEL DISCUSSION by (A) country, (B) level of experience, and (C) specialty of the respondents and learners.

Multimedia Appendix 14
Summary of correct responses for individual topics for the levels 3 and 4 outcomes questionnaire before and after the launch of (A) touchMDT and (B) touchPANEL DISCUSSION.

Multimedia Appendix 15
Summary of correct responses for individual topics for the level 5 outcomes questionnaire before and after the launch of touchMDT and touchPANEL DISCUSSION.

Multimedia Appendix 16
Mean number of correct responses for the level 5 outcomes questionnaire before and after the launch of touchMDT and touchPANEL DISCUSSION by (A) country, (B) level of experience, and (C) specialty of the respondents and learners.

Multimedia Appendix 17
Patient characteristics from patient records data (level 5) submitted by responders and learners for the touchPANEL DISCUSSION activity.

Multimedia Appendix 18
Combinations of measures used to determine patient risk for recurrence from patient records data (level 5) submitted by responders and learners for the touchPANEL DISCUSSION activity.

Multimedia Appendix 19
Measures used to determine patient risk for recurrence reported by respondents and learners in the level 5 patient records questionnaire.

Multimedia Appendix 20
Ki-67 use and Ki-67 index scores reported by respondents and learners in the level 5 patient records questionnaire (Figure S1); adjuvant or planned adjuvant therapy reported by respondents and learners in the level 5 patient records questionnaire (Figure S2); changes in self-reported confidence in (A) levels 3 to 4 questionnaires and (B) level 5 questionnaires following participation in the touchMDT and touchPANEL DISCUSSION activities (Figure S3); unmet educational needs identified by the touchMDT and touchPANEL DISCUSSION learners (Table S1).

References


50. Understanding health research. UK Research and Innovation. URL: https://tinyurl.com/4u32dw35 [accessed 2024-01-26]


56. Larkin A, Schrand J, Le A. Success of CME at improving knowledge of clinical use of real-time CGM. Diabetes 2021;70(Supplement_1):614P [FREE Full text] [doi: 10.2337/db21-614-P]


Abbreviations

Akt: protein kinase B
CME: continuing medical education
EBC: early breast cancer
HCP: health care professional
SDM: shared decision-making
touchIME: touch Independent Medical Education
Development of Therapeutic Alliance and Social Presence in a Digital Intervention for Pediatric Concussion: Qualitative Exploratory Study

Kiarah M K O’Kane¹, BA; Thalia Otamendi², PhD; Noah D Silverberg¹, PhD; Esther Choi¹, BA; Veronik Sicard³, PhD; Roger Zemek³,⁴, MD; Katherine Healey³, MSc; Olivier Brown⁵,⁶, BA; Lauren Butterfield⁷, BSc; Andra Smith⁵, PhD; Gary Goldfield⁶, PhD; Rachel Kardish⁶, MSc; Bechara J Saab⁵, PhD; Andrée-Anne Ledoux⁵,⁶,⁸*, PhD; Molly Cairncross⁹*, PhD

¹Department of Psychology, University of British Columbia, Vancouver, BC, Canada
²Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada
³Children’s Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
⁴Department of Pediatrics and Emergency Medicine, University of Ottawa, Ottawa, ON, Canada
⁵School of Psychology, Faculty of Social Sciences, University of Ottawa, Ottawa, ON, Canada
⁶Department of Neuroscience, Carleton University, Ottawa, ON, Canada
⁷Mobio Interactive, Singapore, Singapore
⁸Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada
⁹Department of Psychology, Simon Fraser University, Burnaby, BC, Canada

*these authors contributed equally

Corresponding Author:
Molly Cairncross, PhD
Department of Psychology
Simon Fraser University
Robert C. Brown Hall 7321
8888 University Drive
Burnaby, BC, V5A 1S6
Canada
Phone: 1 778 782 7458
Email: molly_cairncross@sfu.ca

Related Article:
This is a corrected version. See correction statement: https://formative.jmir.org/2024/1/e59722

Abstract

Background: Despite the promising benefits of self-guided digital interventions for adolescents recovering from concussion, attrition rates for such interventions are high. Evidence suggests that adults can develop therapeutic alliance with self-guided digital interventions, which is in turn associated with intervention engagement. However, no research has examined whether adolescents develop therapeutic alliance with self-guided digital interventions and what factors are important to its development. Additionally, social presence—the extent to which digital encounters feel like they are occurring in person—may be another relevant factor to understanding the nature of the connection between adolescents and a self-guided digital intervention, though this has yet to be explored.

Objective: This qualitative study explored the extent to which adolescents recovering from concussion developed therapeutic alliance and social presence during their use of a self-guided digital mindfulness-based intervention. Additionally, this study aimed to determine factors important to adolescents’ development of therapeutic alliance and social presence with the intervention.

Methods: Adolescents aged between 12 and 17.99 years who sustained a concussion were recruited from 2 sites: a pediatric emergency department up to 48 hours after a concussion and a tertiary care clinic over 1 month following a concussion to capture adolescents who had both acute and persisting symptoms after concussion. Participants (N=10) completed a 4-week
mindfulness-based intervention delivered through a smartphone app. Within the app, participants listened to audio recordings of mindfulness guides (voice actors) narrating psychoeducation and mindfulness practices. At 4 weeks, participants completed questionnaires and a semistructured interview exploring their experience of therapeutic alliance and social presence with the mindfulness guides in the intervention.

**Results:** Themes identified within the qualitative results revealed that participants developed therapeutic alliance and social presence by “developing a genuine connection” with their mindfulness guides and “sensing real people.” Particularly important to the development of therapeutic alliance and social presence were the mindfulness guides’ “personal backgrounds and voices,” such that participants felt more connected to the guides by knowing information about them and through the guides’ calm tone of voice in audio recordings. Quantitative findings supported qualitative results; participants’ average score for therapeutic alliance was far above the scale midpoint, while the mixed results for social presence measures aligned with qualitative findings that participants felt that the mindfulness guides seemed real but not quite as real as an in-person connection would.

**Conclusions:** Our data suggest that adolescents can develop therapeutic alliance and social presence when using digital interventions with no direct human contact. Adolescents’ development of therapeutic alliance and social presence with self-guided digital interventions can be bolstered by increasing human-like qualities (eg, real voices) within interventions. Maximizing therapeutic alliance and social presence may be a promising way to reduce attrition in self-guided digital interventions while providing accessible treatment.

(JMIR Form Res 2024;8:e49133) doi:10.2196/49133

**KEYWORDS**
adolescent; concussion; digital therapeutics; eHealth; mHealth; mindfulness; mobile health; social presence; working alliance

**Introduction**

Concussions are one of the most common injuries among adolescents, responsible for nearly 50,000 pediatric emergency department visits in Canada each year [1]. While most adolescents who sustain a concussion recover within 1 month, approximately 30% develop persistent postconcussion symptoms (PPCS) that continue longer than 1 month after their injury [2-4]. PPCS are associated with a variety of negative outcomes in adolescents, including reduced quality of life [5]. Despite the prevalence of PPCS and the detrimental effects that they have on the quality of life, few accessible evidence-based treatments exist for acute and persistent symptoms of concussion [6].

One promising intervention for those with acute or persisting concussion symptoms is mindfulness-based intervention (MBI) [7,8]. Most MBI research has focused on interventions delivered in-person. However, traditional in-person therapies are inaccessible for many due to financial barriers and sociodemographic disparities in access to health care. Moreover, a higher incidence of concussion in rural settings [9] means that many patients may have difficulty accessing specialized concussion care clinics, which are typically concentrated in urban settings [10]. Beyond geographical distance alone, specialized clinics often have long waitlists and operate solely during work and school hours, further complicating access to care. Additional barriers for adolescents to accessing treatment result from reliance on parental support for transport to and funding for in-person treatments. Finally, during infectious disease outbreaks such as the COVID-19 pandemic, in-person treatments may be severely limited or not available at all due to public health regulations. Taken together, in-person treatment for acute or persisting concussion symptoms in adolescents may not be readily accessible.

By contrast, self-guided digital interventions offer a potential solution to disparities in access to care for concussions in adolescents. A total of 95% of adolescents in Canada own a mobile digital device [11], highlighting that almost all youth in Canada have good access to digital care regardless of socioeconomic status. Digital interventions can be used anywhere and are a more affordable alternative to in-person treatments [12]. Some digital MBIs have also been shown to be effective in treating a range of conditions such as chronic pain, depression, anxiety, and stress [13,14], even when controlling for expectancy effects with active control comparison groups [15]. These conditions are also secondary consequences of concussion [2,16], which can further complicate the recovery process. Given digital MBIs’ effectiveness for treating these secondary consequences of concussion, they may have use for concussion recovery itself as well. However, attrition rates are high, and adherence suboptimal to digital interventions, with attrition rates as high as 52% in self-guided digital MBIs [12,17,18]. If patients do not adhere to treatment, they are less likely to benefit from it as intended. Thus, it is imperative to understand the factors underlying the large variation in self-guided MBI attrition rates.

One factor known to be essential for adherence and positive patient-reported outcomes following clinician-delivered behavioral interventions is the therapeutic alliance. Therapeutic alliance refers to the collaborative relationship and affective bond between a therapist and their patient [19,20]. The therapeutic alliance comprises 3 dimensions: the agreement on therapeutic goals, the agreement on therapeutic tasks, and the emotional bond between a therapist and their patient [21]. Particularly important to the conceptualization of therapeutic alliance is the collaborative nature of the relationship between a therapist and patient; this collaboration is integral to the success of psychotherapeutic treatments [19]. In face-to-face treatments, a strong therapeutic alliance is associated with positive treatment outcomes [20] and lower attrition [22]. Therapeutic alliance may also be relevant to patient adherence in the context of self-guided digital interventions. Self-guided
digital interventions lack the direct contact with human therapists found in in-person treatments. However, research suggests that individuals often respond to digital technology as if it was another person [23]. As such, researchers have explored whether individuals may anthropomorphize—or attribute human-like qualities to—digital interventions themselves and therefore develop therapeutic alliance with the interventions. Indeed, initial evidence from interventions for adults focused on addressing a range of health behaviors and outcomes [24-31] suggests that individuals can develop therapeutic alliance when using self-guided digital interventions by feeling a connection to the intervention or digital social actors within it. Intervention format varied across studies; interventions included audio recordings of human voice actors [30], human-like digital avatars [25], fully automated text-based “conversations” with an intervention [28,29], and intervention content delivered through multimedia modules [24,26,27,31,32], highlighting that therapeutic alliance can develop with self-guided digital interventions in a variety of contexts. Additionally, evidence from quantitative studies of self-guided digital interventions indicates that therapeutic alliance is positively associated with patient engagement [27,32] and therapeutic outcomes [31,32]. This suggests that similar to in-person treatments, feeling socially supported is important to adherence in self-guided digital contexts as well. However, none of these studies sampled adolescents. Adolescents’ experiences of therapeutic alliance with self-guided digital interventions may differ from those of adults, given that adolescents engage with digital technology in unique ways from adults [33,34]. As such, whether and how adolescents develop therapeutic alliance when using self-guided digital interventions remains to be explored.

The nature of therapeutic alliance in self-guided digital interventions may differ from that developed during face-to-face treatments, given the lack of real-time interaction with a human therapist. For example, previous research found that while participants felt emotionally connected to and anthropomorphized their self-guided digital intervention, at other times, they felt that the intervention was a thing rather than a real person [29,30]. Holter’s relational model of “making come-alive” and “keeping un-alive” [28,29] can provide theoretical context as to why therapeutic alliance in self-guided interventions may be unique from therapeutic alliance developed in in-person therapy. The model suggests that individuals engage in 2 types of relational processes when using self-guided digital interventions: “making come-alive,” or thinking of a digital intervention as a social actor with human-like qualities, and “keeping un-alive,” or viewing the intervention as an inanimate object without real emotions. Individuals’ dynamic interplay between “making come-alive” and “keeping un-alive” produces a therapeutic alliance that at once both involves emotional closeness to a digital intervention and emotional distance through recognition that the program is not a real human entity. Social presence may be a suitable construct for further understanding the therapeutic relationship and the extent to which individuals “make come-alive” or “keep un-alive” within self-guided digital interventions. Stemming from the communications and education literature, social presence is the extent to which a digital encounter feels like it is occurring in-person and has a sense of human connection to it [35-37]. Higher social presence is related to more trust in a web-based communicator [38] and greater relationship satisfaction, attachment, and perceived benefit of caregiving robots [39]. This suggests that social presence may approximate how much individuals experience a bond with social actors within a digital intervention. Social presence is also associated with satisfaction with, enjoyability of, and intention to use a mobile app or website [40-42]. Given the importance of these factors for adherence to treatment [43], attending to the social presence in self-guided digital intervention design may maximize adherence.

Social presence has not yet been examined in the context of a clinical intervention, and adolescents’ development of therapeutic alliance in the context of self-guided digital interventions is not well understood. This study sought to address these gaps in the literature and identify what factors may be relevant to adolescents’ development of therapeutic alliance and social presence in self-guided digital interventions. The aim of the present study was to determine if therapeutic alliance and social presence can be developed with audio-recorded mindfulness guides in a self-guided digital intervention among adolescents with acute or persisting concussion symptoms. Second, we aimed to determine what aspects of the intervention were important to the therapeutic alliance’s and social presence’s development.

Methods

Study Design
A qualitative exploratory design was used in the study. Since adolescents’ experiences of social presence and therapeutic alliance in clinical self-guided digital interventions are not yet well understood, we chose to use qualitative methods to explore the nuances of these experiences. Qualitative methods are well-suited for exploring participants’ experiences with digital applications, including what aspects of the application met their needs and what could improve their experience with the application. This was a planned secondary data analysis from an open-label pilot trial of a digital MBI for adolescents post concussion. The study was an open-label pilot trial because both the research team and participants knew that participants would be receiving a digital MBI. The pilot trial was a small-scale study and centered around the first stage of intervention development. The primary aim of the pilot trial was to gain information on the feasibility, credibility, and satisfaction of patient users with the digital MBI. The reporting of the study followed the Journal Article Reporting Standards for Qualitative Research (JARS-Qual) guideline [44].

Research Team and Reflexivity

Interviews were conducted by KMKO, OB, and RK. KMKO conducted interviews with the PPCS cohort (adolescents experiencing PPCS) and OB and RK with the acute cohort (adolescents recovering from acute concussion). KMKO has a bachelor’s degree in psychology and was a directed studies student at the University of British Columbia at the time of the interviews. They were trained and supervised while conducting the interviews by MC, a postdoctoral fellow at the University of British Columbia at the time of data collection. MC holds a
PhD in clinical psychology. OB has a bachelor’s degree in psychology and was a PhD candidate in experimental psychology at the time of the interviews. RK has a bachelor's degree in neuroscience and mental health and was a master’s candidate at the time of the interview. They were trained and supervised while conducting the interviews by AAL, a scientist at the Children’s Hospital of Eastern Ontario Research Institute. AAL holds a PhD in experimental psychology.

Interviewers were specifically trained to conduct interviews in a nondirective manner and were given feedback from supervisors during practice interviews. Given the self-guided nature of the intervention, the research team remained open to the possibility that participants would not develop a therapeutic alliance or social presence with their mindfulness guides. If participants did not develop a therapeutic alliance or social presence when using the app, the research team planned to change the intervention during subsequent stages of intervention development based on participant feedback in this study to better support future users. As such, interviewers explored whether or not participants experienced therapeutic alliance and social presence with curiosity and openness to participant feedback. The participants and interviewers were unknown to each other before the study. At the beginning of the interviews, participants were made aware that the interviewers were research assistants affiliated with the research team that conducted the study. They were told that the researcher was interested in learning about their experiences while using the mindfulness app.

Qualitative data were analyzed by KMKO and TO. At the time of data analysis, KMKO had approximately 2 years of experience conducting research about patient experiences with digital interventions. Based on their professional background and familiarity with the therapeutic alliance and social presence literature, KMKO brought certain assumptions to data analysis. For example, KMKO expected that participants would likely develop some degree of therapeutic alliance and social presence with the mindfulness guides given the human-like qualities present in audio recordings of the guides (eg, using real human voices). Additionally, because KMKO conducted interviews with the PPCS cohort, they had familiarity with the experiences of the PPCS cohort before beginning data analysis. Due to KMKO’s awareness of their preconceived expectations, they exercised caution during data analysis and engaged in constant peer debriefing with TO to ensure that findings were indeed based on participants’ accounts. TO is a Rehabilitation Sciences PhD student at the University of British Columbia who is trained in qualitative methodologies and has a background in psychology. TO was not versed in the therapeutic alliance, social presence, or digital intervention literature and did not conduct any of the interviews. As such, TO had limited assumptions about the findings going into data analysis.

Participants

The sample included 10 adolescents aged between 12 and 17.99 years who were either recovering from acute concussion (acute cohort) or were experiencing PPCS (PPCS cohort). The sample included both the acute cohort and the PPCS cohort because the digital MBI was designed both to help adolescents acutely recovering from concussion and those with persisting symptoms. As such, one of the primary aims of the pilot trial was to understand whether the digital intervention met the needs of both populations. In this study, we did not expect the acute cohort and PPCS cohort to differ with respect to their experiences of therapeutic alliance and social presence. Therefore, the 2 cohorts were collapsed into 1 overall sample for data analysis.

The acute and PPCS cohorts were recruited from 2 separate sites. Participants in the acute cohort (n=7) were recruited from the emergency department within 48 hours of sustaining a concussion. Participants in the PPCS cohort (n=3) were youth who sustained a concussion between 1 and 12 months before participation in the study and continued to experience concussion symptoms in at least 3 categories, as defined by the International Classification of Diseases, Tenth Revision criteria for PPCS [45]. They were recruited from a tertiary care concussion clinic. Additional eligibility criteria for the study included that participants were proficient in English and had access to a smartphone or tablet with an internet connection. Participants were excluded if they had a severe chronic neurological developmental delay resulting in communication difficulties, previous psychiatric hospitalization, previous diagnosis of a severe psychiatric disorder, such as schizophrenia (acute cohort only), or if participants were currently involved in other MBIs (PPCS cohort only). The acute and PPCS cohorts differed in size due to challenges associated with recruiting participants experiencing chronic concussion symptoms. In particular, given that recruitment was carried out during the COVID-19 pandemic, anecdotally, adolescents with PPCS faced greater barriers to care than usual, such as increased family stress and restricted access to the tertiary care concussion clinic where recruitment was carried out due to public health regulations.

Participants’ demographic information is summarized in Table 1. A total of 10 adolescents (mean age 15.93, SD 1.86 years) participated in our study. Overall, 7 participants identified as female, and 3 participants identified as male. Participants are described using pseudonyms to protect their identities.
Table 1. Participant demographic characteristics of adolescents recovering from concussion. Participant names presented in this table are pseudonyms to protect the participants’ anonymity.

<table>
<thead>
<tr>
<th>Cohort and participants</th>
<th>Sex</th>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acute</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sophia</td>
<td>Female</td>
<td>17.22</td>
</tr>
<tr>
<td>Julian</td>
<td>Male</td>
<td>16.62</td>
</tr>
<tr>
<td>June</td>
<td>Female</td>
<td>13.94</td>
</tr>
<tr>
<td>Ella</td>
<td>Female</td>
<td>14.26</td>
</tr>
<tr>
<td>Olivier</td>
<td>Male</td>
<td>16.94</td>
</tr>
<tr>
<td>Marie-Claude</td>
<td>Female</td>
<td>12.63</td>
</tr>
<tr>
<td>Gabrielle</td>
<td>Female</td>
<td>14.73</td>
</tr>
<tr>
<td><strong>Persistent postconcussion symptoms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alex</td>
<td>Male</td>
<td>17.86</td>
</tr>
<tr>
<td>Ada</td>
<td>Female</td>
<td>17.48</td>
</tr>
<tr>
<td>Kay</td>
<td>Female</td>
<td>17.62</td>
</tr>
</tbody>
</table>

**Ethical Considerations**

This study’s procedures were reviewed and granted ethics approval by the Behavioral Research Ethics Board at the University of British Columbia (H20-00120) and the Research Ethics Board at the Children’s Hospital of Eastern Ontario Research Institute (20/72X). Participants capable of consenting provided their informed consent, and parents also provided consent on their children’s behalf. Participants who were unable to consent provided assent instead, and their parents provided informed consent on their behalf. Participants were able to ask the research assistant questions that they had about the study procedures before providing their consent. With respect to study data, all participants were assigned a unique participant ID number that was linked to their data; as such, all study data were deidentified and stored in the secure platform, Research Electronic Data Capture (REDCap) [46,47]. Participants in the acute cohort were compensated for their participation with a CAD $10 (US $7.40) gift card and received a letter stating that they completed 10 hours of volunteer work, while those in the PPCS cohort were given a CAD $15 (US $11.11) gift card. Compensation differed between the 2 cohorts because they were recruited from separate sites, with data collection for each cohort led by teams at 2 different institutions. As such, the cohorts were compensated in accordance with the norms of the given institution leading each cohort’s data collection.

**Procedure**

Prospective participants for the acute cohort were approached in the emergency department. A member of the research team briefly described the study. Interested prospective participants were called by telephone by a research assistant for the eligibility screening and consent or assent process the next day.

Prospective participants for the PPCS cohort were approached at web-based psychoeducation sessions at a tertiary care concussion clinic. A research assistant provided prospective participants with information about the study and gave all participants a link to a survey on the secure platform, REDCap.

The survey allowed patients to provide consent to be contacted by the research team if they were interested in participating in the study. Interested prospective participants were called by telephone by a research assistant for the eligibility screening and consent or assent process.

Upon confirming participants’ eligibility and receiving informed consent, participants were emailed a unique link to register and download the intervention through REDCap. The intervention was developed by the research team and is hosted on the AmDTx app. The app was downloaded on participants’ smartphone or tablet, along with instructions for use. A research assistant also called participants by telephone to help instruct them through the app download process and answer any technical questions they had. Concurrently, participants were emailed links to complete baseline measures through REDCap. All data were collected and stored in a secure database on REDCap.

Participants then engaged with the intervention at their own pace over the course of 4 weeks, following which they received a link to complete outcome measures assessing social presence and therapeutic alliance on REDCap. After participants finished the outcome measures, they completed a 30–45–minute semistructured interview through Zoom (Zoom Video Communication) video call with a member of the research team. Audio of the interviews was recorded for transcription.

**Intervention**

During this pilot phase, participants completed 4 weeks of the 8-week treatment. The recordings of psychoeducation and mindfulness practices in the app were narrated by 2 professional mindfulness guides, Ruby and Brian. Ruby and Brian both have substantial mindfulness training and teaching experience. Moreover, Ruby (who led all the psychoeducation) had experience working with adolescents with concussion. The MBI was modeled based on mindfulness-based stress reduction [48] with meditations and content specifically designed for adolescents with concussion. The psychoeducation topics covered in the MBI involved pain acceptance, stress
management, and nonjudgement toward concussion symptoms. Mindfulness practices included mindful breathing, body scans, and mindful movement. Participants listened to the audio recordings of the mindfulness guides narrating the psychoeducation sessions and mindfulness exercises while using the app. Participants were encouraged to participate in the intervention for 15 minutes per day, for 4 days a week. The content was the same for all participants. See Multimedia Appendix 1 for images of the intervention content.

Throughout treatment, participants also received weekly SMS text messaging support from an “SMS text messaging coach,” who was a trained member of the research team. The SMS text messaging coach’s role was to support participants with any technical issues and problem-solve any barriers to engagement, such as having difficulty finding time to complete their mindfulness practices. SMS text messages sent to participants followed a highly templated SMS text messaging protocol. For a copy of the SMS text messaging protocol, see Multimedia Appendix 2.

Measures

Therapeutic Alliance

Therapeutic alliance was measured using the Working Alliance Inventory adapted for Guided Internet Interventions (WAI-I). The WAI-I is an adaptation of the Working Alliance Inventory–Short Form Revised designed for use with digital interventions. It is a self-reported, 12-item measure assessing participants’ alignment on tasks and goals with a digital intervention and their bond with their mindfulness guides during the intervention [49]. In this study, items were modified to reflect the MBI at hand (eg, “The goals of the app-based mindfulness program are in line with my goals”), including replacing the word “psychologist” in the original measure with “mindfulness guides” (eg, “The mindfulness guides who support me in the app-based program are really interested in my well-being”). Each item was assessed using a 5-point Likert scale ranging from 1 (never) to 5 (always). WAI-I total scores were calculated by summing items out of a possible 60 points. The WAI-I is a reliable measure with good construct, convergent, discriminant, and external validity [49].

Social Presence

In order to evaluate distinct aspects of social presence, 2 measures were used to capture how much digital encounters feel like they are occurring in-person and the ways in which digital encounters have human-like qualities. The first measure was adapted from Nowak and Biocca’s self-reported, 6-item Perceived Social Presence (PSP) scale [36]. The measure assessed how much participants felt like their mindfulness guides were real people and how much it felt like they were “with” their guides while using the app (ie, “To what extent was this as if you were in the same room with your mindfulness guide?”) [36]. We modified the wording of items to reflect that participants were reporting about their experiences with the mindfulness guides, as opposed to the wording “partner” used in the original measure. Items were measured on a Likert scale ranging from 1 (no extent) to 7 (full extent). The measure was scored by summing items out of a possible 42 points. The PSP scale has good internal consistency, convergent validity, and discriminant validity [36,50].

The second measure was the self-reported, 17-item Gunawardena Social Presence Indicators (SPI) scale [51] which assessed the human-like qualities of the audio-recorded mindfulness guides in the intervention. Participants were presented with 5-point bipolar items consisting of contrasting adjectives (ie, “personal-impersonal”) [51] and were instructed to circle the number that best reflected their perceptions of their mindfulness guides’ interpersonal qualities for each item. The measure was scored by summing items out of a possible 85 points. Higher scores reflected more negative perceptions of mindfulness guides. The SPI scale demonstrates good convergent validity [52,53].

Semistructured Interview

A semistructured, qualitative interview was developed by the research team to better understand participants’ experience with the app. Before the study’s commencement, the interview was pilot-tested with a 15-year-old nonparticipant to obtain feedback about understandability and was also reviewed by expert clinicians and researchers for content. The interview covered topics such as participants’ experience of therapeutic alliance and social presence with the mindfulness guides while using the app, as well as changes they would make to the app to improve those experiences (ie, “Can you tell me about your connection with your mindfulness guides? Is there anything you would have liked to change about your connection with your mindfulness guides?”). Other topics explored in the interview included participants’ satisfaction with the app content and ease of use and their satisfaction with support received from their SMS text messaging coaches, though data gathered from these questions is outside the scope of this study (for more information, see [54]). For a copy of the interview guide, see Multimedia Appendix 3.

Data Analysis

Quantitative Analysis

Participant demographics were summarized using descriptive and frequency statistics, as appropriate. Scores on the WAI-I and the PSP scale represent the degree to which participants developed therapeutic alliance and experienced social presence when using the app. As such, scores above the scales’ floor suggest participants developed some degree of therapeutic alliance and social presence. The scale midpoint was used for the SPI scale to indicate whether participants viewed their mindfulness guides more positively, neutrally, or negatively. We chose this method of examining our quantitative results because traditional quantitative data analysis was not appropriate for our data, given our small sample size. The quantitative results are used to supplement the main, qualitative results stemming from the semistructured interview by demonstrating the degree to which participants did or did not develop therapeutic alliance and social presence when using the intervention.
Qualitative Analysis

Qualitative interviews were audio recorded and transcribed verbatim using NVivo 12 software (Lumivero). Following transcription, interviews were analyzed by KMKO and TO using conventional content analysis [55]. The researchers first read each interview multiple times to familiarize themselves with the data. Transcripts were then open-coded, codes were grouped into meaningful clusters, and these clusters were used to form overarching categories that related to social presence and therapeutic alliance. Categories were discussed, refined, and finalized in consultation with MC. Content analysis allows for a mix of deductive and inductive analysis [56]. Deductive analysis was used to code transcripts based on understandings of social presence and therapeutic alliance gained from previous literature, while inductive analysis was used to explore how social presence and therapeutic alliance developed and were important in the novel context of self-guided digital interventions. Representative quotes were chosen to supplement the written findings.

Results

Quantitative Results

A total of 9 (90%) participants completed all quantitative measures. Descriptive values for the quantitative questionnaires are presented in Table 2. Participants’ mean therapeutic alliance score on the WAI-I was far above the WAI-I scale floor, suggesting that in general, our participants did develop therapeutic alliance with the mindfulness guides. The mean score for the bond subscale of the WAI-I was particularly high, suggesting that participants especially developed a sense of bonding with their guides, relative to their sense of agreement with guides on tasks or goals in the intervention.

Participants’ mean social presence score on the PSP scale was far above the scale floor, suggesting that participants developed a social presence with their mindfulness guides. Participants’ average score on the SPI scale was slightly above its scale midpoint—with lower scores indicating greater social presence. This result indicates that participants had fairly neutral impressions of their mindfulness guides’ interpersonal characteristics and the social presence experienced with the guides.

Table 2. Descriptive values for questionnaire outcomes from adolescents recovering from concussion using a digital mindfulness-based intervention.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Minimum-Maximum</th>
<th>Mean (SD)</th>
<th>Scale midpoint</th>
<th>Scale range</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAI-I&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>38.00-59.00</td>
<td>47.00 (6.80)</td>
<td>24.00</td>
<td>12.00-60.00</td>
</tr>
<tr>
<td>Task and goal agreement</td>
<td>20.00-39.00</td>
<td>30.33 (6.34)</td>
<td>16.00</td>
<td>8.00-40.00</td>
</tr>
<tr>
<td>Bonding with guides</td>
<td>12.00-20.00</td>
<td>16.67 (2.55)</td>
<td>8.00</td>
<td>4.00-20.00</td>
</tr>
<tr>
<td>PSP&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>11.00-39.00</td>
<td>24.40 (7.55)</td>
<td>17.50</td>
<td>6.00-30.00</td>
</tr>
<tr>
<td>SPI&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score&lt;sup&gt;d&lt;/sup&gt;</td>
<td>19.00-58.00</td>
<td>35.89 (10.63)</td>
<td>34.00</td>
<td>17.00-85.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>WAI-I: Working Alliance Inventory for Guided Internet Interventions.  
<sup>b</sup>PSP: Perceived Social Presence scale.  
<sup>c</sup>SPI: Social Presence Indicators Scale.  
<sup>d</sup>A lower score indicates a more positive social presence.

Qualitative Results

Developing a “Genuine Connection”

When asked about their mindfulness guides, many participants readily described developing an important emotional bond with their guides. For example, Gabrielle said that their connection with the guides “didn’t feel artificial... it felt like it was a genuine connection.” Olivier expressed, “I like that they kind of got what the listener was feeling... they had pretty good insight into various different things that... I was feeling.” Sophia also indicated feeling a sense of safety with their guides, stating, “I don’t know how to describe it. [The connection is] like feeling safe with someone.” This was echoed by Ella, who stated that listening to the guides “was just kind of like a comfortable situation, and you didn’t really feel like there was anything wrong.”

The participants further elaborated on how this connection with the guides helped them progress through the mindfulness program. Alex noted, “It just makes you feel like they were trying to help you the whole time. Like they want you to accomplish your goals... you can definitely tell that they’re like guiding you through.” Olivier highlighted that the guides also normalized challenges that arose during the program, citing 1 example where the guides said, “If you’re thinking that you’re bad at mindfulness, it’s normal for your thoughts to wander and all that. [The guides] just kind of understood.” This experience was shared with Kay, who said, “I would lose track of what I was doing, and they’re like, if you lose track like it’s okay, just come back to what we’re doing, think about this.” And I’m like,
“Okay, thank you.” The use of certain adjectives when describing the guides, such as “understanding” (Ella, Julian, and Olivier), “caring” (Ada and Olivier), “helpful” (Alex and Sophia), and “friendly” (Alex and Kay), further exemplified the participants’ feeling of being taken care of by the guides, which facilitated their progression through the program.

**Sensing “Real People”**

The use of human-like adjectives to describe the mindfulness guides indicates that participants viewed the guides as real people. Indeed, participants elaborated on the perceived realness of the guides. For example, Gabrielle mentioned that Ruby “felt like... a real person, like really genuine.” Ada also contrasted the realness of the guides to the absence of realness in automated computer voices; they said, “This is a person and like, again, not an automated voice message... it was at one point another person on the other side of the screen, it’s not like Siri talking back to you.” One participant even stated, “I would be surprised if [you] told me they weren’t real people” (Kay). Some participants additionally assigned a particular identity role to their guides, saying that their guides felt like “teachers” (Olivier), or “a parental figure” (Gabrielle); beyond feeling like real people, the mindfulness guides also felt like particular kinds of people for some participants.

In addition to feeling like their guides were real people, participants also noted that their experiences with the guides felt like real-life interactions. For example, Ada offered that it “felt like someone was talking in my ear when I was like, lying down and things... it definitely felt like there was like somebody there.” Alex highlighted that “it feels like they’re actually talking to you.” Gabrielle said “It felt like they were like right in front of me, too. So like, presence.” Multiple participants noted that the interactions with the guides felt private and personal, with both Alex and Ella describing that their mindfulness sessions with the guides felt like a “one-on-one talk” and Olivier saying that sessions “just felt like more like a conversation that they’re having with me, and it just felt like as though they’re talking to me rather than just talking and letting someone hear.”

As participants elaborated on their guides, however, tensions in their accounts were palpable when describing their ability to develop feelings of presence and connection despite the guides not actually being present. Ella offered, “This is probably going to sound weird, but [they were] understanding even though you couldn’t talk to them. It kind of felt like they were listening.” When asked to what extent it felt like the guides were present during the intervention, Sophia answered, “I feel like they’re with me, but not with me at the same time, because I know they’re not with me.” Ada noted that contextual factors were relevant to whether they felt like their guides were present or not, stating, “I don’t think I felt [like the guides were talking in my ear] when I was walking, because when I’m walking, I’m like, nine times out of ten having to pay attention to my areas around me... But when I was lying down, it was definitely there.” Kay also stated, “I wasn’t connected, but I wasn’t not connected.” The dissonance in the participants’ narrative accounts alluded to the sense that participants could not help but feel a connection with their guides but also did not want to deceive themselves into thinking that these connections were real—or at least, quite as real as an in-person connection would be. Indeed, Sophia said that “it felt like I was on a phone call with someone,” suggesting that while their interactions with guides did not necessarily feel like they were happening in-person, their relationship with the guides still felt more interactive than the completely noninteractive connection they actually had.

**Personal Backgrounds and Voices**

The mindfulness guides’ backgrounds and tone of voice emerged as factors key to the development of participants’ bond and sense of presence with the guides.

Getting to know the guides allowed the participants to “trust them” (Olivier) by getting the sense that the guides “knew what they were talking about” (Olivier). Some participants trusted Brian due to his personal experience with concussion. As Alex noted, Brian’s experience of having concussions in the past made him seem very “personable,” as participants felt that it was not “just someone like talking to you about it, it’s someone who’s actually experienced [concussion]” (Alex). Others trusted Ruby because of her professional and educational background, for example, because she mentioned her degree in psychology. Ada noted, “Obviously [Ruby] knew about concussions. And I respected that.” The sharing of the personal or professional backgrounds of the guides allowed participants to feel like the guides had “either been through this (ie, a concussion experience) or they’ve worked with other people who’ve been through this,” (Kay) which allowed them to trust the trajectory and goals of the intervention.

Most participants also expressed that the guides’ voices were important for helping them get to know the guides and making them feel real. Ada captured this sentiment by sharing, “I definitely think that you just kind of get to know a person by the way that they speak.” The most common descriptors of the guides’ voices were “calm” or “calming” (Sophia, Ella, Olivier, Marie-Claude, Sophia, Ada, and Kay) and “soothing” (Julian, Olivier, and Ada). When probed about why this tone of voice was important, participants shared that a calming voice allowed them to feel like they were being cared for, and this had important implications for the development of trust in the guides. For example, Sophia elaborated, “I think them being calm with their voices made me closer to feel like they’re actually here to help me instead of being rushed and stuff.” Thus, participants highlighted that factors that allowed them to get to know their mindfulness guides (ie, personal backgrounds and tone of voice) were beneficial for the development of their therapeutic alliance and sense of social presence, which in turn had benefits for their engagement with the app.

**Discussion**

**Overview**

This study aimed to determine whether and how therapeutic alliance and social presence could develop in the context of a self-guided digital MBI for adolescents post concussion. Qualitative and quantitative findings revealed that adolescents did develop therapeutic alliance through their use of the intervention, with qualitative findings providing insight into the use of the intervention.
factors important for its development; participants readily described experiencing a bond with the mindfulness guides in the intervention. While participants’ average score on the SPI scale was slightly above the scale midpoint, this may be explained by nuances uncovered with qualitative findings, which highlighted the tension in how the participants experienced social presence with their guides (eg, it felt real, but not quite as real as in-person connections). Indeed, this tension aligns with Holter’s relational model [29], suggesting that individuals’ connections with their mindfulness guides in this digital context involved both “making come-alive” and “keeping un-alive.” This, along with participants’ high average score on the PSP scale, suggests that participants did develop social presence during their use of the intervention. The guides’ personal backgrounds and the tone of their voices emerged as particularly important factors in the development of therapeutic alliance and social presence in a digital context. Additionally, participants stated that their connection with the guides and sense of the guides’ realness helped maintain their engagement with and interest in the intervention. Results extend previous findings that adults can develop therapeutic alliance with self-guided digital interventions [24-31] by demonstrating that adolescents also develop therapeutic alliance in self-guided digital contexts. Overall, findings highlighted the potential feasibility and use of maximizing therapeutic alliance and social presence to bolster participant engagement with self-guided digital interventions.

Since the onset of the COVID-19 pandemic, clinicians have identified the need for accessible, engaging, and effective digital interventions [57,58]. Digital interventions circumvent a number of accessibility barriers that limit adolescents’ access to concussion care. Therefore, they represent a promising treatment option for both acute and persistent concussion symptoms. However, numerous studies have noted particularly high attrition rates for self-guided digital interventions, which researchers have suggested are because of their limited human contact and lack of social support [12,17,18]. Our findings demonstrate that even in self-guided digital interventions, participants can develop therapeutic alliance and a sense of social presence in the program with audio recordings of mindfulness guides. This suggests that, despite the absence of a human for participants to directly interact with, adolescents can still experience being socially supported through self-guided interventions. Thus, it is important to consider ways of maximizing the human support participants feel during self-guided digital interventions to bolster participant engagement.

In self-guided contexts, factors that anthropomorphize digital guides—in other words, make them feel more human—emerged as particularly important to the development of therapeutic alliance and social presence among our sample. Participants highlighted that learning about the guides’ backgrounds (either their experiences of concussions or their background in psychology) helped them to feel connected to guides and trust that the content of the intervention would help them work toward their goals—both key aspects of therapeutic alliance. The guides’ voices also emerged as a salient factor that made guides feel like real people to participants, demonstrating that the “realness” of guides’ voices was key to the development of social presence. Indeed, when asked whether guides’ voices sounded like those of real people, every participant except 1 (Sophia) said yes. Thus, factors that increased participants’ ability to anthropomorphize their guides stood out as a potential mechanism important to the development of both therapeutic alliance and social presence in self-guided digital interventions. Our data align with quantitative findings that individuals experience greater social presence when digital conversational agents are characterized by higher levels of anthropomorphic cues (eg, human-like voices) [59]. Potentially, a higher level of anthropomorphic cues in self-guided digital interventions better enables individuals to “make come-alive” [28,29].

**Recommendations**

Our findings have several implications for researchers developing self-guided digital interventions for mental health concerns. Given that anthropomorphizing guides emerged as particularly important to the development of both therapeutic alliance and social presence in our sample, we recommend multiple strategies to increase the human-like qualities of audio-recorded guides in self-guided digital interventions. First, providing information about the guides’ personal connection with the topic of focus of the intervention (eg, concussion knowledge in this study) will promote how relatable and trustworthy guides seem to participants. Our findings suggest that guides either having personal experience with the topic of the intervention or having a background in psychology (and thus, seeming knowledgeable to participants about the intervention) may both be promising ways for mindfulness guides to encourage connection with participants and seem more real. Second, using real human actors to record audio for guides’ voices, as was done in our intervention, will seem more realistic and promote a greater bond with guides compared to using computerized voices. Guides may benefit from training in clinical microskills (eg, conveying warmth, genuineness, and empathy) [60] to ensure that their tone of voice is calming and soothing, both of which were identified by our participants as important attributes in guides’ voices for therapeutic alliance and social presence. Additionally, guides can facilitate participants’ sense of connection with them by anticipating and addressing potential challenges that may arise for participants. Participants in our sample reported that this helped them feel understood by the guides. Lastly, 1 key addition participants in our study suggested to increase their sense of connection with and humanness of guides was to show pictures of the guides’ faces. Indeed, previous research suggests that elements of a digital interface’s design—for example, providing identity cues like names and photos of the communicators within the interface—increase social presence [38,61]. Therefore, having an image of the voice actors who are the mindfulness guides may better allow participants to develop a narrative of who their guides are, impacting both the connection they feel to their guides (therapeutic alliance) and the extent to which the guides feel like real people (social presence).

**Limitations and Future Directions**

Given our small sample size, our findings are limited with respect to the quantitative analyses we were able to conduct. While it was a secondary focus of our study, having a larger sample may have allowed us to conduct formal quantitative
analyses as opposed to the descriptive analyses presented in this paper. Future research can use quantitative methods with a larger sample to formally explore the independent and interactive roles that therapeutic alliance and social presence play in the development of self-guided digital interventions.

Future research may also benefit from building upon this study’s results and previous research on adults [28] by exploring whether individual differences impact the extent to which adolescents experience therapeutic alliance and social presence during self-guided digital interventions and whether they benefit from experiencing them. Adolescents may differ in how much they desire and benefit from human support in digital interventions. For example, therapeutic alliance and social presence may be less important to the engagement of participants who are more introverted and potentially desire human support in therapeutic contexts to a lesser extent. Research can also examine whether sharing characteristics with guides (eg, race or ethnicity and gender) or not impacts how much participants develop therapeutic alliance and social presence with their guides. Potentially, participants may relate to guides who share their characteristics more, thus developing more of a connection with them.

Conclusions
This study found that adolescents with concussion develop therapeutic alliance and social presence in self-guided digital interventions and identified factors key to their development (eg, tone of voice and personal backgrounds). Our study is one of the first to examine the role of social presence in a clinical context with adolescents and demonstrate the use of social presence for encouraging adolescent engagement with self-guided digital interventions. Future research should focus on maximizing therapeutic alliance and social presence for digital intervention development by increasing the number of human-like qualities that guide the intervention process. This may address key concerns about self-guided digital interventions—high attrition rates and low participant engagement—to maximize their effectiveness. Doing so will allow for the development of digital interventions that are more accessible and less resource-intensive than in-person treatments, as well as being engaging for users. Ultimately, understanding the development of therapeutic alliance and social presence within self-guided digital interventions aids in the development of accessible and engaging digital treatments—one digital connection at a time.

Acknowledgments
We would like to thank the research teams and clinicians at the University of British Columbia, the GF Strong Rehabilitation Centre Adolescent Complex Concussion Clinic, and the Children’s Hospital of Eastern Ontario Research Institute for their dedication to the project. We thank Dr Jonathan Greenberg and Dr Ana-Maria Vranceanu for reviewing and providing feedback on the intervention.

Data Availability
The data analyzed during this study are not publicly available in order to protect the privacy and confidentiality of participants, given the small sample size. However, data are available from the corresponding author on reasonable request.

Conflicts of Interest
AAL reports being one of the authors of the mindfulness-based intervention for concussion protocol used in this study; she did not receive any financial benefit. AAL received funding from the Ontario Brain Institute Neurotech Early Research and Development (NERD) program for a separate component of the study. AAL has no other conflict of interest to disclose. MC reports being the CEO and chief scientist of Mobio Interactive and a major shareholder, owning approximately 23% of the company at the time of manuscript submission. RZ reports being an investigator on competitively funded research grants from the Canadian Institutes of Health Research, the Ontario Neurotrauma Foundation, the Physician Services Incorporated Foundation, the Children’s Hospital of Eastern Ontario (CHEO) Foundation, the Ontario Brain Institute, the Ontario SPOR Support Unit, the Ontario Ministry of Health, the Public Health Agency of Canada, Health Canada, and the National Football League Scientific Advisory Board. RZ is a clinical research chair in pediatric concussion from the University of Ottawa. RZ is the scientific director and minority shareholder of 360 Concussion Care (a learning health system and network of interdisciplinary concussion clinics in Ontario). VS receives royalties for the sales of a concussion book (Commotions Cérébrales) by Flammarion Québec. NDS has received research operating funds from multiple granting agencies (Canada Foundation for Innovation, Canadian Institutes of Health Research, Mitacs, Ontario Brain Institute, US Department of Defense, WorkSafeBC, VGH+UBC Hospital Foundation) for research related to traumatic brain injury diagnosis, prognosis, and treatment. He has received speaker fees for providing continuing medical education on these topics. He serves as chair of the American Congress of Rehabilitation Medicine’s Brain Injury Special Interest Group Task Force on Mild TBI (unpaid). He has served as an expert panel member for the Living Concussion Guidelines and as an external reviewer for other clinical practice guidelines on concussion/trumatic brain injury (unpaid). He has provided expert testimony and medical-legal consulting in the past 5 years (<10% of total income). The remaining authors have no conflicts of interest to report.
Multimedia Appendix 1
Images of intervention content.
[DOCX File, 434 KB - formative_v8i1e49133_app1.docx]

Multimedia Appendix 2
Texting protocol.
[DOCX File, 30 KB - formative_v8i1e49133_app2.docx]

Multimedia Appendix 3
Semistructured interview.
[DOCX File, 23 KB - formative_v8i1e49133_app3.docx]

References
11. Table 22-10-0115-01 smartphone use and smartphone habits by gender and age group, inactive. Statistics Canada. Ottawa, ON URL: https://doi.org/10.25318/2210011501-eng [accessed 2024-02-14]


Abbreviations

JARS-Qual: Journal Article Reporting Standards for Qualitative Research

MBI: mindfulness-based intervention

PPCS: persistent postconcussion symptoms

PSP: Perceived Social Presence

REDCap: Research Electronic Data Capture

SPI: Social Presence Indicators
WAI-I: Working Alliance Inventory adapted for Guided Internet Interventions

©Kiarah M K O’Kane, Thalia Otamendi, Noah D Silverberg, Esther Choi, Veronik Sicard, Roger Zemek, Katherine Healey, Olivier Brown, Lauren Butterfield, Andra Smith, Gary Goldfield, Rachel Kardish, Bechara J Saab, Andrée-Anne Ledoux, Molly Cairncross. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
**Low-Fidelity Prototype of a Sensor-Dependent Interaction Platform: Formative Evaluation With Informal Caregivers of Older Adults With Cognitive Impairment**

Nikita Sharma¹,², PhD; Karen Grotenhuijs¹, MSc; J E W C van Gemert-Pijnen¹, PhD; Harri Oinas-Kukkonen², PhD; L M A Braakman-Jansen¹, PhD

¹Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, Netherlands
²Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

**Abstract**

**Background:** Unobtrusive sensing technologies developed for monitoring deviant behaviors in older adult care require integration with an interaction platform to facilitate the flow of information between older adults and their caregivers. However, the continuous monitoring capabilities generate a considerable amount of data that must be interpreted, filtered, and personalized before being communicated to the informal caregivers based on their specific care needs and requirements.

**Objective:** For the effective implementation of unobtrusive sensing solutions (USSs) in the care of older adults with cognitive impairment, we aimed to explore the expectations and preconditions regarding the implementation of USSs from the perspective of informal caregivers. Subsequently, we designed and evaluated a low-fidelity prototype of an interaction platform for its conceptual workflow and usability, incorporating persuasive system design features based on the needs and requirements of informal caregivers.

**Methods:** Overall, 6 informal caregivers of older adults with cognitive impairment living alone participated in this qualitative interview study. We explored the expectation and preconditions regarding implementation through open-ended questions and conducted a formative evaluation (usability study with a think-aloud approach) to evaluate the conceptual workflow and used persuasive system design features in the interaction platform. Overall, a combination of inductive and thematic analyses was used to analyze the interviews.

**Results:** The results of this study present both positive and negative outcome expectations regarding the implementation of USSs, highlighting benefits such as objective decision-making and peace of mind and concerns about information overload and the potential substitution of human contact. Strategic information communication agreements between informal and formal caregivers were deemed crucial for the successful implementation of USSs in care. Overall, informal caregivers had a positive experience with the low-fidelity prototype of the interaction platform, particularly valuing the personalization feature.

**Conclusions:** In conclusion, to achieve successful implementation, a holistic design approach is necessary, and equal consideration should be given to the personalization-privacy paradox to balance users' needs and privacy.

*(JMIR Form Res 2024;8:e53402) doi:10.2196/53402*

**KEYWORDS**

older adult care; informal caregivers; cognitive impairment; sensing solutions; information communication platform; low-fidelity; lo-fi prototype
Introduction

Background

The increase in the older adult population (≥65 years) imposes significant challenges on the organization and functioning of the current health care infrastructure worldwide [1]. It demands the active involvement of different stakeholders including informal caregivers, formal caregivers, general practitioners, technology developers, policy makers, and government organizations to maintain continuous care [2]. Primarily, informal caregivers are perceived as responsible for organizing and ensuring on-time care for older adults, which impacts their physical, financial, emotional, and social well-being [3,4]. Moreover, with the emergence of cognitive impairment or comorbidities, the care process becomes more complex and challenging for informal caregivers [5].

To support informal caregivers in delivering on-time care, sensor-based solutions, specifically those that are unobtrusive or device free (ie, do not demand direct involvement or attention from older adults), are being developed [6]. Studies have shown that unobtrusive sensing solutions (USSs) are in demand and appear to be useful among informal caregivers of older adults with cognitive impairment due to their 24/7 monitoring capabilities, which provide real-time insights into the health of care recipients [2,6]. A USS comprises 3 main units: a sensing unit responsible for collecting data from the care recipient, a computing unit responsible for analyzing the obtained sensing data, and a communicating unit that communicates the output of the computing unit to the informal caregivers to enable monitoring at a distance [7].

Over the past decade, there have been notable advancements and successful endeavors to facilitate the development of unobtrusive and ubiquitous sensing technology [8]. For example, Wi-Fi channel state information (CSI; as a sensing unit) can be used for monitoring physical activities (falls, sitting, hand gestures, etc), physiological activities (heart rate and breathing rate), and behaviors (sleeping patterns, personal hygiene, etc) [9-11]. Significant growth in the computing unit through the use of advanced machine learning methods (such as deep neural networks, generative adversarial networks, etc) to improve privacy, reliability (minimizing false alarms), and computing time is also evident [12,13]. However, efforts from the IT domain (communication unit), particularly in the direction of developing and designing interaction platforms adhering to the information communication (IC) needs and requirements of informal caregivers (or other stakeholders), are limited [14].

Designing an interaction platform according to the preferences of informal caregivers can assist in prioritizing and optimizing their care plans, thereby reducing the care (information) load [7].

In our previous multimethod study encompassing a survey and interviews with informal caregivers of older adults with cognitive impairment, diverse needs and requirements for the following 4 distinct care scenarios were explored: falls, nocturnal unrest, agitation, and normal daily life [7]. The findings indicated varying information needs regarding the mode, content, timing, and stakeholders involved, contingent upon the care scenario. In addition, these needs were observed to be influenced by the personal circumstances of caregivers and care recipients and the progression of illness in care recipients. Furthermore, to facilitate the designing of an interaction platform (ie, this study) persuasive system design (PSD) features, namely reduction, tailoring, personalization, reminders, suggestions, trustworthiness, and social learning, were identified for the involved care scenarios. One of the limitations that we observed was the lack of proper understanding of USSs among informal caregivers. Due to the technical novelty of the solution, informal caregivers perceived it as a black box, potentially introducing biases in their responses regarding its usefulness and expectations.

Objective

Building upon the findings from our previous study [7], the objectives of this study were two-fold: (1) to explore the expectations (positive and negative) and preconditions regarding implementing USSs in the care of older adults with cognitive impairment from the perspective of informal caregivers after showing them a video prototype of the solution and (2) to design and evaluate a clickable, low-fidelity (lo-fi) prototype of a sensor-dependent interaction platform, incorporating the identified PSD features regarding fall, agitation, and normal daily life care scenarios with informal caregivers.

Methods

Ethical Considerations

The Ethics Committee of the Behavioral, Management, and Social Sciences department at the University of Twente formally approved the execution of this study (request number 230141). Before engaging in the surveys and interviews, participants received both oral and written description elucidating the study’s objectives, methodologies, data collection procedures, storage protocols, and the intended use of the collected data. Subsequently, each participant provided a signed consent form. The participants were also free to withdraw from the study at any stage if they felt uncomfortable.

Finally, the participants were offered a small honorarium as a token of appreciation for their valuable time and contributions.

Study Design: Participatory Development

The study followed the Center for eHealth Research and Disease Management (CeHRes) road map to create a sensor-dependent interaction platform that can communicate the information obtained by USSs to informal caregivers of older adults with cognitive impairment [15]. The CeHRes road map fosters progress toward context-aware sensing and computing by offering early feedback regarding users’ needs and requirements to the designers and developers [15]. For instance, if informal caregivers prefer insights into emergencies only, the algorithm can be trained and optimized accordingly to provide relevant data, avoiding computing overload for the system and information overload for the caregivers. The framework encompasses 5 distinct but intertwined phases: contextual inquiry, value specification, design, operationalization, and summative evaluation (Figure 1 [7]). The description of these
phases, along with their relevance to this study, is provided in **Textbox 1**.

**Figure 1.** The Center for eHealth Research and Disease Management framework. IC: information communication; lo-fi: low-fidelity; USS: unobtrusive sensing solution. *The study by Sharma et al [7].

**Textbox 1.** Description of Center for eHealth Research and Disease Management (CeHRes) phases along with their relevance this study.

<table>
<thead>
<tr>
<th>CeHRes phases and description of relevance to this study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contextual inquiry</strong></td>
</tr>
<tr>
<td>• This phase helped in building an understanding of the prospective users (informal caregivers) and their context (care of older adults with cognitive impairment)</td>
</tr>
<tr>
<td>• The understanding of experiences, expectations, and usefulness of unobtrusive sensing solutions (USSs) among informal caregivers of older adults with cognitive impairment were obtained from a previous study [7]</td>
</tr>
<tr>
<td>• To further advance the findings, we dwelled deeper into informal caregivers’ expectations and preconditions regarding the implementation of USSs after providing them with more detailed information about the functioning and potential benefits of using USSs in the care of older adults with cognitive impairment</td>
</tr>
<tr>
<td><strong>Value specification</strong></td>
</tr>
<tr>
<td>• This phase helped in identifying the needs and values that are important for the intended stakeholders (informal caregivers), which can be translated into the requirements later [15]</td>
</tr>
<tr>
<td>• Information communication design requirements regarding different care scenarios (fall, agitation, and normal daily life) obtained in a previous mixed methods study [7] were used to develop the interaction platform</td>
</tr>
<tr>
<td><strong>Design</strong></td>
</tr>
<tr>
<td>• The primary focus of this study lied within the design phase that involves the agile development and testing of the interaction platform</td>
</tr>
<tr>
<td>• A low-fidelity prototype was developed by using the rapid prototyping technique by the involved researchers’ team [16,17]</td>
</tr>
<tr>
<td>• The prototype was subjected to formative evaluation with informal caregivers, with specific emphasis on evaluating the conceptual workflow and used persuasive system design features</td>
</tr>
<tr>
<td>• A task-based study design [18,19] in conjunction with a think-aloud approach [20] was used for this study</td>
</tr>
<tr>
<td><strong>Operationalization and summative evaluation</strong></td>
</tr>
<tr>
<td>• These phases will be activated when the technology is launched into the market. The results obtained from this study can serve as valuable inputs for these phases, ensuring that the final product meets the needs and requirements of informal caregivers of older adults with cognitive impairment</td>
</tr>
</tbody>
</table>
Participants

Participants were recruited from an already existing pool of candidates who had previously been involved in a similar study [7]. While previous experience with digital care technology was not a prerequisite for informal caregivers, all of them were users of the Caren Platform (a digital care platform; Caren [NEDAP]), implying that they had default experience with digital care technology [21]. Informal caregivers were approached for participation in this study via email. The participants received an invitation, along with an information letter describing the study’s purposes, procedures, and the researcher’s contact information. When an informal caregiver expressed willingness to participate, they were screened based on the following inclusion criteria: (1) providing unpaid care to a person with cognitive impairment who is a relative, friend, or someone else within their personal circle; (2) the person with cognitive impairment is aged ≥65 years; and (3) the person with cognitive impairment lives alone at home. Subsequently, an appointment for the evaluation session was scheduled with the researcher.

Materials: Designing the Lo-Fi Prototype of the Interaction Platform

Video Prototype

Given the novelty of USSs, a lack of awareness among informal caregivers about their working and implementation was observed. Therefore, to educate informal caregivers, a video prototype demonstrating the working, system architecture, and benefits of USSs in the care of older adults with cognitive impairment was created. While the video was largely inspired by previous studies that used Wi-Fi CSI as a technology in USSs for recognizing older adult activities [11], some brainstorming sessions with the research team (composed of eHealth researchers, experts, technology developers, and designers) also occurred to align it with the use case of care of older adults with cognitive impairment.

Overall, the video depicted 3 units of USSs, namely, sensing, computing, and communicating units. The “sensing unit” of the solution showed the working (how) and the manner of data collection (what) through Wi-Fi CSI (as an unobtrusive sensing technology). The “computing unit” of the solution presented the use of artificial intelligence (AI) algorithms for analyzing the collected data. In the “communicating unit” of the solution, the channel for communicating the computed information to the caregivers was presented. The video provides examples of 3 different care scenarios, namely fall incident, agitated behavior, nocturnal unrest, and normal daily life (drinking activity).

To make the video footage realistic, it was recorded in the eHealth house at the University of Twente, the Netherlands [22]. The video had a Dutch voice-over with English subtitles, given that informal caregivers were comfortable with Dutch. The video has a total duration of 3.5 minutes. It was presented to the participants at the start of the interview sessions to ensure that they had the necessary information to answer the questions posed in the interview, thereby promoting more informed responses. Figure 2 shows a simplified overview of the system architecture (as conveyed in the video) of the intended USSs.

Figure 2. Simplified overview of the system architecture of unobtrusive sensing solutions.

![Simplified overview of the system architecture of unobtrusive sensing solutions.](image-url)
Lo-Fi Prototype

Overview

We designed the lo-fi prototype of the interaction platform using Figma software [22]. As the Wi-Fi CSI system is in the early development phase (Technology Readiness Level 2/4) [7, 23], a lo-fi prototype was chosen to gain initial insights from informal caregivers about the communication unit, showcasing the conceptual workflow and main functionalities of the interaction platform. Note that this interaction platform does not intend to change the behavior of the informal caregivers but requires persuasion to form (ie, F outcome) or alter (ie, A outcome) the behavior of informal caregivers for complying (ie, C change) with the information communicated (outcome: F and A; change: C) [24].

The findings from previous studies [2, 7, 25] were used to design the lo-fi prototype of the interaction platform. The studies by Wrede et al [2, 25] demonstrate the value of USSs in continuous and objective monitoring, leading to timely interventions. In particular, informal caregivers found USSs to be helpful in clearly classifying the care scenarios as urgent, nonurgent, and future risk. Further exploration in a multimethod study by Sharma et al [7] (comprising a survey [n=464] and interviews [n=10]) revealed divergent IC needs in different care scenarios (fall, nocturnal unrest, agitation, and normal daily), including mode, content, timing, intended users, feedback to the system for self-learning, and dialogue support. Furthermore, the study also identified 7 PSD features: 3 from primary task support (reduction, tailoring, and personalization), 2 from dialogue support (reminders and suggestions), 1 from system credibility (trustworthiness), and 1 from social support (social learning) for designing the interaction platform. On the basis of these findings, the conceptual workflow of the interaction platform and user interfaces for 3 care scenarios, namely, fall incident, agitated behavior, and normal daily life activities were designed. In addition to these features, a system verifiability feature was added to assess its necessity or impact on the interaction platform [26]. Table 1 presents the used PSD features, their interpreted meaning, and their application in the lo-fi prototype. The following sections provide details regarding the design of the conceptual workflow and user interfaces.

Table 1. Persuasive system design (PSD) features used to design the low-fidelity (lo-fi) prototype.

<table>
<thead>
<tr>
<th>PSD category, feature, and meaning</th>
<th>Application in the lo-fi prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary task support</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Personalization</strong></td>
<td>Providing personalized content</td>
</tr>
<tr>
<td><strong>Reduction</strong></td>
<td>Reducing complex tasks into smaller tasks</td>
</tr>
<tr>
<td><strong>Tailoring</strong></td>
<td>Providing information tailored to the user’s needs</td>
</tr>
<tr>
<td><strong>Dialogue support</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Reminder</strong></td>
<td>Reminding users about the target behavior</td>
</tr>
<tr>
<td><strong>Suggestion</strong></td>
<td>Offering suggestions to facilitate behavior</td>
</tr>
<tr>
<td><strong>Social support</strong></td>
<td>Learning from the experiences and behavior of others</td>
</tr>
<tr>
<td><strong>Trustworthiness</strong></td>
<td>Providing reliable information</td>
</tr>
<tr>
<td><strong>Verifiability</strong></td>
<td>Providing evidence to validate the accuracy</td>
</tr>
<tr>
<td><strong>Social learning</strong></td>
<td>Page for sharing experiences where the informal caregivers can read and react to the experiences of others</td>
</tr>
<tr>
<td><strong>System credibility</strong></td>
<td>Reliability percentage indicator and provision to provide feedback to the system</td>
</tr>
<tr>
<td><strong>Caregiver support and communication page that helps informal caregivers discuss the care plans and provides an option to verify the system’s predictions</strong></td>
<td></td>
</tr>
</tbody>
</table>

IC: information communication.
Design of the Conceptual Workflow

The conceptual workflow of the interaction platform is illustrated in Figure 3. This workflow reflects the logical flow of the interaction platform when personalizing the IC options. It starts from the log-in page, followed by choosing the preferred activities for monitoring; adjusting the communication preferences for the chosen activities; and the home screen, where multiple functionalities of the interaction platform can be checked or adjusted. Note that the feature of choosing the activities to be monitored and adjusting the preference is attributed to the personalization feature of the PSD model.

Figure 3. Conceptual workflow of the interaction platform.

Design of the User Interfaces

User interfaces for fall, agitation, normal daily life, and home screen were designed. As falling is an emergency scenario, informal caregivers expect to receive a direct call (reduction feature), and if they do not respond within 5 minutes, they expect a reminder notification (reminder feature) in their preferred content style (raw, interpreted, or suggestions). Furthermore, the details of the fall incident such as time, location, system’s confidence in prediction, and current state of the care recipient were made accessible to the informal caregivers. In addition, to support informal caregivers, options to obtain suggestions from the system regarding what to do and when to act (suggestion feature) and to directly communicate with formal caregivers were also provided. Finally, as informal caregivers desire a trustable system (with minimum false alarms), an option to provide feedback to the system about its predictions to enable self-learning was also added (trustworthiness feature). Figure 4 illustrates the interfaces for the fall scenario.
On the other hand, as agitation is an acute scenario, informal caregivers expect the system to monitor it for a few weeks and share a report tailored to the concerned stakeholders, that is, themselves or formal caregivers \((\text{tailoring feature})\). Interfaces depicting notification (in the preferred content style), details about agitation behavior (duration, system’s confidence in prediction, and other observations), suggestions from the system \((\text{suggestion feature})\), the possibility to share the report with formal caregivers, and an option to provide feedback to the system for its predictions were designed \((\text{trustworthiness feature})\). Figure 5 illustrates the interfaces for the agitation scenario. The user interface for normal daily life (Figure 3) presented multiple self-care activities (eating, drinking, showering, etc). The informal caregivers can adjust their preferences for the content, frequency, and depth of information they want to receive regarding the selected activity.

In general, informal caregivers demand a centralized care approach, that is, the possibility to access all the relevant care information such as general information, medical information, and communications with other stakeholders in 1 platform [7]. Thus, consistent with this requirement, the home screen contained the following functionalities: observing the present and past situation of the care recipient \((\text{reduction})\), obtaining more detailed information or reports about daily activities, general and medical information about the care recipient \((\text{verifiability})\), an overview of the involved formal and informal caregivers, communication options with the involved formal caregivers, and system credibility (Figure 6). In addition, an option to read the care experiences shared by other caregivers as a part of the social learning feature of the PSD model was added. Finally, options for app settings (adjusting preferences) and information about the organization or team developing the app to show system credibility \((\text{real-world feel})\) were also added.
Figure 5. Interfaces for agitation scenario.
Procedure

The semistructured interviews were conducted via Microsoft Teams with 33% (2/6) of the informal caregivers and in person with 67% (4/6) of the caregivers, based on their preference. The interview guide (Multimedia Appendix 1) was used and consisted of the following sections: (1) introduction (goals, procedures, and informed consent), (2) background information obtained from the participant, (3) video prototype to explore the expectations (positive and negative) and preconditions, (4) formative evaluation of user interfaces, and (5) closing remarks. Upon watching the video, participants were asked whether they had any further questions regarding the systems, and the questions were clarified. This video and explanation were important because, due to the novelty of USSs, informal caregivers are not very aware of this concept or type of sensing solution. Then, their expectations and preconditions regarding implementation were discussed in depth.
Thereafter, a formative evaluation (using a task-based and think-aloud approach) of the designed interaction platform was conducted using 5 tasks (Table 2). In task 1, the informal caregivers were asked to choose the emergency or acute situations of their care recipient that they want to monitor, followed by adjusting the IC preferences for the chosen activities. Similarly, in task 2, the informal caregivers were asked to choose and adjust their IC preferences for the daily life (self-care) activities of the care recipient that they want to monitor in the long term. Here, the use of PSD feature personalization was evaluated. Furthermore, for tasks 3 and 4, a possible sequence of actions in the fall and agitation scenarios was evaluated. Specifically, the ability of the platform to immediately call or notify informal caregivers (reduction), send reminder notifications in case they do not respond (reminder), provide suggestions to support the informal caregivers (suggestion), and maintain a transparent link between the system and caregiver by providing the prediction percentage (trustworthiness) were assessed.

Table 2. Tasks used to evaluate the conceptual workflow and persuasive system design features of the interaction platform.

<table>
<thead>
<tr>
<th>Task description</th>
<th>Feature added and evaluated</th>
<th>Value associated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1: Choose emergency activities and adjust preferences for the chosen activities</td>
<td>Personalization</td>
<td>Every care scenario is different, and thus, informal caregivers should be able to choose which activity they want to monitor (for both emergency and daily life). Furthermore, they should also be able to adjust the preference regarding IC for the chosen activities. Informal caregivers need the flexibility to select and monitor specific activities based on the care scenario. They should also have the option to customize their preferences for IC related to the chosen activities [7,25].</td>
</tr>
<tr>
<td>Task 2: Choose self-care activities and adjust preferences for the chosen activities</td>
<td>Reduction, Reminder, Suggestion, Trustworthiness</td>
<td>During emergencies, informal caregivers expect the following: direct calls or reminders if they are unable to answer, trustworthy and accurate information, and suggestions to ensure timely and appropriate actions [7].</td>
</tr>
<tr>
<td>Task 3: Suppose a fall incident occurred in the home of your care recipient</td>
<td>Reduction, Reminder, Suggestion, Trustworthiness</td>
<td>During acute scenarios such as agitation, informal caregivers expect the following: notification and long-term reports that can be shared with formal caregivers and trustworthy and accurate information along with suggestions to support the care recipient in the right manner [7].</td>
</tr>
<tr>
<td>Task 4: Suppose your care recipient is experiencing agitation</td>
<td>Reduction, Tailoring, Suggestion, Trustworthiness</td>
<td>Informal caregivers desire a centralized care platform, where they can find important care elements immediately, eg, quick or detailed overview of the activities, access to medical records, and connection with caregivers [7].</td>
</tr>
<tr>
<td>Task 5: Explore the features of the home screen</td>
<td>Reduction, Verifiability, Social learning, System credibility</td>
<td></td>
</tr>
</tbody>
</table>

aIC: information communication.

Finally, in task 5, the informal caregivers were asked to explore the home screen to check whether it satisfies their requirement of a centralized care platform and present evidence to validate the provided information (verifiability). They were encouraged to identify, suggest, and reason the functionalities that help them in improving the caregiving process. The sessions were conducted in Dutch by a native Dutch speaker and were audio recorded to facilitate analysis. The duration of each session was approximately 60 minutes. On the basis of the feedback obtained from the first 4 sessions, slight improvements in the design were made and further evaluated in the last 2 sessions.

Data Analysis

The recordings were transcribed verbatim by using the description software, Amberscript. Qualitative analysis was performed by using Atlas.ti [27]. A thematic analysis was performed, based on the six steps by Braun and Clarke [28]: (1) familiarizing with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report. The transcripts were coded using a mixed inductive and deductive approach. An inductive approach was used for exploring the expectations and preconditions, whereas an inductive-deductive approach was used for analyzing the experiences with PSD features. All the transcripts were read by both researchers, NS and KG, in English and Dutch, respectively. Overall, the joint probability of agreement was 75%, followed by an in-depth discussion until consensus was reached regarding all the defined themes.

Results

This section is divided into 2 parts: first, the results related to the expectations and preconditions regarding implementation are presented, and second, the results corresponding to the formative evaluation of the lo-fi prototype including PSD features are presented.

Demographics

A total of 6 informal caregivers (mean age 58.7, SD 2.87 years) of older adults (mean age 85.7, SD 4.18 years) living alone participated in the study. Among the 6 participants, 4 (67%) were women and 2 (33%) were men. All informal caregivers (6/6, 100%) were children of the care recipient and were the primary informal caregiver. All care recipients (6/6, 100%) were living alone. Half of the care recipients (3/6, 50%) had Alzheimer disease, and the other half (3/6, 50%) had cognitive impairment due to other causes or no official dementia diagnosis. All informal caregivers (6/6, 100%) have used
technology previously in the care provision, differing from using communication platforms (Caren platform) or medication dispensers to personal alarm and monitoring systems. They have been providing care for at least the past 2 years, and their care hours ranged from 1.5 to 15 hours. Table 3 provides an overview of the characteristics of the interview participants.

Table 3. Sociodemographic characteristics of informal caregivers.

<table>
<thead>
<tr>
<th>Participant number</th>
<th>Sex</th>
<th>Age of the informal caregiver (years)</th>
<th>Age of the care recipient (years)</th>
<th>Experience in providing informal care (years)</th>
<th>Time spent on informal care (hours/week)</th>
<th>Distance from or time needed to reach the care recipient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>61</td>
<td>87</td>
<td>2</td>
<td>5</td>
<td>30 minutes</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>59</td>
<td>88</td>
<td>2-3</td>
<td>1.5</td>
<td>2 hours</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>60</td>
<td>86</td>
<td>3</td>
<td>8</td>
<td>Next door</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>53</td>
<td>79</td>
<td>2</td>
<td>10-15</td>
<td>1.5-2 km</td>
</tr>
<tr>
<td>5</td>
<td>Male</td>
<td>60</td>
<td>91</td>
<td>20</td>
<td>12</td>
<td>45 minutes</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>59</td>
<td>83</td>
<td>5</td>
<td>2</td>
<td>20 minutes</td>
</tr>
</tbody>
</table>

Expectations and Preconditions

The expectations and preconditions regarding implementation are presented as three broad themes: (1) positive outcome expectations, (2) negative outcome expectations, and (3) preconditions for implementation.

Positive Outcome Expectations

Objective Decision-Making

The informal caregivers indicated that USSs could contribute toward making objective decisions regarding the care of their loved ones. Instead of relying solely on observations of both informal and formal caregivers or on what the care recipient mentions themselves, the system can provide the involved informal and professional caregivers with more objective and in-depth monitoring information. According to informal caregivers, this information not only enables prompt diagnosis of underlying health conditions but also facilitates objective communication among the professional caregivers, care recipient, and themselves. This fosters a shared understanding of the care situation, consensus about the provision of care, and better coordination regarding the response to the (emergency) situations:

- **Participant 2**
  
  It provides the facts, so what she herself isn’t mentioning yet, but what is actually already there, that could be beneficial to support her, to make better choices and to better understand what is going on.

- **Participant 3**
  
  It provides monitoring information, for example, we are now at a stalemate with my father; he should have more help and we need to request that, but he doesn’t want that because he believes he can still manage. It [the monitoring data] can prove that we are right, but it can also prove if he is right. If he is right, then we’ll have some peace for a while, so it indicates such things.

- **Participant 4**
  
  Safer Environments for Independent Living

Informal caregivers expected USSs to contribute to the feeling of safety of their loved one; they believe that the system will notice when a safety risk occurs. Moreover, it was mentioned that the system could give insight into whether it is safe for older adults to live independently at home. All informal caregivers (6/6, 100%) mentioned being interested in receiving information about safety matters, for instance, if the door has been opened or if the gas is on:

- **Participant 1**
  
  She [mother of informal caregiver] will feel safer. Her desire is to continue living at home for as long as possible, but she has concerns about it, like: “yes, I am alone and if something happens to me, well, what should I do then?” And this is a system that detects it [a fall] without her having to do anything. So, if she feels safer, she will also feel calmer, which has an impact on her dementia symptoms.

Providing Peace of Mind to Informal Caregivers

Some informal caregivers indicated that they expect USSs to contribute to their peace of mind and probably to the peace of mind of their care recipient also. They find it reassuring that the system acts as a safety net and alerts them or the care professionals when there is an emergency situation. Furthermore, they indicated that the system could confirm the well-being of their loved one, whereas without such a system, there would be uncertainty and doubt about the situation, and they might be unnecessarily worried about their loved one:

- **Participant 2**
  
  It brings peace of mind. It provides, like, you can’t fully rely on the technology, but knowing that you have an additional safety net, that you are a bit more at ease, and also for the person involved it helps.

- **Participant 3**
  
  I only see reassurance, you know, you receive, you know that everything is fine, but you receive a confirmation that it is indeed going well.

Stimulating Meaningful Conversations

A few informal caregivers indicated that if USSs can collect care-related information, they might be able to spend more meaningful time (personal conversations) with their loved ones. This is because the care component is important and requires a lot of attention, and they overlook the personal or relational aspect, thus impacting their relationship with the care recipient:

- **Participant 1**
  
  Because it’s not constantly asking “how are you doing,” there is an additional aspect behind it. Yes,
you still have to keep asking, but it’s more about showing interest in the person rather than focusing solely on the care component. So, I think there’s more room for the human aspect rather than just the caregiving aspect. [Participant 3]

It can help in relational aspect, I would really appreciate that, because I miss the conversations with my mother; there is always that caregiving component that comes in-between. [Participant 3]

Negative Outcome Expectations

Information Overload

A few informal caregivers also expressed concerns about the possibility of information overload from USSs. They mentioned that the continuous availability of information about the care recipient, enabled by USSs, might lead them to constantly check and monitor every aspect of their loved one’s situation. In addition, informal caregivers highlighted that receiving notifications might trigger panic and worry, particularly if they are unable to respond immediately, even after being aware of it:

At some point, you want to know everything. Especially if you’re worried, then it’s nice to be able to see a lot, yet you can’t do anything with it. [Participant 2]

If I look at myself... I think if I receive such a notification [emergency] then the first reaction is panic, okay that is a strong word, but as I already said: I work in healthcare myself. I see the most terrible things, that doesn't affect me. But when it concerns your own parents, it immediately causes stress. [Participant 4]

Feeling Obliged to Undertake Action

The informal caregivers also mentioned that once they are aware of their loved one’s situation, they cannot ignore the situation and they feel obliged to undertake actions according to the information provided by USSs. However, sometimes, it is simply not possible to take action immediately due to physical distance or other factors. However, some participants indicated not having the urge to immediately act upon the data or being able to filter important information; they suggested it might be problematic for other informal caregivers:

If I see worrying things, then I literally and figuratively have to go there, if I see it, then I have to go there: normally, you wouldn’t, or quickly call, but now you see it, so you feel compelled to go there... [Participant 4]

So for my situation that [information overload] won’t happen so quickly. For my sister, it might be a bigger struggle, as she is less able to distance herself from the situation as it is. I think when she receives detailed information from the system, she may feel the need to intervene, whereas I have less trouble with that. [Participant 6]

Substitution of Human Contact

Although not all informal caregivers expect the system to substitute the human contact of their care recipient, some of them perceived this as a risk from using USSs. They suggested that if USSs are capable of providing comprehensive insights into the health of the care recipient, it could potentially result in reduced or no visits from professional care staff. This is concerning, considering the already existing scarcity of professional caregivers:

It’s simply impossible to find enough staff, and apparently the situation is even worse in home care. So, if you’re going to develop technology to do more, with fewer people, to be more efficient, it means there will be less human contact, and that means less home care visits for my father, while on the other hand, he’s already experiencing so much loneliness. [Participant 4]

I find it a risk that people retrieve all their information from this system and they might start thinking they no longer need the contact moment, while it is actually so important. [Participant 1]

Preconditions for Implementation

Shared Decision-Making

According to informal caregivers, it is necessary to discuss and reach an agreement together with professional caregivers about IC, including which activities should be monitored, what communication strategy should be used, who should receive and respond to the information, and what should be the content of the information. This would be necessary to prevent unclarities, unfulfilled expectations, and unaddressed notifications or follow-ups, as it could otherwise potentially hinder the effectiveness of care provision:

You can benefit a lot from it [USSs] together and I think if you don’t do this together, everyone can get a lot of trouble from it. That’s not what you want. [Participant 1]

I would never fill it [the settings of the system] in alone, I would really do that together with other professional caregivers or informal caregivers. I think you should all agree with each other about how you fill this in and what you expect and so on...this would be a nice moment to put our heads together and make a choice together. [Participant 1]

It could be that you alert three parties simultaneously and one thinks, “hold on, I won’t do anything because the other two will take care of it,” and everyone assumes that of themselves. And then, nobody responds... [Participant 6]

USS as a Supportive Tool

The informal caregivers indicated that it is important to perceive USSs as a supportive tool rather than a tool to replace the human component in care:

They [persons with cognitive impairment] actually require people around them to be present. It’s better for them, otherwise, they will completely withdraw.
Human interaction, maintaining contact with others is extremely important. So that aspect should be preserved. The system should not result in less human contact, as that would further distance individuals with dementia. [Participant 1]

However, they expected the solution to provide concrete data to facilitate the conversations about and interpretation of the situation together with formal caregivers:

Cognitive decline happens slowly and there are some things that we [informal caregivers] can’t point out. Now it [wandering in the house] is starting to happen more and more. Such raw data can be important for such situations, especially if you have to go to the neurologist or something, then you can do a lot of things. [Participant 4]

It was also mentioned that care decisions should not solely be based on the information provided by the system. Instead, it is imperative to engage in discussions with professional caregivers before making definitive decisions:

So it’s a support system and it shouldn’t take over the analysis of the situation. It may give the numbers, but if on that basis it is said of oh, she [care recipient] only needs so much more care time, or this is no longer necessary since she can still handle this task herself. Yes, then we’re going the wrong way. [Participant 2]

Informal Caregivers’ Experiences With PSD Features and the User Interface Prototype

Overview

In general, the informal caregivers indicated being quite positive about the user interface prototype. Most of the screens were reported to be clear and understandable. However, there were also some negative experiences and suggestions for improvement. Overall, the experiences of informal caregivers are presented as three main themes: (1) positive experiences, (2) mixed experiences, and (3) suggestions for improvement.

Positive Experiences

Personalization: Options to Customize Settings

All informal caregivers (6/6, 100%) had a positive experience with the flexibility (not a one-time setup) of adjusting the interaction platform settings to accommodate their dynamic care needs, thereby improving the quality of life for care recipients. Specifically, they appreciated being able to personalize the settings based on their individual circumstances, the evolving condition of their care recipient, and the monitoring scenario at hand (such as emergencies or self-care activities):

I think this is a good thing. The more you can adjust it to fit your and well in this case my father’s needs and lifestyle, the quality of care can be improved. [Participant 4]

We’ll do everything first [make all the settings], and then I’ll figure it out, or change it later. It is nice that I could still make adjustments later on, so that it’s not a one-time set-up. [Participant 3]

That depends: do I live next door, or close by, then it might be sufficient to be the only one being notified. But this should be available, like imagine I’m away for a weekend. The other informal caregivers will temporarily take over, then I will adjust whether or not someone is available. And the professional caregivers should also receive that notification. So, I would like to have this screen [settings] flexible, so that you can set it individually, per day or per time. [Participant 5]

Well, it’s already quite intuitive. It is good that you can click through quickly on different options within each activity and it is indeed stored for future usage. Also, it is nice that you can always come back and adjust things later if needed. [Participant 6]

Tailoring: Provide Information According to the Stakeholder

Tailoring alerts, notifications, and reports based on the intended recipients (formal or informal caregivers) were found to be valuable in the development of the interaction platform. Informal caregivers felt that a formal caregiver may require different information compared to an informal caregiver:

The information to professionals should be sent as per their needs. Of course, it will be very different from what informal caregivers need. [Participant 2]

For example, the raw data obtained from the sensors could provide more meaningful information to the formal caregivers, whereas they found the interpreted data to be sufficient for themselves:

The raw data is more useful for the healthcare professional than data which is already interpreted by the system. I don’t want raw data because that won’t help me, so then I would go for interpreted data. [Participant 3]

Reduction: Being Informed About the Situation Directly

The reduction feature was used in 2 ways: to receive direct calls or notifications in emergency situations, and to provide a quick overview of the current and past activities throughout the day on the home screen. In emergency scenarios such as falls, informal caregivers found the system-generated alerts (via quick calls or notifications) to be valuable, as they have the potential to streamline the communication and facilitate on-time care. This automated approach eliminated the need for caregivers to contemplate whom to contact and bypassed potential delays when reaching out to formal caregivers:

It is about on-time care. I think, if something happens, what do you need to do? Whom should you call to organize care quickly? There must be logical thinking behind it and the system can do it quickly. [Participant 3]

What I find important is that there is an alarm service-like solution, but initially it could be directed straight to the caregiver, a direct signal from the system saying: “here we see a deviation, this is what the system, the technology detects and intervention
may be required here” or “we see a fall, immediate intervention in necessary.” [Participant 6]

Furthermore, all informal caregivers (6/6, 100%) found the possibility to glance at the current and past activities in the day on the home screen to be convenient:

I find this [home screen] quite clear, now, that you can see which activity has already been performed earlier today, but also what is happening at the moment. This is really nice, and at the top, okay so is the situation at the moment. [Participant 1]

Yes, I think this is fantastic, I must say. Specifically, the fact that you do indeed see an interpretation of the situation that the system has apparently determined and everything goes well. [Participant 6]

The informal caregivers were also positive about the functionalities of sending messages, finding contact information about caregivers, and connecting to an electronic client dossier as these would address the issue of having to use multiple systems:

I think it is always desirable to have everything in one place and not having to deal with various different systems again. [Participant 2]

**Trustworthiness: Insight Into the Reliability of the Information Provided in the System**

Most informal caregivers (4/6, 66%) indicated being positive about the system providing a reliability percentage regarding the notification and information. According to them, a reliability score increases their trustworthiness toward the system:

It is still a technology, sometimes false alarms may occur, for example, when she has dropped something and trying to pick it up. Then, that's okay, that system indicates it is reliability percent. I actually like it. It points towards the trustworthiness of the system and also indicates that at times it can miss classify some things. [Participant 2]

When the reliability percentage is high, they sense the urgency and seriousness of the situations and were compelled to take required actions:

A reliability of 80 percent, yes, that did something with me...I thought that I should really take this seriously, like really seriously. [Participant 1]

On the other hand, when the confidence percentage was low, informal caregivers might be slightly relaxed, but they still wanted to ensure the safety of the care recipient. However, they felt that an indication of a low or high confidence percentage might help formal caregivers to organize their care better. For example, they can prioritize their visits depending on the system’s reliability percentages:

For me, it’s fine to read that information, whether it is 50 percent or 80 percent, that doesn’t matter. But I think for professional caregivers that it does matter, because if they receive 6 notifications and one has 30 percent reliability and the other 80 percent. Then they will first go to the one with 80 percent reliability. [Participant 4]

Interestingly, 1 informal caregiver expressed that viewing percentages might be slightly confusing for them to interpret; thus, simple terms such as “very reliable” or “less reliable” can be used:

Now I can’t judge 10 percent or 80 percent or 20 percent or whether that’s right. [Participant 3]

Furthermore, informal caregivers demonstrated a willingness to offer feedback to the system to enhance the reliability of the system’s alerts. However, it was recognized that this responsibility should be shared with other caregivers, particularly formal caregivers, who are also involved in responding to alerts and, thus, can also provide context-aware and detailed feedback to enhance the system’s learning:

The system is self-learning, so I’m actually positive about it. I hope people understand that when they provide feedback, they need to specify what exactly went wrong, so that the system can learn from that. For example, if someone didn’t fall but just lay down on the couch, then this should be adjusted. The system can become smarter by processing more data and thus increase the reliability of notifications. So, it’s important to add more context in order for the system to learn from it. [Participant 5]

**Verifiability: Possibility to View the Electronic Health Record and Connect With Formal Caregivers**

The possibility to view the electronic health records and connect with the concerned formal caregivers was found to be very handy and desirable. These functionalities also support the notion of an all-inclusive platform:

For example, If I want to speak to Mrs. Baker [formal caregiver], I click on Mrs. Baker and she can guide me further. [Participant 1]

I think it is always desirable to have everything in one place and not have to deal with various different systems again, also considering different passwords and identification or authentication as well. [Participant 2]

**Mixed Experiences**

**Reminder: Receiving a Reminder in Case of a Missed Emergency Call**

Mixed experiences were reported regarding the reminder that was received in case of a missed emergency call. Some informal caregivers found it useful, and others thought that reminders were not necessary for them, as this would be more useful for formal caregivers, depending on their personal situation. One informal caregiver experienced the reminder as confronting:

I would like a care professional to receive such a reminder when she falls, because I am always at a distance. [Participant 1]

If I see this message, and realize I’ve missed the emergency call, then I feel like I should have been more attentive, then I would like to have the information quickly and in a concise format, without...
Suggestions: Receiving Suggestions Regarding What Actions to Take

The use of suggestions along with alerts or notifications was found to be debatable in older adult care. On one hand, informal caregivers found suggestions to be valuable in situations such as emergencies, where they panic or are unsure about the possible actions to take to facilitate the right care:

We all know what stress and panic can do, in those moments we can sometimes make stupid decisions, or forget the best order of doing things. So, having such a suggestion can serve as a helpful guide. [Participant 1]

I feel that falling is different from agitated behavior. Falling means immediate danger, while agitated behavior often arises in the context of the dementia process that people experience. In such cases, it would be helpful to receive tips on what to do. [Participant 2]

On the other hand, some informal caregivers felt that suggestion were unnecessary and subjective to the care experiences of the informal caregivers:

I think many people would appreciate it. You see, I’ve been working in healthcare for many years, so I’m familiar with these things. I believe there are many people who would benefit from receiving suggestions on what to do in certain situations. While I may quickly come up with solutions based on my experience, this is not the case for everyone. Thus I think many people would find it supportive. [Participant 1]

Moreover, informal caregivers expressed concern that if suggestions are system generated, they will be generic, which could potentially limit their thinking to the provided suggestions only, thus losing the personal touch in care and inducing the feeling of annoyance:

I find this terrible, very annoying. Because I’m already stressed out, and then I get those too obvious suggestions that say “do this, do that.” My stress levels are already high and then I read something stupid...no thank you. Very irritating... [Participant 2]

Overall, while the usefulness of suggestions differs from person to person, it would be valuable to have such an option for those who are willing to receive it.

Suggestions for Improving the User Interface Prototype

Overview

Overall, the conceptual workflow of the prototype of the interaction platform was assessed positively by the informal caregivers. They indicated that most of the screens were clear and understandable. However, they provided some suggestions about screens or connections that were perceived as less logical or where improvements could be made.

Improvements in Conceptual Flow

Informal caregivers highlighted that some choices regarding the notification settings were repetitive or unnecessary, which made the flow unclear or redundant. Specifically, in Figure 3 (third screen [agitation] and fourth screen [fall incident]), the option to share the respective information with the formal caregiver was given; however, informal caregivers already mentioned their choices to share or not share the information with formal caregivers while adjusting their preference (Figure 2). Consistent with an iterative process, this was adjusted for the next (last 2) interview sessions:

Here, I again have the choice if I want to share with a care professional. But if that happens again, then I wonder if I have set it up correctly in the settings. So does this still appear on my screen? In the beginning, you make a choice about sharing information with a care professional, and here that comes up again, so it’s kind of redundant. [Participant 3]

This is what I don’t understand. If I let the notifications go to the home care professionals for this situation, then I should not have to fill this in [choice for the content of the notification]. [Participant 4]

Furthermore, informal caregivers indicated that it was inconvenient to immediately receive the option to provide...
feedback to the system in case of an emergency notification as they mentioned, at that moment, they were not thinking about the feedback and were probably not the right person to provide the information. It was suggested to send a reminder to provide feedback at a later moment. Moreover, there should be an option to receive more details about the situation:

> Provide feedback on this notification, yes that can be useful, but it has to be at a later moment. You don’t do this in the notification itself, but you can add at a later moment what the issue was and whether the notification was accurate. [Participant 2]

> For this, I would appreciate a reminder. I don’t necessarily enjoy receiving a lot of notifications all the time, but specifically for this purpose, yes. It’s about helping each other and helping the system learn, and thus improving the care. And I think when I’m actually there [at loved one] or when I come from there, then I might forget that. So, a reminder would be helpful, but it would be good to have a choice in the type of notification. [Participant 3]

Informal caregivers also suggested that feedback should be provided to them after they received a notification, so that they know that someone handled the situation and what actions have been taken:

> I think that is a bit of a gray area, so you received or made a notification, but what happens with it? That I would expect to receive feedback on. [Participant 5]

> Now I still have the feeling like I have to go there because I don’t know if it [notification] has been received and if someone is going there. [Participant 1]

**Improvements in Visual Design**

The informal caregivers suggested to include a clear visual indication when a deviation in behavior was noticed by the system, for example, a warning sign. In addition, an informal caregiver mentioned it would be more useful to express reliability in words instead of percentages, as this might be easier to interpret. Finally, informal caregivers suggested that the prototype could be improved by providing information in a more visual manner and including more graphs, images, and pictograms, as this could make it easier to interpret the information they were looking for:

> At a glance, I can see that everything is going well...but then [in case of deviation] could have a different color like red, and for yourself there could be an exclamation mark or warning sign to indicate that this is not optional information but something that needs to be looked into because it is not as it should be. [Participant 5]

**Discussion**

**Principal Findings**

In this study, informal caregivers showed a significantly positive attitude toward using USSs driven by AI algorithms for providing care to home-dwelling older adults with cognitive impairment. However, a previous study that explored care recipients’ perspectives regarding AI in health care revealed hesitancy, primarily driven by worries related to safety, privacy, and autonomy [29]. This divergence could be attributed to factors: the difference in the study population and the potential lack of knowledge about USSs (in general, technological care solutions) among the previous study’s participants. The previous study by Richardson et al [29] focused on care recipients’ perspectives, whereas this study involved informal caregivers who might have a different perspective, as USSs will be monitoring the care recipient and it does not concern informal caregivers. Furthermore, many people have limited knowledge about AI algorithms and view AI as a “black box” [30]. Previous studies suggest that educating and engaging individuals about AI can enhance their trust in AI and contribute toward its successful implementation in health care [29,30]. In this study, USSs were explained using a video prototype and additional verbal explanations, which most likely increased the participants’ awareness of AI use. However, individual differences in understanding AI might also influence their positive and accepting attitude toward USSs.

USSs rely on AI algorithms to predict the behavior of older adults with cognitive impairment, which might not function flawlessly and could misclassify certain behavior patterns. Therefore, care providers need to be cautious and should not become overly reliant on USSs, as it may lead to incorrect care choices [31]. This would present an ethical issue regarding accountability, as it prompts the question of who is responsible and to what extent [31]. Educating caregivers about its use, capabilities, and limitations might help to overcome this risk. In addition, the risk of bias when using AI in health care should also be considered [31,32]. If the training data predominantly represent a specific population (sex, age, ethnicity, etc), they might create biases [32]. This risk could be mitigated by ensuring representative and inclusive training data sets, that is, including data from a wide range of individuals with different demographic characteristics and backgrounds when training the AI algorithm [32].

Furthermore, the informal caregivers recognize the value of USSs as a supportive tool in the care of home-dwelling older adults with cognitive impairment [2]. Specifically, USSs can facilitate appropriate care decision-making, contributing to their peace of mind while also creating a safe environment for their care recipients. However, they also expressed some concerns regarding the possible information overload, substitution of the human aspect in care provision, and overinterpretation of data. To mitigate these issues, they acknowledged the importance of setting up (make agreements about the monitored activities and strategies regarding communication) the solution together with other stakeholders (specifically, formal caregivers). This is consistent with the CeHRes road map, that is, for successful implementation of an eHealth technology, it is important to consider the perspectives and needs of the different stakeholders involved [15,33]. Moreover, it can be said that by combining the strengths of technology with the insights and expertise of caregivers, a more comprehensive and effective care approach and implementation could be achieved.
In addition, informal caregivers experienced the lo-fi prototype and the use of most PSD features elicited in a previous study as positive [7]. Particularly, participants valued the possibility to personalize the settings and change them to their preferences at any given moment in time. The use of the personalization feature, as suggested in the PSD theory, has the potential to enhance the usefulness of eHealth technologies such as USSs [7]. However, before making personalized solutions, the personalization-privacy paradox should be considered carefully [34]. This paradox demands right balance between offering personalized experiences and safeguarding user privacy [34,35]. This balance might be achieved by implementing robust privacy measures, obtaining informed consent, being transparent about data use, and providing users with control over their data as described in the ethical guidelines issued by the European Commission for trustworthy AI [36] and the European Health Data Space regulation [37]. Moreover, regarding the PSD features reminders, suggestions, and social learning, the experiences were slightly mixed. These findings emphasize the importance of a user-centered design approach, as the preferences of each individual can vary depending on the care situation, personal circumstances, and preferences in IC [15]. For example, the travel distance to the care recipient could influence their choice regarding whether they want to receive an emergency call when a fall incident occurred.

**Implications for Future Studies and Practices**

For the successful implementation of a complex eHealth intervention such as USSs for home-dwelling older adults with cognitive impairment, a holistic design approach is required [15]. It is important to consider the perspective of different key stakeholders such as informal and professional caregivers, care recipients, and care organizations and secondary stakeholders such as health insurers, governments, and technology businesses while designing and implementing such solutions. In future studies, it would be interesting to perform the next design iteration by using the results of this study as a starting point. Gradually, a high-fidelity prototype of the user interface could be created and evaluated with different stakeholders in the care of older adults with cognitive impairment. Furthermore, the creation of personas would be helpful in studies, as different types of end users may have different needs and requirements [38]. The personas could be based on characteristics such as caregiving experience, educational level, or need for cognition [38]. In addition, it would be interesting to explore the ethical implications of implementing a smart monitoring and communication system for home-dwelling older adults, which could be performed using in-depth interviews with involved stakeholders.

Before the implementation stage, a business modeling approach can be used. It provides insight into how value is generated and delivered to customers, which should be considered to bring eHealth technology to the market [39]. In addition, the implementation of such a solution requires guidelines and agreements (about how to work with such a system) within organizations and at the government level. Caregivers should be educated about how to interact with the system and interpret and communicate the data. In addition, it is essential to consider regulations such as Medical Device Regulation to determine whether USSs will be categorized as medical devices. Medical Device Regulation offers provisions to address privacy and security concerns, especially regarding medical devices that collect and process personal health data [40]. Furthermore, as USSs use AI, it is crucial to take into account the new AI Act proposed by the European Union (EU). This act aims to regulate the use of AI in EU countries, ensuring better conditions for the development and use of AI technologies (EU AI Act, 2023) [41]. In this act, different rules will apply to different risk levels, with USSs probably falling under the category of high risk, and they will be subjected to a high degree of regulations [41]. This might have consequences for the extent and manner in which AI is applied.

**Strengths and Limitations**

The key strength of this study is rooted in its methodology, specifically the adoption of a participatory development process that actively engaged informal caregivers during the early development stage of the interaction platform. The designed lo-fi prototype of the interaction platform provided an overview of the overall conceptual workflow and PSD features in different care scenarios to informal caregivers. While lo-fi prototyping is cost-effective, it can be seen as an opportunity for rapid development with end users in the early stages of development. Specifically, a task-based study design in conjunction with a think-aloud approach was used for this study. The task-based study design facilitated a realistic evaluation of user interactions with interfaces in various care scenarios [18,19], and the think-aloud approach provided direct access to users’ thoughts, perceptions, expectations, experiences, and decision-making processes during their interactions with the interfaces and PSD features [20]. By combining these 2 approaches, a comprehensive understanding of potential issues (in the used PSD features) and areas for improvement (in the conceptual workflow) was obtained, which could be incorporated into subsequent design iterations. In addition, the use of a video prototype assisted participants in comprehensively understanding the proposed USS. This, in turn, facilitated the development of more concrete and well-informed themes regarding the implementation outcomes and preconditions necessary for the successful integration of novel USSs.

Along with strengths, this study also has some limitations, which should be considered when interpreting the results. First, all the informal caregivers (6/6, 100%) who participated in this study had previous experience with technological interventions in care provision (eg, Caren platform) and had previously participated in related studies. While it is important to acknowledge that the findings of this study may not be fully generalizable to participants with no previous experience in digital care technology, the overall growth in digital literacy is noteworthy and holds promise for the realization of the study’s findings. Second, most informal caregivers (4/6, 67%) who participated in this study reported that their care recipient had received a formal diagnosis of cognitive impairment. However, some informal caregivers also expressed their own opinions regarding the indication of cognitive impairment in their care recipients. Given the scope of this study, which aimed to explore the perspectives of informal caregivers, their opinions hold high value in this study [42-44].
Third, it is important to note that this study was conducted in the Netherlands, which may limit the generalizability of the results to international health care infrastructure. Different countries have diverse regulations and policies regarding older adult care; thus, different expectations and preimplementation conditions regarding USSs can be imagined. Finally, it is worth mentioning that data saturation was not reached in this study, as new information was obtained from all interviews. This indicates that there may be additional themes that were not fully explored, suggesting that the results of this study may not be exhaustive. However, it is important to recognize that the design process was iterative, and during the evaluation of the lo-fi prototype, the aim was to further enrich the platform, making data saturation less critical for this stage of development [45].

Conclusions
Overall, informal caregivers of older adults with cognitive impairment had positive expectations regarding the implementation of USSs. They expect the use of such a system to contribute to care decision-making and to provide insight into the situation of the care recipient. However, information overload and loss of human aspect were perceived as risks. To successfully implement a USS, good communication and agreements among informal caregivers, formal caregivers, and the care recipient are needed, thus necessitating a holistic approach in the development and implementation process. Informal caregivers were quite positive about the lo-fi prototype of the user interface and the application of PSD features; however, there were also mixed experiences and suggestions for improvement regarding the conceptual flow and visual design of the prototype. Personalization of the settings of the prototype was perceived as highly valuable. The results of this study, especially the identified concerns, should be considered in the further development and implementation of USSs for home-dwelling older adults with cognitive impairment.

Acknowledgments
This study was funded by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant (814072) for the ITN ENTWINE informal care. The authors would like to thank Prof Dr Paul Havinga, Jan Hendrik, and Dr Duc Le Viet for their valuable feedback.

Data Availability
In this study, some personal details about the participants were obtained, constraining our ability to make quantitative data available for any secondary analysis. However, upon request, a summary of the qualitative data (after anonymization) can be provided.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The semistructured interview script used for qualitative data collection.

References


the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Community Members’ Perceptions of a Resource-Rich Well-Being Website in California During the COVID-19 Pandemic: Qualitative Thematic Analysis

MarySue V Heilemann1, PhD; Jianchao Lai2, MSW, PhD; Madonna P Cadiz2, MSW; Jocelyn I Meza3, PhD; Daniela Flores Romero4, BA; Kenneth B Wells4, MPH, MD

1School of Nursing, University of California, Los Angeles, Los Angeles, CA, United States
2Department of Social Welfare, Luskin School of Public Affairs, University of California, Los Angeles, Los Angeles, CA, United States
3Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States
4Research Center for Health Services and Society, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
MarySue V Heilemann, PhD
School of Nursing
University of California, Los Angeles
700 Tiverton Avenue
Box 956919
Los Angeles, CA, 90095-6919
United States
Phone: 1 310 206 4735
Fax: 1 310 206 3241
Email: mheilema@sonnet.ucla.edu

Abstract

Background: To address needs for emotional well-being resources for Californians during the COVID-19 pandemic, the Together for Wellness/Juntos por Nuestro Bienestar (T4W/Juntos) website was developed in collaboration with multiple community partners across California, funded by the California Department of Health Care Services Behavioral Health Division federal emergency response.

Objective: This qualitative study was designed to explore and describe the perspectives of participants affiliated with California organizations on the T4W/Juntos website, understand their needs for web-based emotional health resources, and inform iterative website development.

Methods: After providing informed consent and reviewing the website, telephone interviews were conducted with 29 participants (n=21, 72% in English and n=8, 28% in Spanish) recruited by partnering community agencies (October 2021-February 2022). A 6-phase thematic analysis was conducted, enhanced using grounded theory techniques. The investigators wrote reflexive memos and performed line-by-line coding of 12 transcripts. Comparative analyses led to the identification of 15 overarching codes. The ATLAS.ti Web software (ATLAS.ti Scientific Software Development GmbH) was used to mark all 29 transcripts using these codes. After examining the data grouped by codes, comparative analyses led to the identification of main themes, each with a central organizing concept.

Results: Four main themes were identified: (1) having to change my coping due to the pandemic, (2) confronting a context of shifting perceptions of mental health stigma among diverse groups, (3) “Feels like home”—experiencing a sense of inclusivity and belonging in T4W/Juntos, and (4) “It’s a one-stop-shop”—judging T4W/Juntos to be a desirable and useful website. Overall, the T4W/Juntos website communicated support and community to this sample during the pandemic. Participants shared suggestions for website improvement, including adding a back button and a drop-down menu to improve functionality as well as resources tailored to the needs of groups such as older adults; adolescents; the lesbian, gay, bisexual, transgender, and queer community; police officers; and veterans.

Conclusions: The qualitative findings from telephone interviews with this sample of community members and service providers in California suggest that, during the COVID-19 pandemic, the T4W/Juntos website was well received as a useful, accessible
tool, with some concerns noted such as language sometimes being too “professional” or “clinical.” The look, feel, and content of the website were described as welcoming due to pictures, animations, and videos that showcased resources in a personal, colorful, and inviting way. Furthermore, the content was perceived as lacking the stigma typically attached to mental health, reflecting the commitment of the T4W/Juntos team. Unique features and diverse resources, including multiple languages, made the T4W/Juntos website a valuable resource, potentially informing dissemination. Future efforts to develop mental health websites should consider engaging a diverse sample of potential users to understand how to tailor messages to specific communities and help reduce stigma.

(JMIR Form Res 2024;8:e55517) doi:10.2196/55517

KEYWORDS
adaptation; humans; pandemics; mental health; COVID-19; health resources; California; psychological; stigma; digital; prevention; public health; emotions; website; qualitative research

Introduction

Background

The COVID-19 pandemic brought unexpected difficulties related to activities of daily living for people worldwide. In addition to affecting physical health, COVID-19 also threatened emotional health. Diverse groups were impacted, including people of different ages [1-4], gender identities [5-9], races and ethnicities [10,11], and geographic locations [12-15]. Those who faced financial concerns [16,17] or lost their jobs [8,18] and those caring for children at home [5] struggled with a variety of additional pressures. For workers [19,20], including health care workers [21,22] and community-based service providers [23-26] in the United States, attempts to serve clients or patients were confounded by pandemic-related challenges such as concerns about infection, reductions in staffing, and transitions to remote care, making their jobs even more complex and challenging.

As stay-at-home orders proliferated, people began to look for ways to strengthen their ability to cope emotionally during the pandemic. Many turned to media outlets such as television, radio, and social media for news and information [27]. Unfortunately, the news often led to increased fear and worries about the COVID-19 virus, illness, death, loss of jobs, economic concerns, and more. Many used Twitter to read the views of others and express their own negative sentiments about the pandemic [28]. Social media use contributed to experiences of stress [27], whereas engagement in self-care activities such as being able to access and use personal support resources helped protect against mental health distress [29]. Consequently, researchers [27,30] called for creative developments to connect individuals with social support and mental health services. To overcome stigma and other barriers, researchers and developers turned to web-based digital tools to make resources for coping and information on emotional well-being more accessible.

In this context, the Together for Wellness/Juntos por Nuestro Bienestar (T4W/Juntos) website was developed in collaboration with multiple community partners across California, funded by the California Department of Health Care Services Behavioral Health Division federal emergency response, to directly address needs for free web-based emotional well-being resources for Californians during the pandemic. The purpose of this paper is to report the findings of a qualitative study on the perspectives of a sample of participants affiliated with California organizations who engaged with the T4W/Juntos website.

Development of the T4W/Juntos Website

The T4W/Juntos website was developed as part of the Federal Emergency Management Agency and Substance Abuse and Mental Health Services Administration crisis counseling contract with California. The goals for T4W/Juntos were developed with a multidisciplinary team of researchers, clinicians, digital resource development experts, and staff from community-based agencies in California. Goals centered on creating inclusive and accessible resources that would provide evidence-informed and evidence-based information to Californians to ease the stress experienced during the pandemic [31].

Meetings were held via Zoom (Zoom Video Communications) with the large collaborative team (4 to 18 members per meeting) to maximize input from community members. The community and study team members made decisions collaboratively about which types of resources to include on the website. The priority was to feature resources that facilitated learning about COVID-19 or offered ways to address anxiety and stress (eg, web-based meditations, breathing exercises, and direct links to warmlines and hotlines), strengthen resilience, cope with grief due to a recent loss, connect with other people (such as through web-based support groups), or support social justice (eg, antiracism and reducing hate crime). Resources included links to web-based toolkits, websites, videos, web-based applications, articles, and downloadable pamphlets. Some resources were available in multiple (up to 10) languages. Community partners’ emphasis on using neutral, nonclinical language to increase comprehension and relatability and reduce stigma led to a monitoring of the length and complexity of messaging for the website. In response to the community partner prioritization of videos to engage users, the team created videos of community members speaking about the website’s purpose and features in English and Spanish [31].

Prior Work

A previous paper related to T4W/Juntos described the process of website development [31]. A second paper described the results of an analysis of quantitative data from an electronically administered web-based survey that were collected at 2 time points (approximately 6 wk apart) from English- and Spanish-speaking adult participants. Of the 366 eligible participants, 315 (86.1%) completed the baseline survey and...
93 (61.3%) completed the follow-up survey, with baseline results showing substantial diversity in gender, gender identity, and race and ethnicity and 32.7% (103/315) having moderate depression or anxiety (2-item Patient Health Questionnaire or 2-item Generalized Anxiety Disorder score of ≥3) [32,33]. Significant predictors of baseline website engagement were Hispanic versus other race or ethnicity and COVID-19–related behavior changes. The use of the T4W/Juntos website during the month before the follow-up survey was significantly associated with a pretest-posttest reduction in depression (2-item Patient Health Questionnaire score), and greater website engagement at baseline predicted reduced hotline use before follow-up [34]. An analysis of short qualitative answers that 199 (63.2%) out of 315 participants typed into textboxes in response to open-ended questions in the previously described web-based survey led to insights into safety concerns and fears during the pandemic and perceived benefits from and suggestions for improving the website [35].

Research Aims
With the goal of supplementing the quantitative results, the aim of this qualitative study was to describe the perspectives of a diverse subset of participants associated with various California community organizations who completed the baseline surveys regarding their experiences with the T4W/Juntos website. We also focused on participants’ needs for website resources that could support emotional well-being for themselves, their families, their clients, or their community. Finally, we sought insight to inform iterative website development in the future.

Methods
Recruitment
The larger sample described previously was recruited during the pandemic through invitations that were sent primarily by email, while stay-at-home orders were in effect, from 11 community partner agencies throughout the state of California to their affiliated community members with information about the website, its purpose, and the research study. Each of the 315 participants who consented and completed the baseline survey was given the option to indicate their interest in participating in a potential future telephone interview by clicking a box at the end of the survey. In total, 73.9% (233/315) of the participants clicked on the box to indicate their interest in being interviewed. Inclusion criteria were being aged ≥18 years, having access to the internet, having already completed the baseline web-based survey in English or Spanish, and agreeing to provide contact information. Using convenience sampling, participants who spoke English or Spanish were contacted via telephone by research staff approximately 2 weeks after completion of the baseline survey, starting with those who were the first to finish the survey, to offer an interview, confirm availability, and set a date and time for the interview. After 15 interviews were conducted, purposive sampling was used to maximize diversity in race and ethnicity, gender, and age. The final sample comprised 29 participants. The interviews (in English or Spanish) were conducted between October 2021 and February 2022.

Ethical Considerations
This study was reviewed and approved by the University of California, Los Angeles, Institutional Review Board of UCLA’s Human Research Protection Program (20-002163-AM-00008). After reading the web-based consent document, each participant clicked to give consent at the time of enrollment in the larger survey study, which included consent to a future potential interview. Our team only contacted individuals who agreed to be contacted for interviews using the contact information they provided. Participants reconfirmed their approval to participate and be audio recorded at the time of the interview. To protect the privacy and confidentiality of participants, the list of the names of the participants and their assigned codes was kept in a password-protected file available only to the principal investigator and project director. Their confidential contact and personal information were kept separate from all other data. Any potentially identifying information was deidentified on the transcriptions of audio-recorded interviews, including any names or descriptors that could possibly identify a participant; all names were changed to code numbers that were used instead of names by the researchers during data analysis. Participants received a US $25 e-gift card after completing the interview.

Data Collection
Demographic data were retrieved from the baseline survey for each of the 29 interview participants. A semistructured interview guide in English and Spanish that was developed by a multidisciplinary team was subsequently used by 2 research team members to conduct all interviews via telephone. Interview questions were designed to explore participants’ perceptions of any aspect of the T4W/Juntos website; gain insight into participants’ needs for support in relation to the resources available via the website for themselves and their families, clients, or communities; and obtain guidance on further development of the website. Audio recordings of interviews in Spanish were professionally translated into English, and all interviews were professionally transcribed verbatim and checked for accuracy. As already noted, identifiers were removed, and code numbers were used instead of names to label transcripts and organize the data.

Data Analysis
Demographic data were analyzed for frequencies using Stata/MP (version 17; StataCorp LLC) [36] for the sample of 29 participants. For the thematic analysis of the 29 transcripts, the study team was guided by a modification of the 6-phase process outlined by Braun and Clarke [37,38]. First, the study team familiarized themselves with the data in all transcripts. Second, the team engaged in initial coding using techniques from grounded theory methodology to enrich our approach [39]. Thus, most codes were developed using the gerund form of verbs, known as process codes, to heighten our focus on the actions taken by participants, as shown in the data [39,40]. To create process codes, coders used heuristic questions to ask What is happening here? and What are they doing here? This allowed coders to get closer to the participants’ point of view while reducing the tendency to prematurely project their own interpretations onto the data [39]. In the third and fourth phases, coders scrutinized the first 12 coded transcripts to identify the
most frequently occurring and significant codes and, through discussion and debate, identified a total of 15 overarching codes. Then, the data from all 29 transcripts were imported into ATLAS.ti Web (version 22.1.5; ATLAS.ti Scientific Software Development GmbH) [41] and coded based on the 15 overarching codes. In the fifth phase of analysis, data reports were created using ATLAS.ti Web [41] based on each of the 15 overarching codes. These were exported to Microsoft Excel (Microsoft Corp) so that the data in each code group could be further examined. Using constant comparison, we sifted, sorted, combined, and collapsed the data in the 15 groupings to form 4 themes, each with a central organizing concept that provided a clear definition of the theme [38]. We continued to compare data with data to develop the properties for each of the 4 themes. Finally, in the sixth phase, each theme was named, and its properties were refined. With a focus on the research aims, the research team then produced a written report interpreting the meaning of each theme.

The overall process of data collection and analysis was influenced by the team’s commitment to social justice and to the goal of understanding the data of each participant while considering their context. Thus, at various points during the research process, each member of the 5-member analysis team engaged in dialogue together and in individual writing of reflexive memos to name any judgments (positive and negative) or concerns that were felt while engaged in the research process, with the goal of reducing the influence of bias on the collection and interpretation of data [38,39].

Results

Participant Characteristics and Sample Demographics

The demographics of our sample of 29 participants are presented in Table 1. Of the 29 participants, 16 (55%) voluntarily shared that they were employed in peer support, hospice care, or health care sales or at a community agency doing health-related work. A total of 72% (21/29) of the interviews were conducted in English, and 28% (8/29) were conducted in Spanish. The duration of the Spanish interviews ranged from 22 to 72 (mean 32, SD 16.61) minutes, and that of the English interviews ranged from 15 to 85 (mean 38, SD 15.49) minutes.
Table 1. Demographics and depression and anxiety scores of community participants in California who were interviewed during the COVID-19 pandemic (N=29).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>21 (72)</td>
</tr>
<tr>
<td>Spanish</td>
<td>8 (28)</td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>46.3 (13.7)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Some high school or lower than high school</td>
<td>2 (7)</td>
</tr>
<tr>
<td>High school graduate or equivalent (ie, GED(^a))</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Some college</td>
<td>9 (31)</td>
</tr>
<tr>
<td>College graduate</td>
<td>12 (41)</td>
</tr>
<tr>
<td>Graduate school (eg, JD(^b), Master’s, PhD(^c), and MD(^d))</td>
<td>5 (17)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>20 (69)</td>
</tr>
<tr>
<td>Man</td>
<td>7 (24)</td>
</tr>
<tr>
<td>Other gender not listed (2-spirit)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Prefer not to state</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Black or African American or African</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Hispanic or Latino (Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race)</td>
<td>10 (34)</td>
</tr>
<tr>
<td>Southeast Asian (Vietnamese, Filipino, Laotian, Thai, Indonesian, and Cambodian)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>White or European</td>
<td>7 (24)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Straight</td>
<td>21 (72)</td>
</tr>
<tr>
<td>Gay or lesbian</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Bisexual or pansexual</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Other sexual orientation not listed</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>GAD-2(^e), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>0-4 (none to minimal)</td>
<td>27 (93)</td>
</tr>
<tr>
<td>5-9 (mild)</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>PHQ-2(^f), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>0-4 (none to minimal)</td>
<td>26 (90)</td>
</tr>
<tr>
<td>5-9 (mild)</td>
<td>3 (10)</td>
</tr>
</tbody>
</table>

\(^a\)GED: General Educational Development.  
\(^b\)JD: Juris Doctor.  
\(^c\)PhD: Doctor of Philosophy.  
\(^d\)MD: Medical Doctor.  
\(^e\)GAD-2: 2-item Generalized Anxiety Disorder scale. A GAD-2 score of 3 is the recommended cutoff point for identifying possible cases of generalized anxiety disorder.  
\(^f\)PHQ-2: Patient Health Questionnaire–2. A PHQ-2 score of 3 is the recommended cutoff point for identifying possible cases of depression.
Qualitative Thematic Analysis Results

Thematic analysis of the qualitative data led to the identification of four themes: (1) having to change my coping due to the pandemic, (2) confronting shifting perceptions of diverse groups on mental health stigma, (3) “Feels like home”—experiencing a sense of inclusivity and belonging in T4W/Juntos, and (4) “It’s a one-stop-shop”—judging T4W/Juntos to be a desirable and useful website.

Theme 1: Having to Change My Coping Due to the Pandemic

Overview

Participants shared that, during the pandemic, they had to change the way in which they coped with daily life stressors. This was represented by 5 properties: increased use of technology to connect with others on the internet, intentionally identifying self-care tips and techniques, coping by helping other people, relying on in-home socialization, and drawing on spiritually oriented coping (Figure 1).

Increased Use of Technology to Connect

Participants found that their use of technology increased during the pandemic, and they had to learn to accept their increased reliance on the internet to be connected in various ways. For example, they used technology to connect with information on a variety of topics and be able to accomplish work for their jobs. They used technology to connect with other people for social reasons; this included using Zoom to connect with friends, family, and their faith communities. They also joined web-based support groups and community groups where they could engage in dialogue with others. One participant described how they felt more “comfortable” having conversations on the internet:

I think I’ve become more dependent on the internet. I also found that not having to deal with people face to face most of the time makes me feel more comfortable. Honestly, it’s easier for me to have a chat on the computer than in real life.

Participants said that they relied on technology to meet their therapeutic needs more than before the pandemic. They found web-based resources to be “easier to forward” and share with others. They described resources on the internet to be “more documented” and viewable. They reported how they learned to click links to use web-based resources, which, for many, was a new behavior.

Intentionally Identifying Self-Care Tips and Techniques

Participants helped themselves by seeking out practical approaches, including self-care tips and techniques. This meant that they were using technology to meet their therapeutic needs, something they had not necessarily done before. They used warmlines, crisis lines, and teletherapy to meet their needs. They learned about meditation and breathing exercises, which they found especially desirable because the stress of the pandemic was experienced as personally difficult. However, participants found it challenging to find “accurate” resources. This put them on a quest to find “reliable” web-based resources they could use to reduce stress. Participants explained that they continued searching on the internet even if their immediate need was resolved because they wanted to have resources ready just in case they needed them in the future. In addition, some used art to self-soothe during the pandemic, whereas others sought “self-improvement” strategies.

Coping by Helping Other People

When asked to say more about how they handled their own stressful experiences during the pandemic, participants repeatedly spoke of helping other people in their personal lives and on the job. It seemed that helping other people was itself a strategy they used to cope. A participant spoke of others who felt “invisible” and as if other “people don’t respect them” and how difficult it was for them because “they’ve lost their purpose because they can’t go to work or can’t do the job they used to”
do. They noted how important it was to share with others that “there’s hope...that there are people out there trying to make a difference, trying to help, trying to listen.” Participants’ efforts were extended to various types of people, including family members, friends, and coworkers, and those who self-identified as health workers reported that they helped both individual clients and families. Experiences of helping others stood out to them; they felt a sense of “satisfaction” from their helping work. One participant said the following:

And the reason I liked it is—and the reason is the feedback, the participation and everything, is actually one of the ways that it makes me feel like not only I’m sharing a resource, but I’m sharing a resource that I know is good—I hate sharing things that I know are not good—and I wanted to like what I’m doing here.

Relying on in-Home Socialization

Finding others to fulfill social needs and desires was mainly limited to whoever was in the home. Participants relied on their family members or roommates for dialogue, socialization, and friendship during the pandemic. However, pets also played an important role as they provided “joy, stress relief, and companionship.” Many described the importance of going on walks with their dogs as it brought about daily exercise and also could open up dialogue with neighbors, which was highly valued at this time of social isolation. One participant shared how meaningful it was to live with their 2 dogs and son. They said the following:

If I had to live by myself, I don’t know how I would get through this. I really mean it. I’m being perfectly honest.

Drawing on a Spiritually Oriented Approach to Cope

With few options for socialization, participants shared that they turned to spiritually oriented routes for coping. They turned to “God” and relied on their faith to keep them going. One participant stated that they could not imagine how anyone could “get through” the pandemic “without God.” For some, listening to religious radio programs filled a crucial need in their daily lives during the pandemic. Others reported using prayer or reading scriptures. A participant described how religion provided “guidance” on daily life:

[Faith in God] helps me first and foremost, that helps me not to have fear. I think that a lot of people now are controlled by fear. And so that helps me, y’know; that strengthens me and gives peace to my heart. I feel secure with my health habits, with my diet, well, because I’m connected to God, and because I get my health practices from the Bible. So, I feel that all around, my mental, my physical, my emotional, all my wellbeing, I dedicate that to God for his blessing. And so, y’know, I think that’s the biggest part of it.

Theme 2: Confronting a Context of Shifting Perceptions of Mental Health Stigma Among Diverse Groups

Overview

When discussing T4W/Juntos and its purpose in helping with emotional well-being, participants were concerned about the context during the pandemic related to public views on mental health. They raised the issue of shifting views on mental health stigma within different communities and how it impacts people. They reported how stigma differs based on age or generation, race, or ethnicity and how the T4W/Juntos website would be received within a context of overarching stigma in various communities. They also suggested ways to reduce stigma. This is reflected in 3 properties (Figure 2).

Figure 2. Theme 2 from qualitative analysis of interview data from community participants in California during the COVID-19 pandemic (N=29): confronting a context of shifting perceptions of mental health stigma among diverse groups.

Experiences With Mental Health Stigma Vary Across Generations

Mental health stigma was perceived as an issue for all generations. Participants explained that stigma itself was the backdrop that set the stage for how the T4W/Juntos website would or would not be received. This could have an impact on whether community members would embrace the website.
Participants explained that there was a difference in how older and younger generations experienced mental health stigma. Older adults who grew up during a time when mental illness was considered “bad” and “dangerous” were described as rejecting the possibility of seeking treatment for their mental health problems. They described a dual process in which the older generation felt too stigmatized to seek mental health support, but at the same time, they perpetuated the stigma surrounding mental health within their communities. Older adults were perceived as having negative beliefs about mental health concerns, sometimes viewing them as a personal “weakness” or a result of lack of religious practice (e.g., “devil’s work”) rather than a psychological condition. They were seen as discouraging other people from reaching out for such treatment, help, or resources. One participant shared how older people spoke about mental health treatment:

Looking for help, like, with a therapist...they shouldn’t be sharing their opinions. That’s what I’ve heard, “Why would you see a therapist, if they’re not God?”...that’s what I’ve heard. Like, “Why should you go around telling them your problems?”

Isolation was perceived as a major contributor to older adults’ mental health issues, especially for those who lived in residential settings. This raised concerns during the pandemic because participants reported that resources tailored to older adults living in such settings were not available, especially for those who were “losing loved ones” due to COVID-19. Participants worried about older adults “not being able to see their families” during the COVID-19 pandemic and how they would cope as stigma could be a barrier to obtaining the help they needed.

Younger generations were perceived as having grown up with greater awareness of mental health and, therefore, were affected by stigma in a different way. They were thought to be “more empowered” to openly discuss mental health issues. In particular, participants noted a more welcoming conversation about mental health on social media among younger generations, including during the pandemic. The differences between generations were described by one participant as follows:

I feel like that [the older generation] was like, “We’re not telling anyone our business.” And “This is family business, keep it to yourself.” Whereas like the mid-30s and maybe late 20s, they’re like, “You know what? Let’s talk to somebody, let’s get help, let’s like—We’re not going to suffer in silence.”

Stigma Is Experienced in Unique Ways in Different Racial and Ethnic Communities

Mental health stigma was perceived as experienced variably based on race and ethnicity, which had implications for the context of the pandemic. Participants shared examples of how Black, Latine, Asian, and other minoritized communities faced more mental health stigma in general compared with other communities due to a combination of societal and cultural factors. One Latine participant said the following:

Because I grew up culturally Latino, so, there is a huge stigma around mental health where you couldn’t just say, “Oh, I’m feeling anxious,” or “I’m feeling a little depressed.”

Another participant said that, in their Asian American community, mental health was “heavily stigmatized” and people “don’t tend to like to ask for help.”

The words participants heard being used to stigmatize mental health or people with emotional challenges, such as “crazy” and “weak,” were similar across racial and ethnic minoritized groups. While all people were seen as actively avoiding being labeled as having a mental health issue, those from minoritized communities were perceived as especially cautious because such negative mental health labels could be used as “leverage” against them. Therefore, a winning strategy used by community-based health workers in our sample was to “give them the information without having to use that word [mental health].” Others described avoiding being labeled by addressing physical rather than psychological symptoms. This provided a perspective for addressing mental health by taking care of physical health. One participant explained this as follows:

I think if the focus is not so much on mental illness but mental wellness, mental health, and that connection between the mind and the body, and that it’s all important, and addresses the person as a whole.

Recommendation on Effective Ways to Combat Mental Health Stigma

Participants recommended providing accessible educational resources on mental health to the public. One suggested that stigma could be reduced if mental health was discussed in the same way in which health providers engaged in “teaching someone how cancer works.” Other suggestions included the strategy of individuals openly sharing their personal stories of mental health struggles to dismantle stigma and encourage help-seeking behaviors. One participant explained that we need “to recognize that we all have trauma, and to set the example ourselves.”

In terms of sharing the T4W/Juntos website and other resources, participants suggested that, rather than just directing people where to go, we should share our experiences honestly. One participant recommended saying things such as “This happened to me and I went here to get help. That helped me a lot because I did this” or “I also went through the same situation.”

Participants encouraged efforts to create safe and supportive spaces, especially for marginalized individuals who may face additional stressors, such as “LGBTQAI+ students,” adolescents, and older adults. Participants reported that they would feel more comfortable recommending a warmline where individuals could connect with trained volunteers, therapists, or peers who could listen and provide support rather than “an Excel [sheet] of resources.” While participants endorsed efforts such as the T4W/Juntos website, they recommended investing in “more intersectional conversations” where leaders “that really represent a community” could share things such as “this is me, and this is what I’m going through. This is what I do to deal with it. This is where I go for help.” This approach was perceived to increase “comfort” and “acceptance,” unlike “faceless” things, because it could reduce stigma and negative perceptions of help seeking.
Theme 3 “Feels Like Home”: Experiencing a Sense of Inclusivity and Belonging in T4W/Juntos

Overview

Participants provided robust reports of feeling welcomed and included when visiting the T4W/Juntos website. This sentiment is reflected in 4 properties: feeling included due to welcoming tone and visuals; feeling included due to the substantial, diverse, and quality resources on T4W/Juntos; many languages making T4W/Juntos “more accessible”; and recommending ways to increase inclusivity on T4W/Juntos (Figure 3).

Feeling Included Due to Welcoming Tone and Diverse Visuals

The images and overall tone of T4W/Juntos gave participants a sense of belonging when using the website. They described the site’s imagery as “cheerful,” “friendly,” “bright,” “happy,” “fun,” “calming,” and “light and airy.” One participant explained the following:

...the color scheme and the font, it’s just very inviting and not intimidating. Cause I think finding like health resources or mental health services or any of these topics, they’re very heavy. So, having a page that’s bright and makes it simple and has the cute little icon next to each topic makes it a little more digestible.

Other participants focused on T4W/Juntos’ esthetics. One participant indicated that it was “well-balanced” with “just enough seriousness.” Another participant highlighted the site’s “high production value” in terms of visual and auditory content. The quality of the content was valued, including the mix of both cartoon and real images, the “scenery” in graphics, and the quality of the spoken Spanish in videos.

Participants noted the importance of diversity in T4W/Juntos’ images in helping them feel included. They appreciated seeing the “authentic representation” of various ages, gender identities, abilities, sexualities, and races and ethnicities, among other characteristics, in “these beautiful faces” they saw on the website. Participants indicated that T4W/Juntos “feels like home” because the images were specifically representative of California’s population. One participant stated the following:

I feel like it covered populations and community members across California who would be possibly using the website. And also, just showing that diversity. So, I think that creates a welcoming environment as well if people can see themselves represented in some capacity on the website, especially on the front page.

Unlike other mental health websites that participants described as judgmental or exclusionary, participants felt that T4W/Juntos was not overly “clinical” or “bashing you with some mental illness stigma.” Furthermore, T4W/Juntos’ diverse representation was different from that of other sites where “only one type of person” was represented, which meant that visitors to T4W/Juntos would not “feel like they’re an outsider,” as one participant succinctly explained:

...if there’s nobody on the website that I can identify with, maybe it doesn’t...doesn’t tally to me kind of thing. There’s plenty of opportunity, I think, for anyone to feel like they fit [on T4W/Juntos].

Feeling Included Due to the Substantial, Diverse, and Quality Resources on T4W/Juntos

The quality, quantity, and variety of resources present on the T4W/Juntos site greatly contributed to participants feeling welcome and included. Some expressed appreciation in broad strokes, noting that there were resources for “every ethnicity” and “different ages” and that the T4W/Juntos team “took many different things into account.” Other participants valued the inclusion of resources for specific groups, such as African American individuals, American Indian and Alaska Native individuals, the “LGBTQIA2+ community,” parents, children, and people living with disabilities. Knowing that the resources were intentionally selected for T4W/Juntos mattered. One participant noted that “…the thought put into making [T4W/Juntos] usable or worthwhile to a number of different communities was made and paid attention to.” Another interviewee found it “refreshing to realize that the [T4W/Juntos]
project...had equity kinda built from the top up.” One participant encapsulated this by saying the following:

> It just felt like I was coming to a buffet, a big place to finally like heal, y’know? It was like “Oh, I don’t even have to eat this. There’s like a little bit of that, more of that.” And there’s really—the variety of choices, and the way that it was put, it was very inviting. Also, it was very welcoming, and I left feeling satisfied, but also, I left—like there was stuff that I could share with people. And I did.

“This Is Kinda Cool”: Many Languages Make T4W/Juntos “More Accessible”

Linguistic accessibility was another important aspect of the T4W/Juntos site that made participants feel included. Because of California’s multicultural population, participants believed that T4W/Juntos needed to be offered in multiple languages, especially Spanish and Vietnamese, for it to be considered “culturally appropriate” and widely accessible. One participant stated that “I love that the website already has a few options in different languages...Just thinking about the different audiences [will] make it that much more accessible.” The creation and availability of a Spanish translation contributed to another participant’s feelings of inclusion:

> So, it made me welcome and then it also made me understand more, like I said, because it was in Spanish and English.

Others noted the need for content in more languages; some specified a desire for content in Vietnamese, Indigenous languages (broadly), and Mixtec due to the large population of Mixtec-speaking migrant workers in California.

In addition to being offered in multiple languages, many participants lauded the site for being written in “everyday” and “plain” language that was “easy to understand” and “basic.” Another participant expressed their thought process upon first hearing about T4W/Juntos:

> I was like, “This is kinda cool. Let’s check this out.” It wasn’t something like, “Oh, wait, this is way beyond my expertise, or this is something I don’t fully understand.”

Although many found T4W/Juntos easy to understand, a few respondents said that the language was inaccessible or overly professional. For instance, one person commented that the “writing [in T4W/Juntos] was too academic,” and therefore, it “wasn’t an easy read. It was like reading a textbook or a law book.” Others noted that terms such as “resilience” or “anxiety” made the site feel overly clinical and not intended for the average user.

There was concern that some users with low computer literacy may not be able to use the site. For example, one participant cautioned that some Spanish speakers may not necessarily know the word for “link” in English or Spanish (enlace). Furthermore, several participants who worked with immigrant communities indicated that many in these groups cannot read or write in Spanish or English, which precludes them from making use of the T4W/Juntos site.

**Recommending Ways to Increase Inclusivity on T4W/Juntos**

Some participants advocated for changes or additions that would further widen the net of inclusivity. For example, participants suggested adding information on mental health symptoms and treatment options; trauma and its potential effects; support for basic needs such as housing, rental assistance, and financial support; information about civic engagement (eg, how to register to vote and contacting local officials); and recreational activities such as art classes and book clubs.

Although many participants found the wide breadth of representation on T4W/Juntos to be quite impressive, some wanted even more diverse visual representation. Some perceived a few groups to be conspicuously absent on the site, such as the lack of representation of older adults. One participant stated the following:

> ...older adults have really had a hard time with isolation and access and I don’t really see older adults represented, at all—at all, at all, at like, at all, in this whole thing. Not just the graphics, but the people you’ve chosen to be on the little videos, the images, the content, there’s nothing about older adults, that I found.

Another participant specified the need for more youth representation:

> ...make sure that we have as many opportunities for youth to be able to see themselves talking, working with each other, reaching out, but really knowing that we’re all here for them. And I think that was one of the things...that I would really suggest, is that opportunity.

Other suggestions included adding more representation of men of color, including “African American men, Latino men and Armenian men,” as well as police officers and military veterans as these groups tend to avoid seeking treatment due to stigma against mental health. Others suggested that representation be enhanced with more images or voices of people with disabilities: individuals from the lesbian, gay, bisexual, transgender, and queer community; and actual community members.

**Theme 4 “It’s a One-Stop-Shop”: Judging T4W/Juntos to Be a Desirable and Useful Website**

**Overview**

Participants judged the website as a hub that brought many things together in one place, making it a “one-stop-shop.” For this reason, most described it as a desirable website. Their perceptions reflect 4 main ways in which they experienced the website: perceiving T4W/Juntos as trustworthy, being equipped with a “first step” tool to use and share, finding navigation to be simple and clear, and easily accessing useful information (Figure 4).
Perceiving T4W/Juntos as Trustworthy
Participants valued the website due to the reliable information on it that had been curated from credible sources, which was able to combat misinformation. One participant reported the following:

*I feel like it’s done by UCLA, UC Davis, all names that I really trust...it’s a name that people recognize and that you can trust. So, I have no reservations whatsoever about this website...I know that when I scroll all the way at the end with the different sponsors or collaborations made it legit.*

Seeing the real faces and hearing the real voices of community members in the videos plus the logos of respected agencies on the website enhanced the sense that it was “legitimate.” Trustworthy information was especially desirable to combat the misinformation they reported hearing about in many communities.

Being Equipped With a “First Step” Tool to Use and Share
Participants felt equipped because the website gave them tools necessary not just for learning about resources but also for sharing with clients, friends, and family members. One participant noted how T4W/Juntos was something they had been searching for but had not found:

*And it just was primarily what it felt like, a resource portal. And so, it felt like the right door to go get help, rather than the wrong door. So that’s a good way of putting it. It felt like I had finally opened the right door that I’d been looking for.*

As a tool, T4W/Juntos was judged as helpful because it allowed participants to gain access to needed mental health resources during the pandemic, with options so that clients could start “where they want to” with the goal of receiving help. They liked having access to specialized information that addressed a wide range of topics, such as resources on grief for pregnant women or up-to-date COVID-19 information. They found it desirable because it was designed for the “average person,” so it was useful as a “first step” even for those with no previous knowledge about mental health resources.

Navigating Is Simple and Clear
Participants perceived T4W/Juntos to have appealing features such as inviting colors and fonts, which made the content “more digestible.” The technology functioned smoothly, including the hyperlinks and videos. One participant said the following:

*And it’s really easy to go on there, navigate, and look for information. And it’s also a good way to empower the clients I work with, so they can go and do their own research about any resources they may need in regards to mental health.*

The process of searching for information was clear even when working with groups lacking digital literacy, which they noted was required for some websites. One participant noted the following:

*Honestly, this is one of the pages that I remember my mom and I—even though we were dazed and emotionally exhausted—we were able to understand and get the information because it was all so simple. That was the only thing I can tell you. Despite everything that happened, we saw that if we were going from one place to another looking for information or trying to analyze, sometimes it was very elaborate. We needed something like they say in English, “short and sweet,” not fancy or too negative. Something within the positive and informative things but without being research papers that we had to be reviewing and analyzing, because we didn’t have the capacity to do that. We needed simple and easy to understand information.*

Easily Accessing Useful Information
The ease of using T4W/Juntos to access information was a valued feature. Participants found the website to be a user-friendly “one-stop-shop,” a place where they could find plenty of useful resources to choose from all at once. They preferred this to having to type specific topic words into a search engine to find needed resources individually. Some especially endorsed the feature that allowed them to receive immediate active help through a direct crisis number, whereas others favored the option to receive informational help by downloading materials to read. While participants overall considered T4W/Juntos to be easy to use, some suggested making the pages more “scroll-friendly...like Instagram,” and another participant suggested adding an “emergency exit” button so that users could quickly switch to a different site if needed for safety reasons.
Discussion

Principal Findings

The qualitative findings from telephone interviews as a complement to quantitative surveys suggest that, during the COVID-19 pandemic, the T4W/Juntos website was well received by both community health service providers and community members in the interview sample as a useful, accessible tool, with some concerns noted such as language sometimes being too “professional.” Our findings further suggest that the pandemic catalyzed significant changes in the way people coped, which fueled a shift to digital solutions when other options were suddenly off-limits due to stay-at-home orders. Our participants tended to their own emotional well-being personally and assisted their friends and families, and some also engaged in trying to help clients or patients as well. Their pivot to reliance on technology during the pandemic ranged from finding new techniques for soothing stress to connecting with others in a meaningful way via the internet to meet socialization and support needs. Roommates were crucial for socializing, as was also suggested by Shigeto et al [42], because social distancing limited social contact. Similar to other studies [43-45], our participants found pets to provide companionship and effective ways to cope with the isolation of stay-at-home orders. Notably, as other researchers found, being able to help other people during the pandemic in and of itself gave participants a mood boost [43,46]; this made the T4W/Juntos website even more valuable because participants could share it with others.

Similar to the findings of other studies [47-49], our participants shared that mental health stigma and taboo attitudes had often thwarted attempts to access needed mental health care, and this was especially the case for those from ethno-racially minoritized communities and older adults. However, participants did not perceive the T4W/Juntos site as invoking stigma, judgment, or condescension. They were particularly cognizant of the efforts of the development team to create a site that was neither intimidating nor shaming. They found it to be a digital space that successfully communicated that someone was out there trying to help others in a world that was otherwise shut down due to COVID-19. The collaborative approach to the development of the website may have been why the written text and verbal communication in the site’s videos were described as an example of a positive way to talk about mental health.

To combat stigma through a website, input from potential users, such as our participants, is crucial for design enhancement. As already noted, during the development of T4W/Juntos, input from various members of diverse California communities addressed the making of the website, including the goal of reducing stigma related to mental health [31]. Efforts to reduce stigma require sensitivity to the language used; with T4W/Juntos, we intentionally used neutral, nonclinical language so that experiences such as stress, anxiety, depression, and grief were addressed as normal aspects of life that many of us deal with [31]. Our participants recommended featuring pictures, animations, and videos to showcase resources in a personal way that is colorful and inviting without stigma and that reflects the commitment of the T4W/Juntos team. In addition, several short videos in English and Spanish were created to introduce each section of the T4W/Juntos website with the goal of making users feel more comfortable with the topics; volunteers from diverse California communities served as relatable actors who were filmed during the pandemic via Zoom in their homes. The diverse representation is likely why participants said that the website felt comfortable. We also included links to active warmlines and hotlines so communication with an actual human being was possible through telephone and texting [31,34,35]. However, consistent attention to making these links convenient and prominent on the website is needed over time to maintain a steady focus on reducing stigma. Additional efforts could be made in the future to link to more and different venues available on the internet where diverse community leaders talk about their own emotional health concerns or share what they have found helpful. In addition to the immediate sense of welcome, participants found the extent and variety of the content on T4W/Juntos (ie, the plentiful links to various resources) to communicate supply rather than unmet demand. The sense of options for resources was understood as high accessibility, which somehow also reduced stigma. Participants seemed to relish what they perceived to be a bounty of ready resource links, including content in multiple languages. This, during the pandemic, was appreciated because it was a time when avenues for the typical sources of useful or desirable material were severely reduced.

The sense of belonging reported by participants suggests the profound impact of a culturally competent design in enhancing user engagement with and experience on the website. For example, the colorful look of T4W/Juntos was developed in collaboration with community members [31]. The decisions to give T4W/Juntos an upbeat feel, feature relatable people from diverse California communities in the videos, and provide options in diverse languages were all made collaboratively with community partners. The team’s intentional efforts and commitment to convey diversity resonated with participants, making the website “feel like home.” Furthermore, as the partnering organizations behind the creation of the website were clearly listed on the site for the purpose of transparency regarding who was behind the website, participants perceived it to be a trustworthy tool.

Overall, our findings highlight that the T4W/Juntos website functions as a comprehensive, inclusive, and user-friendly platform for coping with mental health challenges, particularly during the pandemic. It was a web-based “one-stop-shop” due to the culmination of several integrated features that generated positive regard. However, despite its many strengths, there were also suggestions for improvements to the website to further enhance inclusivity. As was suggested by some participants, more resources, pictures, stories, and testimonials are needed to reduce stigma, specifically for older adults; adolescents; lesbian, gay, bisexual, transgender, and queer communities; police officers; and veterans. In terms of functionality, certain adjustments were requested, including a “back” button and a drop-down menu for a better user experience.
Limitations

Some aspects of this study were restrained due to the pandemic. For example, we were only able to interview English- and Spanish-speaking adults. In addition, we used convenience sampling with recruitment based on invites from clients and community partners of staff and providers of community partner agencies. While this approach resulted in a diverse sample, it included community-based health and wellness workers and is not necessarily representative of California residents. Nonetheless, during the COVID-19 pandemic, these voices were extremely valuable and garnered important insights. While 55% (16/29) of the participants indicated that they worked in community-based support or health care, we did not collect specific data on employment status or occupation. We can only assume based on education and other factors that approximately half of our sample were community members not employed in health care. Thus, future research should systematically collect employment data as context for participants’ level of familiarity with health-related resources.

Relatively, the COVID-19 pandemic put limits on potential participants’ ability to engage in a study when they were dealing with other worries. Thus, our sample was diverse in some ways but could have been more reflective of California’s population. For example, we were successful in recruiting 7% (2/29) of participants who were Southeast Asian; however, no participants self-identified as being from South Asian or East Asian communities despite the large numbers in California. Future research should expand recruitment efforts to be inclusive of the many subgroups in the state to bring insight from a more diverse sample.

Conclusions and Future Implications

Our results complement the findings of the quantitative evaluation that showed engagement in website use and an association with reduced depression over time [34]. The results underscore the value of collaborating with members of the target community to have a meaningful impact when designing a digital tool for the public. Specific partner website design suggestions to include videos; language accessibility; diverse representation; and colorful, cheerful visuals contributed to the positive reception of this website. The findings suggest that, while T4W/Juntos has been effective in addressing diverse needs, there are ongoing opportunities to maximize inclusivity and user experience.

First, future studies on mental health website development would benefit from engaging with a diverse sample of the target group and conducting pilot tests to learn more specifically what accessibility means to potential users. Second, the results showed that mental health stigma continues to be an issue, especially among minoritized communities. Hence, resources tailored to such groups must consider what stigma looks like to members of each group and how to address it in the specific context of minoritized communities so that valuable information about mental health will be received and accepted. Finally, participants indicated that the T4W/Juntos website was useful for their personal needs, sharing with loved ones, or incorporating into their work with minoritized communities. The website’s unique features—especially its diverse representation and availability in multiple languages—make it a valuable addition to the mental health resource landscape, and thus, it may be recommended for dissemination throughout the state of California, especially when including input from other diverse populations.

Acknowledgments

The authors would like to acknowledge the community partner agencies in the state of California who collaborated in the development of the Together for Wellness/Juntos por Nuestro Bienestar website and assisted in disseminating information that led to the recruitment of our sample, including African Communities Public Health Coalition, Boat People SOS, Cal Voices, California LGBTQ Health and Human Services Network, Healthy African American Families II, Health Education Council, National Alliance on Mental Illness California, United Parents, and Visión y Compromiso, and our collaborators at Chorus Innovations and University of California, Davis. Funding was obtained from a Federal Emergency Management Agency and Substance Abuse and Mental Health Services Administration grant (which was managed by the California Mental Health Services Authority through the California Department of Health Care Services for COVID-19 crisis intervention prevention; 700-FEMA-2021-UCLA), a California Health Care Foundation grant (award G-31202), and an award from the California Department of Health Care Services (award 21-10458).

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

MVH, KBW, and DFR conceived and designed the study. MVH analyzed and interpreted the data and drafted and revised the manuscript. JL, MPC, and JIM analyzed and interpreted the data and made substantial contributions to drafting and revising the manuscript. DFR collected and participated in the analysis and interpretation of the data. KBW also critically revised the draft for important intellectual content. All authors granted final approval for the version to be published.
Conflicts of Interest
None declared.

References


Abbreviations

T4W/Juntos: Together for Wellness/Juntos por Nuestro Bienestar

Edited by A Mavragani; submitted 14.12.23; peer-reviewed by C Wafern, D Levine; comments to author 15.01.24; revised version received 23.01.24; accepted 31.01.24; published 25.03.24.

Please cite as:

Heilemann MV, Lai J, Cadiz MP, Meza JI, Flores Romero D, Wells KB
Community Members’ Perceptions of a Resource-Rich Well-Being Website in California During the COVID-19 Pandemic: Qualitative Thematic Analysis
JMIR Form Res 2024;8:e55517
URL: https://formative.jmir.org/2024/1/e55517
doi:10.2196/55517
PMID:38526558

©MarySue V Heilemann, Jianchao Lai, Madonna P Cadiz, Jocelyn I Meza, Daniela Flores Romero, Kenneth B Wells. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Best Practices in Evolving Privacy Frameworks for Patient Age Data: Census Data Study

Colin Moffatt¹, BSc, MRes, MSc, PhD; Jonah Leshin¹, BA, MA, PhD
Datavant, San Francisco, CA, United States

Corresponding Author:
Jonah Leshin, BA, MA, PhD
Datavant
44 Montgomery St 3rd floor
San Francisco, CA, 94104
United States
Phone: 1 415 520 1171
Email: jonah@datavant.com

Abstract

Background: Over the previous 4 decennial censuses, the population of the United States has grown older, with the proportion of individuals aged at least 90 years old in the 2010 census being more than 2 and a half times what it was in the 1980 census. This suggests that the threshold for constraining age introduced in the Safe Harbor method of the HIPAA (Health Insurance Portability and Accountability Act) in 1996 may be increased without exceeding the original levels of risk. This is desirable to maintain or even increase the utility of affected data sets without compromising privacy.

Objective: In light of the upcoming release of 2020 census data, this study presents a straightforward recipe for updating age-constrained thresholds in the context of new census data and derives recommendations for new thresholds from the 2010 census.

Methods: Using census data dating back to 1980, we used group size considerations to analyze the risk associated with various maximum age thresholds over time. We inferred the level of risk of the age cutoff of 90 years at the time of HIPAA’s inception in 1996 and used this as a baseline from which to recommend updated cutoffs.

Results: The maximum age threshold may be increased by at least 2 years without exceeding the levels of risk conferred in HIPAA’s original recommendations. Moreover, in the presence of additional information that restricts the population in question to a known subgroup with increased longevity (for example, restricting to female patients), the threshold may be increased further.

Conclusions: Increasing the maximum age threshold would enable the data user to gain more utility from the data without introducing risk beyond what was originally envisioned with the enactment of HIPAA. Going forward, a recurring update of such thresholds is advised, in line with the considerations detailed in the paper.

Introduction

A person’s age is a singular piece of information. It is linked inextricably to an individual, incrementing ceaselessly throughout life and unable to be modified (despite some having tried [1]). Age is a fundamental piece of information that we routinely use to describe or categorize a person. Age values commonly occur in health data, either directly or more usually as implied by the patient’s date of birth (often aggregated by the year of birth). Thus, age is a useful piece of information when looking to match 2 records pertaining to the same individual [2], and, therefore, within an anonymized data set containing person-level information, the presence of an age value contributes to the risk of reidentification.

With microdata records, a rule of thumb is that the relative amount of reidentification risk that any value contributes is inversely proportional to the number of people in the population who share that value. For example, the sex value “male” is low risk as almost half the population shares that same value. This extends to a combination of values from several fields. The
When working with data sets with US patients, the US decennial census data contain group-level counts that can be incorporated into risk calculations. For example, one may use the census to look up the size of the population of all female patients aged 90 years and older (the census value will serve as an approximation due to shifts in population over time). Note that it is the absolute count of patients that is most relevant, rather than the proportion of a population subgroup relative to the full population.

Age values are not represented equally across the range of all ages. It is natural and expected for age cohorts to reduce in size as a function of age due to increased rates of mortality. While historical events such as epidemics of malign diseases can affect cohorts differently and birth rates over time along with emigration and immigration play a role in cohort population size, in general, population counts are decreasingly frequent in older people. Hence, there are, for example, fewer nonagenarians (individuals aged between 90 and 99 years) than there are octogenarians (aged between 80 and 89 years). Thus, especially in the population of very advanced age people, the reidentification risk contributed by age increases year-on-year as the number of individuals who share the same age reduces. To mitigate this within a data set where sensitive information is also present, it is sensible to constrain age values to some maximum. The HIPAA (Health Insurance Portability and Accountability Act) of 1996, in its Safe Harbor approach to the deidentification of data sets [3], gives a maximum age of 90 years to which all higher values should be lowered.

The risk of reidentification, or “disclosure risk,” is related to the amount of utility that a data set of personal information contains. As the risk is reduced, so is the value of the data set in terms of its usefulness. Aggregation or constraining, while reducing risk, will inherently reduce the granularity of information the data contain, since these methods effectively introduce error to the raw data values, and the quality of any statistical inference based upon the data will suffer as a result. Thus, there is a trade-off between reducing the disclosure risk of a data set and maintaining a level of utility that is sufficient to address questions of interest.

Enabling more granular year-of-age information is beneficial for studying health outcomes in the older populations. Through improved diets, better medical care, and generally healthier lifestyles, people are living longer. In addition to these environmental factors, a spike in birth rates following World War II into the early 1960s has also contributed to the high proportion of individuals aged over 60 years currently living in the United States [4]. With the growth of the older population in the United States and abroad [5], the development of efficacious drugs and treatments for an aging population is becoming increasingly important [6].

This prompts the question of whether the age threshold given by HIPAA is still appropriate. While 90 years of age may have been a good choice 2 and a half decades ago, is it still a good choice today, or could it be raised, and data set utility improved as a result? In consideration of the ways in which an age value contributes to the disclosure risk of a data set, we make the case that this threshold can be increased without reducing disclosure risk.

Methods

Overview

Data from the previous 4 decennial US population censuses were acquired [7] and investigated. As HIPAA was enacted in 1996, presumably, the age-constraining guidance of age of 90 years was based upon data from the most recent census, which was in 1990. At the time of writing, the most recent census is that of 2010, so that is used in support of updated guidance through comparison to the corresponding 1990 census counts.

Analyses used data from the census’s population “PCT12” tables that consist of population counts grouped by combinations of state, sex, and single year of age (at the census year). Using these counts, we can constrain subsets of these 3 variables while aggregating across others. For example, we can sum the population counts across all states for all ages of 92 years and older, grouping by sex (ie, one sum for male and another for female individuals).

To illustrate trends over time, we performed similar computations on data from the 1980 and 2000 censuses. Having obtained these counts, we then compared them across different census years and generated illustrative plots. Findings are presented in simple graphs along with distributions across the United States illustrated by choropleth maps.

All analysis was performed in the computational language R (R Foundation for Statistical Computing). Given the small data storage and compute costs and the public availability of the data, we were able to perform all the analysis on personal computers.

Ethical Considerations

The exploration of reidentification risks associated with varying specificities of age information for older individuals carries certain ethical implications. Although the data under study were publicly available, using it in combination with proprietary data sets and record linkage technologies may elevate the risk of reidentification. The analysis herein was carried out in a technical environment that did not contain any data sets to which the census data could be linked.

The degree of privacy protection afforded to any proprietary data set must be assessed in the context of possible linkages to readily accessible data sets. In fact, HIPAA stipulates those linkages to “reasonably available information” be accounted for in statistical determinations of privacy risk [3].
This analysis does not require review from an institutional review board as per the US Department of Health & Human Services regulations for the protection of human participants (45 CFR 46.104(d)) [8]. In particular, this is “Research that only includes interactions involving…survey procedures…” and meets the criterion that “The information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects.”

**Results**

**Overall US Population and Sex Considerations**

The 1990 census showed that 1.02 million people were aged 90 years or older (Figure 1). In the 2010 census, this had risen by approximately 84% to 1.87 million. The increase was around 40% from 1990 to 2000 (a similar figure to the increase between 1980 and 1990) but smaller at 30% from 2000 to 2010. An increase was evident in both male and female individuals, although their increases were neither similar nor uniform.

The figure of 1.02 million people who were aged 90 years or older was presumably the number that was deemed acceptable for HIPAA when it was enacted in 1996. In 2000, the census data began to include numbers for individuals aged between 90 and 99 years, so we could look up at which age, in 2010, the population was at least this size. With reference to Figure 2, we see that at the age of 92 years and above, the count in 2010 exceeded the required 1.02 million, but at the age of 93 years, it was below this number.

Since the numbers of male and female individuals at any age are far from equal (Figures 1 and 2), it can be argued that it is more sensible to consider the sexes separately, especially since sex values are normally present within a health data set. As there are fewer male than female individuals, presumably, the number of male individuals at least 90 years of age in the 1990 census (n=244,000; the blue dashed line in Figure 2) is a sufficiently large equivalence class size for either sex. In 2010, to achieve the same number, an age of 92 years would be required for male individuals, but for female individuals, it would be 95 years (since the number of people aged 96 years or older falls short by a few thousand).

**Figure 1.** Male and female individuals (in 1000s) at least 90 years of age over the previous 4 decennial US population censuses.
Figure 2. Male individuals, female individuals, and individuals of both sexes (in 1000s) at and above the given ages in the 2010 census. The dashed lines show the numbers of individuals who were aged 90 years and older in the 1990 census.

State Level Considerations
A state-level examination of age is also worthwhile since state data are often included in health data sets. Results can best be illustrated by choropleth maps. Figure 3 shows shading based upon the proportion of people aged 90 years and older in the 1990 census. Alaska had the lowest proportion of people (0.7 per 1000) aged 90 years and older, followed by Nevada (1.8) and Utah (2.4), with the Midwest (Iowa=7.2) and Northeastern regions generally having higher values, along with Florida (4.9). Figure 4 shows the numbers of those aged 90 years or older in 2010 compared with 1990. Nevada has seen the greatest increase (411%), followed by Hawaii (402%), with all states having shown an increase of at least 134% (Nebraska) and those in the Midwest generally seen a smaller increase.

With regard to absolute counts, among all states, Alaska had the smallest number of nonagenarians in 1990 with 358 (Wyoming was second with 1527). In 2010, although Alaska still had the smallest such population, the number increased by a factor of just over 4 to 1438 (Wyoming was second with 2899). If we assume that the decline in population by individual age year in Alaska follows a similar trend to that of the overall US population (Figure 2), it stands to reason that a maximum age cutoff of 93 years would maintain a level of anonymity at least equivalent to what the age-constraining threshold of 90 years achieved in 1990.
Figure 4. Choropleth map of the percentage change in the number of people aged 90 years or older between 1990 and 2010 by state. Alaska and Hawaii are shown (not to scale) to the left of the contiguous states.

Discussion

Recommendations for the Present Day

The increase in longevity seen in the US population over recent decades means that constraining ages to 92 years would achieve at least as low a level of disclosure risk, based on the 2010 census, as 90 years did when it was introduced in HIPAA in 1996. This is desirable to implement as it effects an improvement in the utility of the data set if one accepts that keeping numbers “round” is an insufficient reason for continuing to use 90 years. Furthermore, if constraining is done differentially for the sexes, as is sensible if sex information is present in a data set, since female individuals tend to live longer, then female individuals’ ages can be constrained to 95 years, based on the 2010 census, and achieve the same level of risk protection as male individuals in 1990.

The increases in the proportions of the “oldest old” have not been uniform across the United States, with those underrepresented in 1990 going some way to “catch up” in 2010. HIPAA’s Safe Harbor method allows state values to be retained (along with 3-digit zip codes), so presumably constraining age to 90 years in 1990 was sufficient to have low enough risk levels in each state individually. Based upon the subsequent changes in Nevada and Alaska, there are strong grounds for raising the threshold to 93 years.

In addition to states, Safe Harbor also references 3-digit zip codes. This geographic unit consists of all zip codes with a fixed first 3 digits. Safe Harbor stipulates that 3-digit zips are permissible, provided that the area has a population greater than 20,000 according to the latest available census. It is worth noting that this threshold is based on the dynamic count of current census data rather than an underlying list of static 3-digit zips that are not permitted. This approach is consistent with our recommendation of basing age constraints on current population counts rather than on a static age threshold.

Our recommendation for raising the age-constraining threshold above 90 years is ultimately predicated on the assumption that population counts that combine age with other variables permitted by Safe Harbor would be at least as great as they were at the time HIPAA was enacted. For example, in 3-digit zips with populations above 20,000, we would assume that the count of individuals aged 93 years exceeds the corresponding count from 1990.

Future State

With the release of a new census every 10 years, one may perform an analysis of this sort and provide updated recommendations. It is important, however, to consider the implications of making such updates over time. One consequence of decennial updates is that a reader of the data set must be aware of the age threshold that was used in order to properly analyze the data. Even as age thresholds are likely to increase over time, some health data sources will be late to adopt due to operational constraints. However, in the event that the age threshold decreased, data sources would be required to modify their age-binning procedures.

The emergence of privacy-preserving technology has enabled aggregate-level counts to be computed with statistical privacy guarantees [9,10]. When using counts that were computed through a privacy-preserving method that may alter the true count value, one must understand the quantitative framework used in the computation and modify the methodology accordingly. For example, when using a population count to derive an age-constrained value, one might use the lower end of a 1-sided 99% CI associated with the purported population count.

Due to these technology considerations, special care will be needed when using population counts from the 2020 census. In 2020, population counts were computed using differential privacy [11], which introduces an element of variance into the results. In larger geographic regions, the reported population count of a group with a given combination of variables such as...
Further Considerations

It is routine for personal identifiable information to be replaced by a pseudorandom key that serves as a patient ID, enabling data sets to be linked without exposing an individual’s personal identifiable information. Such linking facilitates the creation of longitudinal data sets, which introduce additional considerations into privacy risk assessment.

For certain fields of a transient nature, such as state of residence or disease diagnosis, the presence of new values over time may enable an attacker to more effectively triangulate a patient’s information, elevating the risk of reidentification. A similar phenomenon can occur with a maximum age cutoff: if a data set from 5 years ago shows an individual to be 89 years old, yet a current data set shows the same individual (as identified by the same “linking token”) to be 90 years old, it is simple enough to deduce that they are actually 94 years old but have been constrained to 90 years.

Another consideration is that changes to the threshold pose a challenge for naive patient matching strategies that may use age or year of birth value as a component of a patient ID used to match patients across data sets (for example, using year of birth in combination with a hash of first name, last name, and address). It is advisable to instead include a patient’s actual (nonbinned) year of birth as part of the hashing input.

In practice, the only instance where current census data are referenced in HIPAA’s Safe Harbor criteria is the list of regions determined by the first 3 digits of a zip code. Any such region with a population under 20,000 is deemed to pose an unacceptable reidentification risk and is not permitted to be part of a Safe Harbor data set [15]. Any change to the Safe Harbor criteria, such as a dynamic age-constraining threshold, would need to come from the US Department of Health and Human Services.

Conclusion

Age is a critical variable when it comes to health data. While having a maximum age threshold is valuable for protecting patient privacy, such a threshold limits the granularity of insights into patient outcomes for the older population. Therefore, the threshold should be increased to the extent that it is possible to do so while staying below HIPAA’s original level of permitted risk.

The decennial release of updated demographic statistics is a sensible time to reevaluate accepted thresholds for variable constraints such as age. When using the census as a reference, thoughtful consideration must be given to both biases in the data collection as well as the use of privacy-preserving technologies.

Acknowledgments

The authors would like to thank Daniel Barth-Jones for his review and helpful insight around regulatory considerations and the use of differential privacy in the census.
Data Availability
The data sets generated during and/or analyzed during this study are available from the US Census Bureau repository [7].

Conflicts of Interest
None declared.

References
7. Explore census data. United States Census Bureau. URL: https://data.census.gov/ [accessed 2023-01-04]
15. Krzyzanowski B, Manson SM. Twenty years of the health insurance portability and accountability act safe harbor provision: unsolved challenges and ways forward. JMIR Med Inform 2022;10(8):e37756 [FREE Full text] [doi: 10.2196/37756] [Medline: 35921140]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
Adapting mHealth Interventions (PrEPmate and DOT Diary) to Support PrEP Retention in Care and Adherence Among English and Spanish-Speaking Men Who Have Sex With Men and Transgender Women in the United States: Formative Work and Pilot Randomized Trial

Albert Y Liu1,2, MD, MPH; Cat-Dancing Alleyne1, BA; Susanne Doblecki-Lewis3, MD; Kimberly A Koester2, PhD; Rafael Gonzalez1, BA; Janie Vinson1, BA; Hyman Scott1,2, MD, MPH; Susan Buchbinder1,2, MD; Thiago S Torres4, PhD

1Bridge HIV, San Francisco Department of Public Health, San Francisco, CA, United States
2Department of Medicine, University of California, San Francisco, San Francisco, CA, United States
3Department of Medicine, University of Miami Miller School of Medicine, Miami, FL, United States
4Instituto Nacional de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

Corresponding Author:
Albert Y Liu, MD, MPH
Bridge HIV
San Francisco Department of Public Health
25 Van Ness Avenue, Suite 100
San Francisco, CA, 94102
United States
Phone: 1 628 217 7408
Fax: 1 415 431 7029
Email: albert.liu@sfdph.org

Abstract

Background: A growing number of mobile health (mHealth) technologies are being developed to support HIV preexposure prophylaxis (PrEP) adherence and persistence; however, most tools have focused on men who have sex with men (MSM), and few are available in Spanish. To maximize the potential impact of these tools in reducing gender and racial/ethnic disparities and promoting health equity, mHealth tools tailored to Spanish-speaking people and transgender women are critically needed.

Objective: The aim of this study is to adapt and tailor 2 mHealth technologies, PrEPmate and DOT Diary, to support daily PrEP adherence and persistence among Spanish-speaking MSM and English- and Spanish-speaking transgender women and to evaluate the feasibility and acceptability of these tools.

Methods: PrEPmate, an interactive, bidirectional, text messaging intervention that promotes personalized communication between PrEP users and providers, and DOT Diary, a mobile app that promotes self-management of PrEP use and sexual health through an integrated electronic pill-taking and sexual activity diary, were previously developed for English-speaking MSM. We conducted 3 focus groups with 15 English- and Spanish-speaking transgender women and MSM in San Francisco and Miami to culturally tailor these tools for these priority populations. We then conducted a 1-month technical pilot among 21 participants to assess the usability and acceptability of the adapted interventions and optimize the functionality of these tools.

Results: Participants in focus groups liked the “human touch” of text messages in PrEPmate and thought it would be helpful for scheduling appointments and asking questions. They liked the daily reminder messages, especially the fun facts, gender affirmations, and transgender history topics. Participants recommended changes to tailor the language and messages for Spanish-speaking and transgender populations. For DOT Diary, participants liked the adherence tracking and protection level feedback and thought the calendar functions were easy to use. Based on participant recommendations, we tailored language within the app for Spanish-speaking MSM and transgender women, simplified the sexual diary, and added motivational badges. In the technical pilot of the refined tools, mean System Usability Scale scores were 81.2/100 for PrEPmate and 76.4/100 for DOT Diary (P=.48), falling in the “good” to “excellent” range, and mean Client Satisfaction Questionnaire scores were 28.6 and 28.3 for
PrEPmate and DOT Diary, respectively (maximum possible score=32). Use of both tools was high over the 1-month pilot (average of 10.5 messages received from each participant for PrEPmate; average of 17.6 times accessing the DOT Diary app), indicating good feasibility for both tools.

Conclusions: Using a user-centered design approach, we culturally tailored PrEPmate and DOT Diary to support daily PrEP use among Spanish-speaking MSM and English- and Spanish-speaking transgender women. Our positive findings in a technical pilot support further testing of these mHealth interventions in an upcoming comparative effectiveness trial.

(JMIR Form Res 2024;8:e54073) doi:10.2196/54073

KEYWORDS
preexposure prophylaxis; PrEP; Spanish-speaking; Latino; transgender; men who have sex with men; mobile health; mHealth; HIV prevention; HIV; technology; formative; development; mobile technology, mobile app; text-messaging; SMS; app; application; USA; United States; health equity; mHealth tool; tool; acceptability; self-management; pilot; support

Introduction
Preexposure prophylaxis (PrEP) with a daily pill of Truvada (tenofovir disoproxil fumarate/emtricitabine) [1-6] or Descovy (tenofovir alafenamide/emtricitabine) [7] has demonstrated high efficacy for HIV prevention, and these are the most commonly prescribed PrEP regimens in the United States [8]. Although more people are initiating PrEP nationally each year [9], retention in PrEP care has been poor. In clinical practice, 37% to 62% of patients discontinue PrEP by 6 months [10-13], with higher discontinuations among people of color [4,14-16], youths [15,17], and the publicly insured (Medicaid) [18]. Reasons for discontinuation include low self-perceived risk, difficulty taking a daily pill, and cost or insurance lapses [17,19,20]. As many individuals who stop PrEP continue to be exposed to HIV, with a number of seroconversions occurring after PrEP discontinuation [10,13,19,21-23], the lack of PrEP persistence has emerged as one of the critical challenges in PrEP implementation [24]. Strategies to support ongoing PrEP use are critically needed to address HIV disparities and maximize the public health impact of PrEP.

Mobile technologies are increasingly being used to support preventive health behaviors [25-29]. Sexual and gender minority individuals may particularly benefit from mobile health (mHealth) interventions that can increase health-related communication and care access and help address stigma to reduce health disparities. Our team has developed 2 innovative mHealth strategies to support PrEP retention in care and adherence among gay, bisexual, and other men who have sex with men (MSM). First, PrEPmate is an interactive, bidirectional, text messaging intervention between PrEP users and their providers that supports PrEP use through its personalized communication component [30]. In 2016, we completed a randomized control trial and demonstrated increased retention in care and PrEP adherence among young MSM in Chicago [31]. Second, DOT Diary is a mobile app that promotes self-management of PrEP use and sexual health through an integrated electronic pill-taking and sex diary that delivers real-time feedback on the level of PrEP protection. The app includes a sexual risk calculator that provides personalized feedback on the overall level of HIV risk on and off PrEP and gamification components to increase app engagement. In 2018, we completed a pilot clinical trial evaluating DOT Diary among young MSM, demonstrating higher levels of feasibility, acceptability, and adherence [32].

Although each mHealth approach has demonstrated high acceptability or preliminary impact in clinical trials, these technologies were originally developed for English-speaking MSM. In the United States, Hispanic/Latino people are disproportionately affected by HIV, accounting for 29% of new diagnoses in 2021 [33], and there is currently a gap in access to HIV prevention interventions tailored for Spanish-speaking populations. To prepare for broad-scale implementation, these mHealth tools need to be tailored to Spanish-speaking MSM and English- and Spanish speaking transgender women, 2 populations who are at risk for HIV acquisition and could benefit from additional PrEP support. In this study, we conducted a user-centered design study to elicit preferences to tailor both interventions to support daily PrEP use among Spanish-speaking MSM and English- and Spanish-speaking transgender women in the United States. We then built English and Spanish versions of PrEPmate and DOT Diary into scalable technology platforms and conducted a technical pilot to optimize functionality of both mobile tools. This work was conducted in 2 geographically diverse sites in the United States.

Methods

Study Population
Eligible participants were cisgender men or transgender women who were aged 18 years or older, were HIV-negative by self-report, were currently taking PrEP or discontinued PrEP within the past year, owned a smartphone, were willing and able to receive SMS text messages and download the DOT Diary app to their phone (technical pilot only), spoke Spanish (for Spanish focus groups, technical pilot) or English (for English focus groups, technical pilot), and were willing to turn on their video during the focus group (for virtual focus groups).

Participants were recruited across 2 locations/sites: the San Francisco Bay Area in California (Bridge HIV within the San Francisco Department of Public Health) and the greater Miami area in Florida (University of Miami). Recruitment methods included clinic-based recruitment; online and social media strategies (eg, Facebook, Instagram, Craigslist, Grindr, Scruff); distributing posters, flyers, and palm cards advertising the study; and direct outreach at local venues, including community-based organizations and community events. Additionally, we recruited
from former study participants who consented to be contacted about future research.

**Interventions**

Development and prior testing of PrEPmate and DOT Diary in English-speaking MSM have been described previously [31,34-35]. Briefly, PrEPmate is a text messaging–based intervention grounded in the information, motivation, behavioral skills theory of behavior change [36]. PrEPmate promotes personalized communication between patients and clinic staff through interactive weekly “check-in” messages asking participants how PrEP is going, allowing staff to identify patients needing more help in taking PrEP, and customized daily pill-taking reminder messages [30]. Additionally, the platform provides daily pill-taking reminders in the form of fun facts, affirmations, quotes, and jokes and supports 2-way communication between patients and PrEP navigators, including reminders for upcoming clinic visits. As part of onboarding, participants are provided links to key information about PrEP (PrEP Basics) [37] and video testimonials of peers taking PrEP.

Dot Diary is a mobile phone app that integrates an electronic pill-taking and sex diary and delivers real-time feedback on PrEP protection. Using the self-management model [38] to increase self-efficacy and patient empowerment, participants log daily PrEP pill-taking and sexual behaviors in the app, which then provides real-time feedback on the level of protection achieved from PrEP (high, medium, low) and customized instructions on doses of PrEP needed to achieve or maintain high protection (Figure 1). Participants can also view a weekly and monthly calendar and a summary page displaying the number of doses in the past 30 days and proportion of sex acts covered by PrEP.

**Figure 1.** Screenshot of the DOT Diary app.

Prior to initiating focus groups, all content in both tools was translated into Spanish through a multistep process. First, a professional translation service was contracted to conduct the initial translation. Our study team provided the translation services with guidelines and examples of language preferred by the community. After receiving the translated text, several native Spanish-speaking research staff reviewed the translations for cultural appropriateness and sensitivity and revised the language to ensure a community-focused perspective. As there are regional differences in Spanish language colloquialism and slang, staff discussed and agreed upon language that could be more broadly acceptable and meet local community standards across different regions. To tailor content for transgender women, additional content was developed and added by research staff, including research staff who identify as transgender.

**Focus Groups for Adapting Interventions**

We conducted 3 focus groups with Spanish-speaking MSM and English- and Spanish-speaking transgender women (1 each) in San Francisco and Miami (15 participants total). Eligible participants were consented in person or online using an information sheet prior to initiation of the focus group. Focus groups were conducted using a Health Insurance Portability and Accountability Act–compliant videoconferencing program (Zoom) or in person in a private room. Using a discussion guide, we elicited preferences on intervention content (eg, tailored text messages, dosing reminders), language, layout, usability, and
functionality, first for PrEPmate, then for DOT Diary with a focus on tailoring these tools for Spanish-speaking MSM and English- and Spanish-speaking transgender women. Screenshots of the app and text message examples were shared using the share screen function in Zoom. At the end of the group, participants completed a brief postgroup questionnaire on demographics, technology use, and initial impressions of both mHealth PrEP support tools. Focus groups were led by 2 research staff members trained in qualitative methods and user-centered design principles and fluent in English or Spanish (for Spanish focus groups). Groups lasted 90 minutes to 120 minutes and were audio recorded and professionally transcribed and translated (for Spanish focus groups) verbatim.

Guided by user-centered design principles [39], we used rapid qualitative methods to analyze the focus group data [40]. After each focus group, key findings from the discussion were summarized in a debrief note. Data captured in this template were focused on information needed to refine and adapt the interventions for our priority populations. Upon receipt of transcripts from the focus groups, members of the team reviewed the transcripts and captured additional details to ensure completeness. These data were used to refine the app before the initiation of the technical pilot.

**Technical Pilot to Assess Usability and Acceptability of the Adapted Interventions**

After finalizing revisions to PrEPmate and DOT Diary, we optimized the functionality of PrEPmate and DOT Diary in a 1-month pilot among 21 MSM and transgender women daily PrEP users recruited in San Francisco and Miami. Eligible participants attended an in-person or online screening visit where they were consented in person or via Docusign and completed a baseline survey to assess demographics, sexual practices, and use of technology. Eligible participants were enrolled, randomized 1:1 to receive either PrEPmate or DOT Diary, and provided access to the intervention. Participants completed in-person or online surveys at baseline and 1-month visits and a brief qualitative exit interview on the acceptability of the intervention at the 1-month visit only. At baseline, participants responded to questions on demographics, sexual behavior, PrEP use, and mobile phone and app use. At the 1-month visit, participants completed the System Usability Scale (SUS) [41], a 10-item instrument that evaluates usability and acceptability of the interventions (scores ranging from 0 to 100), and the Client Satisfaction Questionnaire (CSQ-8) [42], an 8-item instrument that provides scores ranging from 8 to 32. We compared SUS and CSQ-8 scores according to the study arm using t tests. Participants were also asked to rate the components of both interventions, complete questions on social benefits and harm, and complete the PrEP Adherence Self-Efficacy Scale (PrEP-ASES) [43]. Participants were reimbursed US $75 for each visit completed. We evaluated app analytics (including logins to the app and use of different app components) for both interventions. Zoom audio recordings of interviews were transcribed verbatim. Interview content was summarized on a debrief template based on the interview guide, and key quotes were also extracted. Additional team members reviewed the transcript of the audio recording and compared it with the debrief report to ensure completeness, and the original interviewer reviewed and validated any changes. Content of the debrief template for each interview was added to a Microsoft Excel file to create a matrix, and rapid qualitative methods were used to analyze interview data [40]. Each participant debrief template was entered in a single row, and each column addressed different domains and feedback on different components of the interventions. Common themes in each column were then synthesized as key findings.

**Ethical Considerations**

All study procedures were approved by the University of California San Francisco Institutional Review Board (IRB# 20-30256). Focus group participants were provided an information sheet in English or Spanish describing study procedures and risk and benefits of study participation. Technical pilot participants provided written informed consent prior to initiation of study procedures. Focus group transcripts were transcribed in a way to remove any identifying information. Survey data for both phases only contained de-identified data and were coded by a subject number. A Business Associates Agreement was established with both PrEPmate and DOT Diary developers to ensure confidentiality and security of data. Focus group participants were paid US $75 for completing the focus group, and technical pilot participants were paid US $75 for completing each visit (US $150 total).

**Results**

**Findings From the Focus Groups**

Demographics of the 15 focus group participants are shown in Table 1. Mean age was 37.9 (SD 9.4) years; 9 (56%) self-identified as transgender woman, and 6 (38%) identified as cisgender men. A substantial proportion self-identified as Latino(a) or Hispanic (6/14, 43%) and reported currently taking PrEP (11/15, 73%), having health insurance (11/15, 73%), having a primary care provider (13/15, 87%), and using iOS (7/15, 47%) or Android (6/15, 40%) mobile phones.
Table 1. Sociodemographic characteristics and preliminary acceptability of DOT Diary and PrEPmate among English and Spanish-speaking men who have sex with men and transgender women focus group participants in San Francisco, California and Miami, Florida (N=15).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Transgender woman</td>
<td>9 (56)</td>
</tr>
<tr>
<td>Cisgender man</td>
<td>6 (38)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>37.9 (9.4)</td>
</tr>
<tr>
<td><strong>Age (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>3 (20)</td>
</tr>
<tr>
<td>30-39</td>
<td>5 (33)</td>
</tr>
<tr>
<td>≥40</td>
<td>7 (47)</td>
</tr>
<tr>
<td><strong>Race/ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Latino/a/x or Hispanic</td>
<td>6 (43)</td>
</tr>
<tr>
<td>White</td>
<td>5 (36)</td>
</tr>
<tr>
<td>Black</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Native American</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>PrEP use, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Prior PrEP use</td>
<td>4 (27)</td>
</tr>
<tr>
<td>Currently taking PrEP</td>
<td>11 (73)</td>
</tr>
<tr>
<td><strong>Health insurance (yes), n (%)</strong></td>
<td>11 (73)</td>
</tr>
<tr>
<td><strong>Primary care provider (yes), n (%)</strong></td>
<td>13 (87)</td>
</tr>
<tr>
<td><strong>Mobile phone operating system, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>iPhone (iOS)</td>
<td>7 (46)</td>
</tr>
<tr>
<td>Google/Android</td>
<td>6 (40)</td>
</tr>
<tr>
<td>Mobile phone without smartphone features (such as easy internet and email access)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Nokia smartphone (eg, E62, E71x)</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>Mobile phone plan, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Contract billed monthly</td>
<td>13 (81)</td>
</tr>
<tr>
<td>Shared plan with someone</td>
<td>2 (13)</td>
</tr>
<tr>
<td><strong>Mobile phone plan includes data/internet use, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Set number of included text messages each month</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Unlimited text messaging</td>
<td>14 (93)</td>
</tr>
<tr>
<td><strong>Mobile phone temporarily disconnected (past 12 months), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>11 (73)</td>
</tr>
<tr>
<td>Once</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Twice</td>
<td>1 (7)</td>
</tr>
<tr>
<td>I don’t know</td>
<td>1 (6.7)</td>
</tr>
<tr>
<td><strong>Mobile phone lost or stolen (past 12 months), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>12 (80)</td>
</tr>
<tr>
<td>Once</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Twice</td>
<td>2 (13.3)</td>
</tr>
<tr>
<td><strong>Would use DOT diary, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes, definitely</td>
<td>9 (60)</td>
</tr>
<tr>
<td>Yes, I think so</td>
<td>5 (33)</td>
</tr>
</tbody>
</table>
For PrEPmate, participants liked the “human touch” in the messaging and felt like someone cared about them, although a few participants were unclear if the messages were an automatic response from the system versus a reply from a staff member and said they were unsure if engaging with PrEPmate would lead to communicating with a person. They recommended explaining the messaging system and who would be responding to messages during the onboarding process. They also thought PrEPmate would be helpful for scheduling appointments and answering questions. Participants liked the daily pill-taking reminder messages, especially the fun facts, gender affirmations, transgender historical facts, and jokes, and thought these reminder messages would be most helpful for those who have not yet established their pill-taking routine. In general, transgender women participants liked the tone and language of the trans messages, although they commented that one message related to “two-spirit identity utilized an outdated term” and recommended removing or rephrasing this message. They also recommended adding additional trans-affirming messages and those on the history of transgender people, as well as messages providing information about COVID-19. Spanish-speaking participants suggested text messages in Spanish, such as “Una pastille al día, te puede salvar la vida” (“one pill a day can save your life”). Regarding frequency of messaging, participants suggested that a text be sent out after an initial 2-week period asking if they would like to continue receiving messages. Participants varied in their message frequency preference, including once a week, twice a week, or once every 2 weeks. Participants felt the links to PrEP facts and brief videos were helpful, particularly the video on how PrEP works in the body that was available in English and Spanish. They recommended additional videos highlighting stories and testimonials from people like them taking PrEP, which could help reduce stigma.

For DOT Diary, participants had generally positive feedback for the app and particularly liked the adherence tracking and protection level features, including the bright colors indicating protection level (red, yellow, green). They also thought the calendar view was clear and understandable and understood what the different symbols represented. Participants felt there were too many questions in the sexual diary and recommended simplifying the information being collected. One participant stated:

...say all I really want out of this, personally, would be like, you know, did I have sex on this date. And then, if it’s relevant to my risk, like maybe what types of acts I engaged in with an option to possibly put a name or more information in a notes field. But like I can see how people would look at this and felt like it was collecting too much information.

In particular, participants expressed concern that entering specific information about a partner and rating them could negatively impact the relationship. Although a participant suggested that it might be helpful if the app could track dosing of other medications, others felt that the app should just stick to reminders and information about PrEP, as adding the tracking of additional medications may overcomplicate the app. Regarding the sexual risk assessment tool component, only a few participants found this would be a helpful tool—possibly those having a larger number of sexual partners with different people on a regular basis. When asked about the sex trends and insights page, some participants felt that it looked too much like a business plan, with too many graphs and being difficult to understand; these participants recommended that the language and appearance be simplified to make it more accessible. Finally, participants appreciated the idea of earning badges:

Yeah. I like the—the way I seem them is I seem the as, you know, achievement, like, for example, Xbox. You do so many things on a particular game, you get an achievement.

This gamification acts as a motivator and affirms that a person is doing well.

Almost all participants reported that they would use PrEPmate (14/15, 93%) and DOT diary (14/15, 93%). For PrEPmate, the daily pill-taking reminder messages were considered the most useful component (13/15, 87%), and the weekly check-in messages were considered both among the most useful (6/15, 40%) and least useful (6/15, 40%) features in PrEPmate. For DOT Diary, the home screen circle indicating level of protection and the ability to track sexual partners and encounters in the diary were considered the most useful components (10/15, 67% for both), and the “badges” were considered the least useful (5/15, 33%).
Summary of Changes Made to PrEPmate and DOT Diary After Focus Groups

Refinements made to our mHealth tools based on feedback from focus groups are summarized in Table 2. For PrEPmate, we revised our onboarding messaging to clarify that, although the reminder and check-in messages were automated, an actual staff member would respond to messages sent by participants in the system. We also refined the language of trans-specific messages and messages in Spanish based on feedback from participants and added additional content to better tailor the app for these populations. Added examples of transgender affirmative messages and the history of transgender people include:

- “We resist, actively, every time we affirm ourselves.”
- “#StayAffirmed. We got this! Time for our pill of the day.”
- “For generations, South Asian hijra (trans) communities have ‘adopted’ young boys who were rejected by or fled from their biological families.”

We added messages with links to additional video content based on focus group feedback. As most participants appreciated having an initial 2 weeks of reminder messages along with weekly check-in messages, we decided to keep this messaging frequency for the technical pilot.

Table 2. Refinements made to PrEPmate and DOT Diary based on feedback from focus groups.

<table>
<thead>
<tr>
<th>App</th>
<th>Refinements</th>
</tr>
</thead>
<tbody>
<tr>
<td>PrEPmate</td>
<td>Clarified onboarding procedures and who will be responding to messages</td>
</tr>
<tr>
<td></td>
<td>Refined trans-specific messages and Spanish language messages based on feedback</td>
</tr>
<tr>
<td></td>
<td>Added links to additional video content</td>
</tr>
<tr>
<td>DOT Diary</td>
<td>Reduced number of questions in sexual diary, removed rating feature</td>
</tr>
<tr>
<td></td>
<td>Removed sexual risk assessment component and simplified sex trends and insights page</td>
</tr>
<tr>
<td></td>
<td>Added motivational badges</td>
</tr>
</tbody>
</table>

For DOT Diary, participants liked the core features of the app, so we kept the adherence tracking and protection level feedback intact, along with the weekly and monthly calendars summarizing dosing and sexual activity. Based on feedback that the sexual diary was too detailed, we reduced the number of questions and took out the partner rating feature, as several participants found that off-putting. To keep the app streamlined and simple, we decided against adding the ability to track other medications. We also removed the sexual risk assessment tool component of the app given the low level of interest in that feature and simplified the sex trends and insights page based on participant input. Based on participant feedback and input from our Spanish-speaking staff, we revised Spanish-language text in DOT Diary to be more culturally appropriate for sexually active MSM. For example, the phrase for “I rimmed them” was changed from “Le di un beso negro” to “Le comi el culo,” and the phrase for “I jerked them” was changed from “Le hice una paja” to “Le corri la paja.” Finally, we added a series of badges that participants could earn through use of the app, which could be viewed in a “badge collection” section, based on focus group feedback that gamification could increase engagement and dosing in the app.

Findings From the Technical Pilot

Overview

From July 2021 to August 2021, 21 participants were randomized to the DOT diary (n=11) and PrEPmate (n=10) arms: 12 (57%) in Miami and 9 (43%) in San Francisco. Table 3 describes the baseline characteristics of the technical pilot participants by study arm. Overall, mean age was 38.3 (SD 7.6) years; of the 21 participants, 12 (57%) self-identified as cisgender man, and 9 (43%) identified as transgender women; and 9 (43%) of the 21 participants tested the tools in Spanish. The majority self-identified as Latino(a) or Hispanic (17/21, 81%), had completed a college education or higher (12/19, 63%), were currently employed (13/20, 65%), lived in their own house or rented a house (17/21, 81%), and did not have a primary partner (15/20, 71%). The mean number of partners was 2.0 (SD 3.1) in the past 3 months; approximately one-half reported anal sex (10/21, 48%), use of alcohol (11/21, 52%), and any substance use (13/21, 62%) in the past 3 months. A total of 12 (12/21, 57%) participants were using PrEP for more than 12 months, and 15 (15/21, 71%) reported a very good or excellent ability to take daily PrEP during the study. Most participants (15/21, 71%) reported using the iOS mobile phone operating system, and 6 (6/21, 29%) reporting using an Android phone. Most participants had a phone contract billed monthly (15/19, 79%) and used 5 or more apps weekly (15/21, 75%). Of the 21 participants, 15 (71%) reported using apps for dating, 13 (62%) reported using apps for health, and 12 (57%) reported using apps to send reminders.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>DOT Diary (n=11)</th>
<th>PrEPmate (n=10)</th>
<th>Total sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Site, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco</td>
<td>5 (46)</td>
<td>4 (40)</td>
<td>9 (43)</td>
</tr>
<tr>
<td>Miami</td>
<td>6 (55)</td>
<td>6 (60)</td>
<td>12 (57)</td>
</tr>
<tr>
<td><strong>Language, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>7 (64)</td>
<td>2 (20)</td>
<td>9 (43)</td>
</tr>
<tr>
<td>Spanish</td>
<td>4 (36)</td>
<td>8 (80)</td>
<td>12 (57)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>37.3 (9)</td>
<td>39.5 (7)</td>
<td>38.3 (8)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisgender man</td>
<td>6 (55)</td>
<td>6 (60)</td>
<td>12 (57)</td>
</tr>
<tr>
<td>Transgender woman</td>
<td>5 (46)a</td>
<td>4 (40)</td>
<td>9 (43)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gay</td>
<td>7 (64)</td>
<td>5 (50)</td>
<td>12 (57)</td>
</tr>
<tr>
<td>Queer</td>
<td>1 (9)</td>
<td>3 (30)</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>2 (18)</td>
<td>2 (20)</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Straight</td>
<td>1 (9)</td>
<td>0</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Race/ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latino(a) or Hispanic</td>
<td>8 (73)</td>
<td>9 (90)</td>
<td>17 (81)</td>
</tr>
<tr>
<td>Black</td>
<td>1 (9)</td>
<td>0</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Asianb</td>
<td>0</td>
<td>1 (10)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>White</td>
<td>1 (9)</td>
<td>0</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Native American</td>
<td>1 (9)</td>
<td>0</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed college or higher</td>
<td>6 (60)</td>
<td>6 (67)</td>
<td>12 (63)</td>
</tr>
<tr>
<td>Less than college</td>
<td>4 (40)</td>
<td>3 (33)</td>
<td>7 (37)</td>
</tr>
<tr>
<td><strong>Annual income (US $), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-29,999</td>
<td>6 (55)</td>
<td>2 (20)</td>
<td>8 (38)</td>
</tr>
<tr>
<td>≥30,000</td>
<td>0</td>
<td>5 (50)</td>
<td>5 (24)</td>
</tr>
<tr>
<td>Not declared</td>
<td>5 (46)</td>
<td>3 (30)</td>
<td>8 (38)</td>
</tr>
<tr>
<td><strong>Employment, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>6 (60)</td>
<td>7 (70)</td>
<td>13 (65)</td>
</tr>
<tr>
<td>Full-time student</td>
<td>1 (10)</td>
<td>0</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>2 (20)</td>
<td>2 (20)</td>
<td>4 (20)</td>
</tr>
<tr>
<td>Disabled</td>
<td>1 (10)</td>
<td>1 (10)</td>
<td>2 (10)</td>
</tr>
<tr>
<td><strong>Housing, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living in someone’s house</td>
<td>2 (18)</td>
<td>1 (10)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Living in own/rent house</td>
<td>9 (82)</td>
<td>8 (80)</td>
<td>17 (81)</td>
</tr>
<tr>
<td>Living in hotel</td>
<td>0</td>
<td>1 (10)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>DOT Diary (n=11)</td>
<td>PrEPmate (n=10)</td>
<td>Total sample</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Live alone (yes), n (%)</td>
<td>4 (36)</td>
<td>3 (30)</td>
<td>7 (33)</td>
</tr>
<tr>
<td>Health insurance (yes), n (%)</td>
<td>6 (60)</td>
<td>7 (70)</td>
<td>13 (65)</td>
</tr>
<tr>
<td>Primary care provider (yes), n (%)</td>
<td>5 (50)</td>
<td>7 (70)</td>
<td>12 (60)</td>
</tr>
<tr>
<td><strong>Primary partner, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>10 (91)</td>
<td>5 (50)</td>
<td>15 (71)</td>
</tr>
<tr>
<td>Yes</td>
<td>0</td>
<td>4 (40)</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Not sure</td>
<td>1 (9)</td>
<td>1 (10)</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Number of partners&lt;sup&gt;c&lt;/sup&gt;, mean (SD)</td>
<td>2.9 (3.9)</td>
<td>1.1 (1.9)</td>
<td>2.0 (3.1)</td>
</tr>
<tr>
<td>Anal sex&lt;sup&gt;c&lt;/sup&gt;, n (%)</td>
<td>5 (46)</td>
<td>5 (50)</td>
<td>10 (48)</td>
</tr>
<tr>
<td>Alcohol use&lt;sup&gt;c&lt;/sup&gt;, n (%)</td>
<td>6 (45)</td>
<td>5 (50)</td>
<td>11 (52)</td>
</tr>
<tr>
<td>Any substance use&lt;sup&gt;c&lt;/sup&gt;, n (%)</td>
<td>7 (64)</td>
<td>6 (60)</td>
<td>13 (62)</td>
</tr>
<tr>
<td><strong>Time on PrEP&lt;sup&gt;d&lt;/sup&gt; (months), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6</td>
<td>3 (27)</td>
<td>3 (30)</td>
<td>6 (29)</td>
</tr>
<tr>
<td>6–12</td>
<td>1 (9)</td>
<td>2 (20)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>&gt;12</td>
<td>7 (64)</td>
<td>5 (50)</td>
<td>12 (57)</td>
</tr>
<tr>
<td><strong>Ability to take daily PrEP during the study&lt;sup&gt;e&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very poor or fair</td>
<td>1 (9)</td>
<td>1 (10)</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Good</td>
<td>1 (9)</td>
<td>3 (30)</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Very good</td>
<td>5 (46)</td>
<td>3 (30)</td>
<td>8 (38)</td>
</tr>
<tr>
<td>Excellent</td>
<td>4 (36)</td>
<td>3 (30)</td>
<td>7 (33)</td>
</tr>
<tr>
<td><strong>Mobile phone operation system, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google/Android</td>
<td>1 (9)</td>
<td>5 (50)</td>
<td>6 (29)</td>
</tr>
<tr>
<td>iPhone (iOS)</td>
<td>10 (91)</td>
<td>5 (50)</td>
<td>15 (71)</td>
</tr>
<tr>
<td><strong>Mobile phone plan, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contract billed monthly</td>
<td>7 (78)</td>
<td>8 (80)</td>
<td>15 (79)</td>
</tr>
<tr>
<td>Prepaid</td>
<td>1 (11)</td>
<td>1 (10)</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Shared plan with someone</td>
<td>1 (11)</td>
<td>1 (10)</td>
<td>2 (11)</td>
</tr>
<tr>
<td><strong>Mobile phone temporarily disconnected&lt;sup&gt;f&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>10 (91)</td>
<td>7 (78)</td>
<td>17 (85)</td>
</tr>
<tr>
<td>Once or twice</td>
<td>1 (9)</td>
<td>2 (22)</td>
<td>3 (15)</td>
</tr>
<tr>
<td><strong>Mobile phone lost, stolen, or broken&lt;sup&gt;f&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>8 (73)</td>
<td>5 (56)</td>
<td>13 (65)</td>
</tr>
<tr>
<td>Once or twice</td>
<td>3 (27)</td>
<td>4 (44)</td>
<td>7 (35)</td>
</tr>
<tr>
<td><strong>Type of app used, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dating (yes)</td>
<td>10 (91)</td>
<td>5 (50)</td>
<td>15 (71)</td>
</tr>
<tr>
<td>Health (yes)</td>
<td>7 (64)</td>
<td>6 (60)</td>
<td>13 (62)</td>
</tr>
</tbody>
</table>
Mean SUS scores were 81.2 (SD 13.1) for PrEPmate and 76.4 (SD 12.2) for DOT diary ($P=.48$), corresponding to adjective ratings in the “good” to “excellent” range [44] (Table 4). The highest score for both mHealth tools was on the item “I thought the system was easy to use” (DOT Diary: 4.9, SD 0.4; PrEPmate: 4.8, SD 0.5). The score on the item “I found the various functions in DOT Diary/PrEPmate were well integrated” was higher for PrEPmate than for DOT diary (4.6, SD 0.7 vs 3.6, 1.4; $P=.09$).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>DOT Diary (n=11)</th>
<th>PrEPmate (n=10)</th>
<th>Total sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Send reminders (yes)</td>
<td>6 (55)</td>
<td>6 (60)</td>
<td>12 (57)</td>
</tr>
</tbody>
</table>

---

*a* One participant also identified as gender queer.

*b* Participant also identified as Native American.

*c* Past 3 months.

*d* PrEP: preexposure prophylaxis.

*e* Assessed at baseline.

*f* In the past 12 months.
Table 4. Mean System Usability Scale (SUS) and Client Satisfaction Questionnaire (CSQ-8) scores among the technical pilot participants in San Francisco and Miami by study arm (DOT Diary or PrEPmate).

<table>
<thead>
<tr>
<th>Questionnaire responses</th>
<th>DOT Diary, mean (SD)</th>
<th>PrEPmate, mean (SD)</th>
<th>Total sample, mean (SD)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I think that I would like to use DOT Diary/PrEPmate on a regular basis.</td>
<td>4.7 (0.5)</td>
<td>4.5 (0.8)</td>
<td>4.6 (0.6)</td>
<td>.53</td>
</tr>
<tr>
<td>2. I found DOT Diary/PrEPmate unnecessarily complex.</td>
<td>2.7 (1.7)</td>
<td>2.4 (1.7)</td>
<td>2.5 (1.6)</td>
<td>.70</td>
</tr>
<tr>
<td>3. I thought the system was easy to use.</td>
<td>4.9 (0.4)</td>
<td>4.8 (0.5)</td>
<td>4.8 (0.4)</td>
<td>.63</td>
</tr>
<tr>
<td>4. I think that I would need the support of a technical person to be able to use DOT Diary/PrEPmate.</td>
<td>2.4 (1.5)</td>
<td>2.4 (1.7)</td>
<td>2.4 (1.5)</td>
<td>.95</td>
</tr>
<tr>
<td>5. I found the various functions in DOT Diary/PrEPmate were well integrated.</td>
<td>3.6 (1.4)</td>
<td>4.6 (0.7)</td>
<td>4.1 (1.2)</td>
<td>.09</td>
</tr>
<tr>
<td>6. I thought there was too much inconsistency between different parts of DOT Diary/PrEPmate.</td>
<td>2.3 (1.1)</td>
<td>1.5 (1.1)</td>
<td>1.9 (1.1)</td>
<td>.19</td>
</tr>
<tr>
<td>7. I would imagine that most people would learn to use DOT Diary/PrEPmate very quickly.</td>
<td>4.7 (0.8)</td>
<td>4.6 (0.7)</td>
<td>4.7 (0.7)</td>
<td>.82</td>
</tr>
<tr>
<td>8. I found DOT Diary/PrEPmate very cumbersome to use.</td>
<td>2.7 (1.4)</td>
<td>1.6 (1.1)</td>
<td>2.1 (1.3)</td>
<td>.11</td>
</tr>
<tr>
<td>9. I felt very confident using DOT Diary/PrEPmate.</td>
<td>4.9 (0.4)</td>
<td>4.2 (1.4)</td>
<td>4.5 (1.1)</td>
<td>.28</td>
</tr>
<tr>
<td>10. I needed to learn a lot of things before I could get going with DOT Diary/PrEPmate.</td>
<td>2.0 (1.4)</td>
<td>2.4 (1.5)</td>
<td>2.2 (1.4)</td>
<td>.63</td>
</tr>
<tr>
<td>Total score (0-100)</td>
<td>76.4 (12.2)</td>
<td>81.2 (13.1)</td>
<td>79.0 (12.5)</td>
<td>.48</td>
</tr>
<tr>
<td>CSQ-8&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How would you rate the quality of DOT Diary/PrEPmate?</td>
<td>3.6 (0.8)</td>
<td>3.6 (0.5)</td>
<td>3.6 (0.6)</td>
<td>.88</td>
</tr>
<tr>
<td>Did you get the kind of support you wanted from DOT Diary/PrEPmate?</td>
<td>2.9 (1.1)</td>
<td>3.4 (0.9)</td>
<td>3.1 (1.0)</td>
<td>.33</td>
</tr>
<tr>
<td>To what extent did DOT Diary/PrEPmate meet your needs for PrEP support?</td>
<td>3.6 (0.5)</td>
<td>3.5 (0.8)</td>
<td>3.5 (0.6)</td>
<td>.84</td>
</tr>
<tr>
<td>Would you recommend DOT Diary/PrEPmate to a friend?</td>
<td>3.7 (0.5)</td>
<td>3.6 (0.5)</td>
<td>3.7 (0.5)</td>
<td>.74</td>
</tr>
<tr>
<td>How satisfied are you with the amount of help (DOT Diary/PrEPmate) you received from DOT Diary/PrEPmate?</td>
<td>3.7 (0.5)</td>
<td>3.4 (1.1)</td>
<td>3.5 (0.8)</td>
<td>.45</td>
</tr>
<tr>
<td>Did using DOT Diary/PrEPmate help you take PrEP daily?</td>
<td>3.6 (0.5)</td>
<td>3.8 (0.5)</td>
<td>3.7 (0.5)</td>
<td>.50</td>
</tr>
<tr>
<td>In an overall, general sense, how satisfied are you with DOT Diary/PrEPmate?</td>
<td>3.7 (0.5)</td>
<td>3.8 (0.5)</td>
<td>3.7 (0.5)</td>
<td>.89</td>
</tr>
<tr>
<td>If you enrolled in the future in a clinical trial to help researchers discover and evaluate treatments, would you use DOT Diary/PrEPmate?</td>
<td>3.6 (0.5)</td>
<td>3.6 (0.5)</td>
<td>3.6 (0.5)</td>
<td>.85</td>
</tr>
<tr>
<td>Total score (8-32)</td>
<td>28.3 (3.2)</td>
<td>28.6 (4.4)</td>
<td>28.5 (3.8)</td>
<td>.87</td>
</tr>
</tbody>
</table>

<sup>a</sup>t test.
<sup>b</sup>Scores range from 1 (strongly disagree) to 5 (strongly agree).
<sup>c</sup>Scores range from 1 to 4.

Mean CSQ-8 scores were 28.6 (SD 4.4) for PrEPmate and 28.3 (SD 3.2) for DOT Diary (P=.87). The highest scores (3.7) were on items “Would you recommend DOT Diary/PrEPmate to a friend?”, “Did using DOT Diary/PrEPmate help you take PrEP daily?”, and “In an overall, general sense, how satisfied are you with DOT Diary/PrEPmate?” The lowest score (3.1) was on item “Did you get the kind of support you wanted from DOT Diary/PrEPmate?”

Most of the participants (60-100%) rated each of the components of DOT Diary and PrEPmate as 5 (very helpful) on a 5-point Likert scale (Figures 2 and 3). For DOT Diary, all participants rated the components “Daily reminder notification to take pill,” “Calendar view: pop-up showing protection level, dosing, and sexual activity on a given day,” and “Home screen indicating level of protection” as very helpful. For PrEPmate, all participants rated the components “Having a text messaging platform to communicate with the study team” and “Daily pill-taking reminder messages” as very helpful. Most participants reported they reviewed the PrEP Basics information (6/8, 75%), and one-half reported they watched the PrEP videos (4/8, 50%).
At the 1-month follow-up visit, the mean PrEP-ASES score was 9.4 (SD 1.6) for DOT Diary and 9.3 (SD 0.9) for PrEPmate (P=.80).

After participating in this study, most participants reported feeling good about helping others or the community (15/16, 94%), being more confident in taking PrEP (14/16, 88%), and having a better understanding of sexual behaviors and risk (14/16, 88%). One participant reported a social harm related to having trouble getting or keeping housing in the DOT diary arm.

**Feasibility of the Tools**

Among 10 participants assigned to the PrEPmate arm, 1 participant withdrew at the enrollment visit, and the remaining 9 participants responded to at least one check-in message. During their study participation, the mean total number of incoming messages sent from participants was 10.5 (range: 4 to 42), and the mean time to the first response for each weekly check-in message was 21 hours (range: 13 minutes to 67 hours). Most incoming messages from participants were in response to weekly check-ins, and some were related to scheduling appointments.

Among 11 participants assigned to the DOT Diary arm, they opened the app, on average, 17.6 (SD 21.4; range: 2-67) times, with an average duration of 2.77 (SD 3.12) minutes per session (range: 5 seconds to 9.49 minutes). Across all participants, the cumulative time spent in the app ranged from 15 seconds to 2 hours and 26.5 minutes, with an average of 34 (SD 43.8) minutes. The most frequently used functions included diary entries (mean 79.9, SD 132.2 accesses) and taking a dose (15.9, SD 21.2 accesses), whereas the least used features were the...
monthly calendar view (mean 3.7, SD 4.5 accesses) and the summary statistics page (mean 1.3, SD 2.1 accesses).

**Exit Interviews**

The follow-up exit interviews were completed by 17 participants approximately 1 month after enrollment (10 in the DOT Diary arm, 7 in the PrEPmate arm). Regarding PrEPmate, most participants found the system straightforward and easy to use and navigate. They found the onboarding process to be clear and fast, and the onboarding links were a good resource for PrEP information. Both Spanish-speaking and transgender participants reported liking the daily fun facts, affirmations, historical messages, and jokes, which were interesting and not too intrusive. The regularity of these daily messages was helpful as pill-taking reminders, as a Spanish-speaking participant said, “A very – very – very accurate reminder,” and was particularly helpful on days when participants were out of their usual routine, such as waking up late or if they were on their day off. Additionally, participants found access to a PrEP navigator helpful, as it provided reminders to get labs completed. As constructive feedback, participants said that more variety and types of messages were needed, including other topics not related to LGBTQIA+ communities, as over time, the messages “start[ed] to get a little stale,” remarked an English-speaking transgender participant. Participants suggested having more information about PrEP efficacy and the window of time they had to take their pill, as well as HIV testing and other HIV prevention strategies. Participants liked the weekly interaction and being checked on by the navigator, although some participants were still unclear whether messages were automated or coming from an actual person and suggested that this be clarified during the onboarding process, with the staff member saying, “I will occasionally reach out to you through PrEPmate.” Regarding DOT Diary, participants appreciated logging their encounters and receiving notifications, and the reminders for taking daily PrEP were especially useful, particularly for those who don’t have an established pill-taking routine. Participants also found it helpful to know their protection levels, which made them feel more confident having sexual encounters; they also liked the feature showing how many more doses were needed to reach full protection levels. Reported cons of DOT Diary included not being able to retroactively log a dose taken earlier, not seeing the little dot added to the calendar that identified an encounter, and not having reminders for an upcoming 3-month lab visit. Several participants commented on the language used in the app. A young transgender woman (age 24 years) in San Francisco felt the pop-up congratulatory message after taking a dose was a bit “cheesy,” saying “it felt like it was someone who was, like 50, trying to relate to me as a youth, which always feels weird.” A 46-year-old Spanish-speaking MSM in Miami was shocked and amused by the Spanish translation of language in the app and felt that the language was sexually empowering but also a bit taboo, while a 41-year-old English-speaking transgender woman in San Francisco found the language used to identify sex acts as “grotesque.” Several participants liked reporting their sexual encounters, saying “it’s a calendar of your sex life,” and liked seeing a visual record of their sexual encounters, as it helped them be more mindful of their sexual activity:

**It’s like a mirror right? Because it perhaps shows you in another way what your activity has been or Okay, I see that perhaps I’m more active on the weekends or anything. So, it gives you like that reflection of what our activity is.**

An MSM in Miami reported the amount of information the app requested was somewhat intrusive, although he said this was not a dealbreaker and may have found it more useful if he had more sexual partners. Several participants recommended including a function to add doses retroactively and to support on-demand or 2-1-1 PrEP dosing. Participants also wanted to be able to customize the alarm ring tone, as some of the alarm tones were low volume and difficult to hear. On the summary data page, one participant suggested adding information about the total number of sexual partners for the month. Finally, most participants found the onboarding experience easy, although 1 participant found the onboarding survey a bit tedious and another participant recommended separating questions about their male and transfemale partners.

**Discussion**

**Principal Findings**

This study describes findings from 2 phases of formative research, including focus groups and a technical pilot, to tailor 2 innovative mHealth daily PrEP support tools for Spanish-speaking MSM and English- and Spanish-speaking transgender women in the United States. In addition to translating the tools into Spanish, key adaptations informed by our formative work included culturally tailoring content and language within our interventions to our 2 priority populations (both PrEPmate and DOT Diary), clarifying that navigators will respond to text messages from participants during the onboarding process (PrEPmate), and simplifying the sexual diary and adding motivational badges to the app (DOT Diary). In a 1-month technical pilot of these tools, both PrEPmate and DOT Diary demonstrated feasibility based on use of the tools as assessed via paradata and good-to-excellent acceptability assessed via self-report.

A growing number of mHealth tools are being designed and evaluated to support PrEP uptake, adherence, and persistence [45]; however, few tools have been specifically designed or tailored for Spanish-dominant individuals in the United States. Several studies suggest that culturally adapted behavioral interventions for Latino men may outperform standard treatment, and culturally and linguistically tailored resources in Spanish to support PrEP use among MSM and transgender women could help address disparities in PrEP use in these populations [46]. Recent studies have highlighted community-centered approaches to developing these tools for Spanish-speaking populations. MacCarthy and colleagues [47] investigated strategies for improving mobile technology–based HIV prevention interventions with Latino MSM and Latina transgender women. Key findings included that requiring smartphone use could reduce participation in low-income participants; variability in participant preferences regarding personalization, frequency, and timing of text messages; recommendation for messages to be sent on the same days and at the same times to help
participants anticipate receiving information; and the importance of recognizing different language literacies and diverse countries of origin. Although the PrEPmate text-messaging intervention can be used on non-smartphones, the DOT Diary intervention does require use of a smartphone. Tracking individuals who were excluded due to not having a smartphone will be important in future studies and implementation to assess the impact of this requirement on equitable outcomes. Regarding message timing, in both PrEPmate and DOT Diary, the timing of messages can be customized and are set to be delivered at the same time each day. Cantos and colleagues [48] used the Assessment, Decision, Adaptation, Production, Topical Experts-Integration, Training, and Testing (ADAPT-ITT) framework to develop an HIV prevention mobile app to increase uptake of PrEP among Latino sexual minority men called SaludFindr. Through in-depth interviews, they found both general barriers that were common to non-Latino groups, as well as Latino-specific barriers, including lack of available clinics to provide culturally concordant care and limited availability of Spanish language information on PrEP.

Although most technology-based HIV prevention tools being developed have focused on MSM, a few mHealth interventions have been tailored specifically for transgender women. In a cross-disciplinary scoping review of mobile technology interventions to improve HIV prevention and care in transgender and gender diverse youth, Sken and Cain [49] highlighted the importance of gender affirmation as a key social determinant of health for transgender youth, and several interventions have been guided by the gender-affirmative framework [50], which centers the interactive process in which a person receives social recognition and support for their gender identity and expression. In this review, behavioral self-monitoring and access to HIV prevention services were the most frequent features across disciplines. Similarly, our DOT Diary app utilizes self-monitoring and management to promote building of a daily pill-taking routine, and content within our interventions is designed to support gender affirmation (eg, trans-specific affirmation messages in PrEPmate). In the TechStep study, Reback and colleagues [51] developed several culturally responsive technology-based interventions (text messaging, web app, and eCoaching) tailored for transgender and gender diverse youth through focus groups and consultation with youth advisory boards in the Adolescent Trials Network, which was subsequently evaluated in a 3-arm sequentially randomized HIV prevention trial. Finally, Morris and colleagues [52] tailored the individualized Texting for Adherence Building (iTAB) intervention through focus groups with transgender and nonbinary individuals and evaluated this approach alone and in combination with motivational interviewing in a large randomized controlled trial among transgender and nonbinary PrEP users. In this study, adherence outcomes as measured by drug levels were mostly similar between the arms, although self-reported adherence was higher in the iTAB + motivational interviewing group versus the iTAB alone group. Together, these studies highlight the importance of adopting a user-centered approach to tailoring mHealth interventions for our priority populations.

The next step for our mHealth tools will be to incorporate changes suggested in our technical pilot exit interviews to further tailor these tools in Spanish and for transgender women. For PrEPmate, we will refine the content and language used in the text messages based on participant feedback and augment the onboarding process to have navigators clearly describe their role in monitoring and responding to text messages. For DOT Diary, we will add a function to retroactively add a dose after the dosing period has passed and refine the language used in the app based on participant feedback. Additionally, through another study, we will adapt and tailor this app to support on-demand and 2-1-1 PrEP use among MSM (IR34MH121139). With both of these tools optimized, we will then conduct a large multisite randomized controlled trial to compare the effectiveness of PrEPmate versus DOT Diary among English- and Spanish-speaking MSM and transgender women through a study funded by the Patient-Centered Outcomes Research Institute.

Limitations

This study had several limitations. First, the number of participants enrolled in the focus groups and technical pilot was small; therefore, the pilot study was not designed to evaluate the efficacy of the interventions. In addition, due to the short duration of the pilot, longer-term feasibility and acceptability could not be evaluated, and participants were only able to test 1 of the 2 interventions. Still, formative work through focus groups and optimizing usability in a technical pilot are important steps in developing and tailoring interventions prior to evaluating their efficacy in large randomized controlled trials. Social desirability may have impacted participant responses in the focus groups and technical pilot; however, we used a computer-assisted self-interview to assess usability and acceptability, we used paradata metrics from PrEPmate and DOT Diary to objectively assess feasibility, and participants in focus groups and exit interviews were reminded that there were no right or wrong answers and were encouraged to provide honest feedback that would help us best improve the tools. Additionally, for both phases of this study, participants were enrolled in San Francisco and Miami, potentially limiting generalizability of the findings. However, both regions are in high-priority Ending the HIV Epidemic jurisdictions and have large Spanish-speaking Latinx populations, and one site is located in the South, the region accounting for over one-half of new HIV diagnoses nationally. In addition, Miami-Dade County is the metropolitan statistical area with the highest rate of new HIV diagnoses in the United States [53]. Despite these limitations, our study had several strengths, including enrolling a substantial proportion of Spanish-dominant participants and transgender women across 2 diverse sites; the user-centered approach incorporating input from our participants into the final design of our interventions; and the mixed methods approach to evaluate feasibility and acceptability in our technical pilot.

Conclusions

Using a user-centered design approach, we tailored the PrEPmate text messaging intervention and the DOT Diary mobile app to support PrEP adherence and persistence for Spanish-speaking MSM and English- and Spanish-speaking...
transgender women in the United States. Preliminary testing in a technical pilot demonstrated high acceptability and feasibility of the adapted versions of these tools and support further evaluation in an upcoming comparative effectiveness trial of PrEPmate versus DOT Diary in a large national US study.

Acknowledgments

This study was funded by a Gilead Investigator Sponsored Research grant IN-US-412-5777. TST was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). We thank all the participants for their helpful insights and feedback in this study.

Data Availability

Survey data for the technical pilot have been uploaded to the Dryad Repository [54]. Focus group transcripts will be made available upon request.

Conflicts of Interest

AYL has received funding for investigator-sponsored research studies from Gilead Sciences and Viiv Healthcare and has led studies in which study drug has been donated by Gilead Sciences. SB has received funding for research studies from GSK. SDL has received funding for research studies from Gilead Sciences, Merck, and Janssen. TST received consulting fees from Gilead Sciences. The other authors have no conflicts of interest to disclose.

References

11. Liu et al JMIR FORMATIVE RESEARCH
13. Shover CL, Javanbakht M, Shoptaw S, Bolan R, Gorbach P. High discontinuation of pre-exposure prophylaxis within six months of initiation. 2018 Presented at: Conference on Retroviruses and Opportunistic Infections; March 4-7, 2018; Boston, MA.


51. mPrEP technical pilot survey data. Dryad. URL: https://doi.org/10.5061/dryad.7d7wm382d [accessed 2024-02-28]
Abbreviations

ADAPT-ITT: Assessment, Decision, Adaptation, Production, Topical Experts-Integration, Training, and Testing
CSQ-8: Client Satisfaction Questionnaire
iTAB: individualized Texting for Adherence Building
mHealth: mobile health
MSM: men who have sex with men
PrEP: preexposure prophylaxis
SUS: System Usability Scale

Please cite as:
Liu AY, Alleyne CD, Doblecki-Lewis S, Koester KA, Gonzalez R, Vinson J, Scott H, Buchbinder S, Torres TS
Adapting mHealth Interventions (PrEPmate and DOT Diary) to Support PrEP Retention in Care and Adherence Among English and Spanish-Speaking Men Who Have Sex With Men and Transgender Women in the United States: Formative Work and Pilot Randomized Trial
JMIR Form Res 2024;8:e54073
URL: https://formative.jmir.org/2024/1/e54073
doi:10.2196/54073
PMID:38536232

©Albert Y Liu, Cat-Dancing Alleyne, Susanne Doblecki-Lewis, Kimberly A Koester, Rafael Gonzalez, Janie Vinson, Hyman Scott, Susan Buchbinder, Thiago S Torres. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
User Friendliness and Perioperative Guidance Benefits of a Cataract Surgery Education App: Randomized Controlled Trial

Rokas Gerbutavicius\textsuperscript{1,2}, MSc, MD; David A Merle\textsuperscript{1,3,4}, MSc, MD; Armin Wolf\textsuperscript{2}, MD; Spyridon Dimopoulos\textsuperscript{1}, MD; Karsten Ulrich Kortuem\textsuperscript{2,5}, MD; Friederike Charlotte Kortuem\textsuperscript{1}, MSc, MD

\textsuperscript{1}Department for Ophthalmology, University of Tuebingen, Tuebingen, Germany
\textsuperscript{2}University Eye Hospital, University of Ulm, Ulm, Germany
\textsuperscript{3}Institute for Ophthalmic Research, Center for Ophthalmology, University of Tuebingen, Tuebingen, Germany
\textsuperscript{4}Department of Ophthalmology, Medical University of Graz, Graz, Austria
\textsuperscript{5}Augenarztpraxis Dres. Kortuem, Ludwigsburg, Germany

Corresponding Author:
Rokas Gerbutavicius, MSc, MD
Department for Ophthalmology
University of Tuebingen
Elfriede-Aulhorn-Str. 7
Tuebingen, 72076
Germany
Phone: 49 7071 2988088
Email: rokas.gerbutavicius@med.uni-tuebingen.de

Abstract

Background: Cataract surgeries are among the most performed surgeries worldwide. A thorough patient education is essential to inform patients about the perioperative process and postoperative target results concerning the intraocular lens and objectives for visual outcomes. However, addressing all relevant aspects and questions is time-consuming. Mobile apps can facilitate this process for both patients and physicians and thus be beneficial. However, the success of such an app depends on its user friendliness and acceptance by patients.

Objective: This study aimed to evaluate the user friendliness and acceptance of a cataract surgery education app on mobile devices among patients undergoing cataract surgery, the characteristics of patients who benefit the most from app use, and the influence of the app on patient satisfaction with treatment.

Methods: All patients who underwent cataract surgery at an ophthalmological practice from August 2020 to July 2021 were invited to participate in this randomized controlled trial. Out of 493 invited patients, 297 (60.2\%) were enrolled in this study. Patients were randomized into 3 different groups. Half of the patients were offered to participate in Group 1 with use of the “Patient Journey” app. However, if they decided not to use the app, they were included in Group 2 (app denial). The other half of the patients were included in Group 3 (control) with no use of the app and with information provided conventionally. The app provided general information on the ophthalmological center, surgeons, cataract, and treatment options. Different questionnaires were used in all 3 groups to evaluate satisfaction with the perioperative process. Group 1 evaluated the app. Demographic characteristics, such as age, gender, and educational degree, were assessed.

Results: Group 1 included 77 patients (median age 69 years). Group 2 included 61 patients, and their median age was higher (median age 79 years). Group 3 included 159 patients (median age 74 years). There was no difference in satisfaction with the perioperative process and clinic between the 3 groups. Almost all app users appreciated the digital details provided for the organization and the information on the surgery. Age did not play a major role in appreciation of the app. Female patients tended to appreciate the information provided more than male patients. Patients who did not have a higher university degree experienced more benefits from the informational content of the app and were the most satisfied with the information. However, male patients and academics were in general more aware of technology and handled the app more easily.

Conclusions: The app showed high user friendliness and acceptance, and could particularly benefit specific patient groups. App users demonstrated a noninferior high satisfaction with the treatment in the ophthalmological center in comparison with patients who were informed about the surgery only conventionally.
Introduction

Changes in Health Care Toward Digitalization

Health care is currently undergoing a substantial transition due to technological progress that reshapes clinical workflows. The progress in digitalization opens the door for new ways to organize, diagnose, educate, and treat patients. In 2023, the number of smartphone end users reached almost 7 billion globally, which equals roughly 86% of the world’s population [1]. Compared to 7 years earlier, these numbers almost doubled (3.7 billion or 49.4% in 2016) [1]. Intriguingly, mobile phone–based apps can provide valuable new platforms to address patients’ and physicians’ needs. In recent years, many apps in health care were developed driven by an increasing public interest in digital health care products. In 2017 alone, 3.7 billion mobile health (mHealth) app downloads were counted worldwide [2], showing a continuous almost exponential growth over the last few years. Health apps on mobile devices, also known as mHealth apps, are global software programs that can be used by patients, health care professionals, or other care givers. In Germany, since the end of 2019, the Digital Health Care Act (Digitales Versorgungsgesetz, DVG) allows app developers to enter a process at a governmental agency to receive reimbursement for the download and use of the app from statutory health insurances. This has given another boost to the development of apps with greater and proven quality [3].

mHealth in Ophthalmology

In March 2020, a review identified 131 ophthalmology-related mobile apps, with 32% of the apps designed for visual acuity testing and screening, 13% designed for eye relaxation exercises, and 12% designed for professional training. The remaining apps included tools to detect color blindness, tools that served as low vision aids, or tools aimed to provide assistance and patient education. Strikingly, less than 5% were documented to have been tested for validity [4].

A recent review by Nagino et al [5] summarized published data on the clinical utility of 48 mobile apps in ophthalmology. Of those, 35% supported clinical ophthalmological examination, 27% intended to detect ophthalmological diseases, 20% supported medical personnel, 10% informed patients about ophthalmological diseases, and 6% were designed to encourage compliance. Only 2 apps reported significant efficacy in treating diseases [5]. An app for glaucoma treatment reminded patients to administer intraocular pressure (IOP) lowering eye drops [6]. Patients enrolled in the study showed higher adherence when being reminded by the app. The IOP lowering effect was however not significant in any of the subgroups [6]. Patients with macular diseases that required regular injections benefited in terms of visual acuity when provided with mobile hyperacutivity home monitoring via an app and discontinued treatment less often [7].

However, most health apps in ophthalmology and other disciplines still lack evidence of clinical effectiveness and thus lack the rational basis to qualify for reimbursement [8,9]. For the vast majority of apps available in digital app stores, no published scientific data or proof of efficacy is available.

On the other hand, digital apps hold great potential value for health care providers in optimizing work processes and flows. Additionally, in light of the looming physician shortage and ever-increasing patient consultations, apps can be valuable tools to offer new ways of interaction between patients and physicians and can add value by improving patient education [10].

Cataract Surgery

Cataract is a very common condition that is highly associated with aging and ultimately occurs in every individual. The initial symptoms include reduced visual acuity and increased glare sensitivity. Depending on the localization of increased opacity, cataracts can be divided into nuclear, cortical, or subcapsular cataracts. In an observational study by Klein et al [11], the cumulative incidence of nuclear cataract increased from 2.9% in persons aged 43 to 54 years at baseline to 40% in those aged 75 years or older. For cortical and posterior subcapsular cataract, the corresponding numbers were 1.9% and 21.8% and 1.4% and 7.3%, respectively. These numbers illustrate the disease burden and high number of patients experiencing the consequences of cataract. Accordingly, cataract surgery is ranked as the most common surgical procedure performed in the European union, with almost 5 million surgeries performed in 2017 alone. Multiple studies have demonstrated gains in visual function and quality of life after surgery [12-14]. The standard technique in Europe is the removal of the opacified lens by phacoemulsification and the implantation of an artificial intraocular lens (IOL). The surgery is mostly conducted under topical or peribulbar anesthesia [15,16]. In complicated cases, general anesthesia or the use of short-acting sedatives may be necessary. Although phacoemulsification was the standardized procedure for a long time, novel approaches that allow better precision in incisions and fragmentation of the lens have emerged. In this light, laser-supported phacoemulsification, such as nanosecond and femtosecond laser, promise better accuracy and less strain on the cornea without any effect on visual outcomes [17-19]. Furthermore, new developments in IOL design have led to a vast selection of different available lenses, such as multifocal or enhanced depth of focus lenses, possibly increasing quality of life [20]. With the obvious plethora of different options, patients frequently feel overwhelmed regarding the choice of the perioperative process and the targeted refraction (monofocal near or distance vision, and multifocal satisfactory near, intermediate, or distance vision). These questions and decisions possibly involve higher out-of-pocket costs for the patient as newer techniques and IOL designs are often not covered by statutory or private health insurances. Furthermore, multifocal lenses are also known to have potential negative aspects due to phenomena such as glare,
halos, and loss of contrast sensitivity. This may lead to unsatisfactory results in some patients and thus require highly accurate refractive outcomes, including limited astigmatism, in order to achieve good function and postoperative tolerance [21]. To ensure good outcomes and patient satisfaction, the treating ophthalmologist must provide sufficient information and time to the patient, provide guidance throughout the process, and carefully select patients suitable for implantation of specific lenses.

**Perioperative Journey of Patients**

As for every intervention or surgery, the patient’s journey is linked to uncertainty and a certain degree of anxiety. For fields other than ophthalmology, mHealth apps seem to be acknowledged by patients in a perioperative setting. In studies with mobile apps, patients felt more taken care of, with a clear focus on the patient’s satisfaction and health. Furthermore, the apps proved to be more cost-effective and efficient in health care services [22,23]. Likewise, a recent study was able to improve the adherence of cataract patients to postoperative management through the use of reminder messages on a mobile app, which also provided links to educational videos online [24].

However, there is no published study on patients undergoing cataract surgery who have been guided by a mobile app throughout the perioperative process as a whole. Therefore, this study aimed to assess the acceptance and satisfaction with a mobile app accompanying the patients in their cataract surgery journey. As cataract patients are mostly in the second half of their life, this investigation addresses the use of mHealth in an elderly population. Additionally, the study also addresses the question of which group of patients benefits the most from app use.

**Methods**

**Study Patients**

All patients who presented to a local ophthalmological practice center and underwent cataract surgery over a period of 1 year (August 2020 to July 2021) were offered to participate in this prospective, single-center, randomized controlled trial. Invitations to the trial were sent by mail to the patients 2 to 3 weeks prior to the pre-examination day. Invitations comprised relevant information on the trial and a consent sheet for participation and data. Out of 493 patients who received an invitation to participate, 297 (60.2%) were willing to be enrolled in this study. Over the duration of the study, of the 297 patients, 181 (60.9%) were operated on 1 eye and 116 (39.1%) were operated on both eyes. Patients who consented were randomized by even and uneven patient numbers issued by the health care information system (software) into 3 groups as follows: (1) Patients with even patient numbers were offered to participate in the interventional group (Group 1) with the use of the mobile app; (2) Patients who did not want to use the app were included in the app deny group (Group 2); (3) Patients with uneven patient numbers were included in the control group (Group 3) without access to the app. Group 1 received personal login details and information on the download and use of the app.

**App Description**

The app used in this trial was the “Patient Journey” app (Versions 4.15.0, 4.26.8, and 4.30.0; Interactive Studios). The app can be installed on all devices with iOS (Apple) and Android (Google) operating systems. The app was free for all participants.

The app uses an adjustable content management system with the possibility of own branding. The app versions used in this study had the following 4 sections: (1) General information on the ophthalmological center and surgeons along with contact information; (2) Information on appointments (surgery, and preoperative and postoperative follow-ups); (3) Information about the disease (cataract), treatment, anesthesia options, and IOL; and (4) Preoperative, perioperative, and postoperative behavioral recommendations and the medication treatment scheme (Figure 1).
In the third section, a video on cataract development, symptoms, and therapy was available. This video was almost 3 minutes long and included real ophthalmologists and patients who were interviewed in a hospital. Furthermore, pictures and text in the app provided information on IOLs, anesthetic procedures, and postoperative care. Questionnaires were included in order to find a suitable IOL and target refraction. Early preoperative instructions comprised information about necessary blood tests, the need to schedule an appointment with the anesthesiologist in the case of surgery under general anesthesia, and the prohibition to directly drive a car after the pre-examination visit owing to mydriasis. One day before the surgery, patients received information about the time, location, and course of the surgery, and in the case of general anesthesia, they received information about the necessity of preoperative fasting. On the surgery date, patients were, if needed, reminded to continue preoperative fasting and instructed about postoperative behaviors, including intake of medications, wearing of an eye patch, taking a shower, performing physical activity, and observing possible symptoms like pain or foreign body sensation. Postoperative care in the app comprised information about scheduled appointments in the practice as well as about common and warning postoperative symptoms, including their management or the necessity to schedule an additional examination. Push notifications could be set to remind the patient to apply anti-inflammatory drops or to keep the appointments.

Study organizers stayed in contact with patients from all 3 study groups on a regular basis from invitation to the study 2 to 3 weeks before the pre-examination visit until 4 to 6 weeks after the first cataract surgery and after the second cataract surgery, if performed. Patients in all study groups communicated with the practice and were informed about the surgery conventionally while attending ophthalmologist appointments or via a phone call. App users additionally received necessary information about the surgery and could schedule appointments or contact the practice over the app. Patients who could not manage to install the app on their own were offered support for the installation and set-up of the app in the practice at the pre-examination visit or via a phone call. See Multimedia Appendix 1 for a demonstration video of the cataract app.

Patient Survey

For the study, 6 different nonvalidated digital questionnaires were developed by study organizers to address the characteristics of the participating patients and their satisfaction with the app and the practice. Several validated questionnaires were taken into account for the development of these questionnaires [10,25-28]. Groups 1 and 2 had to fill out 3 questionnaires, while Group 3 had to fill out 2 questionnaires. Please change to: "All 3 groups were asked to provide answers to questions regarding satisfaction with the practice. Group 1 additionally had questionnaires that addressed the app (installation and handling), previous use of other health care apps, and the highest educational degree. Group 2 was also asked about the reasons
for denying to use the app, as well as about previous use of other health care apps and the highest educational degree. Group 3 provided answers to questions regarding previous experience with other health care apps and the highest educational degree.

All questionnaires were electronically accessible and could be completed on a webpage of the app developer Interactive Studios (Table 1). See Multimedia Appendix 2 for the questionnaires used in the study groups.

### Table 1. Study questionnaires in the course of treatment.

<table>
<thead>
<tr>
<th>Timeline of the questionnaires</th>
<th>Group 1 (app use)</th>
<th>Group 2 (app denial)</th>
<th>Group 3 (control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery pre-examination appointment</td>
<td>• Questionnaire E (satisfaction with app installation and general information about the patient)</td>
<td>• Questionnaire B (reasons for denying app use) and Questionnaire F (general information about the patient)</td>
<td>• Questionnaire F (general information about the patient)</td>
</tr>
<tr>
<td>1-7 days after surgery</td>
<td>• Questionnaire A (satisfaction with the app)</td>
<td>• No questionnaire</td>
<td>• No questionnaire</td>
</tr>
<tr>
<td>4-6 weeks after surgery</td>
<td>• Questionnaire C (satisfaction with the app and practice)</td>
<td>• Questionnaire D (satisfaction with the practice)</td>
<td>• Questionnaire D (satisfaction with the practice)</td>
</tr>
</tbody>
</table>

Patients who could not fill out the questionnaires online on their own were offered personal support in the practice or via a phone call. Demographic characteristics, including age, gender, and prior cataract surgery, were extracted from medical health records available in the practice. App use data of the entire cohort (Group 1) were collected by the app developer and could be accessed online by the study organizers. Individual personal data were not attainable due to data privacy.

The collected data were subjected to descriptive and quantitative analyses using IBM SPSS Statistics 27 (IBM Corp). Statistical significance was reached at $\alpha=.05$.

### Ethical Considerations

This study was approved by the Ethics Committee of Landesaerztekammer Baden-Wuerttemberg (Stuttgart, Germany; approval number: F-2021-004) as a study involving human subjects and followed the ethical considerations of the Declaration of Helsinki. All participants were thoroughly informed about the study, and if approved, they signed an informed consent form allowing research on the gained data. All obtained data were deidentified, and no images allowing reidentification of the subjects were included. Participants did not receive any monetary compensation. No generative artificial intelligence system was used in any portion for manuscript writing.

### Results

#### Demographics of the Groups

The study included 297 patients who underwent cataract surgery. Of the 297 patients, 77 (25.9%) used the app (Group 1), 61 (20.6%) did not want to use the app (Group 2), and 159 (53.5%) were included in the control group (Group 3). In all groups, more female patients were included. The median patient ages were 69, 79, and 74 years, respectively. The mean patient age was significantly higher in Group 2 than in the other 2 groups (Kruskal-Wallis test, $P<.001$). All patients were asked for their highest degree of education. In Group 1, most patients had completed a professional apprenticeship or a higher university degree. In Group 2, only 8 patients had a higher educational degree. This difference between groups approached significance (chi-square test, $P=.05$). The majority of patients had surgery on only 1 eye.
Table 2. Demographics of the patients in the study groups.

<table>
<thead>
<tr>
<th>Study participants</th>
<th>Group 1 (app use) (n=77)</th>
<th>Group 2 (app denial) (n=61)</th>
<th>Group 3 (control) (n=159)</th>
<th>P value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>34 (44.2)</td>
<td>24 (39.3)</td>
<td>71 (44.7)</td>
<td>.77(^b)</td>
</tr>
<tr>
<td>Female</td>
<td>43 (55.8)</td>
<td>37 (60.7)</td>
<td>88 (55.3)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>Median (minimum-maximum)</td>
<td>69 (50-86)</td>
<td>79 (61-88)</td>
<td>74 (47-94)</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>69.2</td>
<td>77.6</td>
<td>72.7</td>
<td></td>
</tr>
<tr>
<td>Experience with health apps, n (%)</td>
<td>20 (25.9)</td>
<td>3 (5.8)</td>
<td>23 (16.3)</td>
<td>.01(^b)</td>
</tr>
<tr>
<td>Highest degree of education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School degree</td>
<td>23 (29.9)</td>
<td>21 (40.4)</td>
<td>63 (45.0)</td>
<td>.06(^b)</td>
</tr>
<tr>
<td>Professional apprenticeship</td>
<td>29 (37.7)</td>
<td>23 (44.2)</td>
<td>51 (36.4)</td>
<td></td>
</tr>
<tr>
<td>Higher university degree</td>
<td>25 (32.5)</td>
<td>8 (15.4)</td>
<td>26 (18.6)</td>
<td></td>
</tr>
<tr>
<td>Eyes undergoing surgery, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.86(^b)</td>
</tr>
<tr>
<td>One eye</td>
<td>46 (59.7)</td>
<td>39 (63.9)</td>
<td>96 (60.4)</td>
<td></td>
</tr>
<tr>
<td>Both eyes</td>
<td>31 (40.3)</td>
<td>22 (36.1)</td>
<td>63 (39.6)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)P value is the probability of rejecting the correct null hypothesis.
\(^b\)P value was calculated using the chi-square test.
\(^c\)P value was calculated using the nonparametric Kruskal-Wallis test.

Characteristics of Group 2 (App Denial)
Patients who dismissed the app were asked for their reasons. Most (41/61, 67%) answered that they were missing a suitable device. Moreover, 9 patients reported no interest or a lack of technological competence for the use of the app. Two patients mentioned insufficient vision. Most (49/52, 94%) of the patients in this group had no previous experience with health care apps. Group 3 had a relatively high percentage of patients (23/141, 16.3%) who had previous experience. Moreover, in Group 1, 26% (20/77) of patients already had an experience with health apps (chi-square test, \(P=.01\)).

Patient Satisfaction With the Ophthalmological Center During Cataract Surgery
After 4 to 6 weeks, patient satisfaction with the ophthalmological center where they had their surgery was evaluated in all groups (Table 3). In Group 1, 65% (46/71) were very satisfied, 34% (24/71) were fairly satisfied, and 1% (1/71) were unsatisfied. Similar responses were provided in the other 2 groups. In Group 2, 67% (40/60) were very satisfied and 32% (19/60) were fairly satisfied. In Group 3, 66.9% (105/157) were very satisfied and 31.2% (49/157) were fairly satisfied.
### Table 3. Comparison of groups regarding satisfaction with the center and possible areas for improvement within the ophthalmological center.

<table>
<thead>
<tr>
<th>Patient satisfaction</th>
<th>Group 1 (app use), n/N (%)</th>
<th>Group 2 (app denial), n/N (%)</th>
<th>Group 3 (control), n/N (%)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient overall satisfaction with the center</strong></td>
<td></td>
<td></td>
<td></td>
<td>.99&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Very satisfied</td>
<td>46/71 (64.8)</td>
<td>40/60 (66.7)</td>
<td>105/157 (66.9)</td>
<td></td>
</tr>
<tr>
<td>Fairly satisfied</td>
<td>24/71 (33.8)</td>
<td>19/60 (31.7)</td>
<td>49/157 (31.2)</td>
<td></td>
</tr>
<tr>
<td>Unsatisfied</td>
<td>1/71 (1.4)</td>
<td>1/60 (1.7)</td>
<td>3/157 (1.9)</td>
<td></td>
</tr>
<tr>
<td><strong>I was extensively counseled and informed</strong></td>
<td></td>
<td></td>
<td></td>
<td>.97&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Agree completely</td>
<td>49/73 (67.1)</td>
<td>37/59 (62.7)</td>
<td>103/157 (65.6)</td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>21/73 (28.8)</td>
<td>20/59 (33.9)</td>
<td>47/157 (29.9)</td>
<td></td>
</tr>
<tr>
<td>Do not agree (partly)</td>
<td>3/73 (4.1)</td>
<td>2/59 (3.4)</td>
<td>7/157 (4.5)</td>
<td></td>
</tr>
<tr>
<td><strong>I was comprehensibly counseled and informed</strong></td>
<td></td>
<td></td>
<td></td>
<td>.89&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Agree completely</td>
<td>51/73 (69.9)</td>
<td>40/59 (67.8)</td>
<td>108/158 (68.4)</td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>19/73 (26.0)</td>
<td>17/59 (28.8)</td>
<td>40/158 (25.3)</td>
<td></td>
</tr>
<tr>
<td>Do not agree (partly)</td>
<td>3/73 (4.1)</td>
<td>2/59 (3.4)</td>
<td>10/158 (6.3)</td>
<td></td>
</tr>
<tr>
<td><strong>I was friendly and attentively treated</strong></td>
<td></td>
<td></td>
<td></td>
<td>.20&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Agree completely</td>
<td>48/73 (65.8)</td>
<td>43/59 (72.9)</td>
<td>118/158 (74.7)</td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>19/73 (26.0)</td>
<td>14/59 (23.7)</td>
<td>37/158 (23.4)</td>
<td></td>
</tr>
<tr>
<td>Do not agree (partly)</td>
<td>6/73 (8.2)</td>
<td>2/59 (3.4)</td>
<td>3/158 (1.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Areas for improvement (multiple answers possible)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.22&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Center’s organization</td>
<td>20/76 (26.3)</td>
<td>23/65 (35.4)</td>
<td>51/165 (31.1)</td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>11/76 (14.5)</td>
<td>8/65 (12.3)</td>
<td>26/165 (25.9)</td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>9/76 (11.8)</td>
<td>3/65 (4.6)</td>
<td>5/165 (3.0)</td>
<td></td>
</tr>
<tr>
<td>Counseling and information</td>
<td>5/76 (6.6)</td>
<td>7/65 (10.8)</td>
<td>10/165 (6.1)</td>
<td></td>
</tr>
<tr>
<td>No suggestions</td>
<td>31/76 (40.8)</td>
<td>24/65 (36.9)</td>
<td>72/165 (43.9)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> P value was calculated using the chi-square test.

Most patients felt counseled and informed comprehensibly (Group 1: 70/73, 95.9%; Group 2: 57/59, 96.6%; Group 3: 148/158, 93.7%) and in detail (Group 1: 70/73, 95.9%; Group 2: 57/59, 96.6%; Group 3: 150/157, 95.5%). Over 90% of patients in all groups completely agreed or agreed in this matter (Table 3). The vast majority (Group 1: 67/73, 91.8%; Group 2: 57/59, 96.6%; Group 3: 155/158, 98.1%) also felt friendly and attentively treated. There was no significant difference in the response to the questions between the 3 groups (chi-square test, P>.05). Overall, in Group 1, patients were significantly more satisfied with the ophthalmological center when they felt extensively (chi-square test, P=.02) and comprehensibly (chi-square test, P=.006) counseled. The same was true for patients in Group 1 who agreed completely to the statement that they were friendly and respectfully treated (P=.001).

Additionally, patients were asked for areas of possible improvement of the center. Overall, in all groups, roughly 40% (Group 1: 31/76, 40.8%; Group 2: 24/65, 36.9%; Group 3: 72/165, 43.6%) had no complaint (multiple answers were possible). In Groups 2 and 3, 35.4% (23/65) and 31.1% (51/165), respectively, mentioned the center’s organization (multiple answers were possible). In Group 1, the percentage was smaller at 26% (20/76). More patients complained about insufficient communication (multiple answers were possible) in Group 1 (9/76, 11.8%) than in Groups 2 and 3 (3/65, 4.6% and 5/165, 3.0%, respectively; Table 3). More patients mentioned a lack of counseling and information in the center (multiple answers were possible) in Group 2 (7/65, 10.8%) than in Groups 1 and 3 (5/76, 6.6% and 10/165, 6.1%, respectively; Table 3). There were no significant differences between groups.

### Results in Group 1 (App Use)

#### Communication Forms

To ascertain the acceptability of this mode of communication, patients were inquired about their willingness to communicate exclusively through the app. Surprisingly, none of the patients expressed agreement for an exclusively digital communication channel. A significant majority (60/77, 78%) of respondents expressed a preference for a hybrid communication mode, involving both the app and traditional telephone channels. An interesting gender-based trend was observed, with men showing a slightly higher inclination toward this hybrid mode (chi-square test, P=.08). A minority (67/77, 8%) of respondents expressed an exclusive preference for the app. Meanwhile, 14% (11/77)
of participants expressed a leaning toward solely using the telephone for communication.

**Software Installation on the Mobile Device**

The vast majority (59/73, 81%) of patients in Group 1 handled the process of installation very well. Only 5 (7%) patients saw it as a burden and described the installation process as troublesome. However, half (38/77, 49%) of the patients needed support to install the app. Significantly more male patients (chi-square test, $P=.03$) and people with a higher degree of education at a university (chi-square test, $P=.004$) were able to install the app without external help.

Android was the most prevalent operating system (54/77, 70%). The app was used 50 days on average, and the average use time was 6 minutes 44 seconds per session. The most read articles concerned the type of IOL that can be inserted (312 views) and the postoperative recommendation on patient’s behavior (275 views). Information on the practicing doctors (264 views), the process of surgery preparation (243 views), and the recommendations for this appointment (205 views) were also of interest. The description of technical diagnostic examinations was of less interest (less than 5 views).

**Satisfaction With the App**

Participants in Group 1 were asked 1 week postoperatively about the overall experience with the app. In general, the users were very satisfied with the experience (very pleased: 37/77, 48%; pleased: 35/77, 46%). Only 3 (4%) patients were unpleased and none of the patients were very unpleased (Figure 2). There was no difference in satisfaction for patients aged under 70 years, those aged between 70 and 80 years, and those aged 80 years or above. The appreciation for the app 1 week postoperatively was greater among patients with a previous vocational education than among patients with a university degree (chi-square test, $P=.07$).

**Figure 2.** Responses of participants in Group 1 (app use; n=77) regarding app satisfaction after 7 days postoperatively with respect to overall satisfaction, handling, understanding, abbreviations and orientation throughout the perioperative process.

![Figure 2](https://formative.jmir.org/2024/1/e55742/figure2.png)

Most patients also agreed that the handling of the app was pleasing (73/77, 95%), and texts and abbreviations were easy to understand (76/77, 99% and 69/77, 90%, respectively). The vast majority (70/77, 91%) also agreed with the statement that the app provides orientation throughout the perioperative process (Figure 2). Patients who received support for the installation were more satisfied with the app within the first 7 postoperative days (chi-square test, $P=.003$).

**Informational Content of the App**

Patients were also asked about the content provided in the app. Almost all respondents stated that the app is clearly phrased (74/77, 96%), provides sufficient data (68/77, 88%), and provides easy access (67/77, 87%) (Figure 3). Over 90% also agreed that the content is trustworthy (71/77, 92%) and useful (72/77, 94%) during the perioperative process. Patients with a lower degree of education found the content significantly more useful (chi-square test, $P=.04$). Women tended to find the content more useful (chi-square test, $P=.08$). There was no variation in content appreciation across different age groups.
Figure 3. Responses of participants in Group 1 (app use; n=77) regarding app content with respect to clarity, extent, access, trustworthiness, and usefulness.

**Design of the App**

Patients were also queried regarding the design of the app. Satisfaction with the graphical representations was expressed by 90% (69/77) of patients. A predominant proportion (74/77, 96%) of patients concurred that the textual layout was easily legible. A substantial majority (69/77, 90%) affirmed the clarity in content presentation (Figure 4). No significant variations were observed based on age group or gender.

Figure 4. Responses of participants in Group 1 (app use; n=77) regarding graphics, text, and clear presentation of the content in the app.
Re-evaluation After 4 to 6 Weeks and Final Suggestions on the Patient’s Journey

Four to six weeks after surgery, patients were asked if the app helped in the whole process. We found that 80% (58/73) of patients stated that the app was helpful. Approximately 20% thought that the app was less helpful (10/73, 14%) or not helpful (1/73, 6%). Patients who had previously used a health app did not appreciate the app significantly more (chi-square test, \( P = .77 \)), but those who had installed the app with help were significantly more satisfied with it (chi-square test, \( P = .02 \)). After 4 to 6 weeks, a higher proportion of female patients expressed appreciation for the app compared to male patients, though the difference was not statistically significant (chi-square test, \( P = .12 \)). Patients with a high degree of satisfaction with the ophthalmological center tended to be more satisfied with the app throughout the cataract surgery journey (chi-square test, \( P = .13 \)). Patients with a lower degree of educational training reported significantly higher satisfaction with the app (chi-square test, \( P = .04 \)), while patients with a university degree reported less satisfaction (chi-square test, \( P = .01 \)).

There were 11 unsatisfied participants, and they were asked why they did not approve of the app. Multiple responses were possible. Five respondents declared that the app did not provide new information. Four mentioned ill functioning of the app or misleading information. Three agreed that the push notifications were disturbing. One patient favored personal contact with a physician.

These respondents were also asked how to improve the app and make it more helpful. Multiple answers were again possible. Four respondents suggested an improvement in the design. Three wanted more personalized functioning of the app, for example, regarding the push notifications. Only 2 proposed an improvement of the informational content and text. Three patients did not provide any suggestions.

Patients who agreed that the app was helpful were also asked how they would improve the app. Eight respondents thought that personalization of the app would be helpful. This mainly concerns the setting of notifications. Four patients desired an improvement of appointment scheduling. Three respondents addressed concerns regarding the font size and contrast of the font and background. Two desired an improvement of the content of the text, and one desired a desktop version of the app.

Discussion

Principal Findings

In this study, 297 patients who underwent cataract surgery were enrolled, and 3 different groups were established. An adaptable mobile app was used by 77 patients in Group 1 for at least 2 weeks prior to the pre-examination day and for 4 to 6 weeks after surgery. Group 2 included patients who denied use of the app, and Group 3 included control patients.

In our study, almost all patients who used the app reported good user friendliness and appreciated the digital data provided for the organization, and the scheduling and information of the surgery. Age did not play a major role in the appreciation of the app when agreeing to use the app, although patients in Group 2 were significantly older. Patients who did not have a higher university degree had the most benefits from the informational content of the app and were the most satisfied with the information. Female patients tended to appreciate the information provided more than male patients. However, male individuals and academics are in general more aware of technology and can handle the app more easily. Nevertheless, other individuals can benefit from the use of the app throughout the cataract surgery journey. App users demonstrated a noninferior high satisfaction with the treatment in comparison with patients who were only conventionally informed about the surgery.

Satisfaction With the Ophthalmological Center

Among all groups, most respondents in this study were very satisfied with the organization and the ophthalmological center where they underwent the surgery. This is in line with the findings of most other studies on the perioperative satisfaction of patients [29-31]. In these studies, outpatients were very content with the information provided, the politeness, the surgical results, and the professional competence.

Cataract Surgery Satisfaction With a Focus on the Information or Content Provided

In our study, almost all enrolled patients felt well informed during the perioperative process, and there was no significant difference between the 3 study groups. Multiple other studies have confirmed that during the process of a surgical intervention, the degree of information provided is a significant predictive factor of satisfaction [31]. For cataract surgery, preoperative information supplied by a video or other digital content can improve knowledge among patients. The app used in our study included such a video. Pager [32] demonstrated in a randomized controlled trial that preoperative display of a video on the phacoemulsification procedure significantly increases patient understanding of and satisfaction with the cataract surgery while decreasing unease. Patients showed these results independent of past cataract surgery and notwithstanding that, in general, patients stated they had already obtained sufficient information beforehand [32]. Another study showed that information levels were higher in patient groups that received additional information from digital animated videos [33]. However, these videos as well as other information should be thoroughly selected by the ophthalmological center as digital or social media information, such as that on YouTube, might be misleading [34].

Consideration of Age and Gender in the Use of Digital Devices

In this study, the average age of participating respondents was unsurprisingly high, with a median age of 69 years in Group 1, 79 years in Group 2, and 74 years in Group 3. Age seemed to play no obvious role in Group 1, which was the interventional user group. No evidence was found that age influenced the use of the app or satisfaction with the app. However, patient age in Group 2 (app denial) was significantly higher, indicating that older patients might be hesitant to use digital devices. Male
patients tend to be more aware of technology than female patients, and a higher educational background favors easier handling of the app. Another study demonstrated that male elderly people had more ease with the use of the internet and digital devices [35]. This is also true for the use of medical apps. Elderly females seem to be more difficult to reach with these innovative new technologies than males [36]. Perceived serenity and a sense of control while using the app, personal innovativeness, self-perceived effectiveness, and service possibilities were aiding factors to animate mHealth use among elderly communities [36,37].

**mHealth and Higher Educational Background**

This study showed that patients with a lower degree of educational training appreciated the content of the app more than those with an academic degree. Patients who do not screen the internet beforehand for consistent information might benefit more, as general understanding and basic knowledge are lower. Studies have evaluated mHealth in diverse ethnic and educational groups, as well as in older and low-income groups. The results showed that mobile devices might be the preferred method of collecting health information from these groups [38]. Patients from low-income backgrounds appreciate the use of mHealth technology to both manage chronic diseases and overall health. As educational and socioeconomic gaps strongly correlate with higher rates of chronic conditions, such as obesity, diabetes, and hypertension, in these communities, mHealth can prove to be an asset [39]. Apps can provide a low threshold source of information and empower people in these communities to improve health outcomes.

**Economic Perspective on mHealth**

The willingness to use a mobile phone and the readiness of a mobile phone are prerequisites for establishing mHealth in clinical practice. The use of mHealth should also benefit and facilitate workflows within medical centers. If the physician’s time invested in the patient’s education on the disease and surgical interventions can be reduced while keeping the knowledge and satisfaction levels high, the center can gain in terms of workflow effectiveness. However, integration in the work process of a clinic can be troublesome, and it has been proven to be a burden for both health care professionals and consumers [3,40]. Development of the cataract education app was encouraged by the desire to reduce the number of common questions raised by patients as well as include additional information so that patients’ visits to the ophthalmologist would be optimized without loss of quality and satisfaction. In this study, patients received all relevant information material concerning the app by mail, could download the app, and could inform themselves about the surgery prior to the appointment for surgery pre-examination. Study organizers believe that this might have helped to reduce the time needed for explanation about the app and the surgery for app users. On the other hand, the time needed might have increased for other patients who did not have prior experience with such health care apps before and needed support for the app setup or standard consultation about the surgery. However, patient consultations could have been more efficient once the app was installed and patients could attend postoperative examinations after being informed over the app. Additional human resources were needed for the preparation and adaptation of the informational content of the app and information about the app, as well as for periodic review of the arrangements of patient appointments and entry into the app database. In the future, complete integration of the app within the software of the ophthalmological center should be pursued, so that all relevant data could be synchronized automatically. The employees of the practice also needed educational courses and an adaptation period in order to learn about the content of the app and be able to support patients with app use. Study organizers found this process to be uncomplicated for employees who owned smartphones and had prior experience with other apps and to be time-consuming for others. Patients in this study were reluctant to shift the communication and scheduling of the ophthalmological center to a solely digital mode. However, from the health economics perspective, mHealth has already been demonstrated to be a valuable and effective tool in health care systems and environments with limited resources [41,42]. We believe that short time investments in the development of the app, preparation of information material, and education of employees can provide long-term benefits for the optimization of workflow in the ophthalmological center.

**Limitations**

The limitations of this study include the inclusion of patients from a single outpatient health care center, analysis of the app for a single disease treatment, and testing of the app solely in elderly patients. The allocation within this clinical study was not blinded; therefore, the outcome ascertainment might have been influenced by the knowledge of this allocation and furthermore might be evident in the median age of the groups investigated. The study was also limited by the fact that the education of app users was not limited to a solely digital form via the app as they also received conventional information.

**Conclusion**

Mobile apps with high user friendliness have the capacity to increase patient knowledge, ensure satisfaction with the treatment, and improve workflows; however, they still face challenges regarding clinical effectiveness, lack of integration in health care delivery, and further validation processes in safety and privacy [9,10]. Further analysis and large multicenter prospective clinical trials are needed to show areas for improvement, prove the potential benefits of apps, and identify specific patient groups that could benefit the most from app use. Future studies could also investigate if apps help to optimize patient management while reducing the duration of consultation, including surveying medical professionals about their experiences with patients using apps. Furthermore, we recommend further developments to synchronize apps with the software of health care centers in order to optimize management and increase the usability of apps.
Acknowledgments
This study was funded by Zeiss, Jena, Germany. The company paid for the fees of the app.

Data Availability
All data are available upon request by email to the corresponding author.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Demonstration video of the cataract app.
[MP4 File (MP4 Video), 84566 KB - formative_v8i1e55742_app1.mp4]

Multimedia Appendix 2
Questionnaires for all groups.
[DOCX File, 27 KB - formative_v8i1e55742_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 449 KB - formative_v8i1e55742_app3.pdf]

References


Abbreviations

IOL: intraocular lens
IOP: intraocular pressure
mHealth: mobile health
Factors Explaining the Use of Web-Based Consultations With Physicians by Young and Middle-Aged Individuals in China: Qualitative Comparative Analysis

Chunyu Zhang¹, PhD; Ning Hu², MSc; Rui Li³, MA; Aiping Zhu⁴, BMed; Zhongguang Yu³, PhD

¹Department of Human Resources, China-Japan Friendship Hospital, Beijing, China
²School of Management, Beijing University of Chinese Medicine, Beijing, China
³Respiratory Centre, China-Japan Friendship Hospital, Beijing, China
⁴Hospital Office, China-Japan Friendship Hospital, Beijing, China

Corresponding Author:
Zhongguang Yu, PhD
Respiratory Centre
China-Japan Friendship Hospital
Yinghua Road 2#
Beijing, 100013
China
Phone: 86 84206468
Email: yzg081892@163.com

Abstract

Background: It was only upon the occurrence of the COVID-19 pandemic that the demand for web-based consultations with physicians grew at unprecedented rates. To meet the demand, the service environment developed rapidly during the pandemic.

Objective: This study aimed to identify the current status of the use of web-based consultations with physicians among young and middle-aged Chinese individuals and explore users' perspectives on key factors that influence its use in terms of optimizing benefits and compensating for disadvantages.

Methods: We conducted semistructured interviews with 65 individuals (aged 18 to 60 years) across China between September and October 2022. The interviewees were selected through snowball sampling. They described their experiences of using web-based physician consultations and the reasons for using or not using the service. Based on the Andersen Behavioral Model, a qualitative comparative analysis was used to analyze the factors associated with the use of web-based physician consultations and explore the combinations of these factors.

Results: In all, 31 (48%) of the 65 interviewees used web-based consultation services. The singular necessary condition analysis revealed that the complementary role of the service and perceived convenience are necessary conditions for the use of web-based consultation services, and user’s confidence in the service was a sufficient condition. Based on the Andersen Behavioral Model, the configuration analysis uncovered 2 interpretation models: an enabling-oriented model and a need-oriented model. The basic combination of the enabling-oriented model included income and perceived convenience. The basic combination of the need-oriented model included complementary role and user’s confidence.

Conclusions: Among the factors associated with the use of web-based consultations, perceived convenience, complementary role, and user’s confidence were essential factors. Clear instructions on the conduct of the service, cost regulations, provider qualifications guarantee, privacy and safety supervision, the consultations’ application in chronic disease management settings, and subsequent visits can promote the positive development of web-based consultations.

doi:10.2196/50036

KEYWORDS

web-based consultation; Andersen Behavioral Model; qualitative comparative analysis; perceived convenience; complementary role; user's confidence; China
Introduction

The term “internet health care service” refers to a closed-loop service that includes health education, medical information inquiry, electronic health files, disease risk assessment, web-based consultation with physicians, electronic prescription, remote consultation, and remote treatment and rehabilitation via the internet and other technological means [1].

The COVID-19 pandemic created a demand for internet health care services at an unprecedented rate [2-4], as patients became reluctant to go to hospitals because of the fear of infection [5,6]. Accordingly, an increasing number of hospitals and internet companies started to, and continue to, venture into the internet health care industry. Reports show that 52% of outpatient departments in Germany have already adopted internet health care services [7]. By June 2022, more than 1700 hospitals in China were providing services using the internet, an increase from 100 in December 2018 [8].

These rapid changes and the quick adoption of internet health care services during the pandemic, however, have impeded the possibility of sufficient analyses on the experience of accessing these services and on how providers can complement the functions of these services to make them more accessible and attractive to users, as well as promote patients’ intention to use the service.

In this study, we only focus on web-based consultations with physicians, which is a core and controversial segment of internet health care services. Some researchers have studied the barriers to and facilitators of web-based consultations and found that perceived convenience, emotional preference, perceived risks, etc, influence behavioral intention [9,10].

However, factors associated with the use of web-based consultations are mixed. Understanding which factors are essential is conducive to optimizing benefits and compensating for disadvantages. Given that young (18-35 years) and middle-aged (35-60 years) individuals are the groups that use web-based consultations the most frequently, we conducted interviews among them to explore the reasons why web-based consultations are used or not used. Then, based on the Andersen Behavioral Model, we applied a qualitative comparative analysis (QCA) approach to analyze evidence from the interviews, to identify how combinations of these interdependent factors lead to the use of web-based consultations with physicians.

Methods

Theoretical Background

The Andersen Behavioral Model, developed by Andersen [11] in 1968, has been widely used to analyze the factors associated with health service use based on 3 dimensions: predisposing, enabling, and need factors [12-14]. Based on the Andersen Behavioral Model, this study also discussed the factors affecting web-based consultations with physicians in China.

QCA Methodology

The use of web-based consultation has complex influences rather than a single effect. QCA has been applied to explore the different combinations of health care interventions because it bridges qualitative and quantitative methodologies [15]. Based on set theory, QCA compares characteristics of the cases in relation to the outcomes by a scoring system. Moreover, QCA has an advantage in analyzing small samples, which usually requires 10 to 80 cases [16,17]. Crisp-set QCA (csQCA) yields binary scores of 0 and 1, indicating “full out” or “full in” in certain conditions [18].

Sample Selection and Data Collection

The semistructured interviews were centered around three broad questions: (1) Do you have experience using internet health care services? (2) If yes, which function do you use and why do you use it? Which function do you never use and why are you reluctant to use it? and (3) If no, why do you never use internet health care services? When describing their experiences, the participants were asked to share examples and not only feelings about internet health care services.

We conducted interviews with residents of provinces in Eastern, Western, and Central China between September and October 2022.

The initial 5 samples were selected by convenience, and they were patients visiting the China-Japan Friendship Hospital. Subsequently, we asked them to recommend 1 or 2 interviewees, such as their friends, colleagues, or relatives, randomly. We repeated this process until the information on why internet health care services were or were not used was saturated. To obtain representative samples, we analyzed the characteristics of former samples and provided detailed requirements with regard to age, location, income, education, and sex for the following samples.

In total, 70 participants were interviewed, and 5 interviews were excluded owing to a lack of information regarding web-based consultations with physicians. The sample size for QCA should be at least $2^k$, where $k$ is the number of conditions [17,19]. The study includes 6 conditions; hence, the sample size should be at least 64. Ultimately, the study included 65 interviews.

Variable Measurement and Calibration

We analyzed the transcripts using a team-based inductive approach. First, the audio data were transcribed verbatim by a third-party company specialized in transcriptions in the Chinese language; once transcribed, the audio recordings were subsequently discarded to protect the participants’ confidentiality. Second, the first round of open coding was conducted using NVivo 12 (QSR International), and we coded the transcripts independently. We discussed and resolved discrepancies and then recoded the data to compile the major themes. Finally, based on the Andersen Behavioral Model, both the conditions and the results were identified by the lead author, reviewed by coauthors, and finalized by the corresponding author (see Table 1). In this study, csQCA was conducted using a program for crisp and fuzzy set with the fsQCA3.0 package (Charles C Ragin and Sean Davey).
Table 1. Variables in the Andersen Behavioral Model of health service use and its measurements.

<table>
<thead>
<tr>
<th>Dimensions of the Andersen Behavioral Model and variables</th>
<th>Measurement condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predisposing</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0 = 18 to &lt;35 years old</td>
</tr>
<tr>
<td></td>
<td>1 = 35-60 years old</td>
</tr>
<tr>
<td>Education</td>
<td>0 = Master’s and doctorate degree</td>
</tr>
<tr>
<td></td>
<td>0.5 = Bachelor’s degree</td>
</tr>
<tr>
<td></td>
<td>1 = Below bachelor’s degree</td>
</tr>
<tr>
<td><strong>Enabling</strong></td>
<td></td>
</tr>
<tr>
<td>Annual income</td>
<td>0 = &lt;CN ¥120,000 (&lt;US $16,804)</td>
</tr>
<tr>
<td></td>
<td>1 = ≥CN ¥120,000 (≥US $16,804)</td>
</tr>
<tr>
<td>Perceived convenience</td>
<td>0 = “Complex conduct procedure” and “late response”</td>
</tr>
<tr>
<td></td>
<td>1 = “Time saving” and “avoiding infection”</td>
</tr>
<tr>
<td><strong>Need</strong></td>
<td></td>
</tr>
<tr>
<td>Complementary role</td>
<td>0 = “Unable to perform physical examination or laboratory test,”</td>
</tr>
<tr>
<td></td>
<td>“compared with face-to-face consultation, less smooth communica-</td>
</tr>
<tr>
<td></td>
<td>“tion,” and “uncovered by social health insurance”</td>
</tr>
<tr>
<td></td>
<td>1 = “Providing primary suggestion” and “handling the minor prob-</td>
</tr>
<tr>
<td></td>
<td>lems”</td>
</tr>
<tr>
<td>User’s confidence</td>
<td>0 = “Doubts about the safety and privacy of the platforms, and about</td>
</tr>
<tr>
<td></td>
<td>the qualification of the doctor”</td>
</tr>
<tr>
<td></td>
<td>1 = There is no doubt</td>
</tr>
<tr>
<td><strong>Use</strong></td>
<td></td>
</tr>
<tr>
<td>Web-based consultation</td>
<td>0 = Never used web-based consultation</td>
</tr>
<tr>
<td></td>
<td>1 = Has experience in using web-based consultation</td>
</tr>
</tbody>
</table>

**Ethical Considerations**

This study was approved by the China-Japan Friendship Hospital (approval 202-ky-032). We asked the participants whether they would be willing to be interviewed over the phone. Once the participants confirmed that they were interested in participating in the study, we made an appointment with them before the interview. At the beginning of the interview, we reviewed a consent form with the participants and obtained their verbal consent to proceed. Interviews were conducted primarily through phone calls because we aimed to reach more residents from different regions across China. All interviews were recorded and transcribed verbatim for data analysis. No compensation was provided for participation.

**Results**

**Participants’ Characteristics**

Participants were recruited from the provinces of Beijing, Shanghai, Guangdong, and Zhejiang in Eastern China (31/65, 48%); Jilin, Henan, and Jiangxi in Central China (14/65, 22%); and Sichuan, Yunan, and Qinghai in Western China (20/65, 31%). In total, 38% (25/65) of the participants were male and 62% (40/65) were female, and the participants’ average age was 35.4 (range 18-51) years (Table 2).
Participants’ Experiences of Web-Based Consultations

In total, 31 (48%) out of 65 participants had experience consulting with physicians over the web. During the COVID-19 pandemic, web-based consultations allowed people to avoid going out and minimized the risk of infection. Although web-based consultations were not always feasible with regard to curing diseases, the participants used them as a prediagnosis tool, which helped them make appropriate decisions regarding what to do next about their potential condition. Participant 1 shared his web-based consultation experience with us:

*My wife was suffering from gallstones. We paid for an appointment with a famous physician to receive advice on the need for surgery. After uploading the results of an exam and consulting with the physician through the internet, we accepted his suggestion and she underwent an operation.*

Factors Explaining the Use of Web-Based Consultations With Physicians

*Necessity Analysis of Individual Conditions*

The first step of QCA is to examine whether a single condition (including its noncollection) is necessary for a complete merger. When the consistency level is greater than 0.8, the condition is considered sufficient for the use of web-based consultations with physicians. When the consistency level is greater than 0.9, the condition is regarded as necessary for the use [17,20,21].

Table 3 shows the test result of the necessary conditions for the use of web-based consultations with physician using the fsQCA 3.0 package. The consistency of “complementary role” and “perceived convenience” exceeded 0.9. Thus, the complementary role of web-based consultations and its perceived convenience are necessary conditions for the use (consistency of 0.968 and 0.935, respectively), followed by user’s confidence (consistency of 0.806), which is a sufficient condition for the use.

---

Table 2. Demographic characteristics of the participants (n=65).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Using web-based consultation, n (%)</th>
<th>Not using web-based consultation, n (%)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=25)</td>
<td>13 (52)</td>
<td>12 (48)</td>
<td>0.302 (1)</td>
<td>.58</td>
</tr>
<tr>
<td>Female (n=40)</td>
<td>18 (45)</td>
<td>22 (55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age range (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 to &lt;35 (n=23)</td>
<td>9 (39)</td>
<td>14 (61)</td>
<td>1.046 (1)</td>
<td>.31</td>
</tr>
<tr>
<td>35-60 (n=42)</td>
<td>22 (52)</td>
<td>20 (48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Annual average income (CN ¥; US $)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;120,000 (&lt;16,804; n=31)</td>
<td>12 (39)</td>
<td>19 (61)</td>
<td>1.917 (1)</td>
<td>.17</td>
</tr>
<tr>
<td>≥120,000 (≥16,804; n=34)</td>
<td>19 (56)</td>
<td>15 (44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern China (n=31)</td>
<td>17 (55)</td>
<td>14 (45)</td>
<td>1.956 (2)</td>
<td>.38</td>
</tr>
<tr>
<td>Central China (n=14)</td>
<td>7 (50)</td>
<td>7 (50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western China (n=20)</td>
<td>7 (35)</td>
<td>13 (65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s and doctoral degrees (n=20)</td>
<td>13 (65)</td>
<td>7 (35)</td>
<td>4.231 (2)</td>
<td>.12</td>
</tr>
<tr>
<td>Bachelor’s degree (n=29)</td>
<td>13 (45)</td>
<td>16 (55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below bachelor’s degree (n=16)</td>
<td>5 (31)</td>
<td>11 (69)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Necessity test of the factors associated with use of web-based consultations with physicians.

<table>
<thead>
<tr>
<th>Condition variable</th>
<th>Use of web-based consultations</th>
<th>~Use of web-based consultations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.484</td>
<td>0.517</td>
</tr>
<tr>
<td>~Age</td>
<td>0.516</td>
<td>0.444</td>
</tr>
<tr>
<td>Education</td>
<td>0.516</td>
<td>0.464</td>
</tr>
<tr>
<td>~Education</td>
<td>0.484</td>
<td>0.492</td>
</tr>
<tr>
<td>Income</td>
<td>0.355</td>
<td>0.611</td>
</tr>
<tr>
<td>~Income</td>
<td>0.645</td>
<td>0.426</td>
</tr>
<tr>
<td>Perceived convenience</td>
<td>0.925 b</td>
<td>0.547</td>
</tr>
<tr>
<td>~Perceived convenience</td>
<td>0.065</td>
<td>0.167</td>
</tr>
<tr>
<td>Complementary role</td>
<td>0.968</td>
<td>0.698</td>
</tr>
<tr>
<td>~Complementary role</td>
<td>0.032</td>
<td>0.045</td>
</tr>
<tr>
<td>User’s confidence</td>
<td>0.806</td>
<td>0.758</td>
</tr>
<tr>
<td>~User’s confidence</td>
<td>0.394</td>
<td>0.188</td>
</tr>
</tbody>
</table>

a~ indicates that a factor does not appear or is "not."
bItalics denote that the consistency exceeded 0.8.

Adequacy Analysis of Conditional Configuration

In operating the truth table, the configuration analysis was applied to reveal the sufficiency analysis of the use caused by different configurations composed of multiple conditions. We set the consistency threshold to 0.8 and the case frequency threshold to 1 and calculated the complex solution, parsimonious solution, and intermediate solution.

As indicated in Table 4, there are 4 paths to promote the “use of web-based consultation with physicians.” Among the 4 combined paths, the unique coverage of S2 and S4 was 0.177 and 0.274, respectively. The unique coverage of S1 and S3 was 0.032. In total, these 4 paths showed strong explanatory power due to the good consistency (0.953) and the relative high coverage (0.661).

Table 4. Configuration analysis of the factors associated with the use of web-based consultations with physicians (solution consistency=0.953; solution coverage=0.661).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Configuration</th>
<th>Need-oriented model</th>
<th>Enabling-oriented model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
</tr>
<tr>
<td>Age</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>c</td>
<td>d</td>
<td>e</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived convenience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complementary role</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User’s confidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Raw coverage</td>
<td>0.145</td>
<td>0.290</td>
<td>0.032</td>
</tr>
<tr>
<td>Unique coverage</td>
<td>0.032</td>
<td>0.177</td>
<td>0.032</td>
</tr>
</tbody>
</table>

aGeneral conditions.
bGeneral conditions do not appear.
cCore condition.
dCorresponding conditions with path do not matter.
eCore conditions do not appear.
Based on the Andersen Behavioral Model [11], we merged the 4 paths into 2 to build a more explanatory model. The first interpretation model is an enabling-oriented model (M1), which includes paths S1 and S2. The basic expression is $M_1 = \text{age} \times \text{income} \times \text{perceived convenience} \times \text{complementary role}$ ($\times \text{education} \times \times \text{user's confidence}$). The basic combination is the enabling dimension including income and perceived convenience. That is, when web-based consultation brings perceived convenience, the relative high-income group will opt for it.

The participants regarded time saving and avoiding infection during the COVID-19 pandemic as the main conveniences brought by web-based consultations, whereas they regarded complex conduct procedures and late responses as inconveniences.

Participant 2 used web-based consultations because of its time saving characteristic. He said the following:

*I have always made appointments with physicians through Haodaifu [an internet health care platform]. I am satisfied with their services because this website informs me about an upcoming appointment beforehand. Meanwhile, the physicians come [for the consultation] on time. The waiting time is not much.*

Participant 3 used web-based consultations to avoid COVID-19 infection. She said the following:

*I get nervous when my little kid feels any discomfort. On the one hand, I am afraid to go to the hospital because of the risk of infection owing to the COVID-19 pandemic. On the other hand, I also get worried about the adverse consequences of delaying [the child’s treatment]. As a result, I usually opt for a web-based consultation immediately, and use it to determine the necessity of an in-person visit.*

Participant 4 complained about the complex procedures that lack instructions. He said the following:

*The registration process is complex. A lot of personal information must be entered before beginning the web-based consultation. Due to a lack of clear instructions, it is difficult to figure out how to begin the service. I attempted to register the system, but it was unable to use the service.*

The second interpretation model is a need-oriented model ($M_2$), which includes paths S3 and S4. The basic expression is $M_2 = \text{complementary role} \times \text{user's confidence} \times \sim \text{income} \times \sim \text{education} \sim \text{perceived convenience}$ ($\times \sim \text{age} \times \text{age} \times \text{education} \times \sim \text{perceived convenience}$). The basic combination is the need dimension including complementary role and user's confidence. That is, regardless of age and education, when web-based consultations are needed, the relative non-high-income group does not care whether it is inconvenient and will opt for it.

In terms of minor problems or primary suggestions, web-based consultations were regarded as complementary to conventional consultations. Participants 3 and 5 said the following, respectively:

*I also get worried about the adverse consequences of delaying [the child's treatment]. As a result, I usually opt for a web-based consultation immediately, and use it to determine the necessity of an in-person visit.*

Some specialties, such as dentistry and ophthalmology, require careful examination through the use of instruments before making the diagnosis. Regarding urgent cases, it would still be better for patients to visit the hospital.

Compared with in-person consultations, web-based communications are less smooth because physicians are unable to observe the patient’s body language and emotions. Some participants mentioned that web-based consultation services cannot perform laboratory tests and physical exams when it comes to diagnosis. Moreover, the costs of web-based consultations are not yet covered by the social health care insurance system. This means that patients will have to bear the cost of internet health care services. Due these reasons, some participants did not regard it as a substitute for in-person consultations. Regarding this, Participants 6 and 7 stated the following, respectively:

*If it is not face-to-face consultation, I am afraid I could not describe [the symptom] clearly and the doctors would misunderstand me.*

*For senior physicians of P Hospital, the cost of a web-based consultation is three times that of a conventional consultation. Meanwhile, the expenditure on web-based consultations cannot be reimbursed by social healthcare insurance.*

Some participants do not have confidence in web-based consultation services owing to privacy, safety, and qualification concerns, as well as problems surrounding web-based diagnosis.

Below is an interview excerpt of Participant 8, who has experience in using text web-based consultations but not video consultations:

*Although I never used video consultations, I am afraid that the system records the whole process automatically. I am worried that the video will be misused without my permission. Meanwhile, it is difficult to confirm the qualification of the doctors providing the service. After visiting the professor in C Hospital for a lung infection, I uploaded the results of a chest CT for further suggestions. I doubted the suggestion made by the professor’s students primarily. Given the busy schedule of the professor, his students had made the initial suggestions, which were later checked by the professor himself. So, I only trust the platforms run by public hospitals.*

**Robustness Test**

To test the robustness, we increased the consistency level from 0.8 to 0.85 and we also decreased it to 0.72. The result showed that the configuration paths after the adjustment were consistent with those before the adjustment, and the coverage and consistency did not change substantially. Therefore, the results were robust.
Discussion

Principal Findings

We examined the current status of the use of web-based consultations with physicians and the factors associated with the service among young and middle-aged Chinese individuals. About half (31/65, 48%) of 18- to 60-year-old residents have experienced web-based consultations. Among the factors associated with the use of web-based consultation, perceived convenience, complementary role, and user’s confidence were found to be the most essential factors.

Optimizing Web-Based Consultations

We found that perceived convenience is a necessary condition enabling participants to use web-based consultations. In this study, time saving and avoiding COVID-19 infection, which are conveniences provided by web-based consultations, promote the participants’ use of it. Participants in our study, similar to patients in other countries, strongly wish to spend less time to access the services, both when making appointments and while waiting for the appointment at the location; they prefer web-based access to appointment scheduling, want SMS text messaging services for reminders, and prefer for physicians to be available during evenings and weekends [22]. The web-based consultation system provided patients with time-saving and convenient solutions for their health care needs across all treatment processes. Patients could make appointments according to their own schedule and do not have to spend time traveling to the appointments. These findings concur with the research done in the United States. Almathami et al [23] conducted a survey in Saudi Arabia and found that saving time would increase the motivation toward the use of web-based consultations.

The COVID-19 pandemic positively influenced web-based consultation use. This is in line with findings of past studies. Studies show that internet health care services enable patients to avoid going out, decrease the time spent at hospitals when patients need to visit hospitals, and minimize infection risks [3,22,24]. Thus, it is not surprising that the COVID-19 pandemic catalyzed the development and use of the service. Although the service cannot fully substitute traditional in-person appointments, various patients were willing to use web-based consultations in the post–COVID-19 era in the long term.

In our research, participants remarked about the unclear instructions hampering the use of the service. Prior studies also find that patients who lack basic internet-related knowledge are excluded from internet health care services [25,26]. The web-based consultation environment requires patients to be well versed in using web-based platforms and electronic gadgets, and the skill levels regarding this vary by patient. Those with low literacy or limited internet-related knowledge are reluctant to use the service. This potential situation was highlighted in prior studies [27,28]. In the future, web-based consultation providers could attempt to assist persons with less accessibility to the platforms by creating intuitive instructions or even providing staff to support these people and explain how they can navigate the service step-by-step. They can also train patients on the use of available technologies prior to them making an appointment. For example, a video on how to book a web-based consultation could be provided on the front page to guide patients.

Focusing on the Needs of Residents

We found that once the participants’ needs were met, they opted to use web-based consultations with physicians.

In the study, although web-based consultations do not have provisions for laboratory tests and physical exams, they serve as a supplement for minor issues and primary suggestions. Meanwhile, some studies reported that web-based consultations improved outcomes in chronic disease management such as diabetes and hyperactivity disorder [29,30]. Considering patients’ preference and need, applying the service in chronic diseases management and subsequent visits may expand its complementary role and benefit patients to a greater extent.

Our results found that some participants did not regard web-based consultations as complementary due to its cost. This is in line with previous studies that found that the cost is a barrier influencing the use of the service, even in high-income countries [26]. Moreover, similar to Germany and the United States, clear regulations about web-based consultations are lacking in China; accordingly, not only do the costs of web-based consultations vary widely, but expenditures on the service are not yet covered by the social health insurance systems [5,31]. Thus, the economic burden on patients may impede their use of web-based consultations.

Users’ confidence is a sufficient factor influencing the use of web-based consultations in the study. The participants use the service if they feel safe. Several participants expressed concerns about the safety and privacy of web-based platforms, as well as the qualification of physicians. This corroborates the findings of prior studies, wherein participants expressed their concern about such safety and privacy issues and believed that the safety and privacy of users should be guaranteed by clear regulations for such services [32,33]. These regulations should ensure that patient data cannot be misused for purposes other than health care or shared without patients’ informed consent. Best practices and standards should also be created to ensure that providers have the relevant qualifications and service quality to provide web-based consultations.

Limitations

Because of the COVID-19 pandemic, our semistructured interviews were mostly conducted through phone calls. This hindered our ability to observe the participants’ body language and nonverbal cues. Nonetheless, we contacted the participants to explain the topic and purpose of the interview. We shared the questions with the participants 1-7 days in advance, which the participants deemed as adequate and reasonable, enabling them to provide more comprehensive information.

Conclusions

In conclusion, the Andersen Behavioral Model represents a profound reflection and exploration of the factors associated with web-based consultation use from the user’s perspective. Additionally, the csQCA offers guidance for optimizing the
benefits of the service. Perceived convenience, complementary role, and user’s confidence are the essential influencers associated with the use of the service. Clear instructions, comprehensive regulations, and appropriate application can promote the positive development of web-based consultations.

Acknowledgments
We would like to thank the interviewees who participated in the study. This work was supported by National Natural Science Fund for Young Scholars of China (72104255), Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS; 2021-12M-1-046), and National Health Commission Human Resources Development Center for Public Hospital Human Resource Research Project (RCLX2215018).

Data Availability
The data that support the findings presented in this study are available from the corresponding author on reasonable request.

Authors' Contributions
CZ and ZY played a significant role in study design, recruitment, data coding, and paper writing. NH was responsible for conducting the statistical analysis and drafting the Methods and Results sections. RL performed all interviews and data coding. AZ contributed to data coding. All authors thoroughly reviewed the paper before submission and granted their approval for publication.

Conflicts of Interest
None declared.

References


Abbreviations

csQCA: crisp-set qualitative comparative analysis
M1: enabling-oriented model
M2: need-oriented model
Factors Explaining the Use of Web-Based Consultations With Physicians by Young and Middle-Aged Individuals in China: Qualitative Comparative Analysis


JMIR Form Res 2024;8:e50036
URL: https://formative.jmir.org/2024/1/e50036
doi: 10.2196/50036
PMID: 38551645

©Chunyu Zhang, Ning Hu, Rui Li, Aiping Zhu, Zhongguang Yu. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
A Novel Electronic Record System for Documentation and Efficient Workflow for Community Health Workers: Development and Usability Study

Harshdeep Acharya 1, MBBS; Kevin J Sykes 2,3, MPH, PhD; Ton Mirás Neira 4, BA; Angela Scott 4; Christina M Pacheco 4,5, MPH, JD; Matthew Sanner 6, BS; Elizabeth Ablah 4, MPH, PhD; Kevin Oyowe 7, ICT; Edward F Ellerbeck 4, MPH, MD; K Allen Greiner 5, MPH, MD; Erin A Corriveu 3, MPH, MD; Sarah Finocchario-Kessler 4,5, MPH, PhD

1Department of Internal Medicine, Saint Peter's University Hospital, New Brunswick, NJ, United States
2Health and Wellness Center, Baylor Scott and White Health, Dallas, TX, United States
3Department of Otolaryngology-Head and Neck Surgery, University of Kansas Medical Center, Kansas City, KS, United States
4Department of Population Health, University of Kansas Medical Center, Kansas City, KS, United States
5Department of Family Medicine & Community Health, University of Kansas Medical Center, Kansas City, KS, United States
6Sanner Software Solutions, Kansas City, KS, United States
7Global Health Innovations, Nairobi, Kenya

Corresponding Author:
Kevin J Sykes, MPH, PhD
Health and Wellness Center
Baylor Scott and White Health
4500 Spring Avenue
Dallas, TX, 75210
United States
Phone: 1 820 0111 ext 214
Email: Kevin.Sykes@bswhealth.org

Abstract

Background: The COVID-19 pandemic added to the decades of evidence that public health institutions are routinely stretched beyond their capacity. Community health workers (CHWs) can be a crucial extension of public health resources to address health inequities, but systems to document CHW efforts are often fragmented and prone to unneeded redundancy, errors, and inefficiency.

Objective: We sought to develop a more efficient data collection system for recording the wide range of community-based efforts performed by CHWs.

Methods: The Communities Organizing to Promote Equity (COPE) project is an initiative to address health disparities across Kansas, in part, through the deployment of CHWs. Our team iteratively designed and refined the features of a novel data collection system for CHWs. Pilot tests with CHWs occurred over several months to ensure that the functionality supported their daily use. Following implementation of the database, procedures were set to sustain the collection of feedback from CHWs, community partners, and organizations with similar systems to continually modify the database to meet the needs of users. A continuous quality improvement process was conducted monthly to evaluate CHW performance; feedback was exchanged at team and individual levels regarding the continuous quality improvement results and opportunities for improvement. Further, a 15-item feedback survey was distributed to all 33 COPE CHWs and supervisors for assessing the feasibility of database features, accessibility, and overall satisfaction.

Results: At launch, the database had 60 active users in 20 counties. Documented client interactions begin with needs assessments (modified versions of the Arizona Self-sufficiency Matrix and PRAPARE [Protocol for Responding to and Assessing Patient Assets, Risks, and Experiences]) and continue with the longitudinal tracking of progress toward goals. A user-specific automated alerts-based dashboard displays clients needing follow-up and upcoming events. The database contains over 55,000 documented encounters across more than 5079 clients. Available resources from over 2500 community organizations have been documented. Survey data indicated that 84% (27/32) of the respondents considered the overall navigation of the database as very easy. The majority of the respondents indicated they were overall very satisfied (14/32, 44%) or satisfied (15/32, 48%) with the database. Open-ended responses indicated the database features, documentation of community organizations and visual confirmation of...
Conclusions: Our database extends beyond conventional electronic medical records and provides flexibility for ever-changing needs. The COPE database provides real-world data on CHW accomplishments, thereby improving the uniformity of data collection to enhance monitoring and evaluation. This database can serve as a model for community-based documentation systems and be adapted for use in other community settings.

(JMIR Form Res 2024;8:e52920) doi:10.2196/52920

KEYWORDS
public health; database; community health worker; social determinants of health; health worker; health workers; CHW; CHWs; community-based; data collection; functionality; develop; development; EHR; EHRs; EMR; EMRs; dashboard; dashboards; health record; health records; documentation; medical record; medical records; equity; inequity; inequities

Introduction

The COVID-19 pandemic added to the decades of evidence that public health institutions are routinely stretched beyond their capacity, particularly in rural areas [1]. Local and state-level health department resource deficiencies reduce their ability to address health disparities [2]. Although the pandemic affected nearly everyone, communities of color, those residing in rural and frontier counties particularly around meatpacking plants, and other vulnerable populations disproportionately experienced COVID-19–related morbidity and mortality in addition to negative economic consequences [3-8]. Community health workers (CHWs) are a cost-effective intervention for achieving health equity [9]. They can establish trust with individuals and families adversely impacted by the social determinants of health (SDoH), sometimes called social risk factors, and effectively connect them to partner organizations that can address their needs.

The Communities Organizing to Promote Equity (COPE) project is an initiative to address health disparities in 20 counties across Kansas, in part, through the deployment of CHWs. This program also includes Local Health Equity Action Teams (LHEATs), which are community coalitions comprised of community residents, organizational leaders, and a cadre of CHWs who were hired for this project [10]. The goal of the COPE project is to mobilize communities to work together and prioritize local health equity issues, develop feasible strategies to address health equity barriers related to SDoH, collaborate with organizations in the community to strengthen access to services, and elevate the community’s voice in public health planning. One of the primary roles of COPE CHWs is developing partnerships with organizations in their respective counties. These partnerships enable CHWs to connect individuals and families with existing resources and services to address their needs. CHWs build community and individual capacity. They identify individuals who need assistance through community events, self-referrals, or referrals from organizations/clinics and actively work with these individuals to develop tailored plans to address their needs. CHWs are active members of their county’s LHEAT and provide insights from clients’ experiences to support the implementation of LHEAT-driven strategies and community-based events.

To document the project’s wide range of both community-based and client-based efforts, we reviewed 7 existing data tracking systems that accommodate CHW workflows, including electronic medical record (EMR) systems and available community-based documentation systems. Some of the limitations of the existing systems included (1) an exclusive focus on medical needs without documenting social risk factors like employment or housing status, (2) community-based platforms that focus on implementing a referral system but fail to track client goals and progress, (3) systems that lack a structure for progress evaluation and reporting, and (4) limitations on customization for specific project needs (eg, could not accommodate tracking of partnerships and events). In these existing systems, CHWs are often required to document in multiple platforms, resulting in fragmented data and increased redundancy, errors, and inefficiency.

Accurate documentation of CHWs’ influence on the community, public health, and health care processes is vital, particularly as states like Kansas consider Medicaid reimbursement models for CHWs [11]. We could not identify an existing platform, system, or database that efficiently captured specific information about partners at community organizations/facilities, events/activities, client demographics and outcomes, and longitudinal assessment of care plans [11]. Thus, we developed a data collection system capable of recording such information through a user-friendly interface with rigorous security measures for storing clients’ protected health information. We are unaware of other data tracking systems covering the full scope of CHW activities. This database includes details on partner organizations supporting medical and social needs, community outreach events, information about client encounters, and progress toward client’s goals. Further, we can customize workflows to account for unique geographic needs and resources. Finally, this system integrates all program activities into a single platform, simplifying program monitoring and evaluation.

In this paper, we describe the COPE database development process and outline the platform’s features and functionality. The lessons learned through this process and the values of this unique database can support further implementation and similar efforts to develop CHW client management systems.

Methods

Database Development

We applied the principles of human-centered design to develop a database covering the varied needs of the COPE project and serving as a data tracking, quality improvement, and evaluation system and data storage on a Health Insurance Portability and Accountability Act–compliant record system, improved client engagement, enrollment processes, and identification of resources.
platform [12]. The human-centered design approach allowed for iterative development informed by a multidisciplinary team of users. We engaged highly experienced CHWs with 20 years of experience (TMN and AS) and a physician well-versed in partnering with CHWs (HA) as the architects of the system’s functionality and user-friendly interface. The team conducted weekly meetings with the program engineers (MS and KO) to iteratively design and refine features. Pilot tests of the database features occurred over several months with 60 CHWs to ensure that the functionality supported their daily use and to identify areas for improvement. The engagement of COPE project management and evaluation staff ensured the pragmatic design of features to support quality improvement feedback to CHWs and facilitate the generation of automated monitoring reports for routine dissemination to community partners.

Three CHW experts mapped the requirements for data collection and created structured proformas indicating the different fields to be captured and their relevance to client outcomes and CHW productivity. For instance, a needs assessment was embedded within the database for CHWs to identify clients’ SDoH barriers, which CHWs would use to create goals for the client and allow them to track client progress toward goal completion, identify organizations providing services or commodities, and document reasons for goals not completed. These features enhanced the ability of CHWs to track clients and link them to critical community resources. Further, a tracking system was incorporated to monitor how CHWs were spending their time and to notify them when contact with a client was needed.

The plans were provided to 2 software developers who engineered the client and server-side applications. The proformas were modified as necessary to improve uniformity and ease of use. Consensus among the development team ensured the capture of valuable data while avoiding overdocumentation. These proformas were used to create mock-up screens envisioned for the database, and the engineering team mapped the pathway from a data entry interface to data transfer and storage. The system was designed to maintain client confidentiality and included user authentication requirements and a data partition to restrict access by county. Only CHWs who were user-assigned to a county had access to the client data within the county. Users must authenticate access through Microsoft’s Azure Active Directory to access the system. In addition to strong password requirements, multifactor authentication was required for each user account to add an extra layer of security. The application also included a time-out mechanism that logged the user out if they were determined to be inactive. The application was built using Microsoft’s Azure cloud hosting and used industry-standard practices for infrastructure—including but not limited to the encryption of data at rest and in-transit, network isolation, and the principle of least privilege for access management policies.

The COPE database team produced a minimum viable product (a beta version with essential components) and then integrated an iterative process incorporating feedback from CHWs into the development of the database at multiple stages (ie, a local development, staging, and production environment), which was sustained through the database implementation process to refine the product. Early development included a staging environment that was deployed to the cloud. The staging environment allowed the team to test the infrastructure and application code safely as new features were introduced to the database. Through this staging environment, CHWs interacting with the database provided feedback (eg, adapting need assessment items, altering information shown in client charts, adding an alert feature), which were integrated with the initial development of the database. Once essential functionality was established, a production environment was created for use by CHWs, supervisors, and administrators. Ongoing feedback via individual meetings, emails, or in-person communications was collected from CHWs and discussed during weekly meetings with the database engineer; once approved, changes were deployed to the production environment. To ensure Health Insurance Portability and Accountability Act (HIPAA) compliance, the production environment uses industry-standard network isolation, data encryption, and access control to protect against breaches and unauthorized use of the application or its data [13].

**Continuous Quality Improvement and Evaluation**

To evaluate CHW performance, we employed a continuous quality improvement process [14]. This process allows for evaluating CHW workflows and accomplishments based on the number of clients enrolled, progress of client status and goals, and productivity measures. The continuous quality improvement plan is completed monthly for CHWs in each participating county. The supervisor provides feedback at the team and individual levels regarding the continuous quality improvement results and opportunities for improvement.

Primary partners in each county receive a quarterly report, which includes information on the number of partnerships entered, events completed, clients served, and individual client progress toward goals. Examples of local successes and challenges experienced by the LIHEAT and CHW teams are also included. This frequent feedback allows each county to reflect on future goals and helps identify trends across the statewide project. Data exports can also be used to evaluate overall project outcomes. The COPE database was designed to be flexible, allowing the team to continue to innovate for future needs. Specifically, these established feedback systems with CHWs and partnering organizations and weekly meetings with the database engineer to implement technical changes supports the ability to adapt any feature, customize workflows, and alter alert systems, thereby maintaining flexibility of the database.

**Data Collection**

A database feedback survey was conducted at 19 months after implementation (May 2022). A link to a 15-item REDCap (Research Electronic Data Capture) survey was emailed to all COPE CHWs and CHW supervisors. Close-ended questions consisted of demographic items (eg, age, education level, primary language) and assessed feasibility of database features and overall satisfaction. Respondents were asked to rank order the perceived importance of various database features. A Friedman test was conducted to determine whether participants had a differential rank-ordered preference for the database features. Open-ended questions assessed individuals’ perspectives toward accessibility and usability of the database.
and database features. Quantitative results were analyzed using SPSS statistics (version 29.0.0; IBM Corp). Qualitative results were analyzed using inductive content analysis [15].

Ethics Approval

The protocol for this study was approved by the University of Kansas Medical Center institutional review board (STUDY00148455: COPE Project). This study was conducted in compliance with standards established by Good Clinical Practice, the International Council for Harmonisation, and the Declaration of Helsinki. Informed consent was obtained from all individual participants included in this study. Survey data were deidentified and stored on the University of Kansas Medical Center’s secure network drive specifically designated for storage of sensitive personal data. Data access will be restricted to those with appropriate institutional review board authorization and limited to principal and coinvestigators, statisticians, and data analysts involved in the analyses. All computer files and systems will be password-protected and accessible only by authorized personnel. The described data were collected as part of quality improvement in which CHW staff were asked to complete a feedback survey; no compensation was required.

Results

Database Development Process

The COPE database development process spanned 5 months and then went live for community pilot testing among COPE CHWs. The addition of auxiliary functions and refining of existing features continued for an additional 5 months. The development involved 613 cumulative working hours of the software developers and approximately 150 person-hours in active discussions with the COPE team. From the user perspective, CHWs each have access to enter client information, create assessments and goals, and update encounters and client status. They can also add and review events and partnerships within the database. CHW supervisors have additional rights to delete events, partnerships, client goals, or encounter information for flexibility purposes. This functionality facilitates quality improvement and provides a pathway for deduplication of records. At launch, the database had 60 active users in 20 counties, with the flexibility to expand.

Database Components and Function

The COPE database is a secure, HIPAA-compliant, comprehensive, electronic, cloud-based application [16]. All user-specific logins are password secured with multifactor authentication and linked to county-specific access. Client documents such as consent forms and other protected health information are stored in a HIPAA-compliant environment [16]. CHWs and supervisors can locate client information, partners, and events pertinent to their daily work by using the web application interface. Data exports are packaged as CSV files to maintain compatibility with analysis software and are only available to system administrators. Exports can include or exclude identifiable data depending on analysis needs and the intended purpose for the data (eg, quality improvement reviews, evaluation analyses).

The database facilitates linkages to organizational partnerships by capturing contact information, facility location, services offered, and service areas. Once the partnership information is entered, users can build queries of organizations filtered by county, name, and services provided to facilitate client referrals. The database also includes data from community events such as the name and location of the hosting organization(s), event’s purpose, intended beneficiary population(s), and number of attendees. Event locations and partner addresses can be geocoded and overlayed with markers of community needs or vulnerability to guide the strategic deployment of resources to high-priority communities. During and after partner community events, the database is utilized for tracking client referrals and CHW engagement.

For client interactions, the system organizes client demographics, insurance status, and SDoH needs assessments (based on modified items from the Arizona Self-sufficiency Matrix and the PRAPARE [Protocol for Responding to and Assessing Patient Assets, Risks, and Experiences] tool; Figure 1) [17,18]. Once the client assessment is performed and entered in the database, client goals and care plans populate under the client’s chart, allowing CHWs to track progress and goal completion while working with clients (Figure 2). CHWs and clients establish time-bound goals and follow them to completion. Protocols were developed to support CHWs as they work with clients to prioritize needs. CHWs can document client referrals from partnering organizations and send referrals out to service organizations. This allows for monitoring the number of referrals, both into and out of organizations. Once applicable goals are completed, CHWs can close the loop with the referring organization or provider.
Figure 1. Screenshot of the staging environment used to train community health workers on entering and updating clients’ needs in the Communities Organizing to Promote Equity database. The data depicted were created for training purposes and are not from an actual client. SDoH: social determinants of health.

Figure 2. Screenshot of the staging environment used to train community health workers on how clients’ needs are populated as goals and tracked for follow-up and resolution in the Communities Organizing to Promote Equity database. The data depicted were created for training purposes and are not from an actual client. CHW: community health worker.

Algorithms and reports are generated to automate alerts for overdue actions and generate population-level summaries for real-time tracking and auditing. For efficient follow-up and priority setting, an automated alerts-based dashboard specific to the user’s account can notify the user of clients needing follow-up and upcoming events. The dashboard also builds a list of clients currently active, personalized to each CHW, including basic demographic and contact information, dates of the last contact, and the client’s status (eg, engaged, referred, discharged) (Figure 3).
Database Usage and CHW Feedback

To date, over 60 CHWs have utilized the database across 20 counties in Kansas. The database contains over 55,000 documented encounters across more than 5,079 clients. Available resources from over 2,500 community organizations and partners have been documented. Nearly all current CHW and CHW supervisors (32/33, 97%) from the 20 COPE counties completed the survey. Of the respondents, 31% (10/32) were in the age range of 24-34 years, 25% (8/32) in the age range of 35-45 years, 25% (8/32) in the age range of 46-56 years, and 19% (6/32) in the age range of 57-67 years. CHW respondents reported a wide range of educational attainment with 9% (3/32) having a high school diploma or GED (general educational development) equivalent, 19% (6/32) some college, 22% (7/32) associate degree, 31% (10/32) bachelor’s degree, 16% (5/32) master’s degree, and 3% (1/32) having a doctorate’s degree. Most CHWs indicated English as their primary spoken language (24/32, 75%), with the remaining 25% (8/32) reporting Spanish as their primary language. CHWs reported spending an average of 7 hours each week on database documentation. Of the respondents, 84% (27/32) indicated that the overall navigation of the database was very easy. The majority of the respondents indicated that they were overall very satisfied (14/32, 44%) or satisfied (15/32, 48%) with the database.

Respondents ranked the following database features in order of most to least important: (1) dashboard (mean 3.26), (2) database alerts (mean 3.42), (3) assessments (mean 3.55), (4) goals (mean 3.97), (5) client demographics (mean 4.23), (6) time tracking (mean 5.68), (7) event charts (mean 5.90), and (8) partnership charts (mean 6.00). There was significant agreement on the rank-ordered preference for the database features ($\chi^2 = 49.5$; $P < .001$) with a Kendall $W$ of 0.228, which indicates moderate agreement between individuals on the preferable ordering of database features.

In their open-ended responses, CHWs described how the database helped build connections with community organizations by establishing them as partners. Documentation of community organization involvement and provision of these data to corresponding partners have translated into increased engagement at community events, increasing the resources
available to community members. Feedback on event and partner data documentation has highlighted the need for improved workflows, specifically for client referrals during community events. Further, insights from CHWs led to the development of client transfer methods so that clients could be securely transferred within the database to another CHW, thereby enabling full access to client status and notes. This allowed clients to continue to progress toward goal completion. Lastly, a database component consistently mentioned by CHWs, further contributing to improved client enrollment processes, was the ability to provide clients with visible confirmation that all consent forms and data are stored in a HIPAA-compliant record system.

Discussion

Principal Results

The robust community-engagement strategies employed by the COPE project require the development of a novel comprehensive database. The COPE database enhances the workflow and facilitates documentation for this innovative project by capturing process and evaluation data regarding individual and community needs and resources, thereby addressing health inequities and adverse SDoH. The data captured by this system extend beyond the conventional medical needs captured by existing EMRs while providing the required flexibility necessary for the ever-changing needs in communities. Finally, it provides a platform to document and quantify community-related information from the needs of individual members to the presence of key partner organizations positioned to meet these needs. The critical role CHWs play in improving community health outcomes and reducing the cost of health care is increasingly recognized [19-23]. However, little has been published on the process or outcomes of CHWs’ work. This project has contributed to the evidence base documenting the impact of CHWs on health inequities. Clients are empowered to complete goals with the assistance of CHWs. A client’s progress is thoroughly documented in the system, thereby establishing successful processes for connecting clients to services and identifying barriers to care. CHW documentation of client referrals (eg, from partnering organizations to service organizations) allows for monitoring of health barriers by demographics and locations and helps to close the loop with the referring partner. The database tracks client status (eg, referred, engaged, discharged), encounters with clients, progression toward goals, and client appointments with partnering organizations. The result is an efficient and effective case management support system to ensure clients receive the right care at the right time. This documentation demonstrates the impact and reach of CHWs, which is key to supporting health policy changes necessary for expanding and sustaining the CHW workforce through Medicaid or Medicare reimbursement models [24].

Data captured in the COPE database will enhance understanding of CHW workflow and CHWs’ ability to bridge gaps in current health systems. These data are critical to support the health policy changes necessary for expanding and sustaining the CHW workforce [25]. A multisectoral coalition with representation from state and local health departments, federally qualified health centers, community-based organizations, CHW leaders, and universities is advocating for sustainable models for the CHW workforce in Kansas through Medicaid and Medicare reimbursement [26]. Medicaid reimbursement for services provided by CHWs is a developing area of health policy and may require documentation systems like that of EMRs; however, these systems are not always applicable for the type of work conducted by CHWs, which may result in an undervaluation of their work and impact. As Kansas advances policy to support Medicaid reimbursement for CHWs, our database is being leveraged to track CHW performance and outcomes, as it is uniquely positioned to support CHW activities conducted outside the clinical setting [27]. Accordingly, this system could benefit community-based organizations or local health departments that employ CHWs.

Limitations

There were limitations to concurrently developing a novel database platform while launching the COPE project. Given the timeline for system development, alternative strategies for tracking CHW activities were needed in the initial months of the project (eg, spreadsheets, forms stored on secure servers). These early data were transferred upon system deployment. Moreover, system customization and refinements were expedited by immediate and iterative feedback, starting with the first CHWs we hired. The database relies on the end user to close the loop with clients when provided with resources and to determine the effect of those resources. These outcome data are also provided to the referring partners. At this point, the database is not linked to an EMR system; however, the capacity to establish this linkage does exist. Interoperability with other systems implemented in federally qualified health centers and hospitals limited widespread implementation in the initial stages of launch. Most EMRs are not able to accommodate bidirectional data movement between the EMR and external databases. Consequently, depending on where a CHW is employed, this lack of EMR interoperability may create a double documentation issue for the end user.

Further, lack of infrastructure may cause potential limitations in the future. Currently, the project lacked support from a software firm, for example, EPIC or eClinicalWorks, which limited the applicability of the system’s interoperability. We anticipate interoperability with large-scale EMRs in the future, which would ideally allow the inclusion of needed documentation and information from CHWs into clinical care encounters, further facilitating the ability for dynamic navigation with CHWs and health care providers. With the expansion of the current systems’ novel features supporting medical and social needs, community outreach events, and progress toward client’s goals, clinic-based health care providers’ desire for closed-loop referral information can be improved to enhance community-centered care [28].

Comparison With Prior Work

We acknowledge there are existing systems with rigorous features to capture essential aspects of CHW work; however, in the interest of avoiding multiple approaches to capturing the comprehensive range of CHW activities along with individual
and project performance measures, we opted to build our own system. We could not identify an existing platform, system, or database that efficiently captured specific information about partners at community organizations/facilities, events/activities, client demographics and outcomes, and longitudinal assessment of care plans [11]. CHWs, the ultimate end users, designed and refined this system to optimize utility and performance. The current database system collaborates with organizations such as health departments, community-based organizations, health insurance providers, and federally qualified health centers in addition to consulting with similar platforms to incorporate feedback into the database design and discuss possibilities for interoperability of 2 or more systems.

Sustainability
The COPE CHW project managers are continuously monitoring the operability of the database, including frequent review of user feedback and routine communication with system programmers to recommend adaptations and quickly resolve any system malfunctions. User feedback has been particularly helpful in the development and initial implementation phases in tailoring database components to meet CHW, supervisor, and administrator needs. Additionally, the database team collects feedback from partnering organizations and potential new clients looking to adopt the system for their organization. Processes are currently under development to streamline CHW feedback so that CHWs can directly send communications in real time to the database team regarding recommended changes or issues.

Acknowledgments
We thank the Communities Organizing to Promote Equity (COPE) project team (the authors and Vicki Collie-Akers, Kristina Bridges, Clarissa Carrillo, Yvonne Chen, Tatiana Darby, Megan Diniz, Jody Hoener, Allison Honn, Milan King, Nadine Long, Daniel Parente, Mariana Ramirez, Mary Ricketts, Skylar Vance, Jennifer Woodward, Kara Knapp). COPE community health workers, Local Health Equity Action Team members across the state, our partnering organizations, and community residents for all of their hard work and contributions to our shared mission of achieving health equity for communities in Kansas. This work was supported by Kansas Department of Health and Environment through a grant from the Centers for Disease Control and Prevention (grant CDC-RFA-OT21-2103).

Data Availability
There are no data to share publicly from this project. Data references in this manuscript are limited to quality assurance and improvement and would not be useful to those outside of the Communities Organizing to Promote Equity project.

Authors’ Contributions
All authors provided substantial contributions to the design of the work and the conceptualization of this project. HA, KJS, TMN, CMP, AS, and SF-K contributed to the interpretation of data for the work and drafted the final manuscript. All other coauthors provided critical revisions and final approval of the manuscript for publication. All authors agree to be accountable for all aspects of the work and assure the accuracy and integrity of the work represented within this manuscript. Generative artificial intelligence was not used in any portion of the manuscript writing. Funding acquisition was led by SF-K.

Conflicts of Interest
None declared.

References


27. Stelzner WC. Department of health and environment, division of health care finance public notice. Kansas Secretary of State, Kansas Register. 2023. URL: [https://tinyurl.com/2eh3e8fu](https://tinyurl.com/2eh3e8fu) [accessed 2023-04-24]


Abbreviations

CHW: community health worker
COPE: Communities Organizing to Promote Equity
EMR: electronic medical record
GED: general educational development
HIPAA: Health Insurance Portability and Accountability Act
LHEAT: Local Health Equity Action Team
PRAPARE: Protocol for Responding to and Assessing Patient Assets, Risks, and Experiences
REDCap: Research Electronic Data Capture
SDoH: social determinants of health

©Harshdeep Acharya, Kevin J Sykes, Ton Mirás Neira, Angela Scott, Christina M Pacheco, Matthew Sanner, Elizabeth Ablah, Kevin Oyowe, Edward F Ellerbeck, K Allen Greiner, Erin A Corriveau, Sarah Finocchario-Kessler. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Web-Based Emotion Regulation Training for Sexual Health: Randomized Controlled Trial

Vinicius Jobim Fischer¹, PhD; Maila Rossato Holz², PhD; Joël Billieux³, PhD; Gerhard Andersson⁴,⁵, PhD; Claus Vögele¹, PhD

¹Institute for Health and Behaviour, Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg
²Conectare Neuropsi, Porto Alegre, Brazil
³Institute of Psychology, University of Lausanne, Lausanne, Switzerland
⁴Department of Behavioural Science and Learning, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden
⁵Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

Corresponding Author:
Vinicius Jobim Fischer, PhD
Institute for Health and Behaviour
Department of Behavioural and Cognitive Sciences
University of Luxembourg
11, Porte des Sciences
Esch-sur-Alzette, L-4366
Luxembourg
Phone: 352 46 66 44 9740
Email: viniciusjfischer@gmail.com

Abstract

Background: Effective emotional regulation (ER) skills are important for sexual function, as they impact emotional awareness and expression during sexual activity, and therefore, satisfaction and distress. Emotion regulation interventions may offer a promising approach to improve sexual health. Web-based emotion regulation may be a therapeutic strategy for men and women with sexual health concerns. Nevertheless, there is a scarcity of intervention trials investigating its effects in this context, much less using the internet.

Objective: This study aims to investigate the effects of a web-based emotion regulation training program for sexual function in both men and women.

Methods: The participants were recruited based on their self-reported sexual problems, which for men was defined by a score of <25 on the International Index Erectile Function (IIEF) and for women by a score of <26.55 on the Female Sexual Function Index (FSFI). The final sample included 60 participants who were randomized to either a web-based emotion regulation training for sexual function or to a waitlist control group. The treatment consisted of an 8-week web-based emotion regulation training for sexual function. The participants were assessed at baseline, post intervention, and the 3-month follow-up.

Results: Of the 60 participants included, only 6 completed all 3 assessment points (n=5, 20% in the treatment group and n=1, 5% in the waitlist control group) after receiving the intervention. At follow-up, there were no significant differences between groups in any measure. Among the intervention completers, large-to-moderate within-group effect sizes were observed between the assessment points on measures of emotion regulation, depression, lubrication, orgasm, thoughts of sexual failure, and abuse during sexual activity. The adherence rate was very low, limiting the generalizability of the findings.

Conclusions: Participants who completed the intervention showed improvements in both sexual function domains and emotion regulation. Nonetheless, due to a high dropout rate, this trial failed to collect sufficient data to allow for any conclusions to be drawn on treatment effects.

Trial Registration: ClinicalTrials.gov NCT04792177; https://clinicaltrials.gov/study/NCT04792177

(JMIR Form Res 2024;8:e50850) doi:10.2196/50850

https://formative.jmir.org/2024/1/e50850

JMIR Form Res 2024 | vol. 8 | e50850 | p.1575

(page number not for citation purposes)
**KEYWORDS**
emotion regulation; internet; sexual health; FSFI; randomized controlled trial; intervention; psychosexual intervention; sexual disorder; sexual dysfunction; internet-based

**Introduction**

**Background**

Sexual dysfunctions involve difficulty in the ability to respond sexually or obtain sexual pleasure and are common and often disabling conditions [1]. Their etiology is multifactorial and encompasses biopsychosocial factors [2]. Throughout the lifespan, it has been estimated that 40% to 45% of adult women and 20% to 30% of adult men in the general population fulfill the criteria for at least 1 sexual dysfunction [3]. As an example, in the last available national survey in the United Kingdom, 51.2% of women and 41.6% of men who had sex in the last year reported at least 1 sexual problem [4].

The cognitive-affective model of sexual dysfunctions [5] posits that both healthy individuals and those with physical or mental conditions respond differently to sexual situations. While men and women without sexual dysfunctions respond to sexual situations with positive affect, expectations of success, and perception of control, those with sexual dysfunctions respond with anxiety, negative affect, and expectations of failure. In addition, preoccupations about erection and disengagement/thoughts about failure in men and lack of erotic thoughts in women have been found to be negatively correlated with sexual arousal. Across both sexes, sadness and disillusion are negatively associated with sexual arousal [6].

Emotion regulation (ER), the process by which emotions are generated, experienced, and used [7], has also been associated with sexual response cycle difficulties (arousal, lubrication, orgasm, pain, erection, and ejaculation) in both men and women [8-11]. Taking into consideration the high prevalence rates of emotion regulation difficulties and the high comorbidity of sexual disorders with other mental disorders [12], studies are needed to address ER interventions for sexual dysfunctions. The few studies investigating the effects of ER interventions have reported positive results, such as a reduction in sexual compulsivity, drug use, HIV risk behaviors, anxiety, and depression [13], and improvements in sexual functioning and quality of life [14].

Learning effective ER skills is important for sexual function, as they impact emotional awareness and expression during sexual activity, and, therefore, satisfaction and distress [15]. There is evidence that adaptive ER strategies are associated with better sexual function and mental health (anxiety and depression) [6] and reduced sensitivity and reactivity to negative stimuli [15].

Psychosexual interventions have been shown to be effective in treating sexual disorders [16]. They combine several interventions to address disorders experienced by both sexes. Generally, they are based on cognitive and behavioral interventions but can also include couple and systemic therapy components [17]. Most often, psychosexual therapies based on cognitive behavior therapy include (1) psychoeducation about sexual function, (2) cognitive challenges to negative sexual attitudes and distortions, and (3) behavioral activation via sex therapy—specific techniques [18]. The benefits of psychosexual therapies not only involve functional aspects but can also improve other domains of life, such as personal well-being and relationship quality and satisfaction [19,20].

Sexual dysfunctions are often perceived as sensitive, embarrassing, and potentially stigmatizing by those concerned. The relative anonymity of web-based approaches may offer an opportunity to attend therapy sessions in an environment that is perceived as safer and more private, while not requiring much investment of time for both the therapist and the client [21-23].

In general, web-based psychological interventions have been found to be effective in treating a range of conditions. The potential benefits of web-based interventions include their anonymity, availability, and convenience of pace/time [24]. In addition, web-based interventions can support active learning via interactive components [25].

To date, only a few studies have investigated the effects of web-delivered interventions for sexual dysfunctions. None of these studies included ER training. In one study, a treatment program for various sexual dysfunctions led to improvements in sexual functioning in 67% (n=39) of the participants, with improvements maintained at 1-month follow-up [26]. Similarly, treatment was found to be superior to a waitlist control condition in a study on internet-delivered sex therapy for erectile dysfunction [27] and in a study comparing web-delivered cognitive behavioral therapy (CBT) to an internet discussion forum control group [28].

Despite the importance of ER for sexual function, web-based ER interventions have, to our knowledge, not been developed and tested to any greater extent in people with sexual dysfunction. Therefore, this study aims to tackle this gap in the literature as the first ER web-based intervention designed to improve sexual function.

**Aim**

This study aimed to investigate the effects of a web-based ER training program for sexual function in male and female participants experiencing sexual dysfunction named TREP3S (Portuguese acronym for Emotion Regulation Training for Sexual Health). The training was provided in the form of an 8-week web-based program with therapist support via email on a secure treatment platform. We expected that TREP3S group participants would show improvements in self-reported outcomes compared to the waitlist control group.

**Methods**

**Overview**

Participants were recruited from the general public via the internet. Advertisements concerning the project and invitations to take part in the study were disseminated using social media.
(Facebook and Instagram) for 4 months, targeting the Brazilian Portuguese–speaking population. Additionally, participants of a previous web-based survey known as the Sexual Health and Emotion Regulation (SHER) study who had volunteered to participate in future studies were also contacted.

The following inclusion criteria were used: (1) access to a computer/tablet or phone with an internet connection; (2) between 18 and 65 years of age; (3) fluent in Brazilian Portuguese; (4) self-reported sexual problems, as assessed in men by a score of <25 on the International Index Erectile Function (IIEF) and in women by a score of <26.5 on the Female Sexual Function Index (FSFI); and (5) in a stable relationship for at least the preceding last 3 months. The exclusion criteria were: (1) medical conditions that could interfere with the intervention, (eg, diabetes, cancer, and cardiovascular issues) or (2) currently receiving psychological or psychiatric treatment.

Figure 1 illustrates the recruitment steps according to the CONSORT (Consolidated Standards of Reporting Trials) guidelines. After providing informed consent on the platform, the volunteers filled out a screening questionnaire to see if they met the inclusion criteria. For those whose responses indicated suitability for participation in the study, a phone contact was established to verify their eligibility and motivation to participate in the study. Following this, the included participants completed a baseline questionnaire.
Two separate block-randomization lists were created via a computer-generated (SealedEnvelope) block-randomization procedure (comprising a block size of 4 for 2 groups) with a 1:1 randomization ratio. The computer-generated sequence was generated by an independent researcher who was not involved in the trial. The actual allocation of participants to the TREPS intervention group or the waitlist control group was conducted by a clinical psychologist who was not involved in the project. Moreover, the researchers did not influence the allocation of the participants.

After randomization, the TREPS intervention group received the emotion regulation training for sexual function for 8 weeks, while the waitlist control group received no intervention. Both primary and secondary outcomes were assessed online on the study website at baseline, end of treatment, and the 3-month follow-up.
Ethical Considerations

All study procedures were approved by the Ethics Review Panel of the University of Luxembourg (ERP-20-029 SHER). The study was also registered on ClinicalTrials.gov (NCT04792177), and the protocol was published [29]. Prior to taking part in this study, the participants provided electronic informed consent. Personal data collected in this study included information about sociodemographic characteristics and sexual and mental health. To ensure confidentiality, the data collected were pseudonymized and a unique identifier was generated so that participants’ identities could not be disclosed. No compensation was offered to the participants.

Measures

Primary Outcome

The primary outcome measures were IIEF and FSFI scores for the male and female participants, respectively. The main reason for using these scores as main outcomes was that both have shown good reliability and validity for the Brazilian-speaking population [30,31]. For the IIEF, a cut-off score of 25 has been found to discriminate between diagnosis and no diagnosis [32], while for the FSFI, the respective score is 26.55 [33,34].

The IIEF is a 15-item, self-administered questionnaire for assessing sexual functioning in men [35]. Answers are given on a 6-point Likert scale. The IIEF encompasses 5 different domains of sexual functioning: erectile function, orgasm function, sexual desire, intercourse satisfaction, and overall satisfaction. Ferraz and Cicconelli [36] translated and adapted the scale to Brazilian Portuguese. Its psychometric properties have been reported by Gonzáles et al [30].

The FSFI is a 19-item questionnaire for the assessment of sexual functioning in women in the domains of sexual functioning (eg, sexual arousal, orgasm, satisfaction, and pain) [37]. Answers are provided using a 5-point Likert scale. Hentschel et al [31] translated and validated the FSFI into Portuguese.

Secondary Outcomes

The secondary outcome measures encompassed questionnaires about sexual function, mental health (anxiety and depression), ER, sexual self-perception, and thoughts during sexual activity. Sexual function was assessed with both the Sexual Quotient (SQ), using both the female [38] and male versions [39]. The SQ is a brief and comprehensive tool comprising 10 items that are answered on a scale from 0 (never) to 5 (always). It addresses general sexual function, stages of the sexual response cycle (desire, arousal, orgasm), and sexual satisfaction. Scores range between 0 and 100, with a score of ≤60 indicating sexual dysfunction.

Depression and anxiety symptoms were assessed via the Patient Heath Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder-7 (GAD-7). Both instruments are frequently used self-report diagnostic tools for assessing mental health disorders. The PHQ-9 [40] is a 9-item screening instrument that also provides an assessment of depression severity. The diagnostic validity of the tool has been established for its Brazilian version [41].

The GAD-7 [42] is a brief self-report measure specifically developed to assess generalized anxiety disorder. It has good reliability, as well as criterion, construct, factorial, and procedural validity. The Brazilian version has satisfactory psychometric properties [43].

The Difficulties in Emotion Regulation Scale (DERS) was used to assess ER. It covers several facets of ER, including difficulties relevant to an individual’s (1) acceptance of emotional responses, (2) ability to engage in goal-directed behavior under distress, (3) ability to control impulsive behaviors when distressed, (4) access to ER strategies, and (5) emotional clarity. Participants rate their degree of agreement with each statement on a scale from 1 (almost never; 0 to 10%) to 5 (almost always; 91 to 100%). The DERS was developed by Gratza and Roemera [44] and validated in Brazil by Miguel et al [45], with its psychometric properties confirmed by Cancian et al [46].

To assess sexual self-perception, the Sexual Self-Schema Scale (SSSS) was used [47]. It consists of 30 items assessing respondents’ perception of themselves as a sexual person compared to others of the same gender and age. Answers are provided using a 5-point Likert scale ranging from 1 (not at all descriptive of me) to 5 (very much descriptive of me). The psychometric properties of its Brazilian version have been found to be satisfactory [48].

To assess sexual thoughts during sexual activity, the Sexual Modes Questionnaire (SMQ) and its Automatic Thoughts subscale [49] were used. This self-report scale consists of 30 items in the male version and 33 items in the female version. Respondents are asked to rate the frequency, ranging from 1 (never) to 5 (always), with which they have experienced specific automatic thoughts during sexual activity. The psychometric properties of its Brazilian-adapted version have been evaluated [48].

Intervention

The intervention consisted of 8 weekly modules delivered via the internet using a secure web-based contact handling system (eg, similar to internet banking) [50]. Each module consisted of a video presentation, texts, and an exercise (homework) to complete during the week. Access to the intervention modules was granted weekly on the same day, and participants were expected to report homework assignments to the therapist until the day before the next module. The participants were able to contact the therapist via the web-based platform and could expect a reply within 24 hours. No face-to-face or telephone contact with the research team was allowed during the intervention (except for technical problems). Table 1 summarizes the contents of the modules.
Table 1. Summary of the intervention modules.

<table>
<thead>
<tr>
<th>Time frame</th>
<th>Module</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Psychoeducation on sexual function</td>
<td>This module covers information about sexuality, the sexual response cycle (desire, excitement, plateau, orgasm, resolution), and the main difficulties that men and women may face in their sexual function, such as erectile dysfunction, premature ejaculation, desire disorders, pain disorders, and anorgasmia. It also differentiates the psychological characteristics of a functional sexual response from a dysfunctional sexual response and clarifies the difference between sexual function and sexual satisfaction [51].</td>
</tr>
<tr>
<td>Week 2</td>
<td>Psychoeducation on emotions and emotional regulation</td>
<td>This module defines emotions, their evolutionary aspects, emotion functions, emotion response cycles, emotion components (physical sensations, thoughts, and behaviors), and the long-term consequences of maintaining an emotional state for a longer period. It also defines pleasant and unpleasant emotions, the relationship between unpleasant emotions and avoidance behaviors, and avoidance strategies (emotional suppression, distraction, and behavioral avoidance) [52-54].</td>
</tr>
<tr>
<td>Week 3</td>
<td>Relaxation strategies: breathing and muscle relaxation</td>
<td>This module goes over the common physiological responses to anxiety and stress (eg, increases in heart rate, respiration, and muscle tension) and teaches 2 relaxation strategies: breathing relaxation and progressive muscle relaxation [52-54].</td>
</tr>
<tr>
<td>Week 4</td>
<td>Cognitive flexibility</td>
<td>This module refers to the rational component of emotion regulation. It aims to conceptualize and enhance cognitive flexibility. Moreover, the triad situation-thought-emotion is explained and detailed through the concepts of what distinguishes thoughts and interpretations, and what automatic thoughts and cognitive distortions are. The identification of negative thought patterns and the most common cognitive distortions related to sexuality are described [52-54].</td>
</tr>
<tr>
<td>Week 5</td>
<td>Nonjudgmental awareness</td>
<td>This module aims to teach participants to experience their emotions in the present moment in a nonjudgmental way. The module also differentiates between experiencing an unpleasant emotion from experiencing an unpleasant emotion resulting from negative beliefs and reactions (snowball effect) [52-54].</td>
</tr>
<tr>
<td>Week 6</td>
<td>Self-acceptance and compassion</td>
<td>This module focuses on 2 psychological concepts necessary for better emotion management: acceptance and self-compassion. These are important in order not to avoid emotional experiences and to diminish self-criticism associated with sexual difficulties [52-54].</td>
</tr>
<tr>
<td>Week 7</td>
<td>Emotion analysis</td>
<td>This module presents a step-by-step flowchart of how to identify emotions when experiencing them. By identifying emotions properly, we can facilitate effective emotion regulation. The flowchart comprises 6 items to pay attention to when identifying an emotion: (1) emotion, (2) event-trigger situation, (3) evaluation/interpretation, (4) physical sensations, (5) previous similar experiences, and (6) behavior [52-54].</td>
</tr>
<tr>
<td>Week 8</td>
<td>Sexual emotional exposures</td>
<td>This module summarizes all the previous modules and suggests a series of sexual experiences where participants pay attention to emotions experienced during the sexual activities (exposure). This gradual approach diminishes the risk of intense emotions and avoidant behaviors. By succeeding with the exposure and obtaining pleasure during the activities, feelings of danger/discomfort diminish, and more adaptive evaluations arise, facilitating the identification and modification of emotional behaviors.</td>
</tr>
</tbody>
</table>

The supporting clinician was a clinical psychologist and a European-certified psychosexologist. The waitlist control group received the same treatment at the end of the 3-month follow-up assessment.

**Data Analysis**

The baseline data present normal univariate distribution according to skewness (−2 to 2) and kurtosis (−7 to 7). The sociodemographic characteristics of the 2 groups were compared using chi-square tests. Next, a 1-way analysis of variance (ANOVA) on outcome variables was performed to examine group differences at baseline. Mann-Whitney U tests were carried out to compare the 2 groups (intervention and control) at baseline, at the end of intervention (T1), and at follow-up (T2). Given the large amount of missing data (over 80%), we did not impute missing data or estimate missing data using mixed model analyses [33]. All participants who received the intervention and completed the 3 assessment points were compiled and analyzed for within-group efficacy estimation. A within-group nonparametric Fisher test was run to assess pre-post changes. Data were considered significant at a P<.05.

For the parametric analysis, the effect size was calculated using Cohen $d$. For the nonparametric analysis, effect sizes were calculated using $\eta^2$ [55]. All statistical analyses were conducted with SPSS software (version 20.0; IBM Corp).

**Results**

**Participants’ Characteristics**

Initially, 60 participants completed the baseline questionnaires and were randomly assigned to one of 2 groups (intervention and waitlist control). Due to the high dropout rate and incomplete measures, we decided to merge the groups to compensate to some extent for the small sample size at post intervention. Table 2 shows the characteristics of the participants who fully adhered to the intervention and completed all the assessments (baseline, end of intervention, and 3-month follow-up).
Table 2. Participants’ sociodemographic characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Intervention group (n=30)</th>
<th>Control group (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>30 (6.81)</td>
<td>31.3 (6.03)</td>
</tr>
<tr>
<td>Relationship duration (years), mean (SD)</td>
<td>7.60 (5.81)</td>
<td>6.66 (5.84)</td>
</tr>
<tr>
<td><strong>Country of residence, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>28 (93)</td>
<td>28 (93)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (7)</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>29 (97)</td>
<td>29 (97)</td>
</tr>
<tr>
<td>Male</td>
<td>1 (3)</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>22 (73)</td>
<td>26 (87)</td>
</tr>
<tr>
<td>Homosexual</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>7 (23)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Asexual</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>None of the above</td>
<td>1 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Relationship status, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>22 (73)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Informal partnership</td>
<td>7 (23)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Civic/formal partnership</td>
<td>0 (0)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>Married</td>
<td>1 (3)</td>
<td>14 (47)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper secondary</td>
<td>8 (27)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Tertiary/university</td>
<td>22 (73)</td>
<td>26 (87)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student (university)</td>
<td>4 (13)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Working full-time</td>
<td>16 (53)</td>
<td>14 (47)</td>
</tr>
<tr>
<td>Working part-time</td>
<td>4 (13)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Housewife/househusband</td>
<td>2 (7)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Owner/independent</td>
<td>3 (10)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Unemployed/jobseeker</td>
<td>1 (3)</td>
<td>2 (7)</td>
</tr>
</tbody>
</table>

**Intervention Adherence**

Intervention adherence was assessed by examining the number of participants who accessed the platform and opened the weekly modules. More participants completed the end-of-intervention and follow-up assessments than those who completed all the intervention modules. Taking into consideration the total number of participants included in the study, adherence to the different modules varied between 14 (n=2) and 76.6% (n=23). The mean number of modules completed within the 8 weeks was 3.2 in the intervention group and 3 in the control group (after they had completed their intervention). Table 3 summarizes the number of participants who read and watched the modules.
Table 3. Adherence to the weekly modules.

<table>
<thead>
<tr>
<th>Module</th>
<th>Group 1: intervention (n=30), n (%)</th>
<th>Group 2: control(^a) (n=20), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23 (77)</td>
<td>16 (80)</td>
</tr>
<tr>
<td>2</td>
<td>14 (47)</td>
<td>14 (70)</td>
</tr>
<tr>
<td>3</td>
<td>9 (30)</td>
<td>13 (65)</td>
</tr>
<tr>
<td>4</td>
<td>9 (30)</td>
<td>9 (45)</td>
</tr>
<tr>
<td>5</td>
<td>7 (23)</td>
<td>4 (20)</td>
</tr>
<tr>
<td>6</td>
<td>5 (17)</td>
<td>3 (15)</td>
</tr>
<tr>
<td>7</td>
<td>3 (10)</td>
<td>2 (10)</td>
</tr>
<tr>
<td>8</td>
<td>5 (17)</td>
<td>2 (10)</td>
</tr>
</tbody>
</table>

\(^a\)When the intervention was offered to the second group, only 20 (67%) of the initial 30 participants completed the questionnaire and were therefore considered as the second intervention group.

An open-ended question section was used for the participants who completed the study, asking for suggestions and comments. A total of 3 topics emerged: (1) self-confidence, (2) influence on our emotional state of how we imagine others are seeing us, and (3) sexual repression during childhood and adolescence and its consequences in adult life. In addition, 1 participant suggested that the video materials should be longer.

Between-Group Analysis

When comparing the intervention group (n=12, 40%) and the waitlist control group (n=16, 80%), statistically significant differences were found at baseline regarding the subscales about automatic thoughts of abuse (intervention: mean 21.47, SD 7.98; waitlist control: mean 15.40, SD 5.40; \(P=0.004\); \(d=0.81\)) and lack of partner attention (intervention: mean 13.20, SD 4.80; \(P=0.03\); \(d=0.61\)). At the end of the intervention (T1), differences were found regarding orgasm capacity and frequency in the intervention group (n=6, 20%) (mean 3.09, SD 2.54) and waitlist control (n=10, 50%) (mean 2.50, SD 1.47; \(P=0.04\); \(\eta^2=0.28\)), as well as automatic thoughts of failure (intervention group: mean 4.43, SD 3.50; waitlist control: mean 9.33, SD 2.61; \(P=0.01\); \(\eta^2=1.08\)).

Multimedia Appendix 1 presents the between-group comparisons at the 3 assessment time points.

Within-Group Analyses

The intervention effect size calculation was conducted by combining all participants who completed the intervention and the 3-month follow-up assessment (n=5). The results indicated large within-group effect sizes for some of the outcome measures related to improvements in ER (DERS total score, \(\eta^2=0.95\); DERS goals, \(\eta^2=0.94\); DERS nonacceptance, \(\eta^2=0.76\)), depression (PHQ-9, \(\eta^2=0.73\)), orgasm (FSFI: orgasm \(\eta^2=1.25\)), lubrication (FSFI: lubrication, \(\eta^2=0.98\)), and failure automatic thoughts in a sexual context (EPA: failure, \(\eta^2=0.95\)).

Medium effect sizes were found on all other factors of the DERS (\(\eta^2\)=ranging from 0.50 to 0.68), anxiety (GAD-7, \(\eta^2=0.43\)), arousal and partner connection in terms of the female version of the SQ (SQf: arousal and connection, \(\eta^2=0.34\)), preliminaries (SQf: preliminaries, \(\eta^2=0.38\)), abuse automatic thoughts in a sexual context (EPA: abuse, \(\eta^2=0.38\)), and sexual self-schema (SSSS: loving/warm, \(\eta^2=0.34\)).

Table 4 summarizes the within-group calculations.

Table 5 shows sexual function symptom scores for each participant who completed at least 2 assessment points.
Table 4. Within-group analysis after the follow-up.

<table>
<thead>
<tr>
<th>Variables and scales</th>
<th>Mean (SD)</th>
<th>P value</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FSFI(^a) sum</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSFI: Desire</td>
<td>2.63 (1.48)</td>
<td>&gt; .99</td>
<td>0</td>
</tr>
<tr>
<td>FSFI: Arousal</td>
<td>2.56 (2.02)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>FSFI: Lubrication</td>
<td>2.68 (2.25)</td>
<td>.38</td>
<td>0.98</td>
</tr>
<tr>
<td>FSFI: Orgasm</td>
<td>2.38 (1.91)</td>
<td>.14</td>
<td>1.25</td>
</tr>
<tr>
<td>FSFI: Satisfaction</td>
<td>2.64 (1.71)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>FSFI: Pain</td>
<td>3.14 (2.54)</td>
<td>.91</td>
<td>0.12</td>
</tr>
<tr>
<td><strong>DERS(^c) sum</strong></td>
<td>99.65 (24.36)</td>
<td>.02</td>
<td>0.95</td>
</tr>
<tr>
<td>DERS: Nonacceptance</td>
<td>15.73 (6.45)</td>
<td>.06</td>
<td>0.76</td>
</tr>
<tr>
<td>DERS: Goals</td>
<td>16.92 (4.65)</td>
<td>.02</td>
<td>0.94</td>
</tr>
<tr>
<td>DERS: Impulse</td>
<td>14.08 (5.39)</td>
<td>.21</td>
<td>0.51</td>
</tr>
<tr>
<td>DERS: Awareness</td>
<td>15.65 (5.06)</td>
<td>.09</td>
<td>0.68</td>
</tr>
<tr>
<td>DERS: Strategies</td>
<td>23.96 (8.12)</td>
<td>.21</td>
<td>0.50</td>
</tr>
<tr>
<td>DERS: Clarity</td>
<td>13.31 (3.79)</td>
<td>.14</td>
<td>0.60</td>
</tr>
<tr>
<td><strong>EPA(^d): Abuse</strong></td>
<td>14.65 (7.10)</td>
<td>.37</td>
<td>0.38</td>
</tr>
<tr>
<td>EPA: Passivity control</td>
<td>7.65 (4.77)</td>
<td>.57</td>
<td>0.23</td>
</tr>
<tr>
<td>EPA: Negative self-image</td>
<td>13.54 (6.06)</td>
<td>.85</td>
<td>0.08</td>
</tr>
<tr>
<td>EPA: Failure</td>
<td>7.46 (3.41)</td>
<td>.03</td>
<td>0.95</td>
</tr>
<tr>
<td>EPA: Lack of partner attention</td>
<td>10.69 (5.15)</td>
<td>.95</td>
<td>0.03</td>
</tr>
<tr>
<td>EPA: Erotic thoughts</td>
<td>12.85 (5.39)</td>
<td>.80</td>
<td>0.10</td>
</tr>
<tr>
<td><strong>SSSS(^e): Direct/outspoken</strong></td>
<td>0 (0)</td>
<td>.85</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>SSSS: Loving/warm</strong></td>
<td>0 (0)</td>
<td>.41</td>
<td>0.34</td>
</tr>
<tr>
<td><strong>SSSS: Reserved/conservative</strong></td>
<td>0 (0)</td>
<td>.80</td>
<td>0.10</td>
</tr>
<tr>
<td><strong>SQf(^f) sum</strong></td>
<td>54.31 (26.73)</td>
<td>&gt; .99</td>
<td>0.01</td>
</tr>
<tr>
<td>SQf: Sexual desire and interest</td>
<td>7.31 (4.36)</td>
<td>.66</td>
<td>0.19</td>
</tr>
<tr>
<td>SQf: Preliminaries</td>
<td>3.27 (1.82)</td>
<td>.37</td>
<td>0.38</td>
</tr>
<tr>
<td>SQf: Arousal and partner connection</td>
<td>5.81 (3.08)</td>
<td>.41</td>
<td>0.34</td>
</tr>
<tr>
<td>SQf: Comfort</td>
<td>6.19 (3.24)</td>
<td>.75</td>
<td>0.14</td>
</tr>
<tr>
<td>SQf: Orgasm and satisfaction</td>
<td>4.58 (2.98)</td>
<td>.81</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>PHQ-9(^g) sum</strong></td>
<td>13.65 (7.94)</td>
<td>.08</td>
<td>0.73</td>
</tr>
<tr>
<td>PHQ-9: Suicide</td>
<td>N/A</td>
<td>.53</td>
<td>0.27</td>
</tr>
<tr>
<td>PHQ-9: Q10</td>
<td>N/A</td>
<td>.45</td>
<td>0.32</td>
</tr>
<tr>
<td><strong>GAD-7(^h): Sum</strong></td>
<td>10.27 (6.13)</td>
<td>.34</td>
<td>0.43</td>
</tr>
</tbody>
</table>

\(^a\)FSFI: Female Sexual Function Index.
\(^b\)N/A: not applicable.
\(^c\)DERS: Difficulties in Emotion Regulation Scale.
\(^d\)EPA: Automatic Thoughts subscale, Sexual Modes Questionnaire.
\(^e\)SSSS: Sexual Self-Schema Scale.
\(^f\)SQf: Sexual Quotient: female version.
\(^g\)PHQ-9: Patient Health Questionnaire-9.
\(^h\)GAD-7: Generalized Anxiety Disorder-7.
Table 5. Sexual function scores of participants who completed at least 2 assessments.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline (FSFI/SQ)</th>
<th>End of intervention (FSFI/SQ)</th>
<th>3-month follow-up (FSFI/SQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.2/86</td>
<td>30.3/80</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>17.6/40</td>
<td>N/A</td>
<td>2/28</td>
</tr>
<tr>
<td>3</td>
<td>11/36</td>
<td>16.3/50</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>12.3/32</td>
<td>N/A</td>
<td>15.9/N/A</td>
</tr>
<tr>
<td>5</td>
<td>20.9/48</td>
<td>19.7/46</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>22.1/58</td>
<td>26.3/62</td>
<td>30.5/70</td>
</tr>
<tr>
<td>7</td>
<td>11.4/32</td>
<td>26.5/56</td>
<td>22.7/56</td>
</tr>
<tr>
<td>8</td>
<td>2.4/80</td>
<td>3.2/84</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>21.4/60</td>
<td>30.5/64</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>20.8/40</td>
<td>31.4/72</td>
<td>18.2/46</td>
</tr>
<tr>
<td>11</td>
<td>22.3/90</td>
<td>22.7/86</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aFSFI: Female Sexual Function Index.
bSQ: Sexual Quotient.
cN/A: not applicable.

Discussion

Principal Findings

This study aimed to investigate the effects of a web-based ER training program for sexual function. The trial faced a substantial dropout rate, but the preliminary results suggest that the TREPS protocol provided some improvements in relation to sexual function, albeit mainly for mental health and ER abilities. Although no significant effects were seen in the main outcome measures after the 3-month follow-up in the controlled between-group analyses, large and moderate within-group effect sizes were found for a range of components in the within-group analyses.

First, based on complete case analyses and the combined sample, large effects were found for the ER assessment as well as the depression and anxiety measures. Internet-based ER interventions for depression and emotional disorders have been shown to be efficacious with moderate-to-large between-group effect sizes [56-58]. The improvements we observed in this study are in line with those reported for face-to-face interventions focusing on mood disorders [59].

Regarding sexuality, moderate-to-large effects were found regarding the sexual components of arousal/lubrication and orgasm but not in other domains or the overall sexual function score. These sexual domains are more susceptible to individual sex practices and may have benefitted most from the intervention. This is plausible, as frustration with partners was frequently reported in participants’ responses to homework activities, and such components are associated with sexual desire, arousal, and satisfaction [60,61]. Regarding automatic thoughts in sexual activity, there were large intervention effects concerning failure and disengagement thoughts and a moderate effect concerning abuse thoughts. These findings are particularly relevant since both aspects have been suggested as the main cognitive components of sexual dysfunction in women [62,63].

Against our expectations, several factors did not allow us to strictly adhere to the published study protocol [29]. The 2 main changes concerned the follow-up, which was reduced from 6 to 3 months, and the conversion of the initially planned waitlist group into a second intervention group, assessed after receiving the intervention and 3 months later. These changes were implemented to counteract the high dropout rate in the intervention group.

Even though high attrition rates in web-based interventions for sexual function have been described in the past [29,63], the dropout rate in this study superseded those previously reported. When designing the study, precautions regarding adherence were made as suggested in the literature, such as a weekly scheduled treatment program, reminders, direct feedback, and positive messaging upon assignment completion [64,65].

Since no feedback was obtained from nonadherers, any conclusions must remain speculative. Our main hypothesis concerns the COVID-19 pandemic. The intervention was carried out during the COVID-19 outbreak, which affected Brazilians to a larger extent than other countries—by October 2021, Brazil had over 21.6 million infections and over 600 thousand deaths [66]. The number of completed modules was low, which is likely to have impacted the overall response to treatment. Among the participants, only 14% (n=7) accessed the last module, and less than 22% (n=11) accessed more than half of the intervention. Other web-based intervention studies focusing on male sexual disorders (erectile dysfunction) have reported similar findings. For instance, in one study, only 8% of the participants reached the final module, with 54% accessing just up to module 4 (out of 7) [28]. In a similar vein, a 70% dropout rate was reported for another web-based CBT program on sexual dysfunction [19].

Given the limited availability of internet interventions addressing sexual problems, further research is needed to understand the reasons behind the observed high dropout rates in the existing literature. As speculations, potential explanations can be made...
related to the nature of the sexual dysfunction, expectations, or outcomes.

Concerning the nature of the problem, given the personal and often taboo nature of sexuality and intimacy, it is reasonable to consider that worries about privacy or social stigma may significantly influence participant adherence to these interventions. Regarding treatment expectations, the concerned population often does not identify sexual problems as biopsychosocial complex problems. This perspective can lead to unrealistic expectations regarding the duration and effectiveness of interventions.

Regarding outcomes, individuals with sexual problems usually have misconceptions about sexual function and response. It is possible that if sufficient improvement is perceived after the psychoeducational phase (usually initial modules) of internet interventions, they might decide not to continue with the intervention.

Although adherence was low in this trial, the treatment remains easily scalable. The web-based intervention can be easily disseminated or updated at almost no cost and, therefore, is much cheaper than traditional face-to-face specialized treatments. Special attention must be paid to improving adherence.

Another important aspect concerns the unequal participation of men and women. While the TREPS training was developed for both men and women, only women participated in the study. This may be attributed to a higher level of interest among women in research studies [59,67] and related to how we recruited the participants. In addition to social media advertisements and contact with participants from a previous questionnaire study who volunteered to participate in future studies, we credited the female sample to a Brazilian female sexual health social media influencer (mainly followed by women), who shared the study on her social media account.

Limitations
First, the high dropout rate precludes any significant conclusions to be drawn from the results. Significant attrition should be addressed as a risk factor in future similar web-based ER training for sexual function. In addition, and in contrast to the potential advantages of web-based interventions (eg, anonymity, ease of access, etc), the high dropout rates in programs of this kind may indicate a serious limitation to the suitability of web-based training concerning sexual health. Second, most participants were highly educated, which affects the generalizability of the results and their scalability. Third, since the inclusion criteria were based on self-report measures, response bias and possible failures to detect interfering medical conditions not known by the participants may have occurred. Fourth, we did not assess participants’ expectations regarding the effects of the training, so it remains unclear as to whether these impacted the results.

Recommendations
For future studies, we recommend some changes in the study design and treatment components. Regarding recruitment, it would be advisable to have a referral system from a specialized sexual or mental health care service. Perhaps the recommendation of the training from a health service could improve adherence. Regarding the intervention, attention should be paid to the points raised by the intervention adherers, such as increasing the length and the amount of information provided per module. Moreover, there are currently no comparisons between web-based ER training for sexual health and face-to-face interventions. For this reason, comparisons between different intervention delivery methods such as face-to-face individual or group format interventions and web-based programs will be important. Future trials would also benefit from the inclusion of partner-related assessments.

Conclusions
Our preliminary results suggest that web-based ER training may be an effective supplementary or alternative treatment for sexual disorders. Completers of the intervention presented improvements both in sexual function domains as well as in emotion regulation. Nonetheless, due to a high dropout rate, this trial failed to collect sufficient data to allow for any conclusions to be drawn on treatment effects.

Despite its limitations, this study presents valuable information about web-based ER interventions for both mental and sexual health treatments. The improvements in ER skills, mental health, and specific sexual domains call for the further development and evaluation of such interventions. Further research is needed to better understand the feasibility and scalability of similar protocols and how male participants would respond to them.

Data Availability
The study data set generated during this investigation is not publicly available due to ethical concerns. Both the data set and the study metadata are available from the corresponding author upon request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Between-group comparisons at baseline, end of intervention, and 3-month follow-up.
[DOCX File, 34 KB - formative_v8i1e50850_app1.docx ]
Multimedia Appendix 2
CONSORT eHEALTH Checklist (V 1.6.2).
[PDF File (Adobe PDF File), 91 KB - formative_v8i1e50850_app2.pdf]

References


Using Smartphones to Enhance Vision Screening in Rural Areas: Pilot Study

Zheng Wang1, MM; John Kempen2,3,4,5, MD; Gang Luo2, PhD

1School of Medicine, Jiaxing University, Jiaxing, China
2Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, United States
3Department of Ophthalmology, Harvard Medical School, Boston, MA, United States
4Sight for Souls, Bellevue, WA, United States
5Eye Unit, MyungSung Christian Medical Center General Hospital, MyungSung Medical School, Addis Ababa, Ethiopia

Corresponding Author:
Gang Luo, PhD
Schepens Eye Research Institute
Massachusetts Eye and Ear
Harvard Medical School
20 Staniford Street
Boston, MA, 02114
United States
Phone: 1 6179122529
Email: gang_luo@meei.harvard.edu

Abstract

Background: While it is treatable, uncorrected refractive error is the number one cause of visual impairment worldwide. This eye condition alone, or together with ocular misalignment, can also cause amblyopia, which is also treatable if detected early but still occurs in about 4% of the population. Mass vision screening is the first and most critical step to address these issues, but due to limited resources, vision screening in many rural areas remains a major challenge.

Objective: We aimed to pilot-test the feasibility of using smartphone apps to enhance vision screening in areas where access to eye care is limited.

Methods: A vision screening program was piggybacked on a charity summer camp program in a rural county in Sichuan, China. A total of 73 fourth and fifth graders were tested for visual acuity using a standard eye chart and were then tested for refractive error and heterophoria using 2 smartphone apps (a refraction app and a strabismus app, respectively) by nonprofessional personnel.

Results: A total of 5 of 73 (6.8%, 95% CI 2.3%-15.3%) students were found to have visual acuity worse than 20/20 (logarithm of minimal angle of resolution [logMAR] 0) in at least one eye. Among the 5 students, 3 primarily had refractive error according to the refraction app. The other 2 students had manifest strabismus (one with 72–prism diopter [PD] esotropia and one with 33-PD exotropia) according to the strabismus app. Students without manifest strabismus were also measured for phoria using the strabismus app in cover/uncover mode. The median phoria was 0.0-PD (IQR 2.9-PD esophoria to 2.2-PD exophoria).

Conclusions: The results from this vision screening study are consistent with findings from other population-based vision screening studies in which conventional tools were used by ophthalmic professionals. The smartphone apps are promising and have the potential to be used in mass vision screenings for identifying risk factors for amblyopia and for myopia control. The smartphone apps may have significant implications for the future of low-cost vision care, particularly in resource-constrained and geographically remote areas.

doi:10.2196/55270

KEYWORDS
vision screening; refractive error; strabismus; smartphone; visual acuity; vision; visual; eye; eyes; screening; mHealth; mobile health; app; apps; application; applications; feasibility; optometry; ophthalmology
Introduction

Visual impairment impacts individuals’ ability to participate in activities of daily living and social integration, disrupting their quality of life and health. To mitigate the impact of visual impairment, regular eye exams are important. However, regular eye exams are still a luxury for many people for various reasons that lead to limited access to eye care.

For instance, in sub-Saharan Africa, there are only 2 ophthalmologists per million people, while the global mean is 32. It is the only region in the world where the prevalence of moderate and severe vision impairment has increased from 1990 to 2020 [1,2]. Even in developed countries, disparity in eye care is a grim reality. According to US Centers for Disease Control, the percentage of American adults who reported not having an eye examination within the past year ranged from 20% to 30% [3]. The number is even higher among older persons residing in nursing homes, 53% of whom had not seen an eye care provider in the past year [4]. Inconvenience or inability to visit hospitals, cost concerns, and the limited scale of the eye care work force are the major reasons [4].

The World Health Organization and the International Agency for the Prevention of Blindness launched the VISION 2020: The Right to Sight initiative, aiming to eliminate avoidable blindness in the world by 2020 [5]. In 2021, the agency recognized that uncorrected refractive error is still a major cause of blindness and vision impairment in the world, affecting 86 million individuals [6], although it is treatable with glasses.

The rates of annual vision screening among school-aged children in developed countries are relatively high, but there are still gaps. Strabismus is one of the main risk factors for amblyopia, but it is usually not included in screening protocols due to a lack of professional skills in school nurses. When a child is flagged for amblyopia based on a stereopsis test, the vision disorder is often already present and may be difficult to treat. In 2002, the American Academy of Pediatrics strongly urged the development of efficient strabismus screening technology for preschool and school-aged children [7,8].

A number of studies and reports have suggested that community eye care programs should use health workers with entry-level qualifications and platforms to help patients in remote areas more easily access professionals; these are the 2 key approaches to the success of public vision programs [9,10]. However, a gap in these approaches is the lack of detailed vision examinations for informing medical decisions. This may cause excessively high false-positive referral rates [11]. To address this gap, we have invented several key technologies for smartphone-based vision testing, which include computer vision and psychophysical methods for measuring ocular misalignment [12,13], refractive error [14], and retinal degeneration [15]. Key features of these apps are that no specially made attachment is needed in order to use them and they run on standard smartphones. With minimal training, laypersons can use the apps to perform vision tests as long as their smartphones are compatible. The accuracy of these apps has been evaluated previously against standard clinical testing methods in hospital settings [12,14]. This paper reports the findings from a school vision screening program that used the apps we developed for the first time. We aimed to demonstrate the feasibility of an app-based vision screening program involving both strabismus and refractive error tests performed by nonprofessionals.

Methods

Vision Testing Apps

The app for strabismus assessment (Figure 1) has been validated and evaluated previously in a clinical setting by comparing it with standard clinical methods, including the modified Thorington test, the cover test with prism, and synoptophores [12]. When the app was used in this study, the automated cover test mode was used, in which one of the subject’s eyes was covered while the computer vision algorithm kept tracking the status of both eyes, and the app automatically took a snapshot of both eyes with a flash as soon as the covered eye was uncovered. This mode allowed the examiners to perform cover/uncover or alternating-cover tests. In this study, the cover/uncover test was performed. The app calculates ocular alignment using the Hirshberg method and gives results in prism diopters (PDs) [13].

Figure 1. Two vision screening apps were used in this study. On the left is the photo screening app based on the Hirschberg test to measure ocular alignment. Here, the tester has zoomed in to check the visualization of the measurement. On the right is the interface of the refraction measurement app, which captures images of the patient’s face with the selfie camera and shows stimuli (the letter E has been chosen here) on the left of the screen according to a viewing distance estimated from the face images.

The app for refractive error measurement (Figure 1) has been validated previously and evaluated in a clinical setting by comparing it with standard clinical methods, including autorefractors and noncycloplegic subjective refraction [14].

https://iformative.jmir.org/2024/1/e55270 JMIR Form Res 2024 | vol. 8 | e55270 | p.1591 (page number not for citation purposes)
The app estimates refractive error by measuring the far point distance for discerning 20/20 Tumbling E letters [14] or the 20/20-equivalent clock dial chart [16]. When Tumbling E letters are used, the result is spherical equivalent (SE) power in diopters (D), and when the clock dial chart is used, the results give spherical power and cylindrical power based on 2 far points for perpendicular lines. In this study, the Tumbling E letter stimuli were used first to determine the SE, and then clock dial chart was used to estimate cylinder power around the SE range.

**Vision Screening Protocol**

Participants first underwent visual acuity testing at a 5-m distance from a standard Tumbling E chart (Figure 2). There are 5 letters on each line from 20/40 downwards. Participants were given 0.02 log units credit for each extra letter read. Participants who had visual acuity worse than 20/20 in at least one eye were tested further for refractive error with the refraction app (Figure 1). Near phoria was measured with the strabismus app (Figure 1) approximately 40 cm from the phone while the participants were instructed to fixate on the phone flash. The vision measurement with apps was performed by personnel without optometry expertise. All vision tests were performed under a noncycloplegic condition.

**Figure 2.** A standard Tumbling E eye chart was used in vision screening of 73 children in Sichuan, China, in 2023.

**Participants**

A total of 73 fourth and fifth graders in Meigu County, Sichuan, China, who attended a charity summer camp in 2023 were screened for vision as a part of the charity program. All children lived locally and were members of the Yi people, an ethnic minority group in China. The per capita gross domestic product in Meigu County is about US $2270, 18% of the Chinese national average. Being in a low-income mountainous area, most of the children had never undergone vision testing due to very limited access to local eye care services. They needed to be taught how to report the orientation of the letter E on the eye chart.

**Ethical Considerations**

The vision screening was one component of the charity summer camp program. Consent for the vision test was included in the summer camp consent form. Participants received free tutoring classes, meals, and school supplies during the 2-week summer camp. Other than that, they were not compensated specifically for participating in the vision screening. This paper reports the results of a secondary data analysis. The vision measurement data, with individual identifiable information removed, were shared by the charity program. This research received approval with exemption of consent from the Institutional Review Board of Massachusetts Eye and Ear. The study was conducted in accordance with the tenets of the Declaration of Helsinki.

**Results**

We identified 5 students (7%, 95% CI 2.3%-15.3%) as having visual acuity worse than 20/20 in at least one eye (Table 1). Measuring refraction with the app showed that 2 students had astigmatism (patient P1, cylinder –1 D in the left eye, patient P4, cylinder –0.5 D in the right eye). One student was found to be hyperopic in both eyes (patient P3, sphere +3.2 D in both eyes), and also was strabismic (72-PD esotropia). This was probably a case of accommodative esotropia [17]. His refractive error was not measured directly at the site, but estimated based on the commonly used near point distance of human eyes (25 cm). According to the app, his near point was at 1.2 m, which corresponds to –0.8 D, where he could not see 20/80 letters at close range and maintained 20/80 vision beyond the near point. Therefore, the estimate of his refractive error was 4 – 0.8 = 3.2 D. According to Mantyjarvi’s [18] findings in a hyperopic population, the amplitude of accommodation is at least 4 D, based on which the estimate would also be 3.2 D. Another student was also found to be strabismic (patient P4, 35-PD exotropia). The 2 students with manifest strabismus were suspected to have amblyopia. The student with the worst visual acuity among the participants (20/100 in the right eye and 20/167 in the left eye) had moderate myopia (patient P5, sphere –3 D in both eyes), and her near-phoria was 7-PD esotropia. This student was given a pair of off-the-shelf glasses (–3 D for both eyes), and her vision in both eyes improved to 20/33.
Table 1. Among the 73 students screened in a rural area in China, participants who had visual acuity lower than 20/20 in at least 1 eye (5 of 73 children) are listed here. Refraction and phoria were measured using the 2 apps shown in Figure 1. The text describes the estimation method.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Visual acuity 20/xx (logMAR&lt;sup&gt;a&lt;/sup&gt;)</th>
<th>Refraction (diopters)</th>
<th>Phoria/tropia (prism diopters)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Right eye</td>
<td>Left eye</td>
<td>Right eye</td>
</tr>
<tr>
<td>P1</td>
<td>20 (0)</td>
<td>25 (0.1)</td>
<td>Sphere –0.5</td>
</tr>
<tr>
<td>P2</td>
<td>17 (–0.07)</td>
<td>26 (0.12)</td>
<td>Sphere –0.8</td>
</tr>
<tr>
<td>P3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>80 (0.6)</td>
<td>80 (0.6)</td>
<td>Sphere +3.2</td>
</tr>
<tr>
<td>P4</td>
<td>33 (0.2)</td>
<td>25 (0.1)</td>
<td>Sphere 0.5; cylinder –0.5</td>
</tr>
<tr>
<td>P5</td>
<td>100 (0.7)</td>
<td>167 (0.9)</td>
<td>Sphere –3.0</td>
</tr>
</tbody>
</table>

<sup>a</sup>logMAR: logarithm of minimal angle of resolution.
<br><sup>b</sup>Refraction for patient 3 was calculated based on the near point acquired by the refraction app.
<br><sup>c</sup>Manifest strabismus, measured without cover.

Figure 3 shows the distribution of the near-phoria of the students, excluding the 2 students with manifest strabismus, who are listed in Table 1. Phoria was under 10 PD in the majority. The median phoria was 0.0 PD (IQR 2.9-PD esophoria to 2.2-PD exophoria). There was 1 outlier, a student with 26-PD esophoria and normal visual acuity. Excluding this outlier, the mean phoria of the group was 0.2 (SD 4.1)-PD esophoria.

**Discussion**

**Summary**

By using 2 mobile apps designed for vision testing, ocular alignment and refraction measurements were successfully and easily incorporated in school vision screening. The testing results provided much richer information than conventional vision screening, which normally does not include the 2 tests. Based on this particular screening, the vision issues among the rural-area, school-aged children were primarily uncorrected refractive error and amblyopia.

**Interpretations**

While myopia prevalence rates are very high in East Asia, including China, where in some economically developed cities, it is even above 80% among school-aged students [19], the majority of the fourth and fifth grade students in this screening had emmetropic vision (uncorrected visual acuity better than 20/20). Ample outdoor activity time and a light burden of study requiring the use of near vision, as we witnessed and surveyed during the summer camp, are likely the main reasons for the large differences between these students and their urban counterparts. Nevertheless, 4% (3/73) of the children mainly
had refractive error (patients P1, P2 and P5 in Table 1). This rate is very similar to the prevalence (4.1%) found in the rural child population in India [20]. In addition, there were 2 students (2.7%) with manifest strabismus (patients P3 and P4 in Table 1). This rate is within the range of strabismus prevalence (2.06%-3.9%) among children in Taiwan in the years 2014 to 2019 [21]. As for the phoria of the nonstrabismic students in this study (mean phoria 0.2-2.0 PD esophoria, SD 4.1 PD; IQR 2.9-2.4 PD esophoria to 2.2-2.0 PD esophoria), it is similar to the normative range found among 879 elementary school students by Lyon et al [22] using the modified Thorington test (mean 1-1PD exophoria, SD 4 PD; IQR 2.0-2.0 PD esophoria to 2-2 PD exophoria).

The consistency with epidemiology findings from other large population-based vision screening studies suggests that the screening results in this study are probably reliable. A difference from previous screening studies, though, is that this screening study used smartphone-based strabismus and refractive error tests, while most vision screening studies were conducted by professionals using conventional optometry methods. For instance, in the screening study by Dandona et al [20], strabismus was assessed using a penlight and refractive error was measured using a streak retinoscope and handheld autorefractor; both measurements were made by ophthalmic professionals. Our study provides a proof of concept that vision screening tests, including for strabismus and refractive error, can be performed using smartphone apps by persons without optometry expertise. A further study to develop the use of this simple and portable approach to mass vision screening is warranted.

Limitations

One limitation of this study was that the vision test results were not verified by clinical gold standard testing due to the nature of the education charity program. However, it should be noted that the main focus of this study was not to evaluate measurement accuracy, as previous evaluation studies have already addressed that aspect. The difference from our previous evaluation is that this study involved a sample that included people with and without eye conditions, while previous studies deliberately selected patients with a wide range of conditions.

Implications

Thanks to their affordability, ubiquity, and validity, smartphone-based vision screening tools have the potential to play a pivotal role in mass vision screening efforts, particularly in areas where access to eye care is limited. These innovative tools may impact eye health care by providing a cost-effective and convenient solution for screening, identifying, and addressing vision issues, especially in underserved communities.

One example of such technology is the Peek Acuity app (Peek Vision), which has made significant strides in the field of mobile vision screening. This app has been deployed to screen tens of thousands of individuals in some studies in Africa [9,23,24]. Its success in large-scale vision screening efforts underscores the scalability and effectiveness of smartphone-based apps. By offering an efficient way to measure visual acuity, the Peek Acuity app has empowered health care providers to reach more people, making it a useful tool for addressing visual impairment.

Another photoscreening app, GoCheck Kids (Gobiquity Inc), is primarily designed to assess for amblyopia risk factors [11,25]. By analyzing eccentric photorefraction red-reflex images, the app does not require patients to respond, and it therefore can be used to examine preverbal children. The app has been tested in clinical trials by comparing it with its counterparts among dedicated screening devices, such as iScreening [26]. It is reported by the company that developed GoCheck Kids that 4 million children have been screened [27].

Considering the successful deployment of these smartphone apps, we expect that a mobile technology–based vision screening approach, operated by personnel with basic training, will be scalable and will complement service programs provided by ophthalmic clinicians and dedicated screening devices (eg, iScreen, Spot Vision Screener). The mobile app approach is particularly suitable for low-resource countries and could also be used for screening in classrooms. Compared to Peek Acuity and GoCheck Kids, the screening apps assessed in this study can measure refractive error and strabismus quantitatively rather than flagging for risk factors (GoCheck Kids) or not measuring these most common conditions at all (Peek Acuity).

Our future investigations will aim to assess the positive and negative predictive values of these apps in large-scale vision screening programs, building upon the feasibility demonstrated in this preliminary study.

Acknowledgments

The phoria testing app was developed with support in part from the National Institutes of Health (NIH; grant SBIR R44 EYO25902 awarded to EyeNexo LLC, which GL cofounded). The refraction testing app was developed in collaboration with EyeNexo by GL’s lab with support in part from the NIH (grant EYO34345), a Harvard Catalyst award (National Center for Advancing Translational Sciences; NIH award UL1 TR002541), and financial contributions from Harvard University and its affiliated academic health care centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the NIH. JK received funding from Sight for Souls and the Massachusetts Eye and Ear Global Surgery Program. We thank Joye Luo for helping with visual acuity measurement in the vision screening.
Data Availability
Data used in this paper may be made available for research purposes if requested through the corresponding author.

Conflicts of Interest
GL owns a patent on ocular alignment assessment and a patent on refraction assessment. GL is a cofounder of EyeNexo LLC, a startup company developing smartphone apps for vision tests. ZW and JK declare no conflicts.

References


Abbreviations

D: diopter
PD: prism diopter
SE: spherical equivalent

©Zheng Wang, John Kempen, Gang Luo. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
The Role of Mental Health Stigma in University Students’ Satisfaction With Web-Based Stress Management Resources: Intervention Study

Sohyun Cho1, MEd; Laurianne Bastien1, MA; Julia Petrovic1, MA; Bilun Naz Böke1, MEd; Nancy L Heath1, PhD
Department of Educational and Counselling Psychology, McGill University, Montreal, QC, Canada

Corresponding Author:
Sohyun Cho, MEd
Department of Educational and Counselling Psychology
McGill University
3700 McTavish Street
Montreal, QC, H3A 1Y2
Canada
Phone: 1 514 398 4242
Email: sohyun.cho@mail.mcgill.ca

Abstract

Background: University students frequently report elevated levels of stress and mental health difficulties. Thus, the need to build coping capacity on university campuses has been highlighted as critical to mitigating the negative effects of prolonged stress and distress among students. Since the COVID-19 pandemic, web-based stress management resources such as infographics and web-based workshops have been central to supporting university students’ mental health and well-being. However, there is a lack of research on students’ satisfaction with and uptake of these approaches. Furthermore, mental health stigma has been suggested to have not only fueled the emergence of these web-based approaches to stress management but may also influence students’ help-seeking behaviors and their satisfaction with and uptake of these resources.

Objective: This study explored potential differences in students’ satisfaction and strategy use in response to an interactive infographic (an emerging resource delivery modality) presenting stress management strategies and a web-based workshop (a more common modality) presenting identical strategies. This study also examined the relative contribution of students’ strategy use and family-based mental health stigma in predicting their sustained satisfaction with the 2 web-based stress management approaches.

Methods: University students (N=113; mean age 20.93, SD 1.53 years; 100/113, 88.5% women) completed our web-based self-report measure of family-based mental health stigma at baseline and were randomly assigned to either independently review an interactive infographic (n=60) or attend a synchronous web-based workshop (n=53). All participants reported their satisfaction with their assigned modality at postintervention (T1) and follow-up (T2) and their strategy use at T2.

Results: Interestingly, a 2-way mixed ANOVA revealed no significant group × time interaction or main effect of group on satisfaction. However, there was a significant decrease in satisfaction from T1 to T2, despite relatively high levels of satisfaction being reported at both time points. In addition, a 1-way ANOVA revealed no significant difference in strategy use between groups. Results from a hierarchical multiple regression revealed that students’ strategy use positively predicted T2 satisfaction over and above strategy use.

Conclusions: While both approaches were highly satisfactory over time, findings highlight the potential utility of interactive infographics since they are less resource-intensive than web-based workshops and students’ satisfaction with them is not impacted by family-based mental health stigma. Moreover, although numerous intervention studies measure satisfaction at a single time point, this study highlights the need for tracking satisfaction over time following intervention delivery. These findings have implications for student service units in the higher education context, emphasizing the need to consider student perceptions of family-based mental health stigma and preferences regarding delivery format when designing programming aimed at bolstering students’ coping capacity.

(JMIR Form Res 2024;8:e50018) doi:10.2196/50018

https://formative.jmir.org/2024/1/e50018

(XSL FO RenderX)
KEYWORDS
help-seeking behavior; help-seeking; mental health services; mental health stigma; mental health; university students; web-based workshop

Introduction
Overview
The overwhelming reports of stress, distress, and mental health difficulties among university students necessitate urgent action to build coping capacity on university campuses. In a recent survey conducted with 96,489 students across 137 universities by the American College Health Association, 51.2% of university students reported moderate psychological distress, while 24% indicated serious psychological distress [1]. As such, building coping capacity on university campuses is critical, given the negative effects of prolonged stress and distress on academic performance, well-being, and daily functioning [2-4].

Recently proposed theoretical models provide insight into the ways in which university students cope with stress and distress. The Health Theory of Coping [5] posits that university students’ approaches to coping with distress and difficulty exist along a continuum ranging from low- to high-intensity strategies that can be classified as healthy (low risk for adverse health outcomes) or unhealthy (high risk for adverse health outcomes). For example, high-intensity unhealthy coping strategies that pose a risk for unintended negative physical, psychological, and social consequences include substance abuse, self-harm, and suicidality. By contrast, high-intensity coping strategies classified as healthy include social support and professional support. Similarly, low-intensity, unhealthy coping strategies include negative self-talk and rumination, while low-intensity, healthy coping strategies include self-soothing and relaxing or distracting activities. This theory further suggests that individuals will move from lower-intensity coping practices (eg, negative self-talk and self-soothing) to higher-intensity practices (eg, suicidal ideation and professional support) proportional to the degree of distress they experience. Regularly engaging in lower-intensity healthy coping behaviors such as self-soothing, positive self-talk, and breathing exercises can enhance one’s future capacity to cope with distress, difficulty, and uncomfortable emotions [5,6].

Previous research has identified lower-intensity healthy coping strategies (eg, progressive muscle relaxation, diaphragmatic breathing, and meditation) that are effective at promoting resilience by reducing distress and increasing well-being in university students [7,8]. Specific cognitive, behavioral, and mindfulness-based approaches have been shown to effectively reduce levels of anxiety, depression, and the physiological stress response [9-11]. These findings emphasize the value of increased availability of lower-intensity resources that promote coping capacity among university students. Additionally, the use of lower-intensity resources to build coping capacity can also decrease the need for more intensive one-on-one therapeutic services, as university mental health services are struggling to meet elevated demands [12-14]. Therefore, providing students with resources and instruction on these accessible, acceptable, and lower-intensity healthy coping strategies is a timely priority for universities [13,14].

Barriers to Building Coping Capacity in University Students
Barriers to optimal mental health and well-being on university campuses extend beyond a lack of resources [14]. In particular, mental health–related stigma among students functions as a barrier in this context [15,16]. Specifically, mental health stigma has been shown to inhibit help-seeking behaviors among university students [17,18]. According to a systematic review of quantitative and qualitative studies conducted by Clement et al [19], mental health stigma exhibited a negative association with help-seeking. Furthermore, young adults’ perceived stigma from others (eg, family members) regarding seeking mental health treatment undermines their willingness and opportunities to seek help [20]. Similarly, a recent systematic review found that the second most commonly reported barrier by university students to help-seeking behavior was mental health stigma [21]. Specifically, students expressed being worried about their family or friends not being able to understand their situation or that they would perceive them in an unfavorable light.

A body of literature also demonstrates that university students’ cultural, racial, or ethnic identities may have an impact on their help-seeking behaviors. Specifically, university students from certain cultures, races, and ethnic groups may be at greater risk of experiencing mental health difficulties [22] and may be less likely to receive mental health treatment [23]. However, a systematic review and meta-analysis conducted by Wang et al [24] found that culture was not a significant contributor to mental health information-seeking behavior. Thus, the intersectionality of cultural, racial, and ethnic identities in the context of mental health resource use is highly complex, with inconclusive findings, and beyond the scope of this study. Interestingly, as reported in a systemic review by Lui et al [21], stigma around mental health was a key contributor to help-seeking for university students. As such, an individual’s perceptions around the degree of mental health stigma that is present in their familial context provide an opportunity to examine the potential role of mental health stigma in influencing university students’ satisfaction with and use of resources for their mental health and stress management.

Taken together, the above literature demonstrates that there is a need for innovative approaches to the delivery of mental health and well-being resources to university students, which accounts for the impact of mental health stigma on help-seeking in the higher education context. Stigma-related barriers to help-seeking behaviors on university campuses have, in part, fueled the emergence of alternative web-based approaches for sharing stress management and well-being resources with students.

https://formative.jmir.org/2024/1/e50018

JMIR Form Res 2024 | vol. 8 | e50018 | p.1598

(page number not for citation purposes)
Emergence and Utility of Web-Based Resources for Stress Management

During the COVID-19 pandemic, web-based resources became central in providing students with stress management support remotely [25] and have been disseminated using different modalities. Particularly, standard in-person workshops were adapted to a web-based format to be delivered on the web. These web-based workshops have gained popularity in the university context, as they allow students to learn and engage with stress management material remotely while still providing them with the opportunity to ask for clarifications from an experienced facilitator. Similarly, mental health and well-being service delivery through interactive infographics has gained popularity in recent years, given their utility for conveying complex information in a concise, visually appealing, and layperson-friendly manner to general audiences [26,27]. Moreover, interactive infographics include features for users to engage with the content (eg, links to guided audio recordings and strategy practice). However, despite their emerging use, research on the use of infographics for promoting mental health and stress management among university students is scarce. Specifically, although user satisfaction and uptake have been highlighted as crucial elements associated with intervention effectiveness in university students [28], to our knowledge, studies have yet to examine students’ acceptability of emerging modalities of resource provision (eg, web-based workshops and interactive infographics). Thus, there is a need for further evidence on students’ satisfaction, uptake, and use of these web-based resources, as well as which factors contribute to satisfaction with these web-based approaches to university mental health resource provision.

Furthermore, investigations into university students’ satisfaction with interactive infographics and web-based workshops for stress management should consider common barriers to students accessing mental health support [15,16]. Given the level of autonomy that is inherent in using web-based mental health resources [29], ensuring adherence to intervention recommendations (eg, at-home practice of strategies taught) is another challenge that may impact web-based resource satisfaction [30,31]. Furthermore, adherence to resources may vary as a function of the modality of web-based stress management resource provision [31,32]. For instance, some self-guided web-based resources (eg, websites) have been associated with lower levels of adherence [32]. Importantly, research has suggested a potential positive association between adherence, which is measured as participants’ frequency of strategy use, and satisfaction with a web-based mental health intervention [33,34]. Thus, there has been a call to monitor and promote adherence (ie, strategy use) in the dissemination of self-guided web-based resources to optimally support user satisfaction with them [35].

This Study

Drawing on the interdisciplinary literature reviewed above, the overarching aim of this study was to explore students’ relative satisfaction and strategy use when presented with a stress management resource delivered through a web-based workshop (a relatively common mode of resource delivery) versus an interactive infographic (an innovative, emerging mode of resource delivery). This study also aimed to explore potential contributing factors to sustained satisfaction with these 2 modalities of resource delivery, specifically the role of strategy use (ie, adherence) and family-based mental health stigma. The first objective was to examine whether university students’ satisfaction with web-based stress management resources would differ as a function of their delivery format (ie, interactive infographics and web-based workshops) and over time. The second objective was to examine whether students’ strategy use would differ as a function of delivery format. The third objective was to explore the relative contribution of students’ strategy use and family-based mental health stigma in predicting their sustained satisfaction with the interactive infographic versus the web-based workshop. Given the exploratory nature of the study objectives, no specific hypotheses were made. Specifically, for the first objective, based on anecdotal and clinical experience in mental health service delivery, it was anticipated that there might be a differential response such that the participants in the interactive infographic group maintain their satisfaction for a longer period of time than participants in the web-based workshop group given the continued ease of accessing the interactive infographic (ie, self-guided and can be accessed anywhere at any time). However, in the absence of a body of literature examining university students’ satisfaction with the 2 modalities over time, no specific hypothesis was proposed.

Methods

Ethical Considerations

This study was approved by the McGill University Research Ethics Board (#21-10-040). Participants who indicated an interest in participating in this study were invited to complete a web-based survey where the first page of the survey was the consent form explaining that participation in this study was optional and completely voluntary, that responses would be confidential, and that participants could choose not to answer any of the questions should they not want to. Each participant was identified on Qualtrics (Silver Lake) using a unique participant ID number associated with their email address; identifiable data (ie, email addresses) were deleted from Qualtrics once data collection was complete. The master list matching participant information to their unique study ID was protected with a password, saved on a password-protected computer, and was only accessible to the principal investigator and graduate student research assistants working on this study. For the web-based workshop, participants were informed ahead of time that they may choose to keep their cameras off for the duration of the workshop and can log in with only their first name or pseudonym to further preserve confidentiality. Participants were compensated CAD $10 (US $7.43) through electronic transfer for each satisfaction survey completed, up to a maximum of CAD $20 (US $14.85) for completing both satisfaction surveys (T1 and T2).

Participants

A total of 168 undergraduate students were recruited during the Winter 2022 semester. Of those 168 students, 3 graduate
students were excluded from the analyses, given evidence suggesting that undergraduate and graduate students have different stressors and coping strategies [36]. In addition, 1 student was excluded for not completing the demographic questionnaire. A total of 164 students were randomized to either the interactive infographic group (n=81) or the web-based workshop group (n=83). After the randomization, 51 were excluded from the primary analyses due to attrition. Thus, the final sample was composed of 113 students (interactive infographic group n=60; web-based workshop group n=53). Of this final sample of 113 students (mean age 20.93, SD 1.53 years), 88.5% (100/113) self-identified as women, 9.7% (11/113) self-identified as men, and 1.8% (2/113) self-identified as nonbinary. The participants self-identified as White (47/113, 41.6%), Asian (44/113, 38.9%), multiple ethnicities (10/113, 8.9%), Arab or Middle Eastern (5/113, 4.4%), Black or African (4/113, 3.5%), and Hispanic or Latinx (3/113, 2.7%). The participants enrolled in diverse faculties, including arts (66/113, 58.4%), science (20/113, 17.7%), dual majors (11/113, 9.7%), nursing (6/113, 5.3%), law (3/113, 2.7%), and others (7/113, 6.2%).

**Intervention Development and Description**

Both interventions (ie, the interactive infographic and web-based workshop) were researcher-developed for the purposes of this study and focused on four main areas of stress management: (1) pause or break, (2) positive awareness, (3) kindness to self, and (4) social connection. These areas came from a review of the literature on stress management programs for university students [7,9,37]. Pause or break draws on principles of mindfulness, which may be defined as paying attention to what we sense or experience in this moment, on purpose, and with nonjudgmental acceptance [38,39]. Positive awareness promotes our ability to notice the positive things that happen to us [40-42]. Kindness to self draws on research in the area of self-compassion [43,44]. Lastly, building social connections draws on research in the area of social connectedness, such that the aim is to enhance an individual’s sense of belongingness with other people, groups, or their communities, as well as to maintain and strengthen these connections over time [45-47].

Both the interactive infographic and web-based workshop included psychoeducation around stress as well as evidence-based, low-intensity healthy strategies for stress management and to build coping capacity [5,7,8]. Specifically, students were provided with clear descriptions of each of the 4 areas of stress management described above, as well as a variety of research-based stress management strategies pertaining to each area using clear text descriptions, images, videos, guided audio recordings, links to relevant websites, and podcasts. The interactive infographic and the web-based workshop (including a resource sheet that was provided to all workshop attendees) contained identical content; only the delivery format differed. Table S1 in Multimedia Appendix 1 provides a detailed outline of the intervention content.

**Procedure**

**Overview**

Participants were recruited on the web through flyers advertising the study. Potential participants were invited to provide consent and complete a brief demographic survey hosted on the web on Qualtrics. Participants were then randomly assigned to either the interactive infographic group or the web-based workshop group. All participants then received an email providing further information about the study procedure based on their assigned condition and were asked to select their preferred time slot (among 3 choices within the same week) to review the interactive infographic or attend the web-based workshop. Participants were not informed of the nature of the condition to which they were not assigned.

Both the interactive infographic and web-based workshop sessions were scheduled to be delivered at least 1 week after the completion of the demographics survey. After their respective interventions, participants completed a web-based satisfaction questionnaire at 2 time points: immediately after the intervention (postintervention: T1) and 2 weeks later to assess their sustained satisfaction and frequency of strategy use over the 2-week period (follow-up: T2). Between T1 and T2, participants were not provided with any specific instructions or guidelines for practicing strategies. Following completion of the study, participants were debriefed regarding the study purpose and design and received all materials from both interventions.

**Interactive Infographic Group**

Participants in the interactive infographic group received access to the interactive infographic (in PDF format) through Qualtrics. Participants were instructed to review the interactive infographic content and practice the embedded stress management strategies for a total of 30 minutes. To guarantee that participants engaged with the interactive infographic for the full duration, a timer embedded in the Qualtrics page only allowed participants to proceed to the next page of the survey once the 30 minutes had passed. To support participants’ continued use of the strategies presented on the interactive infographic between T1 and T2, the interactive infographic was subsequently shared with participants through email.

**Web-Based Workshop Group**

Participants in the web-based workshop group attended a 30-minute workshop through Zoom (Zoom Video Communications). All workshops were delivered by the same research assistant, who followed an oral script. Similar to the interactive infographic group, to support participants’ continued use of the strategies presented during the web-based workshop between T1 and T2, a 1-page resource sheet summarizing the information and strategies presented during the workshop was shared with participants through email.

**Measures**

**Satisfaction**

Participants’ satisfaction with the stress management resources was assessed using an 8-item researcher-developed questionnaire...
based on the new world Kirkpatrick model for program evaluation [48]. Specifically, the questions assessed participants’ (1) reaction (satisfaction, engagement, and relevance) and (2) learning (knowledge, skills, attitude, confidence, and commitment), which correspond respectively to levels 1 and 2 of the new world Kirkpatrick model. For instance, sample items related to (1) reaction included: “I found the infographic/workshop useful for me and I found that the infographic/workshop was presented in an engaging manner,” while sample items related to (2) learning included:

The strategies presented in the infographic/workshop helped me better understand how to manage my stress and improve my wellness and I feel confident in my understanding of the suggested strategies in the infographic/workshop.

Participants responded on a 4-point Likert scale (1=“strongly disagree” to 4=“strongly agree”). The possible sum satisfaction scores ranged from 8 to 32, where a higher sum score denoted greater satisfaction. This measure demonstrated good internal consistency at T1 (Cronbach α=0.82) and at T2 (Cronbach α=0.90).

**Strategy Use**

Participants’ frequency of strategy use was assessed using a researcher-developed single-item measure at T2. Although participants were introduced to numerous strategies within their stress management resource (Table S1 in Multimedia Appendix 1 provides an outline of the strategies taught), this particular item asks participants about their overall use of any combination of the strategies taught. Specifically, this item corresponds to level 3 (behavior) of the new world Kirkpatrick model for program evaluation [48]. Participants were asked to respond to the following item: “Over the past two weeks, how often did you use the strategies presented in the infographic/workshop?” Participants responded using a 4-point Likert scale (1=“never” to 4=“every day”), where a higher score indicated more frequent use of strategies.

**Family-Based Mental Health Stigma**

Perceived mental health stigma from family members (ie, family-based mental health stigma) was assessed using a researcher-developed single-item measure at baseline. Participants were asked to think about their experiences with mental health-related stigma within their immediate (eg, parents and siblings) or extended family (eg, grandparents, aunts, and uncles) and were asked to respond to the following item: “In my family, I feel there is stigma associated with talking about having mental health difficulties.” Participants responded on a 5-point Likert scale (1=“strongly disagree” to 5=“strongly agree”), where a higher score indicated greater family-based mental health stigma.

**Data Analysis**

For the first objective, a 2-way mixed ANOVA was conducted to examine the effect of group (interactive infographic and web-based workshop) and time (T1 and T2) on students’ satisfaction with web-based stress management resources. For the second objective, a 1-way ANOVA was conducted to examine potential group differences in strategy use. Lastly, for the third objective, separate hierarchical multiple regressions were run for the interactive infographic group and the web-based workshop group to examine whether family-based mental health stigma predicted students’ sustained satisfaction with web-based stress management resources at T2, even when controlling for their strategy use.

Results

**Main Analyses**

**Objective 1**

The first objective sought to compare group differences in university students’ satisfaction with an interactive infographic versus a web-based workshop for stress management over time (from T1 to T2). A 2-way mixed ANOVA revealed that there was no significant interaction between group (interactive infographic and web-based workshop) and time (T1 and T2) on students’ satisfaction (Table 1). In addition, there was no significant main effect of group, indicating that, regardless of time (T1 and T2), there were no group differences in students’ satisfaction between the interactive infographic and the web-based workshop; overall, students in each group were highly satisfied at both T1 and T2. However, there was a significant main effect of time, where satisfaction decreased for all students from T1 to T2, regardless of their assigned group.
Table 1. Results of a 2-way mixed ANOVA (group x time) comparing university students’ satisfaction with 2 stress management resource delivery modalities (N=113).

<table>
<thead>
<tr>
<th>Time points</th>
<th>Intervention groupa (n=60), mean (SD)</th>
<th>Web-based workshop (n=53), mean (SD)</th>
<th>Students’ satisfactiona</th>
<th>F test (df)</th>
<th>ηp²</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1 Interaction</td>
<td>27.67 (2.85)</td>
<td>27.89 (3.23)</td>
<td>0.16 (1,111)</td>
<td>.00</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>T2 Main effect of group (between)</td>
<td>25.97 (2.93)</td>
<td>26.40 (3.42)</td>
<td>0.39 (1,111)</td>
<td>.00</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>Main effect of time (within)</td>
<td>N/A b</td>
<td>N/A</td>
<td>36.81 (1,111)</td>
<td>.25</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

aThe possible range of the postintervention and follow-up satisfaction score was from 8.00 to 32.00.
bN/A: not applicable.

Objective 2
The second objective sought to compare group differences in students’ strategy use with an interactive infographic versus a web-based workshop for stress management over the 2-week study period (assessed retrospectively at T2). A 1-way ANOVA revealed that there was no significant difference in strategy use between the interactive infographic group (mean 2.02, SD 0.57) and the web-based workshop group (mean 2.13, SD 0.62; F4,111=1.065; P=.30).

Objective 3
The third objective sought to examine the potential contribution of students’ strategy use and family-based mental health stigma in predicting their sustained satisfaction with each web-based stress management approach at T2. We ran 2 hierarchical multiple regression analyses (1 for each group), where strategy use frequency was entered in step 1 and family-based mental health stigma was entered in step 2. Table 2 presents the results of the hierarchical multiple regression analyses.

The hierarchical regression results revealed that, in the infographic group, students’ strategy use frequency explained 25% of the variance in sustained satisfaction at T2 (F1,58=19.27; P<.001; R²=0.25). Specifically, strategy use frequency (β=.50; P<.001) emerged as a significant positive predictor of students’ sustained satisfaction at T2. When controlling for strategy use frequency, family-based mental health stigma (β=.03; P=.83) did not emerge as a significant predictor of sustained satisfaction at T2 (F1,57=0.05; P=.83, ΔR²=0.00).

In the web-based workshop group, students’ strategy use frequency explained 11% of the variance in sustained satisfaction at T2 (F1,50=6.14; P=.02; R²=0.11). Specifically, strategy use frequency (β=.33; P=.02) emerged as a significant positive predictor of students’ sustained satisfaction at T2. When controlling for strategy use frequency, family-based mental health stigma contributed an additional 10% explained variance in sustained satisfaction at T2 (F1,50=6.42; P=.01; ΔR²=0.10). Thus, family-based mental health stigma (β=-.32; P=.01) emerged as a significant negative predictor of students’ sustained satisfaction at T2. The full model of strategy use frequency and family-based stigma surrounding mental health difficulties significantly predicted sustained satisfaction at T2 in the web-based workshop group (F2,50=6.60; P=.003; R²=0.21), for a total of 21% explained variance.

Table 2. Summary of hierarchical multiple regression for sustained satisfaction at T2 by group.

<table>
<thead>
<tr>
<th>Sustained satisfaction at T2</th>
<th>Interactive infographic (n=60)</th>
<th>Web-based workshop (n=53)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B SE B β P value</td>
<td>B SE B β P value</td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>20.76 1.23 N/A a</td>
<td>22.55 1.62 N/A</td>
</tr>
<tr>
<td>Strategy use frequency</td>
<td>2.58 0.59 0.50 &lt;.001</td>
<td>1.81 0.73 0.33 .02</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>20.56 1.54 N/A</td>
<td>25.04 1.82 N/A</td>
</tr>
<tr>
<td>Strategy use frequency</td>
<td>2.59 0.59 0.50 &lt;.001</td>
<td>1.74 0.69 0.32 .02</td>
</tr>
<tr>
<td>Family-based mental health stigma</td>
<td>0.05 0.25 0.03 .83</td>
<td>–0.75 0.30 –0.32 .01</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Discussion

Overview

Previous literature has noted the promising benefits of web-based stress management resources for university students, including their cost-effectiveness and flexibility [49-52]. However, it remains unclear how readily students accept interactive infographics, a relatively new and innovative resource modality, compared to the widely adopted standard web-based workshops. This study was thus guided by 3 overarching objectives. The first objective was to compare university students’ satisfaction with an interactive infographic versus a web-based workshop over time (from T1 to T2). The second objective was to compare strategy use between students who engaged with an interactive infographic versus a web-based workshop. Building on the first and second objectives, the third objective was to examine whether students’ sustained satisfaction with each of the web-based stress management approaches at T2 could be predicted by their strategy use and family-based mental health stigma.

The findings from this study demonstrated that across time points, university students who received web-based stress management support through the interactive infographic and the web-based workshop did not significantly differ from one another in their satisfaction with the resources. Furthermore, the findings revealed that university students’ strategy use did not significantly differ between the interactive infographic group and the web-based workshop group. Taken together, these findings suggest that when identical psychoeducation content and evidence-based stress management strategies are delivered to university students, their satisfaction and strategy use over time do not differ as a function of whether they received the stress management instruction through an interactive infographic or a web-based workshop. Rather, in this study, students were highly satisfied with both the interactive infographic and the web-based workshop. Since research on the use of interactive infographics for promoting mental health and stress management among university students is scarce, these novel findings suggest that the delivery of stress management support through interactive infographics may be well-received by university students. Thus, the use of interactive infographics within the context of university-wide stress management and mental health programming warrants further investigation, particularly since they are a novel approach and are less resource-intensive than other resources (eg, in-person and web-based workshops). Additionally, as highlighted in the mental health information seeking literature, students are consistently turning to web-based resources for support; thus, there is a need for universities to provide students with new online approaches to resource delivery to enhance their well-being [21]. Offering evidence-based self-directed resources could benefit university students who prefer self-reliance to address mental health difficulties and can mitigate reported challenges such as time constraints and stigma [21,53].

Although many intervention studies measure satisfaction at a single time point [54], the importance of measuring satisfaction over time following intervention delivery has been highlighted, as there is a possibility that acceptability may change during the weeks following exposure to the intervention [55,56]. Indeed, in this study, students’ satisfaction with the interactive infographic and the web-based workshop significantly decreased from T1 to T2. However, it should be noted that, given the small magnitude of these decreases across both groups (ie, less than a 2-point decrease across groups on a scale ranging from 8 to 32), these decreases may not be clinically meaningful. Furthermore, students’ satisfaction remained relatively high across the 2 time points, with mean scores ranging from 25.97 to 27.89 out of 32 for both groups (Table 1). Nevertheless, we propose 3 possible interpretations for the decrease in satisfaction from T1 to T2, even though it remained relatively high over time. First, it is possible that the observed decrease in satisfaction over time was related to students’ initial enthusiasm regarding the strategies presented in the program, which may have been followed by barriers encountered over the two weeks that followed, such as a lack of time to engage in strategy practice [57,58]. This may have, in turn, slightly negatively impacted their satisfaction with the resources. Another potential explanation for this finding is that participants may be less likely to recall resource content with the passage of time, resulting in a less favorable satisfaction rating when followed up with at a later date. Finally, as with any repeated measures design, other unmeasured confounding variables present during the intervention period may explain this time effect as well. Thus, additional studies are needed to examine factors that may influence satisfaction with stress management resources over time to support the sustainability of resource use in the long term.

Furthermore, this study considered that students’ self-directed practice of the strategies presented, along with their family-based mental health stigma, may have an influence on their sustained satisfaction with the interactive infographic and the web-based workshop. Indeed, the frequency of strategy use was a significant predictor of university students’ sustained satisfaction with both modalities at T2. The more frequently participants used strategies over the 2 weeks, the higher they rated their sustained satisfaction with the web-based stress management resource at T2. This finding is not surprising since previous literature on intervention adherence has shown that greater adherence to self-directed web-based interventions is positively related to more favorable outcomes [59]. However, further research into the temporal nature of the relationship between strategy use and satisfaction is needed, as it remains unclear whether increased strategy use leads to greater satisfaction or vice versa. Nevertheless, this finding highlights the degree to which strategy use and sustained satisfaction are intertwined. In the context of stress management intervention design and delivery in university settings, this suggests that importance should be placed on (1) providing students with a variety of strategies such that they can find ones that they like and that work for them, as well as (2) building a community or environment within the university where strategy use is supported and encouraged.

Interestingly, after controlling for the frequency of strategy use, students’ family-based mental health stigma significantly predicted the web-based workshop group’s sustained satisfaction.
at T2, over and above students’ strategy use. Specifically, when students had greater family-based mental health stigma, their sustained satisfaction with the web-based workshop at T2 was lower. Although it was delivered remotely, the web-based workshop was conducted over Zoom in the presence of a live facilitator and other student attendees, unlike in the context of reviewing an interactive infographic. According to a systematic review conducted by Pretorius et al [29], young people (aged 25 years or younger) tend to be concerned about their privacy and confidentiality even when they are engaging in web-based help-seeking. While participating in the web-based workshop, students with higher levels of family-based stigma may have felt negatively due to the internalization of their perceived stigma from others, as suggested by literature [60,61], which may have negatively influenced their satisfaction. Thus, reports of family-based mental health stigma may have had a greater impact on participants in the web-based workshop group relative to participants in the interactive infographic group because of the relatively lower levels of privacy and confidentiality inherent in workshop participation.

In contrast, students’ family-based mental health stigma did not significantly predict the interactive infographic group’s satisfaction at T2 over and above their strategy use. Due to the relatively private nature of engaging with interactive infographics, students who received the stress management resource through this delivery format did not directly interact with the resource providers or other students while engaging with the infographic. Thus, it is plausible that their satisfaction was less influenced by perceived mental health stigma while acquiring stress management knowledge and practicing the strategies. These findings highlight that, even in the context of web-based stress management initiatives, mental health-related stigma may still be an important factor to consider for university students’ satisfaction with specific resource delivery formats.

Limitations and Future Directions
While these findings shed new light on university students’ satisfaction with online stress management approaches and the associations between their satisfaction and family-based mental health stigma, they should be interpreted within the context of several limitations. First, this study did not directly address the impact of university students’ culture, race, and ethnicity on their help-seeking behavior and instead focused on students’ perceived family-based mental health stigma. Future studies may wish to explore potential cultural, racial, and ethnic differences in university students’ uptake of and satisfaction with different modalities for mental health resource provision. Second, women accounted for a large majority of the sample (100/113, 88.5%), which also limits the generalizability of findings. Future studies may benefit from examining the acceptability of web-based resources for stress management with samples that are more diverse in terms of gender identity. Third, this study used a researcher-developed single-item approach for measuring strategy use and family-based mental health stigma. A limitation of this approach is that it is impossible to calculate internal consistency estimates of reliability [62]. However, it has been suggested that there is no difference in the predictive validity of single-item measures and multi-item measures [63]. Furthermore, an advantage of using single-item measures is the increased brevity and thus feasibility of capturing a psychological construct (eg, individuals’ beliefs) in a simple screening [62]. The fact that both variables were found to be significantly related to sustained satisfaction, despite being limited to single-item measures, demonstrates that these single-item measures have the potential for use, and it is a first step for tapping into individuals’ perceptions of family-based mental health stigma. Lastly, this study examined students’ satisfaction with the 2 web-based stress management approaches but did not compare their effectiveness. Although this was an important first step, future studies are needed to evaluate the effectiveness of different formats of web-based resource delivery, such as interactive infographics and web-based workshops, on students’ mental health-related outcomes (eg, stress, mindfulness, and well-being). Additionally, given findings on the impact of family-based mental health stigma on satisfaction among students in the web-based workshop group, future studies may want to examine specific strategies to address or mitigate the impact of family-based mental health stigma in the context of web-based workshop delivery.

Contributions
Despite these limitations, the findings from this study contribute to our understanding of university students’ satisfaction with web-based stress management approaches, in addition to their association with their family-based mental health stigma. Given the impact of stigma on mental health service and resource use [15,64], as well as the paucity of research examining students’ receptivity to interactive infographics for stress management and well-being resource delivery, this study is a first step in demonstrating the relationship between university students perceived mental health stigma and satisfaction with both interactive infographics and web-based workshops. These findings have implications for student service units in the higher education context. Specifically, they highlight the importance of carefully considering student perceptions of mental health stigma and preferences regarding format of delivery when designing programming to support building students’ coping capacity [34,65].

Conclusion
Since the COVID-19 pandemic, web-based stress management resources are increasingly being delivered to university students to support their elevated mental health and well-being needs. This study provides preliminary evidence that students’ satisfaction with web-based stress management resources may not differ as a function of delivery modality (ie, whether the instruction is delivered through an interactive infographic or a web-based workshop). This study thus has implications for future approaches to mental health service delivery in universities, as the results demonstrate students’ high satisfaction with 2 different web-based delivery formats. Moreover, findings emphasize the negative impact of students’ perceived mental health stigma on their satisfaction with web-based workshops. Overall, while both an interactive infographic and a web-based workshop elicited high levels of satisfaction across time points in this study, results highlight the utility of interactive infographics since they are less resource-intensive than workshops, easy to access and distribute, and given that
students’ satisfaction with them is not impacted by family-based mental health stigma.

Data Availability
The data generated or analyzed during this study are not publicly available since the authors documented on the institutional ethics board that the data will not be shared with any other researchers.

Authors’ Contributions
SC made substantial contributions to the study’s methodological design, data collection, data analysis, interpretation of the results, manuscript writing, and editing of the final manuscript. LB, JP, and NLH made substantial contributions to the study conceptualization, design, data collection, data analysis, interpretation of the results, and critical revisions of manuscript drafts and approved the final manuscript. BNB made substantial contributions to the study conceptualization, design, and critical revisions of manuscript drafts and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Table S1. Detailed outline of the intervention content.

References


27. Cho et al JMIR FORMATIVE RESEARCH


©Sohyun Cho, Laurianne Bastien, Julia Petrovic, Bilun Naz Böke, Nancy L Heath. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Development and Evaluation of a Clinician-Vetted Dementia Caregiver Resources Website: Mixed Methods Approach

Jaye E McLaren1, OTD, MA, OTR/L; Dat Hoang-Gia2, RN, CNS; Eugenia Dorisca3, BS; Stephanie Hartz4, LCSW, BCD; Stuti Dang5,6,7, MPH, MD; Lauren Moo1,8, MD

1New England Geriatric Research Education and Clinical Center, Veterans Affairs Bedford Health Care System, Bedford, MA, United States
2Palo Alto Geriatric Research Education and Clinical Center, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, United States
3Bronx Geriatric Research Education and Clinical Center, Veterans Affairs Bronx Health Care System, Bronx, NY, United States
4Eastern Colorado Geriatric Research Education and Clinical Center, Eastern Colorado Veterans Affairs Health Care System, Aurora, CO, United States
5Miami Geriatric Research Education and Clinical Center, Miami Veterans Affairs Health Care System, Miami, FL, United States
6Division of Geriatrics and Palliative Care, University of Miami Miller School of Medicine, Miami, FL, United States
7Elizabeth Dole Center of Excellence for Veteran and Caregiver Research, US Department of Veterans Affairs, Washington, DC, United States
8Cognitive Behavioral Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Jaye E McLaren, OTD, MA, OTR/L
New England Geriatric Research Education and Clinical Center, Veterans Affairs Bedford Health Care System
200 Springs Rd
Bedford, MA, 01730
United States
Phone: 1 781 687 2000
Email: Jaye.McLaren@va.gov

Abstract

Background: About 11 million Americans are caregivers for the 6.7 million Americans currently living with dementia. They provide over 18 billion hours of unpaid care per year, yet most have no formal dementia education or support. It is extremely difficult for clinicians to keep up with the demand for caregiver education, especially as dementia is neurodegenerative in nature, requiring different information at different stages of the disease process. In this digital age, caregivers often seek dementia information on the internet, but clinicians lack a single, reliable compendium of expert-approved digital resources to provide to dementia caregivers.

Objective: Our aim was to create a dementia caregiver resources website to serve as a hub for user-friendly, high-quality, and expert-reviewed dementia educational resources that clinicians can easily supply to family caregivers of people with dementia.

Methods: An interdisciplinary website development team (representing dementia experts from occupational therapy, nursing, social work, geriatrics, and neurology) went through 6 iterative steps of website development to ensure resource selection quality and eligibility rigor. Steps included (1) resource collection, (2) creation of eligibility criteria, (3) resource organization by topic, (4) additional content identification, (5) finalize resource selection, and (6) website testing and launch. Website visits were tracked, and a 20-item survey about website usability and utility was sent to Veterans Affairs tele-geriatrics interdisciplinary specialty care groups.

Results: Following website development, the dementia caregiver resource website was launched in February 2022. Over the first 9 months, the site averaged 1100 visits per month. The 3 subcategories with the highest number of visits were “general dementia information,” “activities of daily living,” and “self-care and support.” Most (44/45, 98%) respondents agreed or strongly agreed that the website was easy to navigate, and all respondents agreed or strongly agreed that the resources were useful.

Conclusions: The iterative process of creating the dementia caregiver resources website included continuous identification, categorization, and prioritization of resources, followed by clinician feedback on website usability, accessibility, and suggestions for improvement. The website received thousands of visits and positive clinician reviews in its first 9 months. Results demonstrate that an expert-vetted, nationally, and remotely available resource website allows for easy access to dementia education for clinicians to provide for their patients and caregivers. This process of website development can serve as a model for other clinical subspecialty groups seeking to create a comprehensive educational resource for populations who lack easy access to specialty care.
KEYWORDS
Alzheimer disease; caregiver education; dementia; interdisciplinary; older adults; virtual resources; website development

Introduction

The proportion of the world population that is aged over 60 years old is projected to almost double between 2015 and 2050 [1], and with it, the prevalence of dementia and the number of families caring for people with dementia will also increase. About 11 million Americans (70% female) currently provide 18 billion hours of unpaid dementia care per year [2], many without formal dementia education or support. These informal dementia caregivers, such as friends and family members, generally do not have the knowledge or expertise to handle the myriad challenges dementia caregiving can present. This contributes to the high rates of caregiver stress and burnout, often leading to institutionalization for the person with dementia [3-5]. While clinicians (broadly defined to include all health care professionals who care for patients [6]) recognize that supporting the family caregivers of their patients with dementia is important, most do not have the time or ready access to the best educational resources for management of the varied aspects of dementia care (eg, activities of daily living, instrumental activities of daily living, safety, and navigating neuropsychiatric symptoms) at all stages of the disease. Therefore, clinicians require a compendium of high-quality, expert-reviewed resources that are easily available to them to supplement their care and to address common caregiver questions throughout the disease progression.

While some health care systems have dedicated dementia specialists or robust community dementia support, the majority do not [7]. Moreover, people with dementia and their caregivers living in rural areas or experiencing low socioeconomic status face additional barriers to accessing specialty dementia care, education, and support [8]. This often leaves caregivers to find dementia education on their own or rely on recommendations from their primary care team. Many caregivers turn to the internet to find information about caring for someone with dementia and for caregiver support. However, web-based resources are not necessarily scrutinized to ensure information accuracy or to consider the health literacy skills, technology skills, and cognitive and sensory skills of the target audience [9,10]. Similarly, because information is placed in varied locations with varying degrees of quality, cost, and access, this search can be confusing for many caregivers, leading to an increase in misinformation and further caregiver strain [11-14]. Currently, there are hundreds of websites offering advice regarding dementia; however, this plethora of choices can often be confusing or inconvenient for busy clinicians or family caregivers to navigate. It is important for clinicians to be able to direct caregivers of people with dementia to high-quality, easy-to-navigate educational content housed in a single central location that has an expert clinician–curated selection of diverse resources.

Our caregiver resources website aimed to provide clinician-vetted, easy-to-access, and dementia-specific education resources that prioritized the caregiver audience and provided a comprehensive overview of the neurodegenerative disease process, with different resources for different stages of dementia. The website development team (WDT) was comprised of 6 Veterans Affairs (VA) telehealth-based clinical dementia experts based at various Geriatric Research Education and Clinical Centers (GRECCs) centers of excellence focused on serving aging veterans. Members of the WDT had combined decades of dementia experience and represented multidisciplinary backgrounds (neurology, geriatrics, nursing, social work, and occupational therapy). The WDT used an iterative strategy to gather input from interdisciplinary geriatrics clinicians with decades of experience caring for individuals with dementia and their caregivers and create a compendium of clinician-approved, web-based dementia care resources.

Methods

Website Creation Overview

The process of creating this website resource, outlined in Figure 1, included 6 iterative steps over the 1.5 years of website development to ensure resource information quality and appropriateness for informal dementia caregivers. Steps included (1) resource collection, (2) creation of eligibility criteria, (3) resource organization by topic, (4) additional content identification, (5) finalize resource selection, and (6) website testing and launch. Figure 1 shows the development of the dementia caregiver resources website and the summary of the sequential steps taken to develop the educational resource website.
**Step 1: Resource Collection**

The WDT leveraged their existing geriatrics contacts among 18 VA tele-geriatrics programs across the country and asked them to share the dementia caregiver educational resources they found most useful in their clinical practice. The only requirements for these resources were that they needed to be directed to a patient or caregiver audience and freely available on the internet remotely. This request yielded a total of 25 different web-based resources. The WDT reviewed the resources, which varied widely in content, addressing not only physical and cognitive changes related to dementia but also psychosocial, emotional, and familial or community impacts throughout the disease process. Additionally, the resources varied in overall quality, availability (some resources were directly available on the web, others required payment or a subscription, some were applicable only in a particular state or region, etc), visual clarity (challenging to read, difficult to print out, etc), length, format, and consideration for audience. The WDT used these 25 resources as their starting point for Step 2.

**Step 2: Creation of Resource Eligibility Criteria**

Based on a thorough review of resources collected in Step 1, the WDT developed specific eligibility criteria for selecting additional resources for the website. Eligibility criteria for potential resources included the following:

1. Remotely and nationally available (not programming specific to one location or region)
2. Easily accessible in printable or sharable format (PDF, Word [Microsoft Corporation] document, video, etc)
3. Appropriate health literacy level for the family caregiver audience
4. Visual and cognitive ease (eg, large and easy-to-read font, good use of white space and images, and a lack of excessive color or low contrast)
5. Open access resources (eg, not commercially sponsored, behind a paywall, or requiring email registration)

International resources were excluded if they provided country-specific data, policy, or programming. Stand-alone resources (eg, PDFs, Word documents, and videos) were given preference over general dementia or caregiving websites, as some websites can be difficult to navigate, share, and print. Preference was given to resources that were not likely to undergo excessive changes or that seemed to require infrequent updating. While this compendium was created by VA clinicians within the United States, the WTD identified resources that were applicable to dementia caregivers, regardless of veteran status.

**Step 3: Resource Organization by Topic**

In parallel to the gathering of resources from the large national VA tele-geriatrics network and the creation of the eligibility criteria, WDT members met bimonthly and also regularly shared candidate resources with other team members through email. Similar resources were grouped into categories and subcategories, identifying important gaps in content. Categories and subcategories were formed with the goal of providing as much education on common dementia-specific challenges as possible while also trying to provide at least 3 resources in varied formats for each section. Using a category and subcategory organization allows clinicians to easily provide the right resource at the right time for a given caregiver’s dementia care needs, as well as easily display the plethora of resource topics that could impact caregivers throughout the disease process, thus offering a potential “roadmap” for caregivers. So, depending on the caregiver’s needs and interests, the clinician can tailor the website experience. Again, this process of categorization was iterative, and as the identification of resources continued, categories and subcategories were edited, added, or deleted to best organize the content. The 10 finalized categories are included in Textbox 1.
Each category had associated subcategories. For example, the category of “behavioral changes” was further divided into subcategories including wandering, sundowning, anxiety, aggression, and repetitive questions. There were 34 finalized subcategories in total. While most categories had a similar organizational scheme, the categories of “comprehensive dementia care guides” and “additional websites” differed. The “comprehensive dementia care guides” category housed resources that spanned multiple topic areas and dementia stages and required more time to navigate due to the amount of information provided. These guides were given their own organizational scheme and were available for clinicians to give to caregivers or family members interested in a more in-depth resource. The “additional website” category housed general dementia websites that were deemed useful by the WDT but were made up of multiple pages requiring further link selection and thus did not meet the original eligibility criteria for simplicity.

**Step 4: Additional Content Identification**

Upon review of the initial 25 resources suggested and the WDT discussions around important categories and subcategories, gaps in various topic areas were identified (eg, brain health, dementia and nutrition, and emergency preparedness). The WDT members then embarked on a broad internet search to find additional resources that fit the eligibility criteria and filled notable gaps in the established categories and subcategories. This included an expedited review of topic areas and dementia education products from esteemed organizations, such as the Alzheimer’s Association, the National Institute on Aging, etc, which helped to organize the thorough list of categories and subcategories. This included an expedited review of topic areas and dementia education products from esteemed organizations, such as the Alzheimer’s Association, the National Institute on Aging, etc, which helped to organize the thorough list of categories and subcategories. The WDT continued to meet regularly to identify, vet, and organize candidate resources using the eligibility criteria and, if appropriate, add them to the compendium. At the end of the content identification process, there were 170 resources that were candidates for inclusion.

**Step 5: Finalize Resource Selection**

All 170 candidate resources were organized into categories and subcategories, with each subcategory housing between 3 and 10 vetted resources. The WDT then individually reviewed all the resources in each subcategory, and each team member ranked their first, second, and third choices based on the eligibility criteria outlined above as well as the overall quality of the resource, suitability to the corresponding category and subcategory, and perceived clinical utility for clinicians and caregivers. WDT members also identified specifically why they liked or did not like each of the resources and if they had any “honorable mentions” among those that were not ranked as their top 3 choices in each subcategory. After all resources were ranked, the team collated the rankings and chose the resources with the most first, second, and third rankings. When there were more than 3 quality options available for a given subcategory, the WDT discussed them as a group and used their clinical judgment to make the final selections. This detailed analysis of resources helped the team decide whether to add or exclude additional resources, which resulted in some subcategories having more than 3 resources. This process resulted in 3-6 quality resources per subtopic area with a variety of media formats (PDF, video, etc). There were 132 resources selected for final inclusion on the website.

Once the list of resources was finalized, the WDT reached out to the authors or organization for each resource to gain permission to display the resource link on the compendium website. The caregiver resources website only provided links to the resource, ensuring all credit for the resource went to the authors or organization. This also ensured that, if there were updates to the resource in the future, the WDT would only be responsible for providing the updated link. All resources received author permission.

**Step 6: Website Testing, Launch, and Evaluation**

The WDT, in collaboration with the VA tele-geriatrics website developer, created the structure, organization, and overall look of the website. The initial target end users of this website are clinicians caring for people with dementia, although some clinicians may share the whole website (as opposed to specific resources) with caregivers. Therefore, additional care was taken when choosing resources to consider caregivers of people with dementia (who are often older adults with age-related sensory challenges). This included being mindful of category and subcategory organization, language used, font size, white space,
photo content, thumbnail images to ensure inclusivity and demonstrate diversity, etc.

A link to the beta version of the website was sent nationally through email to 60 geographically diverse VA tele-geriatrics specialists who attended a meeting in which the website goals were described. The email also solicited feedback on content, visual appeal, usability, and organization. All feedback was discussed by the WDT and led to further modifications to the website’s appearance and organization. The finalized website launched in January 2022 through an email announcement to the larger VA tele-geriatrics interdisciplinary specialty care group (119 members) who were asked to share the website with their colleagues inside and outside the VA network. Additionally, to increase visibility, the WDT also described the website creation process and demonstrated the use of the website in a case-based national webinar to an audience of 117 clinicians.

A 20-item survey containing a combination of quantitative and qualitative questions was established to gather interdisciplinary clinician feedback on the website’s usability and accessibility and to solicit suggestions for improvement. Survey questions and design were informed by a review of literature on evaluation of health care resources and education websites and content [15-20]. The survey was implemented on the Survey Monkey (Momentive Global Inc) platform and disseminated in the same email as the website launch announcement as well as during the national webinar. A request to complete the survey regarding the resource website was sent through email to all members of the VA tele-geriatrics interdisciplinary specialty care group as well as the 117 attendees of the case-based national webinar. Email recipients were also invited to forward the email with the website link and survey to any colleagues they deemed appropriate. Because of this, it is unknown how many people received the survey link. The survey was open from September 14, 2022, to November 1, 2022.

**Ethical Considerations**

This work was conducted as part of the VA Office of Rural Health-funded “GRECC-Connect” program education and evaluation core activities and was determined to be quality improvement by the VA Bedford Healthcare System Institutional Review Board, not human subjects research. Participants were invited to take part in the optional survey through email. Participation was voluntary and no compensation was provided. Individual participation was not tracked, and individually identifiable data were not collected.

**Results**

**Final Website**

Figure 2 shows a section of the landing page for the dementia caregiver resources website and highlights the table of contents and organizational scheme. The website is housed under the VA Geriatric Scholars website [21].

The website includes 10 categories and 34 subcategories. A table of the final website organization, as well as the number of resources available and the number of visits by subcategory in the first 9 months after launch, is outlined in Table 1.

**Figure 2.** A screenshot of the landing page for the dementia caregiver resources website, a remote compendium of expert-vetted, web-based dementia educational resources for clinicians to provide to caregivers of individuals with dementia.
Table 1. Organizational structure and website visit count for categories and subcategories within the expert clinician-vetted dementia caregiver resources website. The caregiver resources website chart shows the finalized categories and subcategories, the number of resources in each subcategory, and the total number of visits for each subcategory over a 9-month period.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Number of resources available, n</th>
<th>Total number of visits over 9 months, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dementia overview</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General dementia information</td>
<td>6</td>
<td>625</td>
</tr>
<tr>
<td>Non-Alzheimer dementia</td>
<td>3</td>
<td>326</td>
</tr>
<tr>
<td>Dementia research</td>
<td>3</td>
<td>235</td>
</tr>
<tr>
<td><strong>Daily activities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities of daily living</td>
<td>4</td>
<td>450</td>
</tr>
<tr>
<td>Medication</td>
<td>3</td>
<td>282</td>
</tr>
<tr>
<td>Activity participation</td>
<td>3</td>
<td>303</td>
</tr>
<tr>
<td>Traveling</td>
<td>3</td>
<td>252</td>
</tr>
<tr>
<td>Driving</td>
<td>3</td>
<td>306</td>
</tr>
<tr>
<td>Sex and intimacy</td>
<td>3</td>
<td>353</td>
</tr>
<tr>
<td>Sleep</td>
<td>5</td>
<td>250</td>
</tr>
<tr>
<td>Nutrition</td>
<td>5</td>
<td>175</td>
</tr>
<tr>
<td><strong>Behavior changes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety and depression</td>
<td>4</td>
<td>323</td>
</tr>
<tr>
<td>Delirium</td>
<td>2</td>
<td>250</td>
</tr>
<tr>
<td>Sundowning</td>
<td>4</td>
<td>315</td>
</tr>
<tr>
<td>Wandering</td>
<td>3</td>
<td>274</td>
</tr>
<tr>
<td>Aggression and agitation</td>
<td>4</td>
<td>336</td>
</tr>
<tr>
<td>Hallucinations and paranoia</td>
<td>4</td>
<td>285</td>
</tr>
<tr>
<td>Repetitive and inappropriate behaviors</td>
<td>5</td>
<td>251</td>
</tr>
<tr>
<td><strong>Safety</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home safety</td>
<td>5</td>
<td>313</td>
</tr>
<tr>
<td>Firearm safety</td>
<td>2</td>
<td>227</td>
</tr>
<tr>
<td>Falls prevention</td>
<td>3</td>
<td>267</td>
</tr>
<tr>
<td>Emergency preparedness</td>
<td>3</td>
<td>233</td>
</tr>
<tr>
<td><strong>Caregivers’ self-care and support</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Care and support</td>
<td>4</td>
<td>492</td>
</tr>
<tr>
<td>Resources and programming</td>
<td>4</td>
<td>274</td>
</tr>
<tr>
<td>Legal, financial, and employment support</td>
<td>5</td>
<td>271</td>
</tr>
<tr>
<td>Goals of care or end-of-life</td>
<td>3</td>
<td>232</td>
</tr>
<tr>
<td>Communication</td>
<td>4</td>
<td>289</td>
</tr>
<tr>
<td>Long-term placement</td>
<td>4</td>
<td>217</td>
</tr>
<tr>
<td>Brain health</td>
<td>4</td>
<td>386</td>
</tr>
<tr>
<td>Telehealth and technology</td>
<td>6</td>
<td>256</td>
</tr>
<tr>
<td>COVID-19 and dementia</td>
<td>3</td>
<td>219</td>
</tr>
<tr>
<td>Comprehensive care guides</td>
<td>2</td>
<td>365</td>
</tr>
<tr>
<td><strong>Additional websites</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Websites</td>
<td>11</td>
<td>266</td>
</tr>
<tr>
<td>Video series</td>
<td>2</td>
<td>244</td>
</tr>
</tbody>
</table>
Website Usage
In its first 9 months after launch, the website averaged 1100 visits per month between February 2022 and October 2022. A data assessment was completed to determine the number of visits each subcategory had over 9 months (Table 1). Website users frequented the subcategories of general dementia information (625 visits), activities of daily living (450 visits), and self-care and support (492 visits) the most often throughout the 9-month assessment period. This information is supported by the evaluation survey, where participants were specifically asked what resources were most useful to them. Many participants ranked the larger categories of dementia overview, daily activities, and caregivers’ self-care and support in their top 3 most used resource categories.

Survey Results
A total of 60 survey responses were recorded, with each question reported receiving a response from 72% (43/60) or more of the respondents. Survey responses came primarily from clinicians in the fields of nursing, pharmacy, medicine, social work, and rehabilitation. The number of years of clinical practice for respondents ranged from 0-5 years to 30+ years, with 65% (39/49) reporting 10 years or more in the field. A total of 75% (37/49) of respondents reported their caseload was comprised of 50% or more patients aged older than 65 years.

All respondents reported that it was likely or very likely that they would recommend the website to a colleague in the future, and almost all respondents (51/52, 98%) stated that it was likely or very likely that they would recommend the website to a patient or caregiver. To further assess the reach of the website, or very likely that they would recommend the website to a colleague in the future, and almost all respondents (51/52, 98%) stated that it was likely or very likely that they would recommend the website to a patient or caregiver. To further assess the reach of the website, participants were specifically asked what resources were most useful to them. Many participants ranked the larger categories of dementia overview, daily activities, and caregivers’ self-care and support in their top 3 most used resource categories.

Survey Results
A total of 60 survey responses were recorded, with each question reported receiving a response from 72% (43/60) or more of the respondents. Survey responses came primarily from clinicians in the fields of nursing, pharmacy, medicine, social work, and rehabilitation. The number of years of clinical practice for respondents ranged from 0-5 years to 30+ years, with 65% (39/49) reporting 10 years or more in the field. A total of 75% (37/49) of respondents reported their caseload was comprised of 50% or more patients aged older than 65 years.

All respondents reported that it was likely or very likely that they would recommend the website to a colleague in the future, and almost all respondents (51/52, 98%) stated that it was likely or very likely that they would recommend the website to a patient or caregiver. To further assess the reach of the website, participants were asked who they had shared the website with. A total of 51% (25/49) of respondents reported that they primarily shared it with family caregivers, 67% (33/49) with colleagues at the VA, and 16% (8/49) with colleagues outside of the VA. A total of 18% (9/49) of respondents stated that they shared directly with their patients and 14% (7/49) with their personal friends and family. Respondents indicated they had shared the resource website by all the following mechanisms: sharing the overall web page link electronically, sharing links to specific resources electronically, downloading and saving specific resources to send electronically in the future, printing out resources and either handing or mailing them to their recipients, and adding links to resources in their care plan.

A total of 98% (44/45) of survey respondents either agreed or strongly agreed that the website was easy to navigate. Similarly, all respondents either agreed or strongly agreed that the resources on the website were useful for their clinical practice. When asked if the website was easy to find, 77% (34/44) agreed or strongly agreed that the website was easy to find. Survey questions allowing for free-text responses yielded further insight regarding clinician’s overall impressions of the utility of the website and allowed for suggestions for website improvement.

When asked for suggestions on how to improve the website, many respondents identified additional resource subcategories that they thought would be useful, such as resources on death and dying, exercise, and speech pathology (specifically cognition, communication, and swallowing). They also requested more resources offered in video format.

While the website received the majority of positive feedback, the limitation most identified by survey respondents was the inability to easily find the website. Participants stated that they had trouble finding the website through search engines such as Google and wished it were more accessible or visible to those that may not have the direct link (Textbox 2).

Textbox 2. Qualitative responses to open-text questions regarding overall impressions of the dementia caregiver resources website and suggestions for improvement.

- The majority of comments regarding overall impressions of the website were positive.
  - “It is amazing and very comprehensive. Great place to start for someone that is just beginning to be a caregiver for someone with dementia!”
  - “Better than Google for finding information.”
  - “It is an excellent resource with a well-rounded variety of well vetted information from clinical experts.”
  - “It’s everything we need in one space!”
  - “Our caregivers are starving for this information.”
- Some respondents described the ways in which they had successfully used the website in their clinical practice.
  - “I facilitate a dementia educational class for caregivers and share website with them. It has helped them access additional information.”
  - “(I) provided link and or printout to caregivers, as part of education for caregivers in order to manage dementia related behaviors and self-care.”
- Some participants described the ways in which they had trouble finding the website.
  - “I couldn’t find it by Googling. I only found it by using the link provided.”
  - “I wish it showed up just from a Google search so I could find it more easily.”
Discussion

Overview

This paper describes the iterative process of creating the dementia caregiver resources website, a compendium of geriatrics clinician-approved resources presented in an organized, user-friendly format, as well as feedback regarding the website’s usability, accessibility, and suggestions for improvement. The 6-step development process can serve as a model for other clinical subspecialty groups seeking to create comprehensive educational resources. The success of the dementia caregiver resources website is evidenced by the number of visits to the website and the positive clinician reviews received within the first 9 months of its launch.

Previous research has found that family caregivers are using a variety of web-based resources to look for education on the diagnosis and progression of dementia, including searching the internet, reading informal journals or news articles, social media, watching TV, etc [22]. However, even with caregivers receiving information from varied sources, most caregivers expect their health care team to provide them with educational recommendations or reliable resources as a starting point. Many caregivers express confidence in their health care team but report that it can be challenging to connect with them due to clinician’s or caregivers’ busy schedules [23-25]. Furthermore, due to the degenerative nature of dementia, which causes a continuous decline in function, caregivers regularly require education on a variety of developing symptoms [26-28]. This points to a gap the caregiver resources website fills in providing vetted resources, all in one space, for clinicians to give to caregivers whenever needed without requiring additional health care visits.

Similarly, ensuring the topic areas and resources chosen would meet the needs of caregivers and families throughout the dementia spectrum was an important criterion during the process of creating and evaluating this web-based resource. The WDT’s process of continuous identification and prioritization of quality resources, along with iterative assessment of knowledge gaps and reorganization of topics and subtopics, allowed for a breadth of content areas that matched caregivers’ information-seeking trends. These trends show caregivers seek information on disease-specific diagnosis and treatment; disease-focused health care services; caregiver support and self-care; general caregiving skills; and, to a lesser degree, financial and legal information and accommodations [28-30]. Similarly, once the website launched, the team received additional confirmation from the evaluation that the content and organizational structure provided on the website were of great use to caregivers and health care teams alike.

Given the lack of geriatrics specialty care and trained geriatricians available, with even less access to specialty care for rural caregivers, the dementia caregiver resources website can provide resources and education to bridge the gap between specialty medical appointments or for those who are unable to see a geriatrics clinician due to location, limited availability, etc. Finally, those involved in dementia care can use the website to easily find appropriate information at any stage of the disease process to address the specific concerns the caregiver is facing at any given time—in other words, the right information at the right time.

Limitations

There were limitations in the creation and dissemination of the dementia caregiver resources website. The goal of this clinical project was to provide quality resources as quickly as possible to clinicians to then give to their informal dementia caregiving audience. Some clinicians reported sharing specific resources with caregivers, while others reported sharing the website in its entirety. While caregiver feedback was not initially gathered (due to the intended end users being clinicians and the general scope of the project), caregivers were key recipients of the educational information, and therefore, not gathering their input could be a potential limitation of the project. The next steps of this project will include an assessment of caregiver perspectives regarding the website. Additionally, the content does not include all available dementia resources, as it is limited to those that the WDT could review and select within the available time frame of the project.

Another limitation of website dissemination was the location of the website. The website was not easily searchable through common search engines such as Google or Bing, and because it was housed within the GRECC-Connect parent website, it was even more difficult to find. Therefore, to disseminate the website, the end user had to be given the direct URL or know where to look within the GRECC-Connect parent website. Additionally, those in highly rural areas may have difficulty viewing certain resources, such as videos, on the website due to decreased access to high-speed internet. However, the WDT made an effort to ensure that each subcategory had options that did not require high-speed internet access, such as PDFs and Word documents. These challenges clearly limit the number of caregivers who will be able to easily access the website’s information, which further decreases the ability to accurately assess the impact of the website in future evaluations. Lastly, because we provide links to resources that reside on other content creators’ websites, there is the possibility of a change in access or content that is out of our control. The WDT has supplied their contact information so that end users can alert the team to difficulties with accessing the resources and so that visitors can suggest additional resources.

Conclusion

This article describes the iterative process of creating and evaluating the dementia caregiver resources website, a compendium of geriatric expert-approved resources for clinicians to provide targeted education to family caregivers of people with dementia. While there is a plethora of web-based dementia caregiving information available, there lacked a single web-based compendium that housed a plethora of quality resources in an organized fashion, so that navigation was not confusing or inconvenient for busy clinicians. The WDT identified and organized the highest quality resources available that provided education on a wide variety of topics impacting people with dementia and their caregivers. Given the growing number of people with dementia and the heavy reliance on family caregivers without access to specialty dementia care, clinicians benefited from this single-site compendium. The
dementia caregiver resource website received thousands of visits and positive clinician feedback, including 98% (44/45) of survey participants agreeing that the website was easy to navigate and all respondents agreeing that the resources were useful, demonstrating it to be a valuable resource for clinical care. Moving forward, there is potential for caregivers to navigate through the whole website without clinician direction to specific resources. Therefore, the next steps include gathering caregiver feedback and perspectives on the website. Clinicians can refer to the free, publicly available dementia caregiver resources website for expert-reviewed, organized, and quality resources [21]. This process can serve as a model for clinical subspecialty groups, creating comprehensive educational resources available for other populations who lack easy access to specialty care. For those developing a web-based resource that is accessible to both clinicians and older adults themselves, it may be beneficial to include experts in universal website design, specifically for older adult end users, from the beginning of the process to ensure a quality user experience for the broadest audience.

Acknowledgments
This work was made possible by the Veterans Health Administration New England Geriatric Research Education and Clinical Center (GRECC) and the VA Office of Rural Health supported GRECC-Connect program. The authors would like to thank Dr William Hung, GRECC-Connect national program lead; Dr Thomas Caprio, GRECC-Connect education core lead; and Leah Walker, social worker for Madison Wisconsin Veterans Health Administration, for their guidance and feedback throughout the website development process. A special thank you to the health care professionals and website users who completed the website survey and shared their thoughts with them. The views expressed are those of the authors and do not necessarily reflect those of the United States Department of Veterans Affairs.

Data Availability
All data generated or analyzed during this study are included in this published article (and its appendices).

Conflicts of Interest
None declared.

Multimedia Appendix 1
Dementia Caregiver Resources website survey. Survey questions and answer choices, percentage of respondents, and number of respondents reported. Survey questions did not require an answer to submit the survey therefore creating variation in the sample size for each question.

References


Abbreviations
GRECC: Geriatric Research Education and Clinical Center
VA: Veterans Affairs
WDT: website development team
Evaluation of an e-Learning Program for Community Pharmacists for Dispensing Emicizumab (Hemlibra) in France: Nationwide Cross-Sectional Study

Valérie Chamouard¹, PharmD; Julie Freyssenge², PhD; Béatrice Clairaz-Mahiou³, PharmD; Felicia Ferrera Bibas⁴, PharmD; Laurie Fraticelli⁵, PhD

¹Louis Pradel Hospital, Groupement Hospitalier Est, Hospices Civils de Lyon, French Reference Center of Hemophilia, Bron, France
²University Claude Bernard Lyon 1, INSERM U1290 Research on Healthcare Performance RESHAPE, Lyon, France
³French Society of Pharmaceutical Sciences, Châtenay-Malabry, France
⁴Provence-Alpes-Côte-d’Azur Unions for health professionals, Marseille, France
⁵Laboratory P2S (Health Systemic Process), UR 4129, University Claude Bernard Lyon 1, Lyon, France

Corresponding Author:
Laurie Fraticelli, PhD
Laboratory P2S (Health Systemic Process), UR 4129
University Claude Bernard Lyon 1
7-11 rue Guillaume Paradin
Lyon, 69372
France
Phone: 33 04 78 78 57 45
Email: laurie.fraticelli@univ-lyon1.fr

Abstract

Background: Since June 2021, patients with hemophilia A with antifactor VIII inhibitors and those with severe hemophilia A without antifactor VIII inhibitors treated with Hemlibra have had to choose between a community or hospital pharmacy. The French reference center for hemophilia developed the HEMOPHAR e-learning program for community pharmacists for dispensing emicizumab.

Objective: This study aims to evaluate the efficiency and safety of this new care pathway by assessing the HEMOPHAR e-learning program.

Methods: The methodology is based on Kirkpatrick’s model for evaluating the immediate reaction of trained community pharmacists (level 1), their level of acquired knowledge (level 2), and their professional practice after 3 months of dispensation (level 3).

Results: The HEMOPHAR e-learning program reached a large audience, with 67% (337/502) of the eligible community pharmacists following it. The immediate reaction was overall satisfying. High rates of engagement were reported with 63.5% (214/337) to 73.3% (247/337) of completed training modules, along with high rates of success with quizzes of 61.5% (174/337) to 95.7% (244/337). We observed that 83.9% (193/230) of the community pharmacists needed less than 2 attempts to pass the quiz of the module related to professional practice, while the other quizzes required more attempts. Advice on compliance and drug interactions were most frequently provided to patients by the community pharmacists.

Conclusions: This study suggests ways to improve the training of community pharmacists and to optimize coordination with treatment centers. This study also reports on the feasibility of switching to a community pharmacy in a secure pharmaceutical circuit, including in the context of a rare bleeding disease.

Trial Registration: ClinicalTrials.gov NCT05449197; https://clinicaltrials.gov/study/NCT05449197
International Registered Report Identifier (IRRID): RR2-10.2196/43091

(JMIR Form Res 2024;8:e54656) doi:10.2196/54656
KEYWORDS
hemophilia; care pathway; emicizumab; Kirkpatrick model; pharmacy; survey; Hemlibra; France; e-learning program; pharmacists; pharmacist; hemophilia A; hospital; HEMOPHAR; methodology; community; engagement; pharmaceutical; rare disease; digital health; intervention

Introduction
While community pharmacists have long served as safety consultants for patients with minor illnesses, the literature highlights the multifaceted and evolving role of the community pharmacists to improve care pathways especially for patients requiring urgent care [1] or for the management of chronic diseases [2-4]. Evidence of the effectiveness of community pharmacists’ interventions exists for the management of chronic diseases [4], vaccination [5], and optimization of transitions of care [6,7]—the latter referring to the movement of patients across institutions, among providers, between different levels of care, and to and from home. The main pitfall that may occur during these transitions is discontinuity in the patient’s care pathway, leading to potential complications [8]. But offering standardized training regarding processes and procedures for all the involved stakeholders has shown to be a facilitator [9].

The transition of care has typically been implemented in France for patients with hemophilia A with or without inhibitors. This disease is a rare bleeding disorder whose treatment relies on factor VIII injections or bypass agents. Chronic treatment for hemophilia A has been available only in French hospital pharmacies, and monthly dispensing was conditional on reimbursement of these drugs [10]. In France, the delivery of drugs to patients’ homes is outstanding and is only legally reimbursable of the training program are effective or appreciated, and which ones need to be improved.

The e-questionnaire is informative to help determine which elements of the training program are effective or appreciated, and which ones need to be improved.

Methods
Content of the HEMOPHAR e-Learning Program
Community pharmacists receive personal log-in codes for the HEMOPHAR e-learning program from the national reference center for hemophilia (Lyon, France). They can take the course at any time on working days. They can also interrupt the program and resume it at a later date without losing their progress through the program. The program comprises four e-learning modules: (1) presentation of the disease, (2) therapeutic management, (3) organization of care, and (4) practice in the community pharmacy (Multimedia Appendix 1). At the end of each module, the community pharmacist can participate in a quiz composed of 4 or 5 questions. Quizzes are mandatory to obtain the certificate of achievement (with at least 80% of correct answers), but this certificate is not mandatory to dispense emicizumab.

Study Design
We designed a cross-sectional study based on the Kirkpatrick model [15] based on 2 closed surveys and an analysis of secondary data. The evaluation of the training program comprised the investigation of 3 information levels.

Level 1, related to “Reaction,” referred to the evaluation of the immediate feedback of community pharmacists who attended the HEMOPHAR e-learning program. Invitations to complete e-questionnaire were sent via email to community pharmacists who completed the training and obtained the certificate of achievement within 48 hours post training. The e-questionnaire was intentionally short and composed of 8 questions, including 5 questions based on a 4-item Likert scale (“not at all satisfied,” “not very satisfied,” “somewhat satisfied,” and “very satisfied”) and 3 closed-ended questions (yes/no). The estimated time to complete this e-questionnaire is approximately 1 minute. This questionnaire is informative to help determine which elements of the training program are effective or appreciated, and which ones need to be improved.

Level 2, related to “Learning,” referred to knowledge acquisition during the program. This level consists of accurately measuring pharmacists to follow the HEMOPHAR e-learning program by systematically sending an email invitation with personalized log-in details, as soon as the patient has chosen the community pharmacy.

More than 1 year after the implementation of this training program, a nationwide study called “PASODOBLEDEM1” was conducted to evaluate the direct reaction of the community pharmacists following the implementation of the HEMOPHAR e-learning program and the use and functionalities of the available training resource, and to evaluate their professional practice at least 3 months after the first dispensation of emicizumab.
the knowledge acquired during the active training modules. In contrast to immediate reactions to training (level 1), this second level allows for the measurement of specific outcomes by collecting data regarding usage targets such as participation rate, completion rate, or time spent on training to supplement participant feedback.

Level 3, related to “Behavior,” consisted of evaluating the professional practice of community pharmacists at least 3 months after the first dispensation of emicizumab, to determine whether they have followed the HEMOPHAR e-learning program. Specific e-questionnaires were elaborated using Likert-based scoring for both trained and untrained community pharmacists dispensing emicizumab. Level 3 consists of determining whether the training program has a direct impact on the participants’ dispensing behaviors. The community pharmacists were informed about the purpose of and estimated time required to complete the survey via email before opting to complete the e-questionnaire. The level 3 e-questionnaire was composed of 13 multiple-choice questions. The estimated time to complete this questionnaire is approximately 8 minutes.

The contents of the questionnaires were initially developed by a multidisciplinary team composed of hospital and community pharmacists and research scientists specialized in methodology and care accessibility. A scientific committee was specifically constituted for the PASODOBLEDEMI I study to challenge and validate the purposes, methodology, questionnaires, and expected outcomes. This committee was coordinated by the national reference center for hemophilia and composed of hospital and community pharmacists, physicians, and the French association of patients with hemophilia (French: Association Française des Hémophiles). Following the recommendations of the scientific committee, the content of the e-questionnaires has evolved by integrating operational and organizational aspects and optimizing literacy. The e-questionnaires were finally tested in real-life conditions on a sample of eligible populations. The content of the questionnaires was finally validated by the scientific committee in July 2022.

Sample Size
Since emicizumab became available in French community pharmacies, from June 2021 to December 2022, a total of 337 volunteer community pharmacists followed at least 1 of the 4 training modules available in the HEMOPHAR e-learning program. The expected minimum sample size for the questionnaire related to the reaction of the community pharmacists (level 1) was intended to encompass the maximum number of the trained participants who obtained the certificate of achievement during the month preceding the study start date. From July to December 2022, in total, 26 of the 337 community pharmacists completed the short e-questionnaire immediately after following the HEMOPHAR e-learning program, representing the Reaction sample (level 1). As the answers were not divergent, we did not consider it appropriate to extend the duration of data collection. At least 3 months after the implementation of the dual circuit of dispensing, 180 community pharmacists completed the e-questionnaire concerning their professional practice. The Behavior sample (level 3) contains community pharmacists who followed at least 1 of the 4 modules in the HEMOPHAR e-learning program (154/337, 45.7%). The expected minimum sample size for this questionnaire was at least 30%, which has been far exceeded. A total of 363 community pharmacists were included in this study, representing 67% of the eligible community pharmacists in France (337/502 active community pharmacies, last updated in September 2022; Figure 1).

Figure 1. Number of community pharmacists included in the 3 levels of the PASODOBLEDEMI I study. *A total of 26 community pharmacists have either completed the Roche training (n=8) or have not completed any training (n=18).

Statistical Analysis
Baseline characteristics were described by counts and percentages for categorical variables and medians and IQRs for continuous variables, mentioned in parentheses in the following order in the Results section: community and hospital pharmacy. Statistical analysis was performed using R (version 4.1.1; The R Foundation) software.
Ethical Considerations
The study complies with reference methodology MR-004 (“related to research not involving the human person, studies and evaluations in the health field”) [16]. This French legislation mentions that the persons involved in the research or their legal representatives (or both) are informed beforehand and individually of any processing of their personal data for the purposes of research for the present methodology (Title II 2.5.1 related to information of participants); therefore, it did not require written consent from the study participants. No incentive was offered to participate in the survey. All the collected data from community pharmacists were nonsensitive and were stored by the study sponsor (Hospices Civils de Lyon) on a secure server for 15 years. The ethics committee of the Hospices Civils de Lyon approved the study (n°2022-06-01, obtained on June 14, 2022). The PASODOBLEDEMI I study was registered in ClinicalTrials.gov (NCT05449197; related to the evaluation of Reaction, Learning, and Behavior).

Table 1. Satisfaction levels recorded immediately after the HEMOPHAR e-learning program (N=26).

<table>
<thead>
<tr>
<th>Satisfaction levels</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall level of satisfaction</strong></td>
<td></td>
</tr>
<tr>
<td>Very satisfied</td>
<td>18 (69.2)</td>
</tr>
<tr>
<td>Somewhat satisfied</td>
<td>8 (30.8)</td>
</tr>
<tr>
<td>Not very satisfied</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not at all satisfied</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>The training content meets expectations</strong></td>
<td></td>
</tr>
<tr>
<td>Very satisfied</td>
<td>17 (67.4)</td>
</tr>
<tr>
<td>Somewhat satisfied</td>
<td>8 (30.8)</td>
</tr>
<tr>
<td>Not very satisfied</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not at all satisfied</td>
<td>1 (3.8)</td>
</tr>
<tr>
<td><strong>Satisfaction with the communication format</strong></td>
<td></td>
</tr>
<tr>
<td>Very satisfied</td>
<td>14 (53.8)</td>
</tr>
<tr>
<td>Somewhat satisfied</td>
<td>11 (42.3)</td>
</tr>
<tr>
<td>Not very satisfied</td>
<td>1 (3.8)</td>
</tr>
<tr>
<td>Not at all satisfied</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>The training is relevant to professional practice</strong></td>
<td></td>
</tr>
<tr>
<td>Very satisfied</td>
<td>19 (73.1)</td>
</tr>
<tr>
<td>Somewhat satisfied</td>
<td>6 (23.1)</td>
</tr>
<tr>
<td>Not very satisfied</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not at all satisfied</td>
<td>1 (3.8)</td>
</tr>
<tr>
<td><strong>Satisfaction with the duration of training</strong></td>
<td></td>
</tr>
<tr>
<td>Very satisfied</td>
<td>13 (50.0)</td>
</tr>
<tr>
<td>Somewhat satisfied</td>
<td>10 (38.5)</td>
</tr>
<tr>
<td>Not very satisfied</td>
<td>2 (7.7)</td>
</tr>
<tr>
<td>Not at all satisfied</td>
<td>1 (3.8)</td>
</tr>
</tbody>
</table>

Results
Level 1: Immediate Reaction After the HEMOPHAR e-Learning Program
Immediate satisfaction with the HEMOPHAR e-learning program was high, with 18 (69.2%) community pharmacists very satisfied and 8 (30.8%) somewhat satisfied (Table 1). Nearly 3 out of 4 community pharmacists were very satisfied because the training content met their expectations (n=17, 67.4%) and was considered relevant to their professional practice (n=19, 73.1%). They were less very satisfied with the communication format (n=14, 53.8%) and the duration of the training (n=13, 50.0%). Overall, 8 (30.8%) community pharmacists planned to follow another learning program. Finally, 96.2% (25/26) of them would recommend the HEMOPHAR e-learning program to their colleagues.
Level 2: Learning and Knowledge Acquired During the Training

A median delay of approximately 9 (IQR 3-21) days was observed between the day the community pharmacists received their connection codes and the day they logged in to the HEMOPHAR e-learning program. Among the community pharmacists who started at least 1 of the 4 training modules, 81.9% (276/337) of them started all the training modules available in the HEMOPHAR e-learning program, and 40.9% (113/276) of them completed it. The training modules related to the presentation of the disease and the one related to therapeutic management reached most of the community pharmacists, with more than 80% of them having started the training modules (Table 2). Once the training modules were started, between 60% and 75% of community pharmacists completed them.

<table>
<thead>
<tr>
<th>Training modules</th>
<th>Cumulated duration of the videos</th>
<th>Started (n=337), n (%)</th>
<th>Completeda (relative to those who startedb), n (%)</th>
<th>Successful quiz completers (relative to those who startedb), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Presentation of the disease</td>
<td>16 minutes 52 seconds</td>
<td>280 (83.1)</td>
<td>214 (63.5)</td>
<td>214 (76.4)</td>
</tr>
<tr>
<td>2. Therapeutic management</td>
<td>22 minutes 51 seconds</td>
<td>283 (84.0)</td>
<td>247 (73.3)</td>
<td>174 (61.5)</td>
</tr>
<tr>
<td>3. Organization of care</td>
<td>22 minutes 25 seconds</td>
<td>255 (75.7)</td>
<td>244 (72.4)</td>
<td>244 (95.7)</td>
</tr>
<tr>
<td>4. Practice in the community pharmacy</td>
<td>15 minutes 46 seconds</td>
<td>230 (68.2)</td>
<td>223 (66.2)</td>
<td>183 (79.6)</td>
</tr>
</tbody>
</table>

aCommunity pharmacists who completed the training modules, with or without attempting the related quizzes.  
bDenominators are the number of community pharmacists who started the training module (n=337).

Community pharmacists usually need more than 1 attempt to pass the quiz at the end of each module (Figure 2). The quiz related to the organization of care was the most successful with a 95.7% success rate associated with the lowest number of attempts. The time spent per quiz varied greatly with 50% of the community pharmacists, who needed 6.8 minutes for the presentation of the disease module, 3.5 minutes for the therapeutic management module, 1.5 minutes for the organization of care module, and 2.6 minutes for practice in the community pharmacy module.

Figure 2. Number of attempts at the quizzes for each of the training modules in the HEMOPHAR e-learning program.
Level 3: Behavior During Professional Practice of Community Pharmacists

The community pharmacists reported that mostly 1 patient chose their pharmacy for dispensing emicizumab (90.9%, 140/154 unique community pharmacies). Nearly half of the community pharmacists (74/154, 48.1%) who responded to the Behavior questionnaire followed the HEMOPHAR e-learning program for more than 3 months, and one-third of the sample (51/154, 33.1%) followed the program for more than 12 months.

During their professional practice, community pharmacists may be faced with organizational constraints. Even though the HEMOPHAR e-learning program covers all aspects of organization and documentation management, community pharmacists have reported that they did not always rely on the documents available (Figure 3). For example, the most used document was the one containing the stakeholders’ contact information (90.3%, 139/156) and the liaison document for emicizumab (85.7%, 132/154), unlike the checklist, which was used for only 78.6% (121/154) of the community pharmacists. Overall, 89.6% (138/154) of community pharmacists reported that they did not observe any discontinuity in the pathway of care of the patient in general. Those who encountered problems of continuity of care reported constraints in terms of personnel, particularly during the holiday season, and those relating to the supply of emicizumab or those due to a prescription delay by the hospital, even though 76% (117/154) of respondents reported that patients systematically anticipate the renewal of prescription before their arrival at the community pharmacy. About 84.4% (130/154) of the community pharmacists reported that they did not need a third party to manage the dispensation until now (autonomy). Regarding the time spent in supplying emicizumab, 80.5% (124/154) of the community pharmacists received the drug in less than 24 hours.

Figure 3. Organizational items reported from the community pharmacists who followed the HEMOPHAR e-learning program (n=154).

Community pharmacists who followed the HEMOPHAR e-learning program have reported the frequency of advice provided when dispensing emicizumab (Figure 4). The most frequent advice provided by the community pharmacists concerned the importance of medication adherence, with 25.3% (36/154) of the trained community pharmacists having provided this advice time systematically at every visit of the patient or their caregiver and 24.7% (38/154) having provided this advice often. Administration of emicizumab was addressed sometimes, according to 48.1% (74/154) of the community pharmacists, as well as the impact of drug interactions when self-medicating (41.6%, 64/154; sometimes). The mode of action of emicizumab was the least discussed topic between patients and community pharmacists. Most pharmacists felt they had sufficient information for dispensing emicizumab and managing their relationship with patients (92.2%, 142/154). However, we observed that 36.4% (56/154) of community pharmacists trained with the HEMOPHAR e-learning program expressed the need for therapeutic education support.
Discussion

The PASODOBLEDEMI I study represents the first nationwide survey to evaluate the systemic impact of the implementation of the dual dispensing circuit, quickly after its implementation, with high institutional expectations. The strengths of the PASODOBLEDEMI I study lie in its holistic approach using the Kirkpatrick evaluation model [15] and the high-quality representative sample, which enhances the validity of the findings, especially for the Behavior sample in evaluating professional practice at 3 months. We report 36% (180/502) representativeness, with more than 30% representation from each French department, reaching 65% in Hauts-de-France and 63% in Auvergne Rhône-Alpes. Representativeness was lesser for Provence-Alpes-Côte d’Azur (18%) and Ile-de-France (21%) but remained consistent with that reported in other surveys [17,18]. The size of the Reaction sample was considered sufficient (n=26) because of the convergence of the satisfaction ratings.

The HEMOPHAR e-learning program reached a large audience with 67% of the community pharmacists being eligible. The immediate reaction was overall satisfying despite mixed feedback related to the format and duration, which will be reviewed for the next version of HEMOPHAR (ie, version 2). High rates of engagement were reported, with 63.5% to 73.3% of training modules having been completed and high rates of success of 61.5% to 95.7% with the quizzes. As it is possible to take the quiz without attending the training modules, we hypothesized that some community pharmacists tested their knowledge to auto-evaluate their need for training. As an example, 76.4% of them succeeded in the quiz related to the presentation of the disease, while 63.5% of them completed the related training module. We also observed that 83.9% of the community pharmacists needed less than 2 attempts to pass the quiz in the module related to professional practice, while the other quizzes required more attempts. Advice on compliance and drug interactions were most frequently provided to patients by the community pharmacists. The mode of action of emicizumab was less frequently mentioned, which may reflect the effectiveness of the hospital education that patients received.

The e-learning format was evaluated as the best way to educate community pharmacists for dispensing emicizumab and helping them to play an important role with this patient community in primary care [19,20] because the e-learning program provides both theoretical knowledge and practical skills for health professionals and educational reinforcement for patients [21]. Comparison with other training methods, such as classroom teaching or lectures, has shown that e-learning is associated with equally satisfactory results [22]. The HEMOPHAR e-learning program represents a strong model for effective training and skill development among community pharmacists, by offering accessibility, flexibility, cost-effectiveness, multimodal learning, tracking and assessment, and updateability. Moreover, community pharmacists can return to the training modules at any time to review the videos or to download useful documents. A complementary approach would comprise affective peer coaching to increase the uptake and effectiveness of a variety of community pharmacist-led enhanced patient care services [23]. Also, the gamification of training programs presents a promising perspective for the advancement of training and development efforts by offering several advantages, including increased learner engagement, immediate feedback, collaborative learning, development of critical thinking skills, and data-driven optimization [24].

The main limitation of the PASODOBLEDEMI I study is the absence of a control group (ie, untrained community pharmacists). We were constrained by regulatory issues from accessing the list of community pharmacies chosen by the patients from among hemophilia treatment centers. So, we first relied on the community pharmacists who attended the HEMOPHAR e-learning program and then publicly invited all the eligible community pharmacists through social networking platforms (LinkedIn, Facebook, Twitter, etc) and academic societies (French society of pharmaceutical sciences, unions for health professionals, etc) to take the survey.

In France, the short training by telephone on the proper use of the product provided by Roche-Chugai to community pharmacists and validated by the French authorities was not sufficient to optimize the dual circuit for dispensing emicizumab. The HEMOPHAR e-learning program is much broader and goes beyond the strict use and storage of the product to increase the role of community pharmacists in the management of this rare bleeding disorder, becoming a new model of care [25]. Other countries have made the same decision in evolving the dispensing circuit such as Australia [26]. If hemophilia treatment...
centers ensure adequate education and training on the technique for subcutaneous injection for patients, community pharmacists are not, to our knowledge, trained to dispense emicizumab. Other initiatives have been developed for enhancing the role of community pharmacists for consultation services [1] and for urgent care [27], which could result in increased confidence and capability in offering the service, with many community pharmacists recognizing and embracing a shift in their roles. An evaluation of patients’ and carers’ satisfaction will be the next step in the PASODOBLEDEMI II study (registered in ClinicalTrials.gov NCT05450640).

The PASODOBLEDEMI I study provides a holistic approach based on the Kirkpatrick model for evaluating the reaction, learnings, and behaviors of community pharmacists after attending the HEMOPHAR e-learning program. With high engagement rates, the community pharmacists benefited from high quality and relevance of the content on autonomy, flexibility, and interactivity, and encompassing the whole national territory. A future certification program will provide recognition for community pharmacists’ educational efforts. The success of the HEMOPHAR e-learning program in the context of the emicizumab dual dispensing circuit may pave the way for other treatments to be made available in community pharmacies, reaching institutional expectations and enhancing the access to treatment for patients with rare diseases.

Acknowledgments
We thank the scientific committee of the PASODOBLEDEMI I study: BC (community pharmacist, copresident of the French Society of Pharmaceutical Sciences [French: Société Francophone des Sciences Pharmaceutiques Officielles]), Yesim Dargaud (National Reference Center for Hemophilia, Lyon), FF (cochair of the Provence-Alpes-Cote-d’Azur Unions for Health Professionals [French: Société Française de Pharmacie Clinique], Marseille), Nicolas Giraud (president of the French Hemophilia Association [French: Association Française des Hémophiles]), Rémi Varin (hospital pharmacist, Rouen), and Fabienne Volot (Expert Center Coordinator, Dijon). The authors thank the team from Roche-Chugai laboratories for their follow-up on the project (Loïc Bergougnoux, Aurélie de Lehvenfehl). This research was funded by the Roche-Chugai laboratories.

Data Availability
The data that support the findings of this study are available from the corresponding author upon request.

Authors’ Contributions
VC, JF, and LF conceptualized the study and carried out the formal analysis and investigation. LF designed the study methodology. VC and LF drafted the manuscript, and JF, FFB, and BC-M reviewed and edited the manuscript. VC and LF acquired the funding. JF and LF acquired the resources for this study. VC and LF supervised the study.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Screenshots of the e-learning program HEMOPHAR.
[DOCX File, 11862 KB - formative_v8i1e54656_app1.docx ]

References
5. Ciliberti R, Bragazzi NL, Bonsignore A. The implementation of the professional role of the community pharmacist in the immunization practices in Italy to counteract vaccine hesitancy. Pharmacy (Basel) 2020 Aug 25;8(3) [FREE Full text] [doi: \text{10.3390/pharmacy8030155} ] [Medline: 32854420]


23. Chamouard et al JMIR FORMATIVE RESEARCH


Edited by A Mavragani; submitted 17.11.23; peer-reviewed by D Dinh; comments to author 25.01.24; revised version received 05.02.24; accepted 07.03.24; published 04.04.24.

Please cite as:
URL: https://formative.jmir.org/2024/1/e54656
doi:10.2196/54656
PMID:38574351

©Valérie Chamouard, Julie Freyssenge, Béatrice Clairaz-Mahiou, Felicia Ferrera Bibas, Laurie Fraticelli. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Acceptability of a Self-Guided Lifestyle Intervention Among Young Men: Mixed Methods Analysis of Pilot Findings

Jean Miki Reading, PhD; Melissa M Crane, PhD; Justin Guan, MS; Ronston Jackman, BS; Maria D Thomson, PhD; Jessica Gokee LaRose, PhD

1Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
2Department of Family and Preventive Medicine, Rush University, Chicago, IL, United States
3Department of Social and Behavioral Sciences, School of Population Health, Virginia Commonwealth University, Richmond, VA, United States

Corresponding Author:
Jean Miki Reading, PhD
Department of Preventive Medicine
Feinberg School of Medicine
Northwestern University
680 N Lake Shore Dr
Chicago, IL, 60611
United States
Phone: 1 3125034870
Email: jean.reading@northwestern.edu

Abstract

Background: Young men are vastly underrepresented in lifestyle interventions, suggesting a need to develop appealing yet effective interventions for this population.

Objective: This study aimed to determine the acceptability of a self-guided lifestyle intervention designed specifically for young men (age: 18-35 years old).

Methods: Semistructured interviews and surveys were completed by 14 men following completion of a remotely delivered, 12-week lifestyle intervention. The intervention included 1 virtual group session, digital tools, access to self-paced web- and mobile-based content, and 12 weekly health risk text messages. We quantitatively and qualitatively examined young men’s experiences with the intervention components of a remotely delivered, self-guided lifestyle intervention targeting weight loss. Data were integrated using convergent mixed methods analysis.

Results: Men were a mean age of 29.9 (SD 4.9) years with a mean BMI of 31.0 (SD 4.5) kg/m². The self-guided aspect was not acceptable, and a majority preferred more check-ins. Participants expressed a desire for a social aspect in future lifestyle interventions. All men found the focus on health risks appealing. A majority of men found the study-issued, Bluetooth-enabled scale acceptable.

Conclusions: Acceptability of the self-guided lifestyle intervention was perceived as suboptimal by young men. The findings highlight the need to add intervention components that sustain motivation and provide additional social support for young men.

Trial Registration: ClinicalTrials.gov NCT04267263; https://www.clinicaltrials.gov/study/NCT04267263

JMIR Form Res 2024;8:e53841 doi:10.2196/53841

KEYWORDS
digital health; gender; weight loss; health behaviors; low touch; obesity; obese; mixed methods analysis; lifestyle intervention; young men; men; effectiveness; digital tools; food intake; diet

Introduction

Young men with obesity during young adulthood have twice the mortality risk of men with a healthy BMI (kg/m²) [1]. Despite heightened risk, young men are underrepresented in lifestyle interventions targeting weight loss [2-5]. Low enrollment among young men may stem from their limited concern about weight gain [6,7] or the absence of lifestyle interventions designed to meet the specific needs of men [8]. Adapting interventions to align with the needs and preferences of young men [9,10] while also raising awareness about the

Gender-specific lifestyle interventions indicate promise for engaging men to lose weight and include different features appealing to men (eg, sports-based or self-guided approach) [9,12-15]. Implementing a self-guided approach appears efficacious in producing initial weight loss and satisfaction among young men [14,16,17] and is consistent with young men’s preferences for convenient interventions [9]. When considering age, young adults demonstrate a preference for interventions that reduce intensity and promote autonomy [18].

In a pilot weight loss trial targeting young men, a self-guided approach, paired with health risk messaging, promoted modest weight loss compared with the control [11]. However, it remains unclear which specific elements are perceived as helpful in supporting weight loss for young men.

To improve young men’s engagement with weight loss, it is critical to adapt behavioral weight loss interventions based on their needs. Guidelines for behavioral intervention development are outlined in the Obesity-Related Behavioral Intervention Trials (ORBIT) model and recommend using an iterative process to reach optimal treatment outcomes for the target population [19]. The ORBIT model specifically notes using mixed methods approaches for defining and refining interventions and for feasibility pilot testing [19]. Prior to a rollout of a behavioral weight loss intervention, it is important to determine the acceptability of the intervention among users, which is often captured quantitatively through measures of satisfaction, attendance, and efficacy [20-23]. Satisfaction is a key construct to consider when developing an intervention, given that higher levels of satisfaction with an intervention are associated with favorable weight loss outcomes [24]. However, qualitative measures of satisfaction with interventions are limited in behavioral weight loss trials [24-26], especially as they relate to young men. As a result, our understanding of young men’s experiences with behavioral weight loss is limited.

Indeed, studies are needed that incorporate both quantitative and qualitative user feedback and experience with the intervention. To that end, we used a mixed methods approach to explore the experiences of young men who completed a 12-week lifestyle intervention designed specifically for this population. The primary aim of this paper was to explore young men’s satisfaction with specific intervention components of a self-guided lifestyle intervention to inform future development of interventions [11].

Methods

Study Design

This study was part of a randomized clinical trial in which 18 men received the intervention and 17 men were allocated to delayed treatment control. We report data for the 14 men who were randomized to the intervention and completed the 12-week follow-up visit, an exit interview, and a survey following the visit to provide feedback on the intervention components (retention rate: 14/18, 78%). The full design, protocol, and exclusion criteria are detailed elsewhere [11].

Ethics Approval

All procedures were approved by the institutional review board at Virginia Commonwealth University (IRB# HM20015458).

Recruitment

Men between the ages of 18 years and 35 years with a BMI of 25 kg/m² to 45 kg/m² were recruited across North America and locally in the greater Richmond, Virginia, area during a 2-month period (January 2021-March 2021) using unpaid recruitment advertisements distributed through email listservs, university postings, and researchmatch.org. Advertisements for the intervention emphasized that the lifestyle intervention was self-guided, included images of men exercising and a health risk message, and described some of the inclusion criteria (ie, BMI, age, men). Interested participants completed online screening via Research Electronic Data Capture (REDCap) to determine initial eligibility and were contacted by a member of the study team to schedule an orientation to learn more about the study and engage in an informed consent process. All participants provided written informed consent.

Sample

Participants for the main trial [11] were eligible if they met the following inclusion criteria: (1) age 18 years to 35 years and (2) BMI between 25 kg/m² and 35 kg/m². Exclusion criteria included (1) medical contraindications to exercise without medical clearance, (2) a diagnosis of type 1 or 2 diabetes, (3) report of a heart condition, (4) a history of anorexia or bulimia nervosa, (5) report of compensatory behaviors in the last 3 months, (6) hospitalization for psychiatric conditions in last 12 months, (7) participation in another weight loss program (8) ≥5% weight loss in the last 3 months, (9) not able to read or speak English, (10) did not possess a mobile device, or (11) lived or resided outside of North America. Only participants assigned to the intervention arm were eligible for the current aims.

Intervention

Men received the 12-week lifestyle intervention that was primarily self-guided and grounded in behavioral self-regulation [27] and health risk messaging guided by the extended parallel process model [28]. The intervention was remotely delivered and included 1 group session delivered via Zoom. The group session was followed by self-paced content accessible through a private intervention website, 12 weekly health risk text messages (automated and nonresponsive), and personalized feedback at baseline and 12 weeks based on assessment data. All participants were provided with a Bluetooth-capable scale. All intervention content included health risk messaging applying the extended parallel process model’s constructs (perceived susceptibility, perceived severity, self-efficacy, response efficacy) [28]. The messaging highlighted health risks specific to young men, including the association between obesity and cardiovascular disease, and evidence-based strategies for facilitating weight loss and mitigating cardiovascular disease risk. The baseline feedback report was delivered via email and included the participant’s current BMI, weight, 5% and 10% weight loss goals, daily calorie goal, and health risk messaging. The 12-week feedback report included the participant’s BMI,
weight, percentage of weight loss achieved during the intervention, and health risk messaging. Each intervention component is described in greater detail in the following paragraphs.

The virtual group session occurred via Zoom at the start of the intervention. The session was 45 minutes and led by a licensed clinical psychologist with expertise in behavioral weight loss treatment. Men were provided with psychoeducation regarding health risks of obesity [29,30] and behavioral self-regulation principles [31]. Additionally, men received training in evidence-based behavior change techniques for managing weight [31] and engaged in small group experiential activities to apply and practice skills to increase self-efficacy. In addition, men received instructions on how to access the website and content covered during the session.

Digital tools included access to a private intervention website with evidence-based content, tools for self-monitoring, and a Bluetooth-capable scale. The intervention website was hosted through a private server and was accessible via web and mobile devices throughout the 12-week intervention. The website offered additional psychoeducation on healthy weight management, diet, and physical activity and behavioral change techniques for making changes to health behaviors. Content was adapted for health behaviors relevant to young men [9], which focused on improving fitness and reducing consumption of alcohol, sugar-sweetened beverages, processed meals and fast food, and foods high in fat. The website also housed links to publicly and commercially available online videos and apps for physical activity, self-monitoring diet, and meal preparation. See Figure 1 for example screenshots of the intervention website.

To reinforce extended parallel process model constructs, 12 weekly health risk text messages were sent throughout the intervention. For example, “Eating out and fast food can put men at high risk for heart disease due to high fat, calories, and sodium. Cutting back on fast food and the meals you eat away from home can lower your risk! Check out meal planning tips on [study website] for ways to reduce your risk.”

**Figure 1. ACTIVATE Private Intervention Website and Features.**

### Measures

At 12 weeks, participants were asked to report the acceptability of a variety of aspects of the intervention on a 7-point Likert scale. Participants rated aspects of the intervention (eg, male only, young adult only, self-guided) they found appealing upon joining the intervention (1=not appealing at all; 7=very appealing). Participants rated intervention components (group session, text messages, website) and certain features (length, frequency, relevance) perceived as helpful with weight loss (1=not helpful; 7=very helpful). Participants also rated intervention components and delivery methods not offered (eg, in-person meetings focused on physical activity, online meetings about diet).

A female PhD student (JMR, student investigator) with extensive interviewing experience conducted 14 interviews by phone in a private setting. On average, interviews lasted 15 (range: 8-29) minutes. The semistructured interview guide included open-ended questions about motivations for joining, satisfaction with the intervention, challenges and successes with the intervention, and recommendations for future interventions (Multimedia Appendix 1). Standardized probes were used to elicit responses about intervention components that were considered helpful or unhelpful for intervention goals. Interviews were audio-recorded and transcribed verbatim.

### Statistical Analysis

Descriptive statistics were computed for each question assessing intervention acceptability. Intervention components were deemed acceptable if the mean satisfaction score for each item was ≥5 (1 point above the central point of the 7-point scale) [32]. Descriptive statistics were computed using SPSS Version 27 (IBM Corp). A directed content analysis was used to qualitatively code the semistructured interviews [33]. The coders consisted of 2 men with education in public health and behavioral medicine (MS...
and BS degrees) who were not investigators of the study. Both coders were trained by the student investigator with extensive experience in qualitative research (JMR). First, 4 transcripts were reviewed by the student investigator (JMR) to develop an initial codebook of primary and secondary codes. Transcripts were then analyzed in batches of 2 to 4 by the 2 coders, with minimal involvement by the student investigator. Interencoder reliability (kappa > 0.80) and fidelity to the codebook were maintained during coding [34]. Incongruent codes were flagged, discussed, and reviewed at weekly meetings with the student investigator. All discrepancies were resolved through group consensus. Although data achieved saturation after 10 coded interviews [35], all data were coded. Coded categories were grouped by intervention components and aspects of participant satisfaction. Subcategories of each pre-identified theme were cross compared for word similarities to identify overlap in responses. Data were coded using NVivo 12.0 (QSR International).

JMR served as the student investigator and conducted this study as part of her doctoral dissertation in Social and Behavioral Sciences. Under guidance of the other senior authors, JMR made all final decisions throughout the study process. She considered her role as student investigator and her gender from conceptualization to the reporting of findings. Many factors, including time and financial constraints, were taken into account when determining who would collect and analyze the data. JMR collected the semistructured interview data due to availability, experience, and direct contact with participants throughout the study. After each interview, JMR documented her assumptions in memos. Given JMR’s vested interest in her doctoral dissertation, she took a minimal role in coding to mitigate potential bias in the qualitative findings. Her involvement in the qualitative data analysis primarily revolved around facilitating discussion on coding discrepancies and providing qualitative training.

A convergent mixed methods analysis [36] was used following standard guidelines [37]. Specific items from the quantitative data were integrated with corresponding themes related to intervention materials, in which qualitative data were embedded into quantitative data to explain acceptability ratings of intervention components.

**Results**

Of the 18 men, 14 completed the 12-week follow-up visit after completing the intervention (retention: 78%). Participants’ mean age was 29.0 (SD 4.8) years, with a mean BMI of 31.0 (SD 4.7) kg/m$^2$, and 29% (4/14) identified as racial and ethnic minorities. Demographic characteristics of the participants who completed the follow-up visit are displayed in Table 1. On average, participants lost –1.8% (SD 2.8%) of their initial body weight at 12 weeks (range: –9.5% to +1.3%). Results are described in 3 thematic areas in the following sections. Exemplary quotes from qualitative interviews corresponding to themes and subthemes are provided in Table 2.

<table>
<thead>
<tr>
<th>Table 1. Sample characteristics (n=14).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>18-25</td>
</tr>
<tr>
<td>26-35</td>
</tr>
<tr>
<td>Race*</td>
</tr>
<tr>
<td>American Indian/White</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Ethnicity</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
</tr>
<tr>
<td>Relationship status*</td>
</tr>
<tr>
<td>Married</td>
</tr>
<tr>
<td>Single</td>
</tr>
<tr>
<td>Living with partner</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Some college</td>
</tr>
<tr>
<td>College graduate</td>
</tr>
<tr>
<td>Postgraduate degree</td>
</tr>
</tbody>
</table>

*Does not sum to 100% due to missing data not reported.
Table 2. Exemplary quotes by corresponding theme and subtheme.

<table>
<thead>
<tr>
<th>Theme and subthemes</th>
<th>Exemplary quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appealing aspects of the intervention or motivations for joining</td>
<td>• “I think what specifically, uh, got me into it was the fact that it targeted young men [sic], my age group.”</td>
</tr>
<tr>
<td>Age and gender</td>
<td>• “The scale was really, really useful, actually, especially starting out because I could see, like, all the different metrics that I had no idea of before.”</td>
</tr>
<tr>
<td>Acceptability of intervention components</td>
<td>• “I already use some of the nutrition tracking apps, but there was a workout app specifically, Fit On is the one that I kind of latched on to and I hadn’t heard of it before. And I use that as my primary source for exercise and kind of coming up with an actual exercise plan. And I use My Fitness Pal a little bit as well to check diet.”</td>
</tr>
<tr>
<td>Bluetooth scale and application</td>
<td>• “The text messages I found helpful [sic], as a reminder of the ultimate reason why I was doing this.”</td>
</tr>
<tr>
<td>Recommended applications</td>
<td>• “I mean, uh, the uh, the text message that we got, it was good accountability, but it was also, that I can see how that could also just be something that you just kind of slough off because we just like, a here’s a fact.”</td>
</tr>
<tr>
<td>Health risk text messages</td>
<td>• “The text messages I found helpful [sic], as a reminder of the ultimate reason why I was doing this.”</td>
</tr>
<tr>
<td>Preferences for an ideal intervention</td>
<td>• “The text messages I found helpful [sic], as a reminder of the ultimate reason why I was doing this.”</td>
</tr>
<tr>
<td>Social aspect</td>
<td>• “Maybe a social component, more so than the accountability part, you know. Just like a shared experience kind of thing.”</td>
</tr>
<tr>
<td>Frequency of contact</td>
<td>• “I would say like having, like, another check-in would maybe be good, like midway through or a couple of check-ins.”</td>
</tr>
</tbody>
</table>

Appealing Aspects of the Intervention or Motivations for Joining

During interviews, a desire to lose weight was the most common reason for joining the study (8/14, 57%). Other reasons for joining included meeting the advertised age and gender demographic (7/14, 50%), gaining knowledge (6/14, 43%), getting in shape (5/14, 36%), and the self-guided aspect (4/14, 29%).

I liked that it was targeted at my age group by my demographic in general. I feel like there are a lot of weight loss programs for other demographics. And this is the first one that I see that civically targeted, you know, and my age group.

Most participants did not report hesitation about joining the intervention. Those who reported a reluctance (3/14, 21%) about joining the intervention were mainly concerned about having enough time. The quantitative results indicate that the majority of participants found the intervention appealing and decided to join because of the emphasis on general lifestyle changes (13/14, 93%), focus on health risk (14/14, 100%), and weight loss (11/14, 79%). The lowest percentage of men rated the self-guided component (5/14, 36%) or minimal in-person contact (5/14, 36%) as reasons for joining. See Table 3.
Table 3. ACTIVATE intervention acceptability items.

<table>
<thead>
<tr>
<th>Question</th>
<th>Overall results, mean (SD)</th>
<th>Percent rating ≥5.0³, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How satisfied were you with the overall ACTIVATE program that you received during the past 12 weeks?</td>
<td>4.2 (1.1)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>How satisfied were you with what you achieved in the ACTIVATE program?</td>
<td>3.9 (.83)</td>
<td>4 (29)</td>
</tr>
<tr>
<td>Would you recommend the ACTIVATE program to other young men?</td>
<td>4.5 (1.6)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>The information I learned in this program would be relevant to other men of my age who want to lose weight.</td>
<td>5.4 (1.2)</td>
<td>12 (86)</td>
</tr>
<tr>
<td>The length of the program was sufficient for a weight loss program targeting men my age (18-35).</td>
<td>4.4 (1.6)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>What parts of the program did you find appealing?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male only</td>
<td>5 (1.4)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>Young adult only</td>
<td>5 (1.1)</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Minimal in person</td>
<td>4.4 (1.7)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>Self-guided</td>
<td>3.8 (1.8)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>Focus on health risk</td>
<td>5.8 (.73)</td>
<td>14 (100)</td>
</tr>
<tr>
<td>Fitness</td>
<td>5.1 (1.1)</td>
<td>10 (71)</td>
</tr>
<tr>
<td>Diet</td>
<td>5.1 (1.1)</td>
<td>10 (71)</td>
</tr>
<tr>
<td>Weight loss</td>
<td>5.4 (1.0)</td>
<td>11 (79)</td>
</tr>
<tr>
<td>General lifestyle changes</td>
<td>5.9 (1.0)</td>
<td>13 (93)</td>
</tr>
<tr>
<td>How much did each of the following help you to lose weight?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group session</td>
<td>3.7 (1.9)</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Intervention website</td>
<td>3.6 (1.6)</td>
<td>4 (29)</td>
</tr>
<tr>
<td>Meal plans</td>
<td>3 (1.2)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Text messages</td>
<td>4.3 (1.8)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>Feedback</td>
<td>4.1 (1.0)</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Scale</td>
<td>5.1 (1.4)</td>
<td>11 (79)</td>
</tr>
<tr>
<td>App to track food</td>
<td>4.0 (1.6)</td>
<td>6 (43)</td>
</tr>
<tr>
<td>App to track physical activity</td>
<td>4.5 (1.6)</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Rating of website/group session</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The skills taught on the website helped me with my weight loss efforts.</td>
<td>4.1 (1.2)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>The website content was motivating to me.</td>
<td>3.4 (1.3)</td>
<td>2 (14)</td>
</tr>
<tr>
<td>The information on the website was relevant to me.</td>
<td>4.1 (1.2)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>The length of the group session was the right amount of time.</td>
<td>4.4 (1.2)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>The strategies taught in the group session were helpful to me.</td>
<td>4.1 (1.3)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>The information in the group session was motivating to me.</td>
<td>4.3 (1.1)</td>
<td>6 (43)</td>
</tr>
<tr>
<td>The information in the group session was relevant to me.</td>
<td>4.6 (1.2)</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Rating of weekly text messages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The messages were motivating to me.</td>
<td>5.0 (1.5)</td>
<td>9 (64)</td>
</tr>
<tr>
<td>The messages suggested strategies that were helpful to me.</td>
<td>4.2 (1.6)</td>
<td>9 (64)</td>
</tr>
<tr>
<td>The messages made me aware of the risks associated with weight gain.</td>
<td>5.4 (1.6)</td>
<td>11 (79)</td>
</tr>
<tr>
<td>The messages made me aware that I am at risk for cardiovascular disease.</td>
<td>4.9 (1.5)</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Preferred additional features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add an online group component for discussion.</td>
<td>5.6 (1.4)</td>
<td>13 (93)</td>
</tr>
<tr>
<td>Add in-person group meetings focusing on diet.</td>
<td>4.9 (1.7)</td>
<td>9 (64)</td>
</tr>
</tbody>
</table>
Acceptability of Intervention Components

Less than one-half (5/14, 36%) of the men rated the overall intervention acceptable (satisfaction >5). For specific components, the majority of the participants in the interviews mentioned the Bluetooth scales (11/14, 79%), recommended apps (9/14, 64%), and text messages (10/14, 71%) were helpful for weight loss.

For the digital tools, participants discussed finding the Bluetooth scale’s features, which included an app to track metrics and progress, helpful. One young man shared his thoughts on the Bluetooth scale:

*The scale was really, really useful actually, especially starting out because I could see, like, all the different metrics that I had no idea of before.*

Quantitative findings indicated the percentage of men who found the recommended apps as acceptable (rating ≥5) was lower than those in the qualitative findings. In particular, less than one-half of men (6/14, 43%) found the recommended applications to self-monitor diet acceptable. Over one-half of men (8/14, 57%) found the recommended applications to self-monitor physical activity acceptable. See Table 3.

A low percentage of men found the meal planning strategies (1/14, 7%) and the intervention website (4/14, 29%) as acceptable (rating ≥5). Less than one-half of the participants found the strategies taught in the group session helpful (5/14, 36%) and the website content motivating (2/14, 14%).

The quantitative data indicate over one-half of men found the text messages motivating (9/14, 64%), helpful (9/14, 64%), and raised awareness of the health risks of weight gain (11/14, 79%) and cardiovascular disease (8/14, 57%).

The qualitative data indicate participants found the weekly text messages served as good reminders. One-half (7/14, 50%) of the qualitative interviews indicate participants found the text messages of benefit. The other one-half (7/14, 50%) either did not find the messages helpful or had mixed feelings about them—some felt the content of the text messages did not have enough variety or were something that could be easily disregarded. One young man shared thoughts on the weekly health risk messages:

*The text message that we got, it was good accountability, but it was also [sic] something that you just kind of slough off because it’s just like here’s a fact.*

Preferences for an Ideal Intervention

Of the participants, 71% (10/14) described a desire for a social aspect to the intervention, and almost all men (13/14, 93%) preferred an online group component for discussion. In particular, 93% (13/14) of men wanted an online group meeting focusing on physical activity. Participants discussed the desire for a message board or ongoing discussion with other participants to help with motivation or accountability throughout the intervention. Over one-third (5/14, 36%) also mentioned a desire for a more personalized experience and to receive more feedback (5/14, 36%). A participant shared:

*Maybe a social component, more so than the accountability part, you know. Just like a shared experience kind of thing.*

The majority (8/14, 57%) of participants wanted more contact from the intervention. For the most part, participants preferred to have midpoint or monthly online group check-ins, as opposed to weekly check-ins.

Discussion

Principal Findings

This convergent mixed methods analysis provides insight into young men’s experiences with a self-guided lifestyle intervention to inform future intervention development for this vastly underrepresented population. We explored 3 key areas: motivations for joining and appealing aspects of the intervention, acceptability of specific intervention components, and preferences for an ideal intervention.

Motivations for joining and acceptable intervention components of a lifestyle intervention were key areas explored in this study. Qualitative and quantitative data surrounding motivations for joining and acceptable intervention components were fairly consistent. First, qualitative data indicated men’s interest in joining the study was related to age and gender. This underscores the importance of designing recruitment messaging specifically for young men as a way to enhance enrollment in lifestyle interventions [38]. Furthermore, the emphasis on health risks emerged as a key motivator for young men to participate in the lifestyle intervention. However, the existing literature on leveraging health risks to prompt health behavior change among men presents mixed findings. Although some findings suggest that fear of health complications [8] and a desire to improve health [16] act as motivators for engaging with weight loss behaviors, other research indicates that young men may prefer to avoid discussions about associated health risks [10]. It is plausible that the use of health risk messaging in recruitment

### Table 3

<table>
<thead>
<tr>
<th>Question</th>
<th>Overall results, mean (SD)</th>
<th>Percent rating ≥5,a, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add online group meetings focusing on diet.</td>
<td>5.4 (1.4)</td>
<td>11 (79)</td>
</tr>
<tr>
<td>Add in-person group meetings focusing on physical activity.</td>
<td>4.6 (1.6)</td>
<td>9 (64)</td>
</tr>
<tr>
<td>Add online group meetings focusing on physical activity.</td>
<td>5.6 (1.3)</td>
<td>13 (93)</td>
</tr>
<tr>
<td>Add in-person group meetings focusing on muscle strengthening.</td>
<td>4.4 (1.6)</td>
<td>9 (64)</td>
</tr>
<tr>
<td>Add online group meetings focusing on muscle strengthening.</td>
<td>4.9 (1.4)</td>
<td>10 (71)</td>
</tr>
</tbody>
</table>

aPercentage of participants rating intervention features acceptable (>5 on a 7-point scale).
The overall acceptability of the self-guided lifestyle intervention among young men was found to be suboptimal. However, young men reported high acceptability of the study-issued scale and companion app to track weight. Additionally, the online delivery method was well-received by men. Therefore, prioritizing remote delivery of lifestyle interventions for participants who are meeting weight loss goals could enhance accessibility, scalability, and convenience [40,41]. It is worth noting that young men perceived the self-guided aspect as somewhat “hands-off” and lacking personalization. Specifically, young men expressed dissatisfaction with the website, which did not offer new weekly content throughout the active intervention period. Collaborating with young men in future interventions to co-design weekly content might render higher appeal and engagement with the intervention.

Young men highlighted a preference for adding a social aspect in future interventions. Specifically, men reported wanting a shared similar experience—which is a reported benefit of using online platforms for sharing weight loss experiences [42]. Recent data underscore the potential for peer support to promote weight loss in a reduced intensity lifestyle intervention, but the majority of sample was adult women [43]. Of note, these preferences for a shared similar experience could be somewhat negated in standard behavioral weight loss interventions, which are predominantly women [2,44]. Qualitative data indicate men do not feel comfortable discussing men’s health issues and weight loss around women [45]. Thus, integrating a private online platform for young men to discuss relevant health issues might be one strategy for improving this population’s engagement with weight management. In addition to the desire for an online social platform, young men also reported a preference for online group sessions related to physical activity. This finding is consistent with formative data suggesting young men have a greater desire for interventions to focus on physical activity than young women, particularly as it relates to peer support and accountability [18]. Although physical activity is less effective at producing weight loss compared with diet alone [46], promoting physical activity upfront as a way to engage men could have a “spill-over” effect on other behaviors such as diet.

Last, incorporating social support via personalized weekly feedback derived from self-monitoring and goal progress is an evidence-based behavioral change technique [47] that might potentially boost motivation and engagement among young men. Despite men performing well in a self-guided lifestyle intervention targeting weight loss [14], young men in this study expressed a preference for monthly check-ins to aid in accountability. Therefore, future endeavors should consider testing different levels of intensity to determine the optimal level of support needed for men to achieve weight loss goals while also addressing time-related barriers faced by young men.

Limitations
This study had several limitations. The sample was mostly non-Hispanic White with a college education. Given the racial and education disparities in both obesity prevalence and enrollment in behavioral weight loss interventions [48,49], more research is warranted to investigate the weight management needs and preferences of young men from marginalized racial and ethnic identities or men without a college education. These data were collected during COVID-19. Rapid shifts toward digital health interventions and the unique context of a global health pandemic may have impacted the findings in numerous ways (eg, desire for weight management, social connection). The treatment-seeking sample might not be generalizable to young men broadly. Moreover, we only interviewed men who returned for their follow-up visit. Thus, important elements for enhancing engagement were potentially missed in the present sample. Interviews were of shorter duration than a typical qualitative interview. However, given the specific goals and deductive design of the qualitative data and pairing of quantitative data, the depth of the participant responses was sufficient in addressing the paper objectives. Men who completed the interviews were slightly older than men who were lost to follow-up (30 years old vs 26 years old). Therefore, more work is needed to understand the preferences and experiences of emerging adult men specifically. Last, this study tested a “bundled” intervention package, which limited our ability to delineate effects of individual intervention components. Future research should apply rigorous factorial designs, such as the multiphase optimization strategy (MOST) [50], to delineate individual and combined effects and acceptability of intervention components to develop an optimized intervention package for young men.

Strengths
This study had several strengths. To our knowledge, this is the first mixed methods study to report young men’s experiences with and the acceptability of a remotely delivered, self-guided lifestyle intervention targeting weight loss. In future work, fully embracing user-centered design will allow us to identify elements that improve the participant experience and related outcomes in lifestyle interventions [51]. Additionally, we had high agreement between coders. Last, we followed standard guidelines for the best practice of integrating the qualitative and quantitative findings [37]. Behavioral weight loss trials can benefit from a mixed methods design—using qualitative and quantitative data to complement inherent weaknesses in each, generate robust findings, and enhance validity [52].

Conclusions
Our findings suggest that the acceptability of a self-guided approach was less than optimal among young men. To improve acceptability, potential enhancements might include incorporating online group sessions focused on physical activity, providing personalized feedback based on self-monitoring and goal setting, or implementing an online platform to foster peer support among young men. Further refinement of this lifestyle intervention is necessary before conducting a large-scale randomized controlled trial. A cost-efficient design, such as MOST [50], could be utilized to determine the most effective individual and combined intervention components offering the greatest clinical benefit while considering practical aspects such as cost and scalability. These interventions could enhance...
engagement among young men by augmenting the self-guided approach to include additional support such as optional online group exercise classes. These classes might be a relevant time to emphasize the importance of physical activity—as it relates to men’s health. Findings also suggest a social component could enhance accountability for men attempting weight loss. More research is needed to expand our understanding of young men’s experiences with weight management over a longer-term follow-up and engagement strategies for reaching emerging adults.

Acknowledgments
The authors would like to acknowledge Ronald Evans, PhD; Autumn Lanoye, PhD; Morgan Meyer, MS; Justin Guan, MS; and Ronston Jackman, BS, for their help executing this trial. We thank the study participants who dedicated time to participate in this study. JMR was supported by National Institutes of Health (NIH)/National Cancer Institute (NCI) training grant T32CA093423 and Virginia Commonwealth University (VCU) Department of Health Behavior and Policy to carry out this trial and received funding support from NIH/NCI grant T32CA193193 to prepare portions of this manuscript and publication costs. Additional resources were provided by the VCU OPT for Health lab (JGL PI) and VCU Clinical and Translational Science Award (CTSA; UL1TR002649). The NIH had no involvement in the study design, collection, analysis, and interpretation of these data.

Conflicts of Interest
JGL discloses grant funding from WW International, unrelated to the current work. The remaining authors have no other competing interests to disclose.

Multimedia Appendix 1
Semistructured interview guide.

References


32. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31 [FREE Full text] [doi: 10.1002/oby.20232] [Medline: 23505165]
Abbreviations

MOST: multiphase optimization strategy
ORBIT: Obesity-Related Behavioral Intervention Trials
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 20.10.23; peer-reviewed by M Young, D Verran; comments to author 04.12.23; revised version received 08.03.24; accepted 11.03.24; published 05.04.24.

Please cite as:
Reading JM, Crane MM, Guan J, Jackman R, Thomson MD, LaRose JG
Acceptability of a Self-Guided Lifestyle Intervention Among Young Men: Mixed Methods Analysis of Pilot Findings
JMIR Form Res 2024;8:e53841
URL: https://formative.jmir.org/2024/1/e53841
doi:10.2196/53841
PMID:38578686

©Jean Miki Reading, Melissa M Crane, Justin Guan, Ronston Jackman, Maria D Thomson, Jessica Gokee LaRose. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Abstract

**Background:** The escalating prevalence of cesarean delivery globally poses significant health impacts on mothers and newborns. Despite this trend, the underlying reasons for increased cesarean delivery rates, which have risen to 36.3% in Portugal as of 2020, remain unclear. This study delves into these issues within the Portuguese health care context, where national efforts are underway to reduce cesarean delivery occurrences.

**Objective:** This paper aims to introduce a machine learning, algorithm-based support system designed to assist clinical teams in identifying potentially unnecessary cesarean deliveries. Key objectives include developing clinical decision support systems for cesarean deliveries using interoperability standards, identifying predictive factors influencing delivery type, assessing the economic impact of implementing this tool, and comparing system outputs with clinicians’ decisions.

**Methods:** This study used retrospective data collected from 9 public Portuguese hospitals, encompassing maternal and fetal data and delivery methods from 2019 to 2020. We used various machine learning algorithms for model development, with light gradient-boosting machine (LightGBM) selected for deployment due to its efficiency. The model’s performance was compared with clinician assessments through questionnaires. Additionally, an economic simulation was conducted to evaluate the financial impact on Portuguese public hospitals.

**Results:** The deployed model, based on LightGBM, achieved an area under the receiver operating characteristic curve of 88%. In the trial deployment phase at a single hospital, 3.8% (123/3231) of cases triggered alarms for potentially unnecessary cesarean deliveries. Financial simulation results indicated potential benefits for 30% (15/48) of Portuguese public hospitals with the implementation of our tool. However, this study acknowledges biases in the model, such as combining different vaginal delivery types and focusing on potentially unwarranted cesarean deliveries.

**Conclusions:** This study presents a promising system capable of identifying potentially incorrect cesarean delivery decisions, with potentially positive implications for medical practice and health care economics. However, it also highlights the challenges and considerations necessary for real-world application, including further evaluation of clinical decision-making impacts and understanding the diverse reasons behind delivery type choices. This study underscores the need for careful implementation and further robust analysis to realize the full potential and real-world applicability of such clinical support systems.

*JMIR Form Res 2024;8:e54109*  doi:10.2196/54109
KEYWORDS
obstetrics; machine-learning; clinical decision support; interoperability; interoperable; obstetric; cesarean delivery; cesarean; cesarean deliveries; decision support; pregnant; pregnancy; maternal; algorithm; algorithms; simulation; simulations

Introduction

Background

The ability to provide care to both women and newborns during delivery is one of the most important aspects of health care and is often used as a metric to assess health care across different countries. Cesarean delivery is one of the most important aspects of delivering babies since it has a considerable impact on the mother’s health and well-being. Despite the increased prevalence of this procedure over the last few years, the reasons behind this trend still remain unclear. Reports suggest that this increment is a global phenomenon, with the rate of cesarean deliveries almost tripling from 6.7% to 19.1% between 1990 and 2014 [1,2]. Research on the impacts of cesarean deliveries has focused on the risk of infection, hemorrhage, organ injury, and complications related to anesthesia or blood transfusion [3,4]. There is also a higher risk of complications in subsequent pregnancies, such as uterine rupture, abnormal placental implantation, and the need for hysterectomy [5,6]. As for the infant, cesarean deliveries can lead to respiratory problems, asthma, and childhood obesity [5]. In light of this, in 2015, the World Health Organization stated that cesarean delivery rates higher than 10% were not associated with a reduction in maternal or newborn mortality, even though other complications could not be fully assessed [7]. In contrast, there is no evidence of the benefits of this procedure for women or babies when there is no clear medical need; therefore, it is paramount to focus on identifying and reducing such cases [2]. It was estimated that in 2018, there were 8.8 million unnecessary cesarean deliveries [8]. It was with this in mind that a committee was established in Portugal with the specific purpose of decreasing the percentage of cesarean deliveries nationwide. One of the policies resulting from this committee’s work was the reduction of government funding per inpatient cesarean delivery episode for hospitals with rates of cesarean deliveries above 25%; as of 2020, the number of cesarean deliveries in Portugal stands at approximately 36.3%, nearing the all-time high of 36.9% in 2009 [9]. Furthermore, studies have shown that several countries could benefit from similar policies [8]. A quantitative analysis estimated that a reduction in cesarean deliveries could save millions of dollars [10] worldwide. Therefore, lowering the proportion of cesarean deliveries can yield health and financial benefits for both institutions and patients alike. With these considerations in mind, we developed a machine learning, algorithm-based support system to assist clinical teams in identifying cases of potentially unnecessary cesarean deliveries. As such, in this paper, we propose to (1) elaborate on how clinical decision support systems for cesarean deliveries can be developed using interoperability standards; (2) understand, based on the data collected, which features have the most significant impact on predicting delivery type; (3) conduct a concise economic analysis to assess the potential financial impact of implementing the proposed clinical decision support tool; and (4) compare the system’s output with clinicians’ responses.

Rationale and Related Work

Regarding the related work, several teams already tackled the potential of predicting the delivery type before birth. We found studies related to predicting a successful vaginal birth after a previous cesarean delivery, such as the work of Lipschuetz et al [11], where a gradient boosting method was used to predict such an event using prenatal data to do so. Grobman et al [12] performed a similar study with a multivariable logistic regression model. Different modalities of data were also used to predict delivery type. Fergus et al [13] introduced a method of predicting delivery type using fetal heart rate signals. Similarly, the work from Saleem et al [14] proposed a method for predicting delivery type using interactions between the fetal heart rate and maternal uterine contraction. Finally, some studies focus on predicting the delivery mode, such as the work of Ullah et al [15], where a boosting algorithm was used to predict a delivery mode with enriched data sets. In addition, Gimovsky et al [16] introduced decision trees to predict cesarean deliveries by physician group with an area under the receiver operating characteristic curve (AUROC) of 0.73. The works of Rossi et al [17] resulted in a 7-variable model with an AUROC of 0.78, and the works of Guedalia et al [18] resulted in a model with an AUROC of 0.82, reaching 0.93 with a first cervical examination. Finally, the works of Meyer et al [19] focused on selecting something suitable for a trial of labor after cesarean delivery with an area under the precision recall curve graph around 0.351. However, to the best of our knowledge, there was no model tested in clinical practice with an interoperable format of communication such as Fast Healthcare Interoperability Resources (FHIR), which tried to not only predict delivery type but also provide support about possibly wrong deliveries, and none with simulation about financial implication, making our paper a potential novelty on different dimensions.

Methods

Materials

Data were retrospectively collected from 9 different public Portuguese hospitals across the country, focusing on obstetric information and encompassing maternal data, various fetal data points, and the method of delivery retrospectively. The inclusion criteria are all mothers with a registered outcome of the pregnancy from 2019 to 2020. There were no exclusion criteria. Each institution used identical electronic health record software, ensuring that the data columns remained consistent.

Clinical Comparison

The clinical comparison was performed by sending questionnaires to clinicians with a relationship with obstetrics to assess 10 patients, with only access to the variables used by the model, and to answer 3 questions for each. The first question was to give a score from 1 to 10 of how likely it was that a
patient would give birth through cesarean delivery, the second question was to select the feature or variable that most influenced the decision, and the third question was to select which feature they would require to make a better assessment. We sent the questionnaire to 20 people and obtained 6 answers, totaling 60 patient assessments. For these 10 patients, we also predicted the delivery type using our model to compare it with the clinicians’ answers. These patients were new and were not seen by the model during the training phase.

Analysis

All null representations were standardized. Data were prepossessed by removing features with high missing rates (>90% overall). The imputation process was performed using the k-nearest neighbor imputation method (for continuous variables) or a new category (NULLIMP) for categorical variables. Weight was categorized into percentiles defined specifically for Portuguese babies [20]. For this study, the birth type was reduced to binary. All assisted birth were merged into vaginal birth, and cesarean delivery remained as the other class. Procedures and diagnoses were also used and were encoded as binary features, and we took the time to analyze each one of them to avoid leakage because there were procedures obviously related to cesarean deliveries and vaginal deliveries. Feature creation was performed through the free-text variable related to the prescribed medication. Medicine names were collected from it and converted into Anatomical Therapeutic Chemical classification group level 4, which represents chemical subgroups. We also created some new features from data in the data set, namely new categories related to the labor and condition of the baby. In addition, data quality issues were addressed, such as impossible values that were transformed into null values. The main variables affected by data quality were BMI or weight and gestational age. The data were split into training and test sets in a 0.75:0.25 manner. From the overall data sets that comprised over 200 columns, only a few columns were selected (see Table 1). We used a mixture of features selected by surveying the literature [21-23] and features with a high correlation with the outcome. The tested models were logistic regression, decision tree, random forest, 3 different boosting methods (as implemented by extreme gradient boosting [XGBoost], light gradient-boosting machine [LightGBM], and scikit learn), and a linear model based on stochastic gradient descent.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Mode$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother’s age (y), mean (SD)$^b$</td>
<td>31.0 (5.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Weight prepregnancy (kg), mean (SD)</td>
<td>65.8 (13.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Weight on admission (kg), mean (SD)</td>
<td>78.6 (14.2)</td>
<td>N/A</td>
</tr>
<tr>
<td>BMI (kg/m$^2$), mean (SD)</td>
<td>25.0 (5.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Previous eutocic delivery, mean (SD)</td>
<td>0.4 (0.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>Previous vacuum-assisted delivery, mean (SD)</td>
<td>0.1 (0.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Previous forceps, mean (SD)</td>
<td>0.0 (0.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Previous cesarean delivery, mean (SD)</td>
<td>0.1 (0.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Fetal presentation on admission, n (%)</td>
<td>19,305 (26.32)</td>
<td>Cephalic</td>
</tr>
<tr>
<td>Bishop score, mean (SD)</td>
<td>5.5 (3.0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Gestational age on admission (mo), mean (SD)</td>
<td>38.9 (1.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Premature rupture of the membrane, n (%)</td>
<td>64,541 (87.99)</td>
<td>No</td>
</tr>
<tr>
<td>Chronic hypertension, n (%)</td>
<td>71,649 (97.68)</td>
<td>No</td>
</tr>
<tr>
<td>Gestational hypertension, n (%)</td>
<td>71,700 (97.75)</td>
<td>No</td>
</tr>
<tr>
<td>Preeclampsia, n (%)</td>
<td>72,104 (98.3)</td>
<td>No</td>
</tr>
<tr>
<td>Gestational diabetes, n (%)</td>
<td>65,876 (89.81)</td>
<td>No</td>
</tr>
<tr>
<td>Gestational diabetes treated with a diet, n (%)</td>
<td>69,162 (94.29)</td>
<td>No</td>
</tr>
<tr>
<td>Gestational diabetes treated with insulin, n (%)</td>
<td>71,942 (98.08)</td>
<td>No</td>
</tr>
<tr>
<td>Gestational diabetes treated with oral antidiabetic drugs, n (%)</td>
<td>71,737 (97.8)</td>
<td>No</td>
</tr>
<tr>
<td>Maternal diabetes, n (%)</td>
<td>72,991 (99.51)</td>
<td>No</td>
</tr>
<tr>
<td>Type 2 diabetes, n (%)</td>
<td>73,233 (99.84)</td>
<td>No</td>
</tr>
<tr>
<td>Presentation at birth, n (%)</td>
<td>68,950 (94)</td>
<td>Vertex presentation</td>
</tr>
<tr>
<td>Delivery, n (%)</td>
<td>39,507 (53.86)</td>
<td>Spontaneous</td>
</tr>
<tr>
<td>Gestational age on birth (mo), mean (SD)</td>
<td>39.0 (1.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Smoking during pregnancy, n (%)</td>
<td>64,871 (88.44)</td>
<td>No</td>
</tr>
<tr>
<td>Alcohol consumption during pregnancy, n (%)</td>
<td>72,360 (98.65)</td>
<td>No</td>
</tr>
<tr>
<td>Consumed drugs during pregnancy, n (%)</td>
<td>73,226 (99.83)</td>
<td>No</td>
</tr>
<tr>
<td>Number of pregnancies (with current), mean (SD)</td>
<td>1.9 (1.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pregnancy type, n (%)</td>
<td>62,656 (85.42)</td>
<td>Spontaneous</td>
</tr>
<tr>
<td>Surveillance, n (%)</td>
<td>71,664 (97.7)</td>
<td>Yes</td>
</tr>
<tr>
<td>Hospital surveillance, n (%)</td>
<td>49,739 (67.81)</td>
<td>Yes</td>
</tr>
<tr>
<td>Pelvis adequacy, n (%)</td>
<td>12,844 (17.51)</td>
<td>Adequate</td>
</tr>
<tr>
<td>Consistency of the cervix, mean (SD)</td>
<td>1.6 (0.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Fetal station, mean (SD)</td>
<td>0.8 (0.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Dilation of the cervix, mean (SD)</td>
<td>1.3 (0.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Effacement of the cervix, mean (SD)</td>
<td>1.2 (1.2)</td>
<td>N/A</td>
</tr>
<tr>
<td>Position of the cervix, mean (SD)</td>
<td>0.6 (0.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>Hematologic disease, n (%)</td>
<td>70,182 (95.68)</td>
<td>No</td>
</tr>
<tr>
<td>Respiratory disease, n (%)</td>
<td>70,131 (95.61)</td>
<td>No</td>
</tr>
<tr>
<td>Cerebral disease, n (%)</td>
<td>72,470 (98.8)</td>
<td>No</td>
</tr>
<tr>
<td>Cardiac disease, n (%)</td>
<td>68,194 (92.97)</td>
<td>No</td>
</tr>
</tbody>
</table>
The evaluation was performed with repeated stratified cross-validation with 10 splits and 2 repetitions, with 2 full cycles of dividing the training set into 10 equal parts and using 9 as the training set and 1 as the validation set. The results are shown in Table 2. The application programming interface (API) for serving the prediction model was developed using FastAPI. We wrote all the code in Python (version 3.9.7; Python Software Foundation).

**Table 2.** Repeated cross-validation (10 × 2) results in the training set with mean AUROC\(^a\) and 95% CI for the best hyper-parameters found for each algorithm. Wilcoxon test was used for comparing with the best performing algorithm.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>AUROC (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGBoost(^b)</td>
<td>0.8809 (0.88-0.8818)</td>
<td><em>(^c)</em></td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.8338 (0.8328-0.8347)</td>
<td>(\leq 0.001)</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.8716 (0.8708-0.8724)</td>
<td>(\leq 0.001)</td>
</tr>
<tr>
<td>AdaBoost(^d)</td>
<td>0.8753 (0.8744-0.8763)</td>
<td>(\leq 0.001)</td>
</tr>
<tr>
<td>LightGBM(^e)</td>
<td>0.8805 (0.8795-0.8815)</td>
<td>.003</td>
</tr>
<tr>
<td>Stochastic gradient descent</td>
<td>0.8704 (0.8695-0.8712)</td>
<td>(\leq 0.001)</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.8752 (0.8744-0.8761)</td>
<td>(\leq 0.001)</td>
</tr>
</tbody>
</table>

\(^a\)AUROC: area under the receiver operating characteristic curve. \(^b\)XGBoost: extreme gradient boosting. \(^c\)Not available. \(^d\)AdaBoost: Adaptive Boosting. \(^e\)LightGBM: light gradient-boosting machine.

**Ethical Considerations**

This study received institutional review board approval from all hospitals included in this study with the following references: Centro Hospitalar São João (08/2021), Centro Hospitalar Baixo Vouga (12-03-2021), Unidade Local de Saúde de Matosinhos (39/CES/JAS), Hospital da Senhora da Oliveira (85/2020), Centro Hospitalar Tamega Sousa (43/2020), Centro Hospitalar Vila Nova de Gaia/Espinho (192/2020), Centro Hospitalar entre Douro e Vouga (CA-371/2020-06_MP/CC), and Unidade Local de Saúde do Alto Minho (11/2021). All methods were carried out per relevant guidelines and regulations. The need for informed consent was waived by the ethics committee.

**Results**

**Descriptive Statistics**

The number of samples varied across the hospitals, ranging from 2364 to 18,177. Distributions of the selected variables are presented in Table 1. The sum of all samples was 73,351. The outcome variable distribution is stated in Table 3.

**Model**

The AUROC is presented in Table 2 for the best hyper-parameters found for each algorithm in the training data. All models used the variables indicated in Table 1.

While XGBoost was the best-performing algorithm, we selected LightGBM [24] because of its speed and lower memory requirements, which we believe are better suited for deployment in a low-hardware environment. The threshold selected for deploying the model was 0.7457, which rendered the metrics in the test set, as shown in Table 4.
Deployment

The purpose of this model is to serve as an API for usage within a health care institution and to act as a supplementary clinical decision support tool for obstetrics teams. For this to happen, a health information system must make the requests to the API. Even though a concrete, vendor-specific information model and input health information system were used, we hope to create a more interoperable clinical decision support system that can be used by every system that acts on birth and obstetrics departments. Therefore, we built it around the HL7 FHIR standard (R5 version) to simplify the method of interacting with the API. This decision, opposed as to using a proprietary model for the data, sits upon the usage of FHIR resources: bundle and observation for request and returning the result as a message through a custom operation called "$predict." It is intended to publish the profiles of these objects to facilitate access to the API using standardized mechanisms and data models. The current build of the profiles can be seen in the published FHIR implementation guide where the current specifications are described in detail [25]. The process is illustrated in Figure 1. We deployed this model in production in a single hospital without a user interface, collecting only the data and predictions for later discussion and analysis. We collected 3231 requests. During this period, 123 (3.8%) alarms were triggered. From this, we tried to understand the level of certainty for the decision and check the difference from the threshold of these alarms. The distance to the threshold for 73 was lower than 0.1 and was bigger than 0.1 for 50 (1.55%) cases.

Figure 1. Deployment and decision mechanism of the model. EHR: electronic health record; FHIR: Fast Healthcare Interoperability Resources.

Clinical Comparison

The median scores given by each clinician are presented in Figure 2. We also predicted the result using our model as stated in Figure 2. The model misclassified only 1 record (ID 4). As for the analysis of missing features for the responders, they were divided into 3 categories: (1) existent in the data set but not included in the model, (2) nonexistent in the data set, and (3) existent in the data set and included but that particular information was not filled for the patient assessed. Out of 60 responders, this rendered a total of 37 (62%) with nonexistent features and 23 (38%) with existent features but no information was provided at that moment. No feature mentioned existed in the data set but had not been included in the model. From the 37 nonexistent features, 14 (38%) were new clinical assessments, 14 (38%) were linked to information from previous births, 5 (15%) were connected to more in-depth information about provided information (ie, a motive for induction), and 4 (11%) were related to the mother’s choice (if she wanted a cesarean delivery). As for feature importance, from the 60 answers, 33 (55%) stated that labor was the most important factor. Further, 9 (15%) stated the number of previous vaginal births, 5 (8%) stated the evolution of weight, and another 5 (8%) stated the number of previous cesarean deliveries as being the

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.8052</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.8223</td>
</tr>
<tr>
<td>Precision</td>
<td>0.9023</td>
</tr>
<tr>
<td>$F_1$ score</td>
<td>0.8605</td>
</tr>
</tbody>
</table>
most important. The remaining 8 (14%) were various features, such as BMI, neuroaxis techniques, gestational age, and weight of the mother. Of all of these, 54 (90%) were in the top 10 features of the model.

Figure 2. Validation data. The color represents the actual birth type. The boxplot represents the median and IQR of the reviewers, and the x-axis represent the patient cases. There were 6 vaginal births and 4 cesarean deliveries. ID 4 represents wrong predictions of the model.

**Potential Financial Impact**

The financial support provided to public hospitals in Portugal is partially tied to the rate of cesarean deliveries. To assess the potential impact of this mechanism on Portuguese public hospitals, we conducted a simulation. We got data for every public hospital for the last 12 months and applied a 3.8% reduction (the rate of warnings triggered in the new data set) and recalculated the rate of cesarean deliveries. The increase in support was calculated by the state-mandated rate as shown in Table 5. With this new rate, we observed that implementing our tool would result in financial benefits for 30% (15/48) of the public hospitals. Specifically, 5 hospitals would begin receiving support instead of no support at all. Further, 3 hospitals would experience a doubling of their financial benefit, while 2 hospitals would see a 50% increase. Furthermore, 1 hospital would receive an additional one-third of financial support. If we assumed that only half of the warnings found in the new data were actually true (1.9%), we found that only 6 hospitals would be benefited: 3 from 0 to 0.25, 2 from 0.25 to 0.50, and 1 from 0.50 to 0.75.

**Table 5.** With this new rate, we observed that implementing our tool would result in financial benefits for 30% (15/48) of the public hospitals. Specifically, 5 hospitals would begin receiving support instead of no support at all. Further, 3 hospitals would experience a doubling of their financial benefit, while 2 hospitals would see a 50% increase. Furthermore, 1 hospital would receive an additional one-third of financial support. If we assumed that only half of the warnings found in the new data were actually true (1.9%), we found that only 6 hospitals would be benefited: 3 from 0 to 0.25, 2 from 0.25 to 0.50, and 1 from 0.50 to 0.75.

**Table 5.** The rule set for state-provided financial support indexed to cesarean deliveries. x is the current payment of a cesarean delivery inpatient episode [26].

<table>
<thead>
<tr>
<th>Rate of cesarean deliveries</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;25%</td>
<td>x</td>
</tr>
<tr>
<td>25%-26.4%</td>
<td>0.75x</td>
</tr>
<tr>
<td>26.5%-27.9%</td>
<td>0.5x</td>
</tr>
<tr>
<td>28%-29.4%</td>
<td>0.25x</td>
</tr>
<tr>
<td>&gt;29.5%</td>
<td>0</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

The first thing to address about this model is the number of biases that we introduced in the model by choice. We joined all vaginal delivery types into a single category (assisted and nonassisted), which introduces a bias since these delivery modes are indeed different. Second, the fact that we want to predict if the delivery type was wrongly chosen, mainly for the case of a cesarean delivery that did not need to be so, is also a bias. We used this approach because the initially collected data did not have the representation of such events. Thus, the biases of possibly wrong delivery types were present in the training data. We attempted to minimize this issue by selecting a threshold that gave the model higher sensitivity than specificity so that only large probabilities would trigger an alarm for human consideration. Parallel to this, we are starting to gather labeled cases, with the help of clinicians to create a better training data set. Furthermore, since the data were collected from different hospitals, differences in the data input can also occur. Even though the health information system is the same, the processes that originate the data and are being used for secondary purposes could introduce several biases in the data. This is an issue that was accepted from the start regarding the mechanism of data collection and model training. Despite this, we reached a model with a very high AUROC (88%, 95% CI 0.8795-0.8815), which is encouraging when compared to the state-of-the-art, which is between 0.73 and 0.82 [16-18]. Moreover, assuming that more...
data are provided, and proper labeling is done regarding the outcome variable (such as a clinical evaluation of needless cesarean deliveries) is added as well, a better model could be developed.

Regarding the preliminary clinical evaluation, it was only possible to obtain an overview of the possible comparison due to the number of responders. Despite that, the results are encouraging, since the model seems to behave better than humans with the data provided. However, this is a biased vision since clinicians in the real world have access to more data and information than the model has. It is encouraging, but caution is advised before more tests and evaluations are done. As for the deployment, future work could be the improvement of the API to map all variables to an ontology such as SNOMED CT or similar, making it easier for every system and person to access it and obtain a suggestion of the delivery type. Finally, we believe the assessment can be improved. A more robust clinical assessment is necessary, as well as a thorough analysis of the impact of the tool in the real world, since we need to create the bridge between the results of the model and how clinical decisions are affected by it. A full cost-effectiveness analysis is also necessary to understand the real-world impact of the model. Further, an interesting result is the fact that 38% of the answers regarding the most important data element missing from the patient record refer to data that are being collected but was missing for that specific patient, raising an important question about data input methodology, interoperability, and quality. If we cannot have access to data when these matter the most, these can become meaningless. Missing data are a problem of biomedical data as a whole. However, when specifically targeted at machine learning usage of this data for predicting something, we did not find any work comparing them with clinicians. However, we did find reporting of similar missing values in obstetrics data [27] and we also found works of a similar nature using machine learning models with robust handling of missing data such as XGBoost [28] to counter this problem. This indicates that our model has the potential to counter the missing data problem as well since LightGBM can also handle missing data natively.

Conclusions

We believe we have developed a robust system capable of detecting potentially incorrect cesarean delivery decisions, which could positively impact real-world medical practice. However, before implementation, several challenges must be addressed, particularly the need for further evaluation of the system’s impact on clinical decision-making and the reasons underlying suboptimal delivery-type decisions. Cesarean deliveries may be performed for various reasons, from a mother’s preference to a decision made by the obstetrics team. This system is not designed to impede medical practice or to highlight flawed decisions, potentially scrutinizing specific professionals. Such caution is necessary when implementing systems like these. While having a high AUROC is beneficial, the real-world impact is another consideration. The assumptions and biases associated with autonomous systems supporting clinical practice must be carefully considered. Nonetheless, the metrics and results we have achieved so far are promising for positively influencing health and economic outcomes.

Acknowledgments

The authors would like to acknowledge the support of Component 5—Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union, framed in the NextGenerationEU, for the period 2021-2026, within project HIPT (41).

Data Availability

The data cannot be publicly available due to privacy and ethical restrictions. The code is available [29].

Conflicts of Interest

None declared.

References


**Abbreviations**

- **API**: application programming interface
- **AUROC**: area under the receiver operating characteristic curve
- **FHIR**: Fast Healthcare Interoperability Resources
- **LightGBM**: light gradient-boosting machine
- **XGBoost**: extreme gradient boosting

©João Coutinho-Almeida, Alexandrina Cardoso, Ricardo Cruz-Correia, Pedro Pereira-Rodrigues. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Mobile Technology Use in Clinical Research Examining Challenges and Implications for Health Promotion in South Africa: Mixed Methods Study

Khuthala Mabetha1, PhD; Larske M Soepnel1,2, PhD; Gugulethu Mabena1, MA; Molebogeng Motlhatlheldi1, BSc; Lukhanyo Nyati1, PhD; Shane A Norris1,3, PhD; Catherine E Draper1, PhD

1South African Medical Research Council/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, School of Clinical Medicine, University of the Witwatersrand, Johannesburg, South Africa
2Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht University, Utrecht, Netherlands
3School of Human Development and Health, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Khuthala Mabetha, PhD
South African Medical Research Council/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, School of Clinical Medicine
University of the Witwatersrand
7 York Rd, Parktown
Johannesburg, 2193
South Africa
Phone: 27 011 717 2382
Email: khuthala.mabetha@wits.ac.za

Abstract

Background: The use of mobile technologies in fostering health promotion and healthy behaviors is becoming an increasingly common phenomenon in global health programs. Although mobile technologies have been effective in health promotion initiatives and follow-up research in higher-income countries and concerns have been raised within clinical practice and research in low- and middle-income settings, there is a lack of literature that has qualitatively explored the challenges that participants experience in terms of being contactable through mobile technologies.

Objective: This study aims to explore the challenges that participants experience in terms of being contactable through mobile technologies in a trial conducted in Soweto, South Africa.

Methods: A convergent parallel mixed methods research design was used. In the quantitative phase, 363 young women in the age cohorts 18 to 28 years were contacted telephonically between August 2019 and January 2022 to have a session delivered to them or to be booked for a session. Call attempts initiated by the study team were restricted to only 1 call attempt, and participants who were reached at the first call attempt were classified as contactable (189/363, 52.1%), whereas those whom the study team failed to contact were classified as hard to reach (174/363, 47.9%). Two outcomes of interest in the quantitative phase were “contactability of the participants” and “participants’ mobile number changes,” and these outcomes were analyzed at a univariate and bivariate level using descriptive statistics and a 2-way contingency table. In the qualitative phase, a subsample of young women (20 who were part of the trial for ≥12 months) participated in in-depth interviews and were recruited using a convenience sampling method. A reflexive thematic analysis approach was used to analyze the data using MAXQDA software (version 20; VERBI GmbH).

Results: Of the 363 trial participants, 174 (47.9%) were hard to reach telephonically, whereas approximately 189 (52.1%) were easy to reach telephonically. Most participants (133/243, 54.7%) who were contactable did not change their mobile number. The highest percentage of mobile number changes was observed among participants who were hard to reach, with three-quarters of the participants (12/16, 75%) being reported to have changed their mobile number ≥2 times. Eight themes were generated following the analysis of the transcripts, which provided an in-depth account of the reasons why some participants were hard to reach. These included mobile technical issues, coverage issues, lack of ownership of personal cell phones, and unregistered number.

Conclusions: Remote data collection remains an important tool in public health research. It could, thus, serve as a hugely beneficial mechanism in connecting with participants while actively leveraging the established relationships with participants or community-based organizations to deliver health promotion and practice.
Introduction

The use of mobile technologies in engaging people and communities to choose healthy behaviors and to improve their health is becoming an increasingly common phenomenon in global health programs [1-4]. Mobile technologies have increasingly become a common mode of delivering telemedicine in sub-Saharan Africa and other low- and middle-income countries (LMICs) given improvements in coverage [5], with the prospects of improving access to information pertaining to health care services in areas where health care systems lag in meeting minimum health standards [6,7]. Mobile phone technology use in clinical studies is a potential avenue for both the follow-up of participants in research [8-10] and the provision of targeted and tailored messaging to facilitate behavior change support. For instance, patients now progressively receive reminders regarding clinic appointments [11], medication uptake [12], health promotion [13,14], medical treatment or diagnosis [15], and accessing information on disease outbreaks and other health-related data [16] through SMS text messaging systems and telephone calls; all these methods have collectively contributed to better health [1,17]. These digital interventions are reinforced by the high level of mobile phone penetration in many LMICs.

Mobile technologies typically exist in the form of internet-enabled devices such as smartphones, tablets, and watches [18]. Mobile phones constitute the most commonly used mobile technologies for cellular communication in LMICs [19,20] and are used as a cellular communication system for making and receiving phone calls, video calls, text and instant messaging, surfing the internet, GPS navigation, and playing games [21]. For the purposes of this study, the focus was particularly restricted to mobile phones, with the form of cellular communication being restricted only to phone calls.

Research conducted in the sub-Saharan African region has shown that individuals who participate in research studies are intermittently reached through the phone, which is the most commonly used method in the region [22]. Among African countries, South Africa has the third highest number of people who are active mobile phone users, providing a robust platform for telemedicine prospects [22]. For example, 97% of households in South Africa have access to a mobile phone [23]. According to the Independent Communications Authority of South Africa 2019 report on the state of information and communications technology, an estimated 82% of the country’s total population have a smartphone [24], with 88% of the population reported to have prepaid mobile cellular subscriptions and 12% of the population reported to have contract subscriptions [24]. Although mobile technologies have shown some success in their effectiveness in health promotion initiatives and in improving health knowledge, the use of mobile phones in delivering these interventions within clinical practice and research raises some concerns [25]. For example, although a large proportion of people own mobile phones, there are low retention rates in clinical research, and this has been largely attributed to challenges of participant reachability and accessibility by phone [22]. For instance, in a clinical trial conducted in Togo, out of an approximate proportion of 13,726 respondents who had provided their mobile number, 80% could not be reached on that number [26].

In addition, research conducted in Côte d’Ivoire among patients who were receiving antiretroviral treatment showed that out of the 7000 patients who were traced telephonically after initially actively participating in the research but were lost at the point of follow-up in the study, only 40% of the patients could be reached on the telephone number that was provided [27]. A randomized controlled trial conducted in South Africa among 400 HIV-positive patients found that only 60.3% of the patients could be reached telephonically after being contacted a maximum of 3 times via telephone on different days and at different times for every phone number that the participant had provided [22]. In addition, another South African study indicated that reachability by phone is a common challenge in the country. This is caused, for instance, by theft and loss of phones, which results in connectivity loss with respondents and loss to follow-up [28]. Although several scholars have investigated the effectiveness of mobile phones in follow-up studies [1,6-8,13,24,27-29], there is a dearth of literature in South Africa that has qualitatively explored the challenges that participants experience in terms of being contactable through mobile technologies. Gaining insight into these challenges will help in developing targeted strategies to facilitate ways for research staff and researchers to remain in contact with participants over time and for follow-up research. The results could also potentially inform the use of mobile technologies for reaching patients in a clinical setting.

The trial is being conducted in Soweto, South Africa. It is the Bukhali preconception health trial, which is nested in the Healthy Life Trajectories Initiative [30]. The aim of the trial is to optimize the physical and mental health of young women, fostering positive dynamic changes in their health for themselves, and where relevant, for future cohorts. In this trial, young women received intervention materials and resources delivered by community health workers, referred to as “Health Helpers.” These materials include health literacy materials, multimicronutrient supplements, health screening and referral conducted in person, nutritional risk and support, and monthly sessions that help with modifying or transforming health behaviors. The monthly sessions were conducted telephonically by the Health Helpers with on-site visits scheduled every 6 months [30-32]. However, there have been challenges in contacting participants for booking and delivering monthly trial sessions because the participants were unreachable on the contact numbers registered for the study. Therefore, this study aimed to explore the challenges that participants experienced...
with regard to being contactable through mobile technologies, using both quantitative and qualitative data.

**Methods**

**Study Design**

A convergent parallel mixed methods research design was applied. It is a research design that involves the collection and analysis of quantitative and qualitative data separately, which are then compared simultaneously to better understand the phenomenon under study [33]. The main presumption of the research design was that both quantitative and qualitative data offer varying information, with the qualitative data providing narrative accounts of the participants and the quantitative data providing scores on selected variables [33]. This design provides a broader insight into the issue being studied by comparing the findings and results to discern whether they confirm or disconfirm each other. Combining both qualitative and quantitative methods as complementary methods has the potential to contribute to broader applicability of the findings in other settings or with other wider populations [33-36]. This method is applicable in this study given that the quantitative data provided insight into data on many Bukhali trial participants, whereas the qualitative data provided an in-depth narrative account of the mobile phone challenges faced by participants and other reasons for being hard to reach.

**Participants and Recruitment**

The study population consisted of young women aged 18 to 28 years who participated in the trial within the precinct of the Chris Hani Baragwanath Academic Hospital, located in Soweto, Johannesburg, South Africa, and were recruited into the intervention arm of the Bukhali trial [30]. Soweto is a largely overpopulated and multilingual area that lies on the outskirts of the greater Johannesburg city. It is characterized by several socioeconomic issues, namely, joblessness, gender-based violence, and food insecurity, as well as factors hindering healthy behaviors [37], which have the potential to influence the mobile phone use patterns of young women.

Overall, the study team contacted participants telephonically to (1) deliver monthly sessions to the participants, (2) book subsequent monthly sessions for the participants, (3) arrange to have supplements delivered to the participants’ homes, or (4) follow-up on the participants who have experienced an adverse event. The mobile phone information of the participants was captured by the study team to enable the team to reach the participants. For the quantitative component of this study, we included all young women who were contacted telephonically between August 2019 and January 2022. These women were contacted to either have a session delivered to them or to be booked for a session (n=363) out of the wider group of Healthy Life Trajectories Initiative participants. This was done to identify young women who were reachable and those who were unreachable. Participants who had (1) in-person contact with the study team (either off-site to deliver supplements or on-site to deliver a session), (2) contact via SMS, or (3) contacted through other means of communication (eg, home visits or instant messaging) were excluded from this study to focus on the extent to which participants could be contacted by phone.

The standard process used by the intervention team to facilitate the intervention involved making several call attempts to reach the participants. All call attempts and contact statuses of the participants were captured on a contact log. At most, 3 call attempts were made and if the participants were still unreachable, the study team generated tracing lists that prompted them to trace these participants by conducting further fieldwork to update the contact details of the participants. For the purposes of this study, we restricted the call attempts to only 1 call attempt given that the study only aimed to provide a cross-sectional picture and not an analysis that focuses on more longitudinal repeated attempts, which was beyond the scope of this study. Thus, participants who were reached at the first call attempt were classified as “reachable.” Participants whom the study team failed to make contact with at the first call attempt were classified as “hard to reach” or “failed to make contact.” In the same period, some participants changed their mobile number once, some did it several times, and others did not change their number. A flowchart indicating the stages of participant recruitment is shown in Figure 1.
For the qualitative component, a subsample of participants (n=20) who were part of the *Bukhali* trial for ≥12 months were recruited telephonically for those who were easily contactable, and recruitment was performed by a fieldwork team for those who were hard to reach. This sample of participants was purposively selected using a convenience sampling technique. A group of 20 participants was selected by drawing them from the same group of young women who participated in the trial between August 2019 and January 2022. The final group of women who participated in the trial would be booked to receive a monthly session either in person, telephonically, or through SMS or to have supplements delivered to them. We then first restricted this group to only young women whom we failed to make contact with, those who had a session delivered to them, and those who were booked for a session. Furthermore, participants who could not be reached after repeated call attempts were excluded, and we then further restricted the sample to those who only had 1 call attempt. This resulted in a group of those who were contactable and those who were hard to reach, and this group was included in the quantitative analysis. Thereafter, 20 young women who were included in the qualitative analysis were selected from the 363 young women who were included in the quantitative analysis. This was done by selecting the record IDs of young women extracted from REDCap (Research Electronic Data Capture; Vanderbilt University), a secure database that stores data for research studies [38]. Qualitative data were collected only for this group of young women given that the trial aimed to deliver an intervention that fosters behavior change. Thus, the lack of contactability or loss to follow-up of this group of participants poses a substantial threat to the effective delivery of the intervention and could potentially have adverse biases on the conclusiveness of the results of the trial, which could have significant effects on the credibility and reliability of the trial.

**Data Collection**

**Quantitative Data Collection**

Quantitative data collection was conducted, and the data were recorded and stored on REDCap [38]. The data included a set of interview-led administered questionnaires that collected baseline data on the demographic characteristics of the participants. The variables that were used in this study to show a descriptive profile of the sociodemographic characteristics of the participants who were contactable and those who were hard to reach included (1) age, (2) family size, (3) highest level of education, (4) vocational activity, (5) type of dwelling unit, (6) times in the past 12 months when family went hungry, (7) phone ownership, (8) contract type, and (9) type of cell phone.

In this component of the study, the first outcome of interest was the “contactability” of the participants. Two groups of young women were of interest in this phase of analysis. These were (1) young women who were contactable and (2) young women who were hard to reach at the first call attempt. The second outcome was the “mobile number changes of participants,” which were captured on REDCap. These number changes were tracked from randomization to the trial. The number changes were categorized as (1) no number change, (2) 1 number change, or (3) >2 number changes.

**Qualitative Data Collection**

Overall, 20 participants were interviewed individually using an in-depth interview guide by 2 female interviewers who were fluent in both English and other South African vernacular languages. The interview schedule was developed by the study team to capture the participants’ reasons for changing their mobile numbers and to discern why staying in contact with participants was challenging. Interview questions focused on young women’s demographic characteristics, questions around...
ownership of a mobile phone or mobile phones, mobile phone use behavior and patterns, and receipt of health information. The interviews were conducted face-to-face with the participants within the precinct of the Chris Hani Baragwanath Academic Hospital in February and March 2022. The interviews lasted between 30 and 40 minutes and were audio recorded. Interview notes were captured during each scheduled interview session to record the participants’ main accounts and nonverbal cues. Before analysis was conducted, the recordings were transcribed to their original form and translated into English, where necessary.

Data Analysis

Quantitative Data Analysis
To statistically analyze the differences between the 2 groups of participants (failed to make contact and contactable), the data were analyzed descriptively at univariate and bivariate levels. At the univariate level, to visually illustrate the differences between the contact status of participants by mobile number change, a graph was generated to compare the proportion or rate of mobile number change of each group in relation to another. Similarly, another graph was generated to identify the most cited and least cited outcomes of the attempted call for participants we failed to make contact with. At the bivariate level, a contingency table was generated to depict the differences in the results of the tabulation of observations at each level of a variable for both groups, that is, cross-tabulation of the 2 groups to show the relationship or differences in the proportions and frequencies of the sociodemographic characteristics of the 2 groups. This table included the Pearson chi-square test of association to statistically identify the independence or association between the 2 groups or categorical variables and accompanying sociodemographic covariates. All data were analyzed using STATA (version 17; StataCorp) [39].

Qualitative Data Analysis
A reflexive thematic analysis approach was used for analysis using MAXQDA software (version 20; VERBI GmbH) [36], which helped in the recording, coding, and interpretation of the transcripts. Reflexive thematic analysis is an interpretive method that enables researchers to establish the outcome of the work rather than following a specific theory [35]. Six steps were used in the analysis. In the first step, data familiarization was conducted by checking the quality of the transcripts against the recordings to understand the data thoroughly, which would make it possible to search for patterns and meanings. This involved reading and rereading transcripts and taking notes. In the second step, initial codes were generated by labeling and organizing participants’ narratives to form a complete meaning. In the third phase, initial themes were generated by sorting initial codes that had been generated into themes and identifying meanings and relationships between the initial codes. In the fourth phase, the themes were reviewed by identifying coherent patterns and ensuring that the generation of themes was supported by sufficient data. This also included collapsing overlapping themes and regenerating and improving the codes and themes. This was done as a collective effort among KM, LMS, and CED. In the fifth phase, themes were defined and named by linking each narrative to an appropriate theme. Finally, in the sixth phase, the narratives of the participants were presented under each generated theme in a concise manner.

Ethical Considerations
Ethics approval was granted by the Human Research Ethics Committee (Medical) located at the University of the Witwatersrand (M190449). All participants provided written informed consent to participate and consented to have the interviews audio recorded.

Although the information provided by the participants was captured on an audio recorder, the data were protected by securely storing them in a password-protected computer in a secure locked cabinet. The identities of the participants were protected through the deidentification of all their personal information and characteristics. The transcripts contained ID numbers that represented each participant’s response.

Results

Quantitative Results
The sociodemographic characteristics of the participants (N=363) are presented in Table 1. Of the 363 participants, 174 (47.9%) were hard to reach telephonically, whereas 189 (52.1%) were easy to reach telephonically. Most participants (133/243, 54.7%) who were contactable did not change their mobile number. The highest percentage of mobile number changes was observed among participants who were difficult to reach, with three-quarters of the participants (12/16, 75%) being reported to have changed their mobile number ≥2 times (Multimedia Appendix 1). Although 58% (101/174) of the participants could not be reached as the attempted call went to voicemail, 29.3% (51/174) of the participants were reported to have taken the call when the attempted call was placed, and only 2.2% (4/174) of the participants took the call but ended it (Multimedia Appendix 2).

Table 2 shows the mobile phone information of both hard-to-reach participants and contactable participants. There was a significant difference in mobile phone ownership among participants who were hard to reach and those who were contactable, as 96.8% (183/189; P=.03) of those who were contactable owned a personal cell phone compared with 91.4% (159/174) of those who were hard to reach. Most of the young women (both contactable and hard to reach) had a smartphone (297/342; ≥80% each; P=.02) and did not have a phone contract (324/342; >90% each). Results indicated no statistically significant differences in these variables among participants who were hard to reach and participants who were contactable.
Table 1. Sociodemographic characteristics of the trial participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Failed to make contact (n=174), n (%)</th>
<th>Contactable (n=189), n (%)</th>
<th>Total (N=363), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-20</td>
<td>70 (40.2)</td>
<td>59 (31.2)</td>
<td>129 (35.5)</td>
<td>.19</td>
</tr>
<tr>
<td>21-23</td>
<td>48 (27.6)</td>
<td>64 (33.9)</td>
<td>112 (30.8)</td>
<td></td>
</tr>
<tr>
<td>24-26</td>
<td>38 (21.8)</td>
<td>51 (27)</td>
<td>89 (24.5)</td>
<td></td>
</tr>
<tr>
<td>27-28</td>
<td>18 (10.3)</td>
<td>15 (7.9)</td>
<td>33 (9.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Family size</strong></td>
<td></td>
<td></td>
<td></td>
<td>.78</td>
</tr>
<tr>
<td>1</td>
<td>8 (4.6)</td>
<td>13 (6.9)</td>
<td>21 (5.8)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13 (7.5)</td>
<td>17 (9)</td>
<td>30 (8.3)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25 (14.4)</td>
<td>23 (12.2)</td>
<td>48 (13.2)</td>
<td></td>
</tr>
<tr>
<td>4-7</td>
<td>95 (54.6)</td>
<td>105 (55.6)</td>
<td>200 (55.1)</td>
<td></td>
</tr>
<tr>
<td>&gt;7</td>
<td>33 (19)</td>
<td>31 (16.4)</td>
<td>64 (17.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Education and socioeconomic characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td><strong>Highest level of education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>97 (27.3)</td>
<td>42 (22.3)</td>
<td>99 (27.3)</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>91 (52.3)</td>
<td>118 (62.8)</td>
<td>209 (57.7)</td>
<td></td>
</tr>
<tr>
<td>Tertiary</td>
<td>8 (4.6)</td>
<td>13 (6.9)</td>
<td>21 (5.8)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>18 (10.3)</td>
<td>15 (8)</td>
<td>33 (9.1)</td>
<td></td>
</tr>
<tr>
<td>Missing&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0 (0)</td>
<td>1 (0.5)</td>
<td>1 (0.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Vocational activity</strong></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Currently enrolled in higher education institution, employed, or currently looking for a job</td>
<td>9 (5.2)</td>
<td>19 (10)</td>
<td>28 (7.7)</td>
<td></td>
</tr>
<tr>
<td>Not enrolled in higher education institution, not employed, or not currently looking for a job</td>
<td>165 (94.8)</td>
<td>170 (89.9)</td>
<td>334 (92.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Type of dwelling unit</strong></td>
<td></td>
<td></td>
<td></td>
<td>.43</td>
</tr>
<tr>
<td>House of brick or concrete block structure on a separate stand or yard or on a farm</td>
<td>122 (70.1)</td>
<td>128 (67.7)</td>
<td>250 (68.9)</td>
<td></td>
</tr>
<tr>
<td>House, flat, or room on your homestead</td>
<td>24 (13.8)</td>
<td>35 (18.5)</td>
<td>59 (16.2)</td>
<td></td>
</tr>
<tr>
<td>Informal dwelling or shack in backyard</td>
<td>15 (8.6)</td>
<td>14 (7.4)</td>
<td>29 (8)</td>
<td></td>
</tr>
<tr>
<td>Informal dwelling or shack not in backyard</td>
<td>7 (4)</td>
<td>3 (1.6)</td>
<td>10 (2.7)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>6 (3.4)</td>
<td>9 (4.8)</td>
<td>15 (4.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Were there times in past 12 months when members of your family went hungry because there was not enough food in the house to eat</strong></td>
<td></td>
<td></td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Yes</td>
<td>68 (39.1)</td>
<td>59 (31.2)</td>
<td>127 (35)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>106 (60.9)</td>
<td>130 (68.8)</td>
<td>236 (65)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Missing refers to missing data or missing values that have not been stored for the education level of some participants (could be attributed to incomplete data entry or no information provided by participants).
<table>
<thead>
<tr>
<th>Failed to make contact, n (%)</th>
<th>Contactable, n (%)</th>
<th>Total, n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Do you have a personal cell phone?</strong></td>
<td></td>
<td></td>
<td>.03 a</td>
</tr>
<tr>
<td>Yes</td>
<td>159 (91.4)</td>
<td>183 (96.8)</td>
<td>342 (94.2)</td>
</tr>
<tr>
<td>No</td>
<td>25 (8.6)</td>
<td>6 (3.2)</td>
<td>21 (5.8)</td>
</tr>
<tr>
<td>Total</td>
<td>174 (100)</td>
<td>189 (100)</td>
<td>363 (100)</td>
</tr>
<tr>
<td><strong>If yes, is it on contract?</strong></td>
<td></td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>Yes</td>
<td>11 (6.9)</td>
<td>7 (3.8)</td>
<td>18 (5.3)</td>
</tr>
<tr>
<td>No</td>
<td>148 (93.1)</td>
<td>176 (96.2)</td>
<td>324 (94.7)</td>
</tr>
<tr>
<td>Total</td>
<td>159 (100)</td>
<td>183 (100)</td>
<td>342 (100)</td>
</tr>
<tr>
<td><strong>Is it a smartphone?</strong></td>
<td></td>
<td></td>
<td>.32</td>
</tr>
<tr>
<td>Yes</td>
<td>135 (84.9)</td>
<td>162 (88.5)</td>
<td>297 (86.8)</td>
</tr>
<tr>
<td>No</td>
<td>24 (15.1)</td>
<td>21 (11.5)</td>
<td>45 (13.2)</td>
</tr>
<tr>
<td>Total</td>
<td>159 (100)</td>
<td>183 (100)</td>
<td>342 (100)</td>
</tr>
</tbody>
</table>

*P<.05 (indicates significant association between participants’ ownership of a mobile phone and contact status).*

**Qualitative Results**

**Overview**

Table 3 shows the sociodemographic profile of the interview participants. The mean age of the participants who were interviewed was 22 (SD 2.94) years. In addition, 50% (10/20) of the participants had a secondary school qualification, 40% (8/20) had not completed secondary education, and only 10% (2/20) had a tertiary qualification at the time of the assessment. Almost all the participants (19/20, 95%) lived with a parent or relative, whereas only 5% (1/20) of the participants lived with her partner and 65% (13/20) of the participants were not in a relationship at the time of assessment.

Textbox 1 presents an overview of the superordinate themes and subthemes generated from the qualitative analysis. As presented in Textbox 1, this section is sorted into the narrative findings, with illustrative quotations provided for each superordinate theme and subtheme.
Table 3. Sociodemographic profile of the interview participants.

<table>
<thead>
<tr>
<th>Respondent number</th>
<th>Current age (years)</th>
<th>Highest level of education attained</th>
<th>Lives with</th>
<th>Relationship status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>25</td>
<td>Secondary school qualification</td>
<td>Siblings and child</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 2</td>
<td>22</td>
<td>Tertiary qualification</td>
<td>Parent</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 3</td>
<td>21</td>
<td>Secondary school qualification</td>
<td>Parents, child, and sibling</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 4</td>
<td>22</td>
<td>Grade 11</td>
<td>Parent, siblings, and child</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 5</td>
<td>28</td>
<td>Grade 11</td>
<td>Parent, sibling, children, and relatives</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 6</td>
<td>20</td>
<td>Grade 11</td>
<td>Parent and siblings</td>
<td>In a relationship</td>
</tr>
<tr>
<td>Participant 7</td>
<td>21</td>
<td>Secondary school qualification</td>
<td>Parent</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 8</td>
<td>26</td>
<td>Grade 11</td>
<td>Partner</td>
<td>In a relationship</td>
</tr>
<tr>
<td>Participant 9</td>
<td>20</td>
<td>Secondary school qualification</td>
<td>Parent, siblings, and relatives</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 10</td>
<td>24</td>
<td>Grade 10</td>
<td>Grandparent, siblings, and children</td>
<td>In a relationship</td>
</tr>
<tr>
<td>Participant 11</td>
<td>22</td>
<td>Grade 11</td>
<td>Grandparent and relatives</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 12</td>
<td>24</td>
<td>Grade 11</td>
<td>Parent, siblings, and relative</td>
<td>In a relationship</td>
</tr>
<tr>
<td>Participant 13</td>
<td>20</td>
<td>Secondary school qualification</td>
<td>Parent, sibling, and child</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 14</td>
<td>21</td>
<td>Currently doing matric</td>
<td>Sibling, parent, and child</td>
<td>In a relationship</td>
</tr>
<tr>
<td>Participant 15</td>
<td>25</td>
<td>Secondary school qualification</td>
<td>Relatives</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 16</td>
<td>20</td>
<td>Secondary school qualification</td>
<td>Parent, siblings, and child</td>
<td>In a relationship</td>
</tr>
<tr>
<td>Participant 17</td>
<td>20</td>
<td>Secondary school qualification</td>
<td>Siblings</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 18</td>
<td>22</td>
<td>Secondary school qualification</td>
<td>Parents</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 19</td>
<td>22</td>
<td>Tertiary qualification</td>
<td>Grandparents and relatives</td>
<td>Single</td>
</tr>
<tr>
<td>Participant 20</td>
<td>24</td>
<td>Secondary school qualification</td>
<td>Sibling, relatives, and child</td>
<td>In a relationship</td>
</tr>
</tbody>
</table>

Textbox 1. Themes focusing on young women’s reasons for being hard to reach.

Superordinate theme and subthemes
- Mobile phone technical issues
  - Coverage issues
    - Electrical unreliability
    - Network connectivity issues
  - Lack of ownership of personal cell phone
  - Unregistered number
- Use patterns
  - Cell phone theft in the community
  - Stalking behavior
- Inconsistent SIM use
- Availability of data and airtime
- Lack of interest in possessing a mobile phone

Mobile Phone Technical Issues
Technical barriers were reported by most participants as a major challenge that affected the efficiency, functioning, and usability of their mobile phones. Key challenges reported by the participants that resulted in inherent limitations in operating their mobile phones included poor battery life, faulty charging system, mobile phone and app crashes, and an inadequate touch screen response. Such challenges resulted in the participants not being easily contactable or contactable at all. Some participants resorted to changing their mobile phones or using alternative communication methods such as using the phones of individuals within their social networks:
I’ve experienced battery challenges and the charging system, ja that’s the only thing that’s been wrong with my phone. I had to charge my phone a certain way and I couldn’t even use it until I finished charging and the battery will finish fast sometimes when I’m trying to do something. [Participant 2, 25 years old, single]

The phone I was using, it lived on life support. I had to use it while it was in a charger at home. The charger also was faulty so sometimes I would think it’s on and it’s off. [Participant 9, 20 years old, single]

It sometimes jams. The battery finishes quickly while you are chatting. So those are the challenges that I face with my phone. [Participant 16, 20 years old, in a relationship]

My phone has problems. Like it just switches off, because it fell once, and now every time it hits hard or something, it just switches off, and it is going to take time for it to switch on, so I wait for it to switch on. [Participant 1, 25 years old, single]

The touch screen doesn’t work properly now because it fell sometime last year. I don’t know if it’s a space issue or just the phone itself because sometimes when I get incoming calls the phone rings but then I can’t touch it for me to be able to pick up the call because of the touch screen problems with the phone. [Participant 13, 20 years old, single]

It has a tendency to freeze if it fell, maybe if I’m receiving a call, I can’t answer it because the phone is frozen, you know. [Participant 15, 25 years old, single]

**Coverage Issues**

Coverage issues such as network and electricity made it difficult for some participants to be contactable. Participants indicated that they could not receive calls or messages or had to change their mobile SIMs because of challenges concerning the connectivity, reliability, and quality of the mobile network. This included signal weakness and network availability issues with some mobile service providers being cited as having general network unreliability or the network being offline in some instances. Participants reported experiencing unreliability of the electric supply owing to frequent power cuts. This greatly affected their ability to use their mobile phones, thus being unable to be in touch with people:

My network becomes very slow, sometimes I don’t even get calls I just get messages that I have a missed call but then it didn’t even ring if you don’t have electricity. It’s been hectic hey, we didn’t have electricity on Sunday. [Participant 2, 22 years old, single]

It has been a while since we have been without electricity but right now, we don’t have it because they said that the substation was on fire. [Participant 7, 21 years old, single]

I first bought the Telkom sim card but then I don’t know what happened with the coverage in Orlando, but it was giving me problems, so I couldn’t get network, so I used my MTN sim card. I think I got it like two years ago. I was not using it as frequently as the Telkom number, so I had to use the MTN number. [Participant 9, 20 years old, single]

Telkom network is problematic, sometimes you want to make a call, it will tell you that there is no service, things like that. Sometimes the sim card does not connect. I will find myself taking it out and putting it back on but still, so I decided to change MTN, then, MTN gave me a problem of taking my airtime every time when I recharged. [Participant 1, 25 years old, single]

**Lack of Ownership of Personal Cell Phone**

Some participants indicated that they were not easily contactable because they do not currently own their own personal cell phone or previously did not own one. The reasons attributed to the lack of ownership ranged from the phone being reported to be broken, stolen, lost, or not having one at all. Despite this, participants used alternative means of communication to try and be contactable. Participants reported using their social networks to try and solve the problem of not having a mobile phone by sharing a phone with their siblings, peers, or partners. However, sharing a phone limited their ability to stay connected with others and to be reachable because their time with the device was limited. In some cases, this was exacerbated by the participants not living close to the people they shared the phone with. In other instances, participants’ social networks would not relay the message back to them. These narratives are reflected in the following excerpts:

I didn’t have a phone for two months, so I spent quite a while without a phone. I used parents’ phones. It was for when people couldn’t get hold of me, they would call my parents for emergencies. [Participant 18, 22 years old, single]

I didn’t have a phone for four months. I was not reachable. I used my boyfriend’s phone. The first number that I applied with is my grandmothers, that was the first one. I did not have a phone at that time, and so I applied with her phone, and they tried to call her, but it did not get through to her. Because her phone, she is a person that does business, it would switch off and she would not find the charger, and then it would be charged after a while, and then I gave them the number of my partner, and then they said that they can’t get me with that one as well. [Participant 10, 24 years old, in a relationship]

I stayed for a long time without a phone. I think it had been two years without having a phone. I used to borrow my friend’s phone if for instance there is someone who had asked for my number, then I would give them my friend’s number, and my friend would tell me that so and so had phoned, they called me because you don’t have a phone. So, when people wanted my number, I would give them my friend’s number. It was not easy for them to get hold of me.
because sometimes I am at my home, and she is at her home. [Participant 1, 25 years old, single]

Other participants further highlighted that they were not contactable because their mobile phone was stolen or had gone missing:

I lost my phone and then I took time to do the sim swap and by the time I decided to do it, they had already given the number to someone else. [Participant 9, 20 years old, single]

I started by using said provider, yes I was using said provider first, and now I am on another provider because my phones got lost. [Participant 11, 22 years old, single]

I had a phone, it was last year, I had another phone and they mugged me in March. [Participant 20, 24 years old, in a relationship]

One participant lost her mobile phone due to theft, but her mobile SIM was still in her possession. As an alternative means of communication, she would insert her SIM card in other people’s phones and reach out to people given that she did not have the financial means to buy a new phone:

I was going to the mall and then I was mugged but then I have a sim card which I did a sim swap, yes so every time I need a phone, I can contact people I wasn’t able to contact. [Participant 12, 24 years old, in a relationship]

**Unregistered Number**

Another factor contributing to the difficulty in reaching participants is the unregistered status of their mobile numbers. Some participants reported that they had lost their phones with the mobile SIM card. Two participants reported that they bought new phones and SIM cards and went to a mobile phone store to get a SIM swap, but the mobile phone store could not perform the SIM swap. One reason for this was because their previous mobile numbers were no longer registered under the participants’ names and had been reassigned to new owners. Another reason was the inability to recall the last 5 numbers called on the SIM, which is a requirement for performing a SIM swap. Consequently, participants had to purchase new SIM cards with new mobile numbers, different from those recorded by the trial:

I had to go this year January to Telkom to do a sim swap because I had applied for jobs with that number. Then when I got to Telkom, they told me that that number was not registered under my name, you see these sim cards that you get on the streets. So, I had to start another sim and so I started afresh with that number. [Participant 17, 20 years old, single]

They took my number in January and then I got this one in February. [Participant 11, 22 years old, single]

I lost the phone and then when I went to do the sim swap, they told me I couldn’t do it because the numbers were wrong, and that I wanted the last five numbers that I last called, and so I was not able to continue with the sim swap. [Participant 7, 21 years old, single]

**Use Patterns**

Although many participants owned a personal cell phone, an inconsistent pattern of mobile phone use was observed in the narratives of the participants, which impacted the ability to contact them. Some participants reported that they did not carry their mobile phones outdoors and only used the device at home. However, the reasons for this behavior were not reported by the participants:

I don’t use it when I am on the road, even when I am going to town or to the mall, I usually leave it behind. [Participant 3, 21 years old, single]

When I go outside, I don’t take it with me, I leave it at home. I use it when I know that I am not going anywhere. [Participant 11, 22 years old, single]

Sometimes I’m at school, and sometimes I’m busy and I left it at home and went to the shops, ja I’m not on my phone every time, I just use my phone only when I do important things. ja. [Participant 14, 21 years old, in a relationship]

I couldn’t go out with it, like I had to use it at home. When I go out, I leave it at home so I had to tell the person that I am meeting that we are meeting at a specific time and place, and I just had to hope they will be there. [Participant 9, 20 years old, single]

Conversely, some participants reported that their communities were not safe and there was a high rate of mobile phone theft in their area. This prompted them to not use their mobile phones in public. This has contributed to participants not being easily contactable given that they use their mobile phones inconsistently:

I don’t carry my phone in the streets. It’s not safe because most of the time there’s people taking other people’s phones, most of the time, like maybe in a week we’ll hear about 2 people, 3, ja so it’s not safe. [Participant 2, 22 years old, single]

I use when I am at home alone, outside there are a lot of thieves, when you are outside and you are using your phone, they take it from you. [Participant 10, 24 years old, in a relationship]

Conversely, another participant indicated that she was no longer comfortable using her phone or taking calls as she had experienced continuous threatening harassment over the phone from an unknown individual. This thus resulted in other people finding it difficult to reach her:

There were people walking me, and I didn’t know where they were getting my number from. They would tell me where I was, meanwhile, I don’t know them and I didn’t see them, then I decided to change my number. [Participant 3, 21 years old, single]

**Inconsistent SIM Use**

For other participants, their inconsistent mobile phone behavior was largely because of the participants owning dual SIM cards.
Participants reported that they often used one mobile number more than the other, whereas another participant indicated that she had to change her SIM cards frequently given some technical issues she was experiencing with the card. Similarly, another participant reported changing her SIM cards frequently because she would lose 1 SIM because of repeatedly changing mobile devices and thereafter had to replace it with another SIM card. Another participant’s inconsistent use of her mobile phone was largely because of her younger sibling who did not own his own personal cell phone. Given this circumstance, she shared her phone with her younger sibling as he frequently took out the participant’s SIM card and put his own for a certain period. This greatly affected the participant’s ability to receive calls or messages, thus making her seemingly unreachable. Other participants indicated that the procedures for performing a SIM swap are often offline every time they went to do one. This has led to the participants’ losing interest in doing a SIM swap and has prompted them to change their mobile number. All these factors have thus contributed greatly to the participants being unreachable:

The sim card used to get lost because I kept changing the phones. I was using the other network’s * number. The 060. So now I am using another network. I don’t know the other network’s number, but this one. [Participant 10, 24 years old, in a relationship]

My younger brother likes using it a lot because he likes logging in on Facebook and chat with his friends and then bring it back. He takes out my starter pack and puts in his own. Maybe they call me, and it goes to voicemail, that is because my sim card is out. [Participant 1, 25 years old, single]

I have two numbers and there is one that I always use and the other one that I use sometimes, because I have two sim cards, and both are in the phone. So sometimes I give people the number that I don’t use every day. [Participant 16, 20 years old, in a relationship]

I used to use two starter packs because I wanted to use Telkom and MTN, and at the time I was using Cell C as well. I had a lot of starter packs, and I changed the numbers a lot. Telkom was giving me problems, and then MTN used to take airtime every time when I recharged. I didn’t know where my airtime was going every time when I recharged the number, so yeah, I used to change the number a lot. [Participant 1, 25 years old, in a relationship]

Because when I do a sim swap, they ask a lot of questions, so I sometimes find that it is better if I change the number. Yes because this time I had to do a sim swap, and they said I should go to Shoprite with my ID and my proof of address, when I got to Shoprite, they said that their machines were offline, and then that is where I gave up, and then I said it is better that I change my numbers, because the numbers of most of the people that I know, I have them, I have written them in my book. [Participant 11, 22 years old, single]

You know what, I was tired of changing number so I decided that you know what, let me stick to this number because a lot of people have this number, and they call me on it, when I give out my CVs, this is the number that I use, so I don’t take it out, this is the one that I am using. [Participant 17, 20 years old, single]

Interestingly, 1 contrary finding from 1 participant’s narrative showed that the participant had the same number for >10 years and decided not to change her number to ensure that she remained contactable:

I’ve had this number for more than 10 years I think, the reason why it was more it made sense to do a sim swap instead of changing numbers it was for convenience and with this phone number I had registered it on a lot of platforms, you know when you apply online when they ask you to leave your details you leave your number, so the first thing that came to mind is that for people to be able to get hold of me it’d better to have the same number and also because had lost contacts as well for me to get back my contacts I had to have the number I was using. [Participant 15, 25 years old, single]

Availability of Data and Airtime

Other participants reported that a major challenge that they face, which has resulted in them not being able to stay in contact with people or being difficult to reach, is that airtime and data costs are very steep, and they lack the financial ability to afford data and airtime and thus cannot purchase it regularly. As a result, there are times when they do not have any airtime or data on their phone and thus cannot make or return any calls or messages received:

When I don’t have money where I am, or I don’t have money at that time. It depends on whether I have data at the time. If in a week my data finishes on Wednesday, then I will buy it on Sunday. There are days when I don’t have data. Maybe for a short while, for maybe two or one week. [Participant 16, 20 years old, in a relationship]

I was supposed to attend the interview today and I was unable to go, because I did not have data and I saw the message on a Friday. [Participant 3, 21 years old, single]

Before I bought a new phone, I used Vodacom, but I just didn’t like it because their data and airtime are expensive. So, the other phone came with an MTN sim, and I carried on with that one. [Participant 18, 22 years old, single]

Lack of Interest in Possessing a Mobile Phone

Some participants reported no interest in possessing a mobile device to stay in contact with their social networks or being contactable on the phone. One of the common reasons for this was that the participants were often in the vicinity of their social networks and could communicate in person as opposed to communicating via mobile phone. Others mentioned that they do not have friends with whom to communicate, whereas others
hardly used their phones and often misplaced it given the lack of importance they give to owning a phone:

Not that important, it’s not that important in fact. Because I don’t spend a lot of time on the phone, so I wouldn’t say that it is important. Eh, they can contact me during the day but sometimes I don’t answer the phone. I think I don’t like it that much to have it all the time, I feel like I can stay without a phone. [Participant 16, 20 years old, in a relationship]

It’s not that important. It’s just to keep boredom off. I don’t care much about the phone. [Participant 10, 24 years old, in a relationship]

I don’t spend a lot of time on the phone, so I wouldn’t say that it is important. I think I don’t like it that much to have it all the time, I feel like I can stay without a phone. [Participant 16, 20 years old, in a relationship]

I am very reckless with my phone. I don’t use my phone often so I just leave it on the table or the cupboard or wherever and then I leave the house and people will call me and not reach me until I started to realise that okay, I need to send my CV somewhere and a phone is necessary to have on you. [Participant 5, 28 years old, single]

One participant indicated that she had previously lost her phone before purchasing a new one months later. The amount of time she spent without a phone resulted in her losing interest in possessing a phone as she had grown accustomed to being a nonmobile user:

I didn’t see the importance of having a phone. I stayed a long time, like I didn’t care about the phone. I stayed for a long time without it, and I didn’t feel like I didn’t have a phone. So, I didn’t see the importance of having a phone, even though I saw other people with phones, but for me, I never had that feeling of not having a phone. I had already gotten used to not having a phone. [Participant 12, 24 years old, in a relationship]

In contrast to these findings, other participants indicated that possessing a mobile phone was important to them and helped them to stay in touch with their social networks, to keep abreast of any job opportunities that may arise, to apply to university or college, and to access emails and the internet. These findings are reflected in the following excerpts:

It’s very important because they need to find me every time I’m needed and most things I do on the phone, so I view it most times. I use my cell phone a lot, like every day. [Participant 2, 22 years old, single]

I use my phone all the time like 10 hours. When I wake up, before I go to bed, and I always provide the alternative number in case my phone is maybe low, or I have lost it. I always put in an alternative number. [Participant 9, 20 years old, single]

It’s important because even the study get hold of me on that phone. When people are looking for me, they get hold of me on that phone. Even now the study helped me because if I didn’t have a phone, I would not be reachable because it has been a while since I have been here. [Participant 19, 22 years old, single]

Discussion

Principal Findings

This study aimed to explore the challenges that participants experience in terms of being contactable through mobile phones in a trial conducted in Soweto, South Africa. This study highlighted that although most participants who are difficult to contact do own a personal cell phone, which is a smartphone, most of them face a myriad of interrelated mobile phone challenges, particularly technical and coverage issues, mobile theft in their communities, and high costs of data and airtime. Consequently, these challenges contribute to participants’ need to change their mobile numbers, which subsequently has the potential to affect their reachability. The novelty of this study was the use of a mixed methods research design. It was important to conduct a study of this nature within the context of Soweto, a resource-constrained setting, which is characterized by a short supply of free public Wi-Fi, unequal mobile coverage, and comparably high data costs [40], among other factors.

This study was descriptive in nature. Therefore, caution should be exercised when interpreting the results of this study as no significant associations were found in the quantitative analysis. The quantitative results of this study showed that a large proportion of individuals owned a mobile phone. This result is in line with previous literature, which has shown that mobile connectivity in South Africa has grown rapidly with a mobile penetration rate of >95%, with 91% of all phones being smartphones [1,17,24]. Although this was also corroborated by the qualitative findings, a key finding was that some participants were mobile phone sharers, whereas others had no phones at all and thus relied on their social networks who had phones and often depended on using these phones for a short while to be in touch with other individuals. Previous literature has shown that the use of mobile phones is an inherent part of people’s lives, which exists in socioeconomic practices and realities [15]. Although South Africa has seen an exponential increase in mobile phone ownership and access, the high cost of mobile phones still represents a formidable barrier that limits mobile phone use [40]. In fact, the lack of income has been reported as the primary factor in the lack of phone ownership in many LMICs [41,42]. This largely surpasses other probable barriers such as the perceived importance of using a mobile phone and technical literacy [41,42].

Given that this study aimed to explore the challenges that participants experience in terms of being contactable through mobile phones, overall, the quantitative and qualitative findings in this study suggest that the socioeconomic context of the participants is a major factor that has contributed to their mobile use behavior and difficulty in reaching them. Thus, the findings of this study uniquely contribute to the existing body of knowledge in various ways. First, although phone ownership has increased rapidly in South Africa, there are some important use gaps that remain pronounced, and these are evident in an
LMIC such as Soweto. To illustrate this, although phone ownership may be predominantly high among both groups included in this study, their socioeconomic backgrounds are complex and create different realities, which subsequently leads to variability in mobile phone use and usability. Thus, daily realities such as unemployment, poverty, and infrastructural barriers have the potential to prevent them from fully using their devices in a sustainable manner. In addition, the challenges that were explored in this study in terms of participants being contactable through the mobile phone show that the experiences of youth are vastly different, and complexities around socioeconomic backgrounds and environmental settings create different realities pertaining to phone possession and use. Although the participants may reside in one central area, the circumstances of young people (phone sharers, phone owners, and non–phone owners) differ quite substantially, not only depending on their age, level of education, or vocational activity but also on factors such as socioeconomic circumstances, family structure, household resources, and livelihood patterns. This aligns with the evidence from the World Bank, which has shown that 50% of South African inhabitants who reside in urban settings live in low-income settings, accounting for approximately 40% of the individuals who constitute the working population and catering to approximately 60% of nonworking citizens who live in cramped family settings [29].

These previous findings align with the qualitative results of this study, which demonstrated that most participants were individuals who were still in early young adulthood (18-25 years), with >90% of them Not in Education, Employment, or Training and only 5% had a mobile phone contract. Previous studies have also indicated a digital divide along gender lines [43]. The study indicated that women are more at risk of poor mobile phone use compared with men and attributed it to economic inequality, namely, women’s lower educational attainment and poor income levels [43]. These findings illuminate the way mobile phone use is interwoven with the daily realities and experiences of the participants, which has the potential to prevent the participants from fully using their devices in a sustainable manner.

Existing research has further shown that individuals who earn a low income have limited connectivity, which is inhibited by disparities in wealth, resources, and other opportunities, presenting a challenge in terms of using mobile technologies [44]. This finding is strongly congruent with the qualitative findings, which revealed that participants faced a myriad of mobile phone challenges, leading to them being hard to reach, which is in line with what the study aimed to explore (challenges experienced in terms of being contactable over the mobile phone). Such structural inequalities that make it hard for participants to be easy to reach include coverage issues, high costs of data, and airtime and safety in their communities. Furthermore, the quantitative results of this study showed that in addition to the fact that most participants only have a secondary school qualification and are Not in Education, Employment, or Training, approximately one-third of these participants have been reported to live in households that are food insecure. Household constraints serve as important factors for people’s access to mobile technology [45]. Another study showed that ownership of a mobile phone necessitates the need to have a phone that is in good working condition (available data and airtime, functional battery, and network signal) and the financial freedom to sustain this without giving up other necessities [46]. All these factors are unevenly distributed, with those living in resource-constrained settings being the most disadvantaged [46]. These findings thus suggest that the mobile phone challenges experienced by the participants are interwoven with their socioeconomic backgrounds, thus making it difficult to maintain mobile technology use.

This study further found that the proportion of those who own phones decreased with each successive age category for both participants who were hard to reach and those who were contactable. These findings are supported by the previous literature, which showed that age is a social determinant of digital inequality [22]. A curvilinear relationship was found between age and digital use in LMICs [22]. For instance, 1 study conducted in India showed that phone ownership was >66% for participants aged 15 to 24 years but only 55% for those aged 25 to 44 years [41]. In addition, a study conducted in Tanzania found that women aged ≤30 years were 17% less likely to remain phone owners (Roessler, P, unpublished data, December 2018).

Mobile technologies have the potential to improve health behaviors among individuals participating in clinical research [38]. For instance, the COVID-19 pandemic has led to an increased need to use mobile technologies for health promotion because of social distancing and nationwide lockdowns [47]. This reliance on mobile technologies means that challenges in reaching these participants could disrupt the continuity of health promotion for both research and clinical practice. Participants could, for example, seek health advice outside of their health care system or not attend their regular health sessions [39] owing to unresolved concerns via telephone. This may thus lead to clinical interventions being highly inefficient. Thus, using mobile phones in any clinical study can inform how these devices could be used in future interventional research. Although this study found that the mobile challenges that participants face are mostly influenced by structural inequalities, it is pivotal to note that addressing and alleviating these structural inequalities is beyond the scope of this study.

However, several feasible solutions exist for the problems encountered in this study. To maximize engagement in mobile health studies, health promotion initiatives can be delivered via in-person consultations or sessions with the participants. These in-person consultations would afford participants the opportunity to have increased personal contact with the study team and to obtain study support and personalized study feedback. This strategy has proved to be successful in a previous clinical trial where personalized care, including allowing participants to share their personal problems and enabling participants to increase contact with the study team or investigators, helped in maximizing engagement [48]. This could include fieldwork to complement the use of mobile technologies and thus ensure participant retention. Furthermore, advisory groups could be created by bringing together the participants and giving them an opportunity to engage in discussions around crafting possible solutions to the mobile technology barriers that were identified.
in this study. In addition, automated health promotion messages could be relayed to the participants’ phones, particularly in instances where calls do go through but are not answered by participants.

**Strengths and Limitations**

The strength of this study was the use of a convergent parallel mixed methods approach, which assisted in exploring the quantitative findings by relating the findings obtained from the qualitative and quantitative analyses. The use of this method thus helped in providing comprehensive insight into the mobile phone behaviors of young women and the factors that contributed to this behavior. A limitation of this study was that the quantitative analysis was only restricted to the first instance (ie, the first call attempt in which some young women were hard to reach and some were easily contactable). Therefore, repeat instances or call attempts for subsequent monthly sessions were excluded. This could have provided a broader longitudinal picture of the time from randomization to infancy of all call attempts and whether any demographic characteristics were associated with these changes. However, this was beyond the scope of this study as the focus was only on providing a descriptive profile of women who were hard to reach while comparing their demographic profile with those who were easily contactable.

**Conclusions**

Despite the availability of mobile technology and network accessibility, there are substantial economic barriers that prevent young adults from fully benefitting from technology. These barriers consequently affect health promotion and behavioral changes. For individuals to remain digitally connected, multiple strategies should be used. For research, remote data collection remains an important tool in public health research and could thus serve as a hugely beneficial mechanism in connecting with participants while actively using established relationships with participants or community-based organizations to deliver health promotion and practice. With increasing phone ownership in LMICs including South Africa, greater accessibility to less expensive or free data networks or digital platforms is needed universally across South Africa to ensure more equitable access to such technologies.

**Acknowledgments**

This study was supported by the South African Medical Research Council and the Canadian Institutes of Health Research. SAN was supported by the Department of Science and Innovation-National Research Foundation Centre of Excellence in Human Development.

**Data Availability**

The data sets generated and analyzed during this study are not publicly available because the trial is still ongoing but are available from the corresponding author on reasonable request.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Contact status of the trial participants by mobile number change.  
[**PNG File, 121 KB** - formative_v811e48144_app1.png]

**Multimedia Appendix 2**

Outcome of the call attempted for participants who were unreachable.  
[**PNG File, 101 KB** - formative_v811e48144_app2.png]

**References**


https://formative.jmir.org/2024/1/e48144


42. Faith B. Maintenance affordances and structural inequalities: mobile phone use by low-income women in the United Kingdom. Inf Technol Int Dev 2018;66-80 [FREE Full text]


**Abbreviations**

LMIC: low- and middle-income country
Abstract

Background: In oncohematology, both the development of the disease and the side effects of antineoplastic treatment often take a toll on patients’ physical and nutritional well-being. In this era of digital transformation, we launched a pioneering project for oncohematologic patients to promote adherence to a healthy lifestyle and improve their physical and nutritional well-being. We aim to achieve this goal by involving doctors and nutritionists through the Nootric app.

Objective: This study aims to assess the impact of the use of eHealth tools to facilitate nutrition and well-being in oncohematologic patients. We also aim to determine the usefulness of physical-nutritional management in improving tolerance to chemotherapy treatments within routine clinical practice.

Methods: We designed a descriptive, observational, longitudinal, prospective cohort pilot study that included a total of 22 patients from March to May 2022 in the Vinalopó University Hospital. The inclusion criteria were adults over 18 years of age diagnosed with oncohematological pathology in active chemotherapy treatment. An action plan was created to generate alerts between the doctor and the nutritionist. In the beginning, the patients were trained to use the app and received education highlighting the importance of nutrition and physical exercise. Sociodemographic, clinical-biological-analytical (e.g., malnutrition index), health care impact, usability, and patient adherence data were collected. Tolerance to chemotherapy treatment and its health care impact were evaluated.

Results: We included 22 patients, 11 (50%) female and 11 (50%) male, ranging between 42 and 84 years of age. Among them, 13 (59%) were adherents to the program. The most frequent diseases were lymphoproliferative syndromes (13/22, 59%) and multiple myeloma (4/22, 18%). Moreover, 15 (68%) out of 22 patients received immunochemotherapy, while 7 (32%) out of 22 patients received biological treatment. No worsening of clinical-biological parameters was observed. Excluding dropouts and abandonments (n=9/22, 41%), the adherence rate was 81%, established by calculating the arithmetic mean of the adherence rates of 13 patients. No admission was observed due to gastrointestinal toxicity or discontinuation of treatment related to alterations in physical and nutritional well-being. In addition, only 5.5% of unscheduled consultations were increased due to incidents in well-being, mostly telematic (n=6/103 consultation are unscheduled). Additionally, 92% of patients reported an improvement in their nutritional habits (n=12/13), and up to 45% required adjustment of medical supportive treatment (n=5/11). There were no cases of grade 3 or greater gastrointestinal toxicity. All of this reflects improved tolerance to treatments. Patients reported a satisfaction score of 4.3 out of 5, while professionals rated their satisfaction at 4.8 out of 5.
Conclusions: We demonstrated the usefulness of integrating new technologies through a multidisciplinary approach. The Nootric app facilitated collaboration among the medical team, nutritionists, and patients. It enabled us to detect health issues related to physical-nutritional well-being, anticipate major complications, and mitigate potentially avoidable risks. Consequently, there was a decrease in unscheduled visits and admissions related to this condition.

(JMIR Form Res 2024;8:e49574) doi: 10.2196/49574

KEYWORDS
Nootric app; oncohematology patient; physical-nutritional well-being; multidisciplinary team

Introduction

Background

Hematological malignancies encompass a heterogeneous group of diseases that have different behaviors, evolution, treatments, and prognoses. However, all of them similarly compromise the patient’s nutritional and physical status. This is because both the development of the disease and antineoplastic treatment can lead to caloric-protein malnutrition, leading to a high prevalence of adverse effects in daily clinical practice [1]. New treatments and conventional chemotherapy lead to toxicity in the gastrointestinal tract, which has a direct impact on the patient’s well-being and survival [2-9]. Medical management is often insufficient to carry out a comprehensive assessment of the patient’s physical and nutritional well-being. Therefore, professional support in this aspect through the application of information and communication technologies (ICTs) in patients’ everyday environment outside the hospital is a useful tool to improve well-being and reduce health care costs [6,10-12]. There is increasing evidence showing that lifestyle interventions can improve symptoms, quality of life, and even overall survival rates for patients with cancer. Digital interventions can help implement physical-nutritional behavior modifications and empower patients through healthy lifestyle education and support [13].

An effective system for patient physical-nutritional monitoring and treatment after nutritional risk assessment appears to be lacking. We identified the need for a standardized system to prevent and treat malnutrition related to these diseases. Currently, there are studies involving mobile apps for nutritional control and support to monitor dietary intake among patients who are hospitalized and face nutritional risks. These apps have demonstrated good acceptance among patients and have the potential to be useful dietary evaluation tools for use in clinical practice. These results suggest that such tools could be extrapolated to the field of oncohematology consultations [14].

The studies available so far confirm that the application of mobile apps, among other appropriately designed digital interventions, can be effective tools in nutritional interventions [2,3,15]. It is estimated that 59% or more of the currently available apps are health-related [2]. Some studies show that mobile app interventions can improve the quality of life of patients with malignant hemopathies by reducing symptoms [16].

However, many of these apps are not developed by nutrition experts, validated by official agencies, or part of routine use in the hospital setting. The use of mobile apps for nutritional interventions to improve dietary patterns, avoid or reduce side effects, and improve patient physical and nutritional well-being is a new challenge currently facing health care.

Therefore, we have launched a pioneering project with the aim of improving and mitigating malnutrition and side effects through proper nutritional and physical well-being monitoring. We aim to deliver this digital nutrition service via the Nootric app, thereby promoting patients’ adherence to a healthy lifestyle.

In this study, we included under the term “oncohematological patient” those who met the inclusion criteria: adults over 18 years of age with a diagnosis of oncohematologic pathology undergoing active treatment. The hematologic malignancies included were mostly lymphoproliferative syndromes and multiple myeloma. We analyzed only nutritional and physical parameters within this group.

Objectives

We evaluated an intervention designed to support oncohematological patients in active treatment. The primary goal was to assess how eHealth tools in nutrition and well-being management impact patients in oncohematology. Additionally, we aimed to determine the usefulness of physical-nutritional management in improving tolerance to chemotherapy treatments within routine clinical practice.

As secondary objectives, on the one hand, we aimed to evaluate the usefulness of the application of a physical-nutritional intervention among oncohematological patients by observing serial cases. On the other hand, we wanted to evaluate the adequacy and acceptability of this app in this group of patients. We aimed to qualitatively evaluate how knowledge guides the reorientation of intervention strategies regarding physical-nutritional well-being in these patients. Finally, we aimed to understand the nutritional and physical requirements throughout each phase of these patients’ treatment, considering the potential implications for their well-being. Measures of engagement with the intervention and semistructured interviews with intervention participants were used to evaluate the feasibility of the intervention.

Methods

Study Design

A descriptive, observational, prospective, longitudinal cohort pilot study was conducted among oncohematological patients in our hospital center (Figure 1).

Recruitment took place between April and May 22. The exposure, follow-up, and data collection period lasted 3 months.
(from May 22 to August 22). The sample consisted of 22 patients and was not divided under any concept at the beginning of the study. During study development, patients who had good adherence were included in the exposed cohort. However, those who dropped out of the study and had a low adherence rate were included in the unexposed cohort. They were not taken into account in the analysis of the results, and no comparative study was performed. To ensure that older adults were not excluded due to the digital gap, we included patients aged above 80 years old, and during the first visit, we encouraged them to continue using the Nootric app.

**Figure 1.** Study flow diagram.

---

**Study Setting**

This study was conducted in the province of Alicante, Spain, at Vialopó University Hospital, which tends to a culturally, linguistically, and socioeconomically diverse population. The participants were recruited from the hospital’s hematology department.

**Participants**

Patients were eligible to participate if they (1) were aged ≥ 18 years, (2) had a documented diagnosis of oncohematological pathology, and (3) were receiving medical treatment for cancer at the time of study initiation. Patients were excluded from the study if they (1) were children, adolescents, or pregnant women; (2) were due for surgery with hospital nutritional treatment; (3) were following nutritional care or hospital treatment; (4) had any acute or chronic condition that the practitioner believed limited their ability to participate in the study; (5) were unable to provide written consent; (6) were not literate; and (7) did not have a smartphone.

**Intervention: OncohematoNootric Program**

**Overview**

In recent years, there has been a growing use of eHealth tools, such as mobile apps, in nutritional interventions with good acceptance and results [2,3]. These technologies have been applied in different health care settings, such as mental health support and chronic disease management, to enable interaction with patients and promote engagement with health care interventions. Ultimately, they aim to increase the acceptability, use, and effectiveness of interventions [17,18]. However, to date, few studies have evaluated the effectiveness of these apps in routine clinical practice within the health care environment. There have been no studies in patients with hematological malignancies undergoing treatment [4].

The OncohematoNootric program involves a continuous approach and follow-up of patients by their physician and nutritionist through the Nootric app over 3 months via face-to-face/telematic consultations, voice calls, and direct chat through the app.
The intervention aims to provide nutritional-physical support tools to oncohematology patients receiving treatment who may develop adverse effects that put their physical and nutritional well-being at risk. The program provides training and information. It also facilitates risk assessment and clinical support, as needed, with the help of a physician-nutritionist alarm system.

**Education Website**

Nootric is a digital nutrition service that creates personalized nutrition plans tailored to the oncohematological patient, featuring recipes compatible with the potential side effects of treatment. It augments cognitive-behavioral therapy and provides guides and challenges that address aspects related to nutrition and its application in daily life. Physicians can monitor patients through the app using a dynamic panel that displays real-time actions. It also includes a chat feature to communicate with a dietitian-nutritionist.

In this study, all patients received dietary and exercise recommendations from professionals. The Nootric app aimed to help patients improve their health and well-being by facilitating behavioral changes.

**Intervention Development and Patient Involvement**

The intervention was designed in conjunction with patients, medical and nutritional health professionals, and professionals with expertise in eHealth and wellness management programs and technologists. No generative artificial intelligence (AI) was used in this study.

For program implementation, a training session on the use of the Nootric app was held with different medical teams. When the candidate patient was selected by the center, the team gave them a patient information sheet, an informed consent form, and an information leaflet. Once the patient agreed to participate in the pilot, they handed over the signed consent form and began to participate. The center registered the patient on the Nootric website by entering a code assigned to the patient. Once this registration was completed, the patient downloaded the app and logged in. Next, the patient completed a series of forms that served as a basis for the dietitian-nutritionist to establish their personalized plan. In addition, a clinical-biological test was performed via a blood test requested by the medical team and carried out at the hospital, after which the results were recorded, and the nutritional counseling intervention was initiated. At the beginning of the intervention, the health care professionals oriented the patient on the use of the Nootric application and emphasized the importance of good nutrition and physical exercise. Each patient received a weekly menu and shopping list, was able to upload photos of their meals during follow-up, and had direct access to an app-based chat with a nutritionist. During the study, the patient was able to contact their dietitian-nutritionist through the app’s chat function to solve nutritional doubts and receive motivational support to increase physical activities and food recording.

During the project, improvements were made to the app to provide better patient care, including adapting 70 menu prescriptions to be compatible with the potential side effects of the treatment, configuring menu items to ensure suitability for the most common side effects, and preparing and adapting informative guides and challenges. Other improvements included sending activity and hydration reminders, optimizing the internal messaging system for medical professionals to exchange information with nutritionists, and making functional changes to facilitate uploading files for medical professionals.

An action plan was created to generate alerts between the physician and nutritionist with all the possible adverse events that patients could present (e.g., hyporexia, weight loss, skin and nail changes, diarrhea, dyspepsia, pain according to a visual analog scale, edema, fatigue, constipation, dysphagia, odynophagia, mucositis, canker sores, nausea, vomiting, diarrhea, insomnia, urinary and bladder problems, anuria, bleeding, flu-like symptoms, fever, xerostomia, rash, and pruritus), along with their severity criteria and the action plan to be followed by the doctor and nutritionist.

When any of these issues were detected, the nutritionist informed the doctor, who carried out an unscheduled telematic/presential consultation with the patient to resolve the issue with the nutritionist’s support. On the other hand, if the issue was detected by the doctor, the latter informed the nutritionist so that the patient could receive support from both. In this way, both professionals were always kept up to date on incidents and procedures. The communication channel between professionals was the Nootric web platform.

For patients who were observed to have low adherence to the program during the pilot, the professionals studied the potential causes and intensified their actions to avoid dropout.

During the follow-up period, a weekly evaluation of the variables under study was carried out by the professionals. Biweekly follow-up meetings were held with the team to track the program and make possible improvements. Patients completed weekly program evaluation forms. After the follow-up study, a satisfaction survey was conducted to qualitatively evaluate patient satisfaction and program usefulness.

A new clinical-biological test was performed to comparatively analyze the results obtained before and after the intervention (Figure 2).
Study Outcomes
The primary outcome of the study was the health impact of the use of an eHealth tool related to nutrition and physical well-being on the oncohematological patient. This was measured by determining the following variables: nº alerts resolved, nº emergency visits, nº unscheduled consultations, nº treatments suspended, nº hospital admissions and improvement in nutritional habits according to the patient's perception. nº patients referred to the hospital’s Nutrition Unit, nº patients requiring adjustment of support treatment, and nº patients with gastrointestinal toxicity (which determines the impact on improving tolerance to treatments). The secondary outcomes included assessing the perceived improvement in the patients’ physical and nutritional well-being, determined through satisfaction and usefulness questionnaires at the end of the intervention. We also sought to assess the feasibility of the intervention, focusing on usability, acceptability, and adherence regarding different intervention components.

Data Collection and Study Procedures
The variables were collected on a form based on Microsoft Office Excel 2021 (Microsoft Corp), with each coded and subsequently exported to the SPSS statistical software (version 28.0; IBM Corp). To describe continuous variables with normal distribution, measures of central tendency were used, such as arithmetic mean. The qualitative variables were presented as frequency and proportion.

The variables collected at the beginning of the pilot study, during follow-up, and at the end were: (1) sociodemographic, including sex, age, level of education, main disease, comorbidities, and medication; (2) clinical-biological-analytical, with the following parameters included in the blood analysis to assess nutritional status and treatment toxicity and direct the professionals’ action plan: hemoglobin, creatinine clearance, calcium, total protein, albumin, vitamin B12, folic acid, ferritin, malnutrition index, total cholesterol, liver enzymes aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, and gamma-glutamyl transferase, cancer medical treatment scheme, food consumption patterns measured through the form provided by Nootric through the app, and BMI; (3) health care impact; (4) usability; and (5) adherence.

To weight the usability of the application, the following user interaction variables were considered: viewing, rating, comments on recipes, uploaded photos, access to the chat, review of a guide or completion of a challenge (framed in cognitive-behavioral training), points achieved (result of the gamification provided by Nootric), and exercises displayed. Adherence was determined by analyzing each patient’s use of the Nootric app during the program application period. For adherence, the percentage of access to the Nootric application was calculated, considering the 120-day access as 100% adherence.

The evaluations performed can be viewed in more detail in Tables 1 and 2 and Figures 3-5. The data extracted from the patients’ medical records included results of the blood analysis, main disease, comorbidities and their medication, cancer medical treatment scheme, and health care impact variables.
Table 1. Variables related to patients’ sociodemographic characteristics, oncohematological diseases, cancer medical treatment received during the study, and adherence to the program during 12 weeks of intervention (N=22).

<table>
<thead>
<tr>
<th>Variable and description</th>
<th>Patients and program adherence (N=22), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socioeconomic level</td>
<td></td>
</tr>
<tr>
<td>No formal education</td>
<td>8 (36)</td>
</tr>
<tr>
<td>Primary education</td>
<td>7 (32)</td>
</tr>
<tr>
<td>Vocational education</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Secondary education</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Higher education</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Comorbidity</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>13 (59)</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>8 (36)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>17 (77)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Oncohematological diseasesa</td>
<td></td>
</tr>
<tr>
<td>Lymphoproliferative syndromes</td>
<td>13 (59)</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Myelodysplastic syndrome</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Chronic myeloid leukemia</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Cancer-directed treatment</td>
<td></td>
</tr>
<tr>
<td>Immunochemotherapyb</td>
<td>15 (68)</td>
</tr>
<tr>
<td>Targeted biological or targeted therapyc</td>
<td>7 (32)</td>
</tr>
<tr>
<td>Program adherence</td>
<td></td>
</tr>
<tr>
<td>Total abandonmentsd</td>
<td>6 (27)</td>
</tr>
<tr>
<td>Dropouts (adherence &lt;7%)</td>
<td>3 (14)</td>
</tr>
</tbody>
</table>

*a*Patients received medical treatment, physical-nutritional recommendations, and close monitoring by their multidisciplinary assistance program. Most of the study population consisted of patients without severe comorbidities (18/22, 80%).

*b*These include rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone, (R-CHOP); adriamycin, bleomycin, vinblastine, and dacarbazine (ABVD); rituximab, bortezomib, cyclophosphamide, adriamycin, and prednisone (VR-CAP); and methotrexate, regimens with melphalan, and 5-azacitidine.

*c*Therapeutic regimens included daratumumab, bortezomib, venetoclax, obinutuzumab, ibrutinib, imatinib, and nilotinib.

*d*These included 3 women and 3 men. Dropouts were considered as users with adherence rates below 7%, and they were not accounted for in the analytical results, body variables, and interactions with the app.
Table 2. Clinical-biological-analytical laboratory assessment.

<table>
<thead>
<tr>
<th>Laboratory parameters analyzed at the beginning and end of the study</th>
<th>Patients who had abnormalities at the beginning of the study (n=13), n (%)</th>
<th>Patients who had abnormalities at the end of the study (n=13), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altered renal function</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Altered blood glucose</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Altered calcium</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Elevated total cholesterol</td>
<td>2 (15)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Altered transaminases</td>
<td>1 (8)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Iron deficiency</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Elevated malnutrition index</td>
<td>5 (38)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Hypoproteinemia</td>
<td>4 (30)</td>
<td>3 (23)</td>
</tr>
<tr>
<td>Vitamin B12 deficiency</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Folic acid deficiency</td>
<td>1 (8)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Anemia</td>
<td>3 (23)</td>
<td>3 (23)</td>
</tr>
</tbody>
</table>

Figure 3. Biological test: Relevant data that were observed at baseline and the end of the study (n=13).
Results regarding usability data, user experience, and adherence rate: A high level of interaction was achieved with patients using the service provided by the nutrition and medical team via the app.

Figure 5. Results of the health care impact: Attendance data and rates.

Ethical Considerations
This study was approved by the Ethics Committee of Vinalopo´ University Hospital on March 30, 2022. All research activities involving human patients in this study have been treated in accordance with the ethical guidelines established by The Organic Law on Data Protection. All necessary approvals were obtained, including for the analysis of the research data.

This study complies with the ethical provisions outlined in the informed consent, and any additional analysis has been conducted in accordance with the existing ethical approvals. Informed consent was obtained from all patients for the conduct of this study and publication of this article, and no compensation was provided. Patients were informed that the doctor would receive information about their progress.

To protect the privacy and confidentiality of the participants, all data collected in this study were anonymized before the analysis. Measures were taken to ensure that the participants’ identifiable details were not disclosed. We used an anonymous identification system that consisted of assigning an alphanumeric code to each patient registered on the Nootric web platform and in the Nootric app. No one else, apart from Ribera’s medical team and Nootric’s team of nutritionists, had access to the data in the Nootric app and website, which are confidential and encrypted.

The original informed consent allows for secondary analysis without additional consent; this includes data collected from the participants’ medical history.

Sociodemographic Variables
Of the 22 included patients, 11 (50%) were female and 11 (50%) were male, with an average age of 70 (range 42-84) years. Among them, 13 (60%) patients were under 65 years and 9 (40%) were over 65 years. The variables regarding...
sociodemographic characteristics, treatment received during the study, and program adherence are described in Table 1.

**Clinical-Biological-Analytical Variables**

A clinical-biological test was conducted for all patients at the beginning of the study, but only 13 (59%) out of 22 patients completed it. On this population, we performed the analysis of the results. Table 2 presents all the analyzed laboratory parameters and the results at the beginning and end of the intervention. Some of the relevant data are shown in Figure 3. Some patients received corticosteroid therapy, which increases the risk of developing steroidal hyperglycemia. These patients benefited from an adapted physical-nutritional plan. There were no patients in the study who presented with alterations in blood glucose levels. Of the 2 (15%) patients who had dyslipidemia at baseline, 1 (8%) patient did not maintain dyslipidemia at the end of the study. Furthermore, 1 (8%) patient presented with iron overload secondary to a high transfusion requirement. Iron chelation therapy was initiated, which triggered a grade 3 hepatotoxicity. After discontinuing it, it improved clinically and analytically to grade 1. Of 6 (46%) patients with malnutrition at baseline, only 1 (8%) patient still had malnutrition at the end of the study. Of the 4 (30%) patients who presented with hypoproteinemia at baseline, 1 (23%) did not have it at the end. Moreover, 1 patient (8%) presented with a folic deficit at baseline and maintained it at baseline despite having received supportive treatment and dietary recommendations.

The clinical interview confirmed that treatment compliance was inadequate. Among 13 patients who had anemia at the beginning of the study, 3 (23%) did not maintain it at the end of the study. Conversely, 3 (23%) of the 13 patients who had anemia at the end of the study did not have it at baseline. None of the causes were caused by vitamin B12, folic acid, or iron deficiency but by myelotoxicity due to targeted cancer therapy.

**Program Adherence and Usability**

A total of 6 (27%) out of 22 patients abandoned the study due to a lack of adherence to the program, attributed to advanced age, insufficient socioeconomic level to ensure proper use of the app, lack of family support for improving adherence, and lack of initiative to establish a change in the physical-nutritional routine. They were not taken into account in the analysis of the results. In addition, users with less than 7% adherence were considered dropouts (3/22, 14%), and they were not taken into account in the analysis of the results. None of these patients were older adults or dropped out due to the digital gap.

Excluding dropouts and abandonments (n=9, 30%), the adherence rate was 81%, established by calculating the arithmetic mean of the adherence rates of 13 patients, much higher than the average rate in well-being. Of the 22 patients, between 7 and 11 (30%-50%) perceived an improvement in well-being, determined by the satisfaction survey conducted at the end of the study. To determine the adherence rate of each patient, the use of the app by the patients was evaluated and the parameters were n° interactions with the app, use of the chat with the nutritionist, n° interactions with the recipes, n° photos of recipes uploaded by the patients, and gamification points earned (Figure 4). Regarding the impact of usability, we obtained an average of 655 impacts per app user (Figure 4).

**Impact on Health Care**

Regarding the impact on health care quality, Table 2 and Figure 5 show data that demonstrates highly satisfactory results. None of the total emergency visits were related to physical and nutritional well-being. No hospitalizations occurred during the study for any cause, including those related to physical and nutritional well-being. No patient had to be referred to the hospital’s Nutrition Unit. Of the total number of medical consultations carried out, only 5.5% (6/103) were unscheduled and none of them were carried out for physical or nutritional issues. A total of 7 patients presented toxicity, among which 5 (71%) were cases of digestive toxicity. Of the 11 patients who required adjustment of supportive or symptomatic treatment due to toxicity, 5 (45%) had digestive toxicity. No treatment was suspended due to physical or nutritional conditions.

The patients showed an improvement in tolerance to chemotherapy treatments since there were no cases of grade 3 or higher gastrointestinal toxicity, defined as complications requiring intravenous support treatment. There were also no hospitalizations, emergency visits, or chemotherapy treatment discontinuations for this reason. A high percentage of patients (12/13, 92%) perceived an improvement in their nutritional habits.

**Impact on Patients’ Perceived Improvements in Physical and Nutritional Well-Being**

At the end of the study, the Nootric team disseminated an anonymous survey to measure satisfaction and usefulness for the patients and medical professionals involved. This enabled us to evaluate the impact on the perceived improvement in the patients’ physical and nutritional well-being. We observed that the users who adhered adequately to the program showed an improvement in this aspect. A total of 12 questionnaires were filled out by patients.

The average satisfaction rating among professionals was 4.8 out of 5, while patients rated their satisfaction at 4.3 out of 5. Table 3 highlights these results.
Table 3. Results of the impact on the perceived improvement in patients’ physical and nutritional well-being. A satisfaction and usefulness questionnaire was carried out for the patients involved at the end of the study (n=13).

<table>
<thead>
<tr>
<th>Question</th>
<th>Score out of 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>To what extent do you feel that the nutritional wellness program has enabled you to improve your eating habits?</td>
<td>4</td>
</tr>
<tr>
<td>Do you think your illness has allowed you to use the program correctly?</td>
<td>4.08</td>
</tr>
<tr>
<td>To what degree would you like to continue using this tool as part of your day-to-day life beyond the pilot study?</td>
<td></td>
</tr>
<tr>
<td>What is the degree of satisfaction with the nutritional wellness program?</td>
<td>4.25</td>
</tr>
<tr>
<td>What is the degree of satisfaction with the nutritional professionals who have assisted you in the program?</td>
<td>4.92</td>
</tr>
<tr>
<td>Do you think the app is easy to use?</td>
<td>4.42</td>
</tr>
</tbody>
</table>

Discussion

Context

Our study is based on a multidisciplinary nutritional and exercise support program for oncohematological patients undergoing active treatment using new technologies. This study provides initial data on the effectiveness of a novel physical and nutritional support program aimed at patients with malignant hematologies (largely represented in our study as lymphoproliferative syndromes and multiple myeloma) receiving targeted cancer treatment. It also provides a detailed evaluation of the implementation, adoption, and overall acceptability of this digital care intervention through a mobile app. The evaluation design has been adapted to the study objectives to provide new data to enable a better estimation of such an intervention’s impact and inform further development of digital care interventions for malignant blood diseases under active treatment.

Principal Findings

We did not observe any hospital admissions or discontinuation of chemotherapy treatment related to the patients’ physical-nutritional well-being, supporting the benefit of the program. Adequate nutritional support was provided to ensure patients’ well-being and mitigate the need for referral to the hospital’s Nutrition Unit. Regarding the impact on physical and nutritional well-being, we observed that users who adhered adequately to the program improved in this aspect.

We observed a reduction in the number of unplanned consultations related to physical and nutritional well-being. In terms of impact on health care quality, the results demonstrated highly satisfactory results. None of the total emergency visits were related to physical and nutritional well-being. Moreover, the intervention received a high satisfaction rating from both professionals and patients.

Regarding other works in the field, there is little information about the best nutritional support for patients with cancer [9,19,20]. Antineoplastic agents are known to be associated with gastrointestinal complications that lead to physical and nutritional repercussions, which can decrease well-being and result in death due to malnutrition [19]. Additionally, early nutritional intervention can improve prognosis and reduce the disease’s complication rate [12,19].

One study used a novel mobile app to assess and evaluate dietary behaviors in 39 oncologic patients. Although 5 patients dropped out prior to the study, the authors concluded that participants who tracked their daily dietary habits using a mobile phone app were more likely to reach their nutritional goals than the control patients. Other studies have used mobile apps to record nutritional status and activity levels in patients with breast cancer or other or other diseases, but none of them are similar to our study [20]. Our study was performed by a multidisciplinary team using both the app and the internet to maintain contact with the patients. Furthermore, the multidisciplinary team tailored each patient’s diet to suit their individual needs throughout their cancer treatment journey, particularly addressing gastrointestinal toxicities associated with active chemotherapy. This underscores the effectiveness of such technologies for integration into clinical practice without compromising the human touch in health care delivery.

Limitations and Strengths

This study highlights the importance of eHealth programs in addressing nutrition and well-being among oncohematology patients, offering significant value in multidisciplinary care management. The use of the Nootric app allowed for improved health care indicators and physical-nutritional well-being, promoting better patient outcomes.

Another notable strength of this study is the finding that over 50% (n=11) of the patients improved their physical-nutritional habits, leading to a considerable enhancement in their perception of well-being.

In terms of limitations, we must point out that this study has a small sample size of 22 patients, which may limit the generalizability of the results. Moreover, we experienced a 27% (6/22) dropout rate due to a lack of adherence to the program, which could have affected the overall results. In addition, not all of those who completed the study completed the clinical-biological tests. Finally, we acknowledge that the 12-week follow-up period might not adequately capture the program’s clinical impact on adherence to healthy habits and improved physical and nutritional well-being.

Conclusion

In conclusion, using targeted eHealth programs for nutrition and well-being among oncohematological patients undergoing active treatment offers significant value in multidisciplinary care management. This is achieved through enhanced interaction between physicians, dietitian-nutritionists, and patients via a...
digital nutrition service, such as the Nootric app. Supporting patients throughout their cancer journey, these technologies serve as valuable tools for integration into clinical practice without detracting from the human aspect of health care. Therefore, implementing projects that leverage new technologies in routine holistic clinical practice for oncohematological patients could prove cost-effective in both the short and long term. By facilitating the early detection of health issues related to physical-nutritional well-being and anticipating potential complications, these initiatives may help reduce unscheduled visits and admissions related to this condition.

Acknowledgments
We extend our gratitude to all the patients and families involved in this study.

Data Availability
Our data adheres to open science availability guidelines for broader research accessibility.

Conflicts of Interest
None declared.

References

Abbreviations

AI: artificial intelligence  
ICT: information and communication technology  
VAS: visual analog scale

©Beatriz Sánchez-Quiñones, Cristina Antón-Maldonado, Nataly Ibarra Vega, Isabel Martorell Mariné, Amparo Santamaria. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Implementation of Video-Based Care in Interdisciplinary Primary Care Settings at the Veterans Health Administration: Qualitative Study

Claudia Der-Martirosian¹²*, PhD; Cynthia Hou¹*, MS; Sona Hovsepian¹*, LCSW; Maia Diarra Carter³*, MPH, MD; Leonie Heyworth⁴⁵*, MPH, MD; Aram Dobalian⁶*, MPH, JD, PhD; Lucinda Leung⁷¹,²,³, MPH, MD, PhD

¹Veterans Affairs Greater Los Angeles Healthcare System Center for the Study of Healthcare Innovation, Implementation, and Policy, Los Angeles, CA, United States
²Veterans Emergency Management Evaluation Center, US Department of Veterans Affairs, North Hills, CA, United States
³Office of Primary Care/Patient Care Services, Veterans Health Administration, Washington, DC, United States
⁴Office of Connected Care/Telehealth Services, Veterans Health Administration, Washington, DC, United States
⁵Department of Medicine, University of California San Diego School of Medicine, San Diego, CA, United States
⁶Division of Health Services Management and Policy, The Ohio State University College of Public Health, Columbus, OH, United States
⁷Division of General Internal Medicine and Health Services Research, Department of Medicine, UCLA David Geffen School of Medicine, Los Angeles, CA, United States

* these authors contributed equally

Corresponding Author:
Claudia Der-Martirosian, PhD
Veterans Affairs Greater Los Angeles Healthcare System Center for the Study of Healthcare Innovation, Implementation, and Policy
16111 Plummer St
Los Angeles, CA, 91343
United States
Phone: 1 818 984 4596
Email: claudia.der-martirosian@va.gov

Abstract

Background: With the rapid shift to telehealth, there remains a knowledge gap in how video-based care is implemented in interdisciplinary primary care (PC) settings.

Objective: The objective of this study was to gain an in-depth understanding of how video telehealth services were implemented in PC from the perspectives of patients and interdisciplinary PC team members at the Veterans Health Administration (VHA) 2 years after the onset of the COVID-19 pandemic.

Methods: We applied a positive and negative deviance approach and selected the 6% highest (n=8) and the 6% lowest (n=8) video-using PC sites in 2022 from a total of 130 VHA medical centers nationally. A total of 12 VHA sites were included in the study, where 43 PC interdisciplinary team members (August-October 2022) and 25 patients (February-May 2023) were interviewed. The 5 domains from the diffusion of innovation theory and the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework guided the development of the 2 study interview guides (provider and patient). We identified themes that emerged across all interviews that were associated with the implementation of video-based care in interdisciplinary PC settings, using directed-content rapid analysis of the interview transcripts. The analysis was guided by 5 a priori NASSS domains: (1) patient condition or characteristic, (2) technology, (3) adopter system, (4) health care organization, and (5) adaptation over time.

Results: The study findings include the following common themes and factors, organized by the 5 NASSS domains: (1) patient condition or characteristic—visit type or purpose (eg, follow-up visits that do not require physical examination), health condition (eg, homebound or semihomebound patients), and sociodemographic characteristic (eg, patients who have a long commute time); (2) technology—key features (eg, access to video-enabled devices), knowledge (eg, how to use videoconferencing software), and technical support for patients and providers; (3) adopter system—changes in staff roles and clinical practice (eg, coordination of video-based care), provider and patient preference or comfort to use video-based care, and caregiver’s role (eg, participation of caregivers during video visits); (4) health care organization—leadership support and access to resources, scheduling for video visits (eg, schedule or block off digital half or full days), and training and telehealth champions (eg, hands-on or on-site training.
for staff, patients, or caregivers); (5) adaptation over time—capacity to improve all aspects of video-based care and provide continued access to resources (e.g., effective communication about updates).

**Conclusions:** This study identified key factors associated with the implementation of video-based services in interdisciplinary PC settings at the VHA from the perspectives of PC team members and patients. The identified multifaceted factors may inform recommendations on how to sustain and improve the provision of video-based care in VHA PC settings as well as non-VHA patient-centered medical homes.

**(Jmir Form Res 2024;8:e52830)** doi:10.2196/52830

**KEYWORDS**
interdisciplinary primary care team members; NASSS framework; nonadoption, abandonment, scale-up, spread, and sustainability; primary care; telehealth; video-based care

**Introduction**

With the rapid expansion of telehealth services since the onset of the COVID-19 pandemic, numerous studies have focused on health care clinicians’ perspectives on telehealth service implementation in primary care (PC). These studies primarily focused on satisfaction [1-4] and barriers to and benefits of telehealth services [5-24]. Regarding barriers to PC telehealth implementation, some clinicians have experienced challenges [5-14], such as technical issues with video and patient portals, privacy or confidentiality concerns, workflow and scheduling changes, low reimbursement rates, improper telecommunication infrastructure, inability to conduct physical examinations, difficulty maintaining the therapeutic relationship with patients, skill and comfort with technology, and access to technology [5-24]. However, clinicians and patients have shared several benefits of PC telehealth implementation [21,25-34], such as reducing infection and communicable disease exposure, eliminating commute time and patients’ transportation expenses, improved medication management, better evaluation of patients’ home environment, continuity of outpatient care, flexibility in scheduling appointments, and effective screening and triaging [5-9,21,25-34].

Despite its barriers and challenges, during the COVID-19 pandemic, video-based care has garnered high levels of satisfaction among health care clinicians and patients [1-4]. However, little is known about telehealth implementation at the national level, its use among interdisciplinary PC team members, and its sustainment beyond the initial phases of the pandemic. To address these knowledge gaps, the diffusion of innovation theory and the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [35], an evidence-based, theory-informed, pragmatic model, are used to help understand the factors associated with the implementation of a technology-supported health care program. The NASSS framework builds on the work of the diffusion of innovation theory [36] and addresses the success of telehealth services, referred to in this study as video-based care. Guided by 5 NASSS domains and the corresponding subdomains, this study identifies the common factors (across patients, providers, and sites) that are associated with achieving fully mainstreamed implementation of video-based care.

The main objective of this study is to understand PC team members’ and patients’ perspectives on the implementation of video-based services within the Veterans Health Administration (VHA) 2 years after the onset of the COVID-19 pandemic. VHA PC is based on a patient-centered medical home model that includes interdisciplinary team members (physicians, nurse practitioners, physician assistants, nurses, social workers, clinical pharmacists, and mental health specialists) who work together to coordinate the provision of care, including video-based care. The VHA is well-suited to examine these issues given its over 2 decades of experience in video-based care [37-40] and its rapid expansion of video-based services in PC at the onset of the COVID-19 pandemic [41].

**Methods**

**Study Setting, Site Selection, and Recruitment**

To have a greater understanding of the interdisciplinary PC team members’ and patients’ perspectives on the use of video-based services at the VHA, we applied a positive and negative deviance approach [42,43] and selected the 6% highest (n=8) and the 6% lowest (n=8) video-using PC sites in 2022 from a total of 130 VHA medical centers nationally. For each of the 16 selected VHA sites, we contacted the medical directors and chiefs of staff through email and shared the study materials, such as the study information sheet and study approvals. We specified in the study information sheet that participation in the study was voluntary, and if they decided to participate, they could withdraw from the study at any time or refuse to answer any question. VHA medical center directors and chiefs of staff from 12 sites (6 low and 6 high) agreed to participate in the study. The 12 study sites represent all 5 US regions (3 West, 3 Midwest, 2 Southwest, 2 Southeast, and 2 Northeast), where 9 are urban and 3 are rural.

For the provider interviews, the inclusion criteria to participate was to be a member of a VHA interdisciplinary PC team at any of the 12 study sites. After receiving approval at the 12 VHA sites, we were then referred to the PC chief of staff at each site, who assisted with the recruitment process by sharing the study information sheet with their PC team members. Interested PC team members (n=53; 3-5 per site) contacted our study team to express that they wanted to participate in a 30-minute interview. Up to 3 follow-up emails were sent to schedule the study interviews. A total of 43 PC staff members (3-4 per site) were interviewed by 2 study members remotely through Microsoft Teams during August-October 2022. All interviews were audio-recorded and transcribed.
For the patient interviews, we began the recruitment process in January 2023 by randomly selecting a total of 120 VHA PC patients (10 per site) using the VHA electronic health records. The study inclusion criteria included (1) at least 1 PC visit during the past 2 years (2021-2022), (2) a valid US postal address, and (3) a phone number. We first mailed the study recruitment letters to all potential study participants, explaining the purpose of the study, that study participation was voluntary, and that all collected interview data would be kept confidential. We also explained that they had the option to opt out by calling a designated study phone number, and if they chose not to opt out, we may contact them (by phone) in 2-3 weeks to schedule a 30-minute phone interview. To minimize the burden on the 2 study members, we mailed the recruitment letters in 2 batches (60 letters in January 2023 and 60 letters in February 2023). We contacted 83 recruited patients by phone (up to 3 phone calls) with the following results: (1) 42 (51%) voice messages (no answer), (2) 7 (8%) wrong phone numbers, (3) 9 (11%) refusals, and (4) 25 (30%) completed phone interviews (2-3 per site) during February-May 2023. The response rate for patient interviews was 33% (\(\frac{25}{83-7}\) \times 100). We stopped recruiting patients after reaching data saturation [44]. All interviews were audio-recorded and transcribed.

**Semistructured Interview Guide**

For this study, 2 semistructured interview guides (provider and patient) with open-ended questions were used. Both interview guides were developed based on 5 of the NASSS domains [35] pertaining to telehealth implementation: (1) patient conditions, illness, and characteristic; (2) technology; (3) adopter system; (4) health care organization; and (5) adaptation over time (see Table 1 for details).
<table>
<thead>
<tr>
<th>NASSS domains and corresponding subdomains</th>
<th>Patient or provider</th>
<th>Open-ended interview questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient condition, illness, and characteristic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A. Visit type or purpose</td>
<td>Patient</td>
<td>• Why and how was VA Video Connect (VVC), (the VA video technology/software), useful for your health conditions?</td>
</tr>
</tbody>
</table>
| 1B. Health condition, illness, or patient type | Provider | • For what types of services/visits, patients, and health conditions is VVC appropriate?  
• Do any of the barriers or challenges to VVC use differ by patient health conditions (eg, chronic care management)? |
| 1C. Sociodemographic characteristics | Patient | • What is your experience with VVC? Probe: What are/were some of the challenges/barriers to using VVC? |
| 1C. Sociodemographic characteristics | Provider | • What are some of the challenges to using VVC?  
• Do any of the barriers or challenges to VVC use differ by patient socio-demographics (eg, older patients, patients experiencing homelessness)? |
| **Technology** | | |
| 2A. Key features of the technology | Patient | • What type of device do you use (or used; eg, iPad, iPhone, Smartphone, laptop) for a video visit? How do you connect to the internet (eg, WiFi, broadband)? |
| 2A. Key features of the technology | Provider | • Do you use (or have you used) any other video app other than VVC (eg, Doximity, FaceTime) to connect with patients at their homes using video? Probe: Can you elaborate your experience with other platforms compared to VVC?  
• What new/ongoing features of technology for VVC is your clinic or medical center using to meet the ongoing needs of patients and staff? |
| 2B. Knowledge about technology | Patient | • Describe the steps (scheduling a VVC appointment, during the appointment, after the appointment) used to connect to VVC? |
| 2B. Knowledge about technology | Provider | • What type of support/and or guidance have you/your PC team received for VVC? |
| 2C. Technical support | Patient | • Did you receive any consults, educational trainings, or assistance from the VA or family members/caregivers for your VVC appointment (for any of the appointment phases)? Any barriers? |
| 2C. Technical support | Provider | • What types of technical support, guidance, or trainings have you, and/or your colleagues, in your clinic received for VVC? Probes: What recommendations do you have for additional staff support, guidance? |
| **Adopter system** | | |
| 3A. Changes in staff roles and clinical practice | Provider | • What were some of the changes in provider roles/workflows and practices, care management, care coordination, team interactions with VVC?  
• How did these changes impact clinical practice? |
| 3B. Provider and patient preference or comfort | Patient or provider | • How do you feel about VVC (vs phone, vs in-person)? Probe: Do you prefer using VVC? Why or why not? Did anyone help with your VVC scheduling, VVC visit, after VVC visits? |
| 3C. Caregiver’s role | Provider | • What are the advantages and disadvantages of using VVC from your patients’ perspectives? Did you or your team provide any consultation/help with setting up VVC visits? |
| **Health care organizations** | | |
| 4A. Leadership support and resources | Provider or leadership | • What changes had to take place in your practice/clinic to implement VVC? |
Study Population
A total of 43 PC team members and 25 PC patients were interviewed remotely. PC team members represented different roles in the PC team, including 16 primary care providers (PCPs), which included physicians, physician assistants, and nurse practitioners; 12 nurses; 3 clinical pharmacists; 3 social workers; 2 mental health specialists (eg, psychiatrists and clinical psychologists); 4 scheduling clerks or supervisors; and 3 health care leadership team members. A diverse group of patients with respect to age, gender, race, ethnicity, and place of residence (rural or urban) were interviewed.

Analysis
We identified emerged themes and factors across all provider and patient interviews that were associated with the implementation of video-based care in interdisciplinary PC settings, using directed-content rapid analysis [45,46] of the interview transcripts. This analysis was based on substantive significance [47], and it was guided by 5 a priori NASSS domains (mentioned above). A structured template was created to summarize data from each interview, with 3 team members (CH, SH, and CDM) revising the template after independently analyzing a single transcript and reviewing the others’ templates for consistency. A summary matrix of the summaries was then used to consolidate all the interview findings. Key points were transposed and sorted into NASSS framework–identified themes, then reviewed and validated by 2 project team members (CH, SH, or CDM).

Ethical Considerations
This study was part of an ongoing quality improvement effort, and hence it was exempt from review by the institutional review board at the VHA Greater Los Angeles Healthcare System. Following study protocols, audio-recorded verbal consent was sufficient for participation in the study. To protect the privacy of study participants, all collected study information was deidentified and kept confidential. Each patient interviewee received a US $30 VHA canteen voucher after completing the interview. Provider interviewees did not receive any monetary incentives since their participation in the study was during the workday.

Results
Overview
In this section, the emerged themes and factors from the directed-content rapid analysis of provider and patient qualitative interviews are discussed in detail. The emerged or identified themes (or identified factors) were common across all patient and provider types. Table 2 displays the study findings for each of the 5 NASSS domains and the corresponding subdomains.
**Table 2. Factors guided by the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework that impact the use of video-based services in primary care settings at the Veterans Health Administration.**

<table>
<thead>
<tr>
<th>NASSS domain and subdomain</th>
<th>Specific scenarios or examples of the appropriateness of video-based primary care</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient condition, illness, and characteristic</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Visit type or purpose | Follow-up visits to address an issue that has already been examined during an in-person visit (eg, discussing side effects from a newly prescribed medication)  
• Follow-up visits for chronic disease management (eg, diabetes and hypertension)  
• Getting a referral to see a specialist  
• Continuity of care for patients who relocate to another state  
• Medication reconciliation  
• Clinical pharmacy (eg, review new prescriptions or respond to questions)  
• Conducting social work assessments (eg, assess patient’s home environment)  
• Video visits are appropriate for mental health counseling  
• Video-based care is a good fit for patients who do not need a physical exam |
| Health condition, illness, or patient type | Established patients (eg, who have been with the same clinician for many years)  
• Patients with mobility issues (eg, homebound or semihomebound)  
• Patients who cannot come into the clinic for various reasons (eg, substance abuse, posttraumatic stress disorder, or anxiety)  
• Recently discharged patients from a hospital stay or urgent or emergency care  
• Patients in palliative care  
• COVID-19–positive patients |
| Sociodemographic characteristics | Patients who live far away or in rural areas  
• Patients who are experiencing homelessness  
• Patients who are working, going to school, or have family obligations  
• Patients who would like to save money on gas and transportation cost  
• Patients who have working or employed caregivers who need to be present during a visit  
• Patients who have access to a private or quiet space with no disruptions |
| **Technology** | |
| Key features of videoconferencing technology | Video-enabled devices for patients and primary care team members (at the office or teleworking)  
• Access to high-speed broadband and stable internet connectivity  
• Robust network coverage at the office  
• User-friendly, simple-to-use videoconferencing software |
| Knowledge about technology | Be able to use a video-enabled device and navigate the videoconferencing software  
• Test the technology before using it for the first time (eg, download the app ahead of time or test the video links) |
| Technical support | IT support for primary care team members at the office or when teleworking  
• IT support for patients  
• Before the first video visit, the IT help desk conducts a test call with patients to ensure everything is set for a videoconferencing visit |
| **Adopter system (staff, patient, and caregiver)** | |
| Changes in staff roles and clinical practice | Involve primary care team members in the video-based care coordination process (eg, if a nurse visit happens, then the primary care provider joins the video-based visit)  
• A video-based care flow that mimics the in-person care process  
• Telehealth protocols for different departments, services, and types of care  
• Staffing flexibility (eg, remote provider who floats or provider coverage for other clinics) |
| Clinician and patient preferences and comfort | Provides flexibility to patients and providers to connect  
• Patient and provider preference or comfort level to use video  
• Patients who are not distracted (eg, they are at home and not in a store or driving)  
• Telehealth is less stressful for some patients |
| Caregiver’s role | The caregiver is present to help with the video visit (eg, troubleshoot technical issues)  
• The caregiver can be part of the video visit if there is a need to participate (eg, provide information)  
• The caregiver can provide emotional support during the video visit |
| **Health care organization** | |

https://formative.jmir.org/2024/1/e52830
### Leadership support and resources

- Leadership support at all levels (eg, primary care chief of staff, clinic, or medical director) and at all service lines (eg, primary care providers, nursing, scheduling, or pharmacy)
- Veterans Health Administration–issued tablets or iPads sent to qualified patients
- Access to a private or quiet office space for providers to conduct video visits

### Scheduling, reminders, or day of video-based appointment

- Scheduling video visits involves multiple steps and need to ensure the patient is ready to connect remotely on the day of the visit
- Schedule or block off remote half or full days (ie, create telehealth clinic days)
- More availability of telehealth appointments compared to in-person appointments
- Timeliness of video visits compared to in-person appointments
- Treat a video-based visit same as an in-person visit (eg, patients should dress appropriately, prepare a list of questions, and have their medication list ready to review)
- Automatically enroll new patients in video-based care

### Trainings and telehealth champions

- Online trainings and resources are available for patients and providers
- Prioritize video-based care trainings for all, including newly hired staff and newly enrolled patients
- Provide hands-on and site- or team-specific training or education for all providers, staff, and patients
- Identify on-site staff and patient telehealth champions to assist with hands-on training or education

### Adaptation over time

#### Capacity to improve

- Improve technical resources, personnel support, hands-on or site-specific trainings, or education about video-based care coordination and scheduling
- Access video links from patient’s portal that is embedded in the electronic health record system

#### Continued access to resources

- Effective and widespread communication about changes in resources and new updates to the videoconferencing software or video-based care management system and options
- Annual video-based care training day for providers, staff, and patients to learn about updates

### Patient Condition, Illness, and Characteristic

#### Visit Type or Purpose

Video-based care was considered most suitable for follow-ups to routine care, particularly for patients with chronic conditions. Seeing new patients and care requiring a physical assessment was preferably done in-person. A patient explained:

> If I need a quick medicine change or something along those lines where I don’t need to be physically examined then I would, yes, use a video call. [Patient 18]

Moreover, video visits allow providers to assess patients’ home environment.

> Video appointments give much more insights to the patients’ home status, their pride, their pets... their current living conditions that we can discuss about. [Leadership 1]

> I do believe the best place to take care of the patient is in their home... so, the more video we could do, I think the better we would be at taking care of Veterans. We could touch them more often. [Physician 15]

Video visits also allowed better medication reconciliation, as patients could just show their bottles through the video camera.

> They’re at home with their pills. They can pull them all out. They don’t have to drag them into the appointment. [Clinical pharmacist 3]

Video-based care is conducive for explaining new medications or laboratory results to patients. Blood pressure readings were cited as being video-appropriate, as patients usually have their own equipment and can be overseen by their clinician while taking their own blood pressure at home. Mental health specialists felt comfortable using the technology for telemental health sessions, stating the proven effectiveness and comparability of video to in-person care.

#### Health Condition, Illness, or Patient Type

For older patients with difficulty ambulating and those who depend on family members for transportation to the clinic, video-based care is a viable option. Similarly, for established patients who have been with the same PCP for many years, and for patients who cannot come to the clinic for a variety of reasons (eg, substance abuse and anxiety), video-based care might be a good alternative. Moreover, video-based care is appropriate for patients in palliative care, COVID-19–positive patients, and recently discharged patients. One patient explained:

> Following up from emergency room visits or hospital stays, those are pretty easy to do over the phone or video because you should be okay. You shouldn’t have any other issues, or if you are having issues, as long as they’re not extreme and you don’t need to go back to the ER. [Patient 17]
Sociodemographic Characteristics

Even though younger patients might be more amenable to video-based care, age is arbitrary when understanding how to use telehealth technology.

We still live in a reality where people think that elderly people are not technologically savvy, and I don’t think that is correct. [Physician 7]

It’s our older population that are not working that I think we see more of them in the virtual appointments than anything. [Nurse 8]

Therefore, age should not be an exclusion criterion, but it often is. A social worker further explained that, at times, it is unrealistic for patients to be expected to come into the clinic when there is no need for them to make that extra effort.

Patients with long commutes to their local VHA clinics appreciated video visits. This allowed them to save time and money on gas and transportation. For patients who are traveling within the continental United States and who need continuity of care, video-based care can be a good option. Similarly, patients who work, have family obligations or go to school and accordingly have a hard time scheduling a clinic or in-person visit found video care to be beneficial. One patient explained:

Sometimes, I’ll set my appointment when I’m picking up my son and I’m waiting outside... and it’s usually a really good time. I’m in my car and it’s very quiet. [Patient 21]

Technology

Key Features of Videoconferencing Technology

Access to video-enabled devices (either a smartphone or laptop), as well as high-speed broadband and stable internet connectivity, is necessary for video-based care. At times, patients living in rural areas lacked the wireless bandwidth to take video calls. Homeless-experienced or low-income patients lacked suitable equipment (eg, a smartphone) for video services, but this issue was ameliorated by the Digital Divide Consult program that provided VHA-issued iPads or tablets for qualified patients. Yet some patients who received iPads still faced barriers to video-based care and did not use them because of internet connectivity issues or a lack of broadband access.

I saw that also happen with one of my Veterans. He wouldn’t show up on his video appointments. And I know he had a VA-issued iPad. [Physician 10]

Knowledge About Technology

Even though video-based care might be more suitable for tech-savvy patients and providers who can use video-enabled devices, it is still important for both patients and providers to conduct at least 1 test call before using the VHA video-based technology for the first time (eg, download the app ahead of time and test the video links).

Technical Support

IT support is available for PC team members at the office or when teleworking. IT support for patients is also available, which can help patients ensure everything is prepared for their upcoming videoconferencing visit. A video test call can be done with the IT help desk or a live web-based chatbot to ensure patients’ devices are compatible with the videoconferencing software, but often these services are underused by patients.

Adopter System (Staff, Patient, and Caregiver)

Changes in Staff Roles and Clinical Practice

There have been challenges in video-based care coordination, as some aspects and processes of in-person visits are not easily transferable to video-based visits. For example, after a nurse conducted the initial video-based check-in (eg, going over vitals and the patient’s medication list), the video call was dropped during the transfer, and the PCP was unable to join the video visit. Creating processes and procedures to manage video-based care coordination among PC team members is key for providing team-based video care.

Additionally, video-based care can provide staffing flexibility. For example, video-based providers float and provide coverage at other clinics. Regarding clinical practice, there is a need to set up video-based care protocols and guidelines for different departments, services, and types of care, so there is guidance as to when video visits are appropriate.

Provider and Patient Preference or Comfort

Video-based visits provide flexibility for patients and providers to connect remotely. Provider and patient preferences regarding video-based care are important factors, given that there were strong opinions among providers against video-based care.

We have some providers that refuse to conduct video appointments... they just don’t like change. [Nurse 5]

Regarding patient preference and comfort coming into the clinic, video-based care offers a less stressful option for patients to connect with their providers. One patient explained:

I was back in the comfort of my personal space where I feel safe, and I can open up and talk better than at an office where you’re not as comfortable because it’s not your space... And I feel that I was able to open up more and really benefit from the mental healthcare than when I go to the office because I’m already like stressed because I’m there and I’m in a hospital environment and I don’t like hospitals at all. [Patient 23]

However, there was a concern with distracted patients during video visits.

Sometimes the patient does not engage enough with you, or they get distracted by the environment. [Physician 2]

They might log on, but they’re busy doing their hair. It’s almost like they’re facing timing of a friend or a family member. [Mental health specialist 03]

Overall, patients preferred occasional in-person visits to feel like they were receiving adequate attention and care. Generally, patients have strong opinions either for or against video-based care.
Caregiver’s Role

In some instances, it is important for caregivers to be present during the visit and be involved in the patient’s care, such as providing emotional support, having a better understanding of the health condition, and sharing information about what the patient is going through. Video-based visits can provide more opportunities for caregivers to get more involved with patient care. One patient explained:

I think it’s wonderful because it gave my wife a better feel for what was going on with me. [Patient 14]

Another patient shared:

I did have one session with my husband because I was trying to get him qualified as a caregiver. So, we had a video appointment for that. It was very convenient and a great option because, at the time, I was bedridden with a back injury before I had back surgery. And he was pretty much taking full time care of me to get me out of bed, showered, dressed, the whole thing. It was very helpful to have that option that we could all meet and have that evaluation done. [Patient 22]

In another instance, a social worker explained:

They’d [adult son] rather just schedule that time to be at their parent’s house and sign into the video than come in person. [Social worker 3]

Health Care Organization

Leadership Support and Resources

Leadership support at all levels (eg, PC chief of staff and clinic or medical director) and from all departments (eg, PCPs, nursing, scheduling, and pharmacy) is key for the successful implementation of video-based care in PC. Additionally, having a designated physical space (eg, a private quiet space with no distractions or interruptions) and computer hardware, such as video-enabled devices, are a necessity for the successful implementation of video-based care.

Scheduling, Reminders, and Day of Video Appointment

Before scheduling a video-based PC visit, schedulers or clerks ask patients a series of questions about the types of devices they own to determine if a video visit is feasible. The scheduler or clerk also sends the patient the video link and creates the follow-up appointment. Before the day of the video appointment, reminders through SMS text messages as well as email are sent. Automatically enrolling new patients for video visits is also an option, which makes scheduling for video-based visits easier. In-person or clinic visits are still the most frequently scheduled across clinics. However, after the provider puts in the video request that the patient agreed to, some schedulers or clerks failed to follow through with scheduling the video-based appointment.

Video clinics are being underutilized, not because of lack of patient interest. It’s being underutilized because schedulers are not calling patients to make the appointment in the video clinic. [Physician 2]

There is also the option to block off half or full days for video visits. Patients shared that video visits were more timely and easier to schedule compared to in-person appointments. Video-based visits should be treated just the same as an in-person visit. As such, patients need to dress appropriately, prepare a list of questions, have their medication list ready to review and find a quiet place for the video call.

Trainings and Telehealth Champions

Video-based care training for PC team members is done through formal web-based training. However, hands-on training or trial-and-error is the best way to learn how to use the VHA video technology.

I think the best way people can walk somebody else through it is if they do it themselves. [Physician 13]

Training should be tailored for each clinic and team member, so each can understand their role in the video-based care process. VHA leadership noted that they combined multiple video training courses into 1 web-based course for new providers. An optional 1-day intensive training course is also available. Ultimately, having a telehealth champion who provides hands-on, personalized training and updates to PC team members can be critical to effectively educating the PC team on how to use video-based care.

We did have a super user in the clinic who was able to go in and show everyone how to do it [ie, use VA video technology]. [Nurse 3]

For patients, training for video-based visits usually falls on the nurses, schedulers, clerks, technicians, or other clinic staff who enroll patients in video-based scheduling for the first time. Having a patient telehealth champion, for example, a veteran sharing how they benefited from a video visit, might be helpful in providing peer-to-peer mentoring and promoting video-based visits. Overall, it seems that patients choose to have more telehealth appointments after being trained.

Once we get it set up for them [patients] and they feel comfortable with the whole [thing]—they’re utilizing it [VA video technology] quite a bit. [Nurse 8]

Adaptation Over Time (Capacity to Improve and Continued Access to Resources)

Suggestions for improving video-based care at the VHA centered around 5 main areas: technological resources, personnel support, training, care coordination, and scheduling. Regarding resources, video-based care should be connected to the patient’s electronic health care records for easier access to the web-based appointment link. Accessibility of video on cellular devices, like using SMS text messages to quickly communicate with patients. There is a need to hire more clerical staff to assist with video-based care scheduling and technical support (eg, a telehealth enrollment coordinator role on the team or a telehealth help center to assist veterans with technical issues). Beyond personnel support, there is a need for more physical space in
clinics, so they could have privacy to conduct telehealth visits. Other suggestions included having more resources to support video technology, such as choosing site-specific internet networks and improving network coverage overall.

Recommendations to improve training focused on having nonclinical staff administer patient education in group and in-person settings. Patients need adequate training to use VHA video technology.

We need to teach them. We need to take the time and put the investment... but too often, we blame the patient, I think inappropriately. [Physician 11]

When asked about care coordination, the video care process should mimic the flow of in-person care while striving for “care management and the same depth of treatment.”

Regarding scheduling, providers asked for separate days designated for video and face-to-face appointments. It is important to have 1 centralized platform or web-based grid where all types of appointments, telehealth and in-person, can be scheduled. Schedulers or clerks should have a script or checklist to ensure that patients understand their video-based care options. Finally, site-specific telehealth management is needed, acknowledging that regional differences and specific site needs may affect each VHA clinic’s telehealth or video adoption and implementation.

**Discussion**

**Overview**

In this study, the 5 NASSS domains that pertained to the implementation of telehealth services guided the qualitative analysis and identified key factors associated with the implementation of video-based services in PC from interdisciplinary team members’ and patients’ perspectives 2 years after the onset of the COVID-19 pandemic. In terms of patient condition, illness, and characteristics, the study findings concurred with previous studies that video-based care is most appropriate for follow-up visits that do not require physical examinations, and in some cases, it might be well-suited for follow-up visits for chronic care management [48,49]. This study corroborated findings from previous studies that video visits are appropriate for semi- or fully homebound patients with mobility issues who benefit from having a caregiver present at their medical appointments [50,51]. Similar to previous studies, convenient access to video-based care was needed for patients with full-time jobs, family obligations, transportation difficulties, and long travel distances for in-person care [5-24].

Regarding age, the study findings illustrated that regional differences and specific site needs may affect each VHA clinic’s telehealth or video adoption and implementation.

Study participants shared many areas of improvement that can help sustain telehealth adaptation over time. Like previous studies [53], the study findings highlighted that it is important for health care organizations to have the capacity to improve technological resources, personnel support, training, and video-based care coordination and continue to provide access to these resources. In addition, the study findings also highlight the importance of improving scheduling processes and platforms to better meet the needs of all providers and patients. Finally, it is important to acknowledge that site-specific telehealth management is needed, since regional differences and specific site needs may affect how each VHA clinic implements telehealth or video-based care in PC settings.

**Limitations**

The major strengths of this study are the following: first, the study included 12 VHA medical centers located in geographically diverse settings. Second, these study sites, including 9 urban and 3 rural, had varying rates of video use, ranging from lowest to highest. Third, the directed-content rapid
analysis, guided by a priori NASSS domains, identified emerged themes or factors across all interviews (43 PC team members and 25 PC patients). Despite this, the study has several limitations. First, this study did not examine the differences between patient types, appointment types, provider types, and sites. Instead, the factors that were identified in this study were common across all patients, all PC team members, and all sites. Future studies should consider which factors differentiate between sites (eg, urban vs rural), provider types (eg, mental health specialists vs PCPs), patient types (eg, male vs female and with diabetes vs without diabetes), as well as appointment types (mental health visit vs PC visit). Second, since this study did not use probability sampling strategies to select the study sites or the provider study sample, the generalizability of the study findings to all VHA PC clinics and providers is limited. Furthermore, even though simple random sampling was used to select the patient study sample, the study’s main purpose was to have an in-depth understanding of the implementation process of video-based care at the selected VHA study sites. Hence, the generalizability of the study findings to all VHA patients is limited. Third, the generalizability of study findings may be limited in non-VHA health care systems for various reasons, including that VHA clinicians do not have the same cross-state licensure restrictions, especially since the passage of the Anywhere-to-Anywhere Act (in May 2018) [54,55], where VHA expanded telehealth services by allowing health care clinicians to treat patients across state lines; and VHA has a capitated payment system, which makes it easier to implement telehealth since it is not subject to third-party payer arrangements [54]. However, recent COVID-19 telehealth waivers have increased non-VHA health care providers’ and clinics’ telehealth capability, such as allowing telehealth services across state lines. Therefore, there are more similarities now between VHA and non-VHA telehealth services than even in the recent past. As such, study findings may still be applicable to non-VHA clinical settings and contribute to the growing evidence base surrounding factors most salient to the successful implementation of telehealth services at PC clinics.

Conclusions

Given that VHA PC is based on a patient-centered medical home model that includes interdisciplinary team members who work together to coordinate the provision of care, including video-based care, we examined the implementation of video-based care in interdisciplinary PC settings from the perspectives of PC team members as well as patients at 12 different VHA health care settings (9 urban and 3 rural). Guided by 5 a priori NASSS domains and the corresponding subdomains, we identified common factors (across patients and PC team members) that were associated with the implementation of video-based care in interdisciplinary PC settings. We identified multifaceted factors that resulted from the qualitative analysis of the collected interview data may help inform recommendations on how to sustain and improve video-based care in VHA PC settings and other non-VHA patient-centered medical homes.

Acknowledgments

This material was supported by the United States Department of Veterans Affairs, Veterans Health Administration, Office of Connected Care (XVA 65-127), Health Services Research and Development (PPO 21-247), and the Veterans Emergency Management Evaluation Center. The views expressed in this manuscript are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government.

Data Availability

The data sets that were generated and analyzed during this study are not publicly available because the collected information might compromise the privacy of the study participants. The data that support the findings of this study are available from the corresponding author (CD-M) on reasonable request.

Conflicts of Interest

None declared.

References


Abbreviations

NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
PC: primary care
PCP: primary care providers
VHA: Veterans Health Administration

©Claudia Der-Martirosian, Cynthia Hou, Sona Hovsepian, Maia Diarra Carter, Leonie Heyworth, Aram Dobalian, Lucinda Leung. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Nudges and Prompts Increase Engagement in Self-Guided Digital Health Treatment for Depression and Anxiety: Results From a 3-Arm Randomized Controlled Trial

Trevor van Mierlo1*, MBA, MScCH, MScBMR, GEMBA, DBA; Renante Rondina2*, PhD; Rachel Fournier1*, MSc
1Evolution Health, Toronto, ON, Canada
2Rotman School of Management, University of Toronto, Toronto, ON, Canada
*all authors contributed equally

Corresponding Author:
Trevor van Mierlo, MBA, MScCH, MScBMR, GEMBA, DBA
Evolution Health
206-90 Eglinton Avenue East
Toronto, ON, M4P 2Y3
Canada
Phone: 1 4166448476
Email: tvanmierlo@evolutionhealth.ca

Abstract

Background: Accessible and effective approaches to mental health treatment are important because of common barriers such as cost, stigma, and provider shortage. The effectiveness of self-guided treatment is well established, and its use has intensified because of the COVID-19 pandemic. Engagement remains important as dose-response relationships have been observed. Platforms such as Facebook (Meta Platform, Inc), LinkedIn (Microsoft Corp), and X Corp (formerly known as Twitter, Inc) use principles of behavioral economics to increase engagement. We hypothesized that similar concepts would increase engagement in self-guided digital health.

Objective: This 3-arm randomized controlled trial aimed to test whether members of 2 digital self-health courses for anxiety and depression would engage with behavioral nudges and prompts. Our primary hypothesis was that members would click on 2 features: tips and a to-do checklist. Our secondary hypothesis was that members would prefer to engage with directive tips in arm 2 versus social proof and present bias tips in arm 3. Our tertiary hypothesis was that rotating tips and a to-do checklist would increase completion rates. The results of this study will form a baseline for future artificial intelligence–directed research.

Methods: Overall, 13,224 new members registered between November 2021 and May 2022 for Evolution Health’s self-guided treatment courses for anxiety and depression. The control arm featured a member home page without nudges or prompts. Arm 2 featured a home page with a tip-of-the-day section. Arm 3 featured a home page with a tip-of-the-day section and a to-do checklist. The research protocol for this study was published in JMIR Research Protocols on August 15, 2022.

Results: Arm 3 had significantly younger members ($F_{2,4564}=40.97; P<.001$) and significantly more female members ($\chi^2=92.2; P<.001$) than the other 2 arms. Control arm members (1788/13,224, 13.52%) completed an average of 1.5 course components. Arm 2 members (865/13,224, 6.54%) clicked on 5% of tips and completed an average of 1.8 course components. Arm 3 members (1914/13,224, 14.47%) clicked on 5% of tips, completed 2.7 of 8 to-do checklist items, and completed an average of 2.11 course components. Completion rates in arm 2 were greater than those in arm 1 ($z$ score=3.37; $P<.001$). Completion rates in arm 3 were greater than those in arm 1 ($z$ score=12.23; $P<.001$). Engagement in all 8 components in arm 3 was higher than that in arm 2 ($z$ score=1.31; $P<.001$).

Conclusions: Members engaged with behavioral nudges and prompts. The results of this study may be important because efficacy is related to increased engagement. Due to its novel approach, the outcomes of this study should be interpreted with caution and used as a guideline for future research in this nascent field.

International Registered Report Identifier (IRRID): RR2-10.2196/37231

(JMIR Form Res 2024;8:e52558) doi:10.2196/52558
KEYWORDS
behavioral economics; digital health; attrition; engagement; nudges; depression; anxiety; mood disorders

Introduction

Background
Since its earliest use in the mid-1990s, digital health promised personalized treatments that patients could access from home. It was anticipated that treatment would have a broad reach, resulting in improved health outcomes and decreased costs [1-3]. Over the past 2 decades, and with increasing consistency, research examining the efficacy of self-guided digital health interventions show evidence of efficacy, especially for those with mental health concerns [4-7].

Although digital health interventions appear to be effective, poor adherence and lack of compliance have remained consistent patterns in research [8-10]. This pattern was first recognized in 2005 and coined The Law of Attrition [11].

As early as 2009, systematic reviews identified poor adherence and lack of compliance as engagement issues that required attention [12]. These issues persist, as demonstrated by a recent meta-analytic review on digital interventions for depression, which showed efficacy but highlighted compliance as a major challenge [13].

Adherence and compliance are complex and rooted in several systemic and individual factors [14-19]. However, it is an important topic as evidence indicates a dose-response relationship, and higher levels of engagement are associated with improved health outcomes [20,21].

Moreover, digital health interventions are becoming increasingly common and accessible. Patients' use of, and trust, in these interventions has been intensified by the COVID-19 pandemic. The use of self-guided digital health interventions for mental health concerns is growing [22,23] because of access barriers such as high cost, stigma, and lack of access due to a shortage of professionals who can meet this growing demand [24,25].

How do we increase engagement in digital health programs to maximize their efficacy?

Behavioral Economics

Overview
Behavioral economics leverages psychological experimentation to develop theories about human decision-making. The field has identified a range of unconscious biases around how people think and feel [26,27].

The utility of behavioral economics is vast. Digital health has leveraged the discipline to investigate how people use digital health programs and to gain insights into the characteristics of people who use them. Several digital health studies have investigated the use of these strategies, including cooperative games and incentives [28], gamification [29,30], serious games [31,32], and positive behavioral support [33,34].

Our Use of Behavioral Economics
In our study, we examined the effectiveness of the nudge theory and behavioral prompts in 2 ad libitum self-guided digital behavior change courses.

Nudge Theory
Nudge theory, popularized in the 2008 book Nudge: Improving Decisions About Health, Wealth, and Happiness [27], leverages indirect, positive suggestions to influence decision-making and behavior.

There is a lack of quality research analyzing the use of nudges in digital health. A 2019 scoping review examined the use of nudges in both web-based and real-world settings in physical activity interventions [35]. Of the 35 publications reviewed, 8 were web-based studies. The authors concluded that although nudging may be an effective approach to promote physical activity, there are large gaps in research, and further studies that are explicitly based on nudge insights are needed.

A 2020 editorial in Personalized Medicine addressed the meaningful adoption of nudges in digital health [36]. The authors acknowledged that using nudges in digital health interventions is rare and advocated for the use of nudges to promote positive behavior change.

Behavioral Prompts
In applied behavioral analysis, behavioral prompts are cues specifically designed to encourage a specific task [34]. In this study, we used 2 types of behavioral prompts anchored in the nudge theory: daily tips and a to-do checklist (Table 1).

Table 1. Example nudges and prompts.

<table>
<thead>
<tr>
<th>Delivery format</th>
<th>Content type</th>
<th>Example text from our study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tip</td>
<td>Directive content</td>
<td>Express yourself by uploading your image</td>
</tr>
<tr>
<td>Tip</td>
<td>Social proof</td>
<td>Many members have similar goals as you. Reviewing other members’ goals can help you reach your goal.</td>
</tr>
<tr>
<td>Tip</td>
<td>Present bias</td>
<td>Feel better sooner by learning from others. Read what others have posted on the community.</td>
</tr>
<tr>
<td>To-do checklist</td>
<td>Directive content</td>
<td>Watch the getting started video</td>
</tr>
</tbody>
</table>
**Directive, Social Proof, and Present Bias Tips**

Directive content, as the name suggests, offers concrete suggestions to members [37]. These types of tips are brief and instructional.

Social proof is derived from the behavioral economics concept where we tend to copy the actions of those around us. These tips speak of our tendency to be swayed by other people’s choices, which we attempt to mirror [38,39].

Present bias is the inclination to prefer a smaller present reward now over a larger reward later. These tips encourage users to perform a task that provides an immediate benefit [40,41].

**Our Use of Behavioral Prompts and Technical Functionality of Tips and to-Do Checklist Items**

An example of a course tool in both Overcoming Anxiety and Overcoming Depression courses is goal setting. Goal setting is an important component of cognitive behavioral therapy (CBT). However, in a self-guided environment, many members are unsure of how to set personal goals. Evolution Health encourages members to review goals set by others as examples.

Figure 1 shows an example of a tip of the day that encourages members to review the goals of other members.

Figure 2 shows an example of the completed to-do checklist item, set goals. In this figure, the item is marked complete by a check box. This signifies that a member has clicked on the item and visited the page.

**Figure 1.** Present bias tip: review other members’ goals.

**Figure 2.** Completed to-do checklist item, set goals.

By clicking on either the tip (Figure 2) or the to-do checklist item (Figure 3), the member is brought to member goals, a course component that allows members to browse various goals set by other members (Figure 3).

Engagement experiments in popular non-health care digital platforms are common. Although they are scientific in nature, they are not typically published, as they are conducted within private companies.

For example, social network sites such as Facebook (Meta Platforms, Inc), LinkedIn (Microsoft Corp), and X Corp. (formerly known as Twitter) generate revenue based on ad revenue derived from page views and the time members spend on their site. In a 2015 presentation, it was revealed that LinkedIn had >400 controlled experiments being conducted per day [42]. Similar studies with an ad libitum population are required for digital health, and this study is an attempt to fill this gap.

We have not observed sufficient evidence in the literature to determine whether nudges and prompts can be strategically applied to increase engagement with and decrease attrition in the courses for depression and anxiety [43]. Furthermore, because of the nascent state of behavioral economics within digital health, we did not find any quantitative benchmarks that would help us determine whether our use of nudges and prompts was successful.
Objectives

Our primary hypothesis (H1) was that members would engage with the tips and to-do checklist. Engagement of the members was determined by the percentage of tips that were clicked on compared to the number shown. Because there is no prior literature, we first needed to establish baseline data. Data were reported as observational and served as a benchmark for future studies.

Our secondary hypothesis was that members would prefer to engage with directive tips in arm 2 versus social proof or present bias tips in arm 3. To assess the preference for engagement, we compared the number of tips clicked in arm 2 versus those clicked in arm 3.

Our tertiary hypothesis was that the addition of tips and a to-do checklist would increase completion rates with course tools. Increased completion rates were determined by comparing completion rates in arms 2 and 3 with those in arm 1 (control group).
In each hypothesis, we assessed whether engagement was influenced by gender or age.

**Methods**

**Overview**

The digital health platform used in this study was managed by Evolution Health. Evolution Health is an evidence-based, self-guided digital health platform that features courses and brief interventions based on behavior change techniques including CBT, stages of change, structured relapse prevention, normative feedback, and harm reduction.

The platform offers interactive courses and quizzes for people with mental health, addiction, and obesity issues. The platform contains a moderated community based on social cognitive theory.

Memberships are available to individuals who register through the organization’s free-to-consumer program and white-label instances that are licensed by employers, insurance companies, employee assistance programs, educational institutions, nonprofit organizations, for-profit health care organizations, and individual therapists.

The research protocol for this study was published in *JMIR Research Protocols* on August 15, 2022 [44]. The International Registered Report Identifier is DERR1-10.2196/37231.

**The Interventions**

The 2 interventions in this study contain self-guided interactive behavior change treatment courses based on best practices, and both have been examined extensively in the literature [9,45-53].

The 2 interventions have undergone several iterations. For example, Overcoming Anxiety was the first intervention noted in *The Law of Attrition* (previously known as The Panic Center) paper by Eysenbach [11]. In that iteration, the course contained a tunnel design with 12 successive sessions. The course now has a gamified free-form matrix design, among other technical and usability enhancements.

Table 2 outlines each course’s current theoretical construct and evidence base. Table 3 outlines the main course components.

<table>
<thead>
<tr>
<th>Table 2. Theoretical constructs and evidence base.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical construct</td>
</tr>
<tr>
<td>Brief intervention</td>
</tr>
<tr>
<td>Cognitive behavioral therapy</td>
</tr>
<tr>
<td>Gamification</td>
</tr>
<tr>
<td>Health belief model</td>
</tr>
<tr>
<td>Motivational interviewing</td>
</tr>
<tr>
<td>Normative feedback</td>
</tr>
<tr>
<td>Social cognitive theory</td>
</tr>
<tr>
<td>Targeting and tailoring</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Main course components.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course component</td>
</tr>
<tr>
<td>Avatar upload</td>
</tr>
<tr>
<td>Course completion certificate</td>
</tr>
<tr>
<td>Course diary (mood tracker or symptom tracker)</td>
</tr>
<tr>
<td>Course worksheets</td>
</tr>
<tr>
<td>Gamified CBTa course</td>
</tr>
<tr>
<td>Getting started video</td>
</tr>
<tr>
<td>Goals exercise</td>
</tr>
<tr>
<td>Moderated community</td>
</tr>
<tr>
<td>Private messaging</td>
</tr>
<tr>
<td>Statistics interface (for corporate clients)</td>
</tr>
<tr>
<td>Tailored depression and anxiety test</td>
</tr>
<tr>
<td>Therapist interface</td>
</tr>
</tbody>
</table>

aCBT: cognitive behavioral therapy.
Ethical Considerations

At registration, all members endorsed a checkbox to confirm that they consented to have their nonidentifiable data used for research purposes and approved the platform’s privacy policy. The participants did not receive compensation for their involvement. The platform is available in several languages, and the English language privacy policy and terms of use are presented in Multimedia Appendix 1. All data collection policies and procedures adhered to the international privacy guidelines [54-56] and the Helsinki Declaration of 1975 [57].

This study was conducted on self-guided treatment for depression and anxiety, but it did not measure clinical outcomes. Although the study participants were randomly assigned to a control or intervention arm, the study was a randomized controlled trial and not a randomized clinical trial [58]. This study did not assess whether the study participants’ engagement with course tools decreased the depressive symptoms, severity of panic attacks, or frequency of panic attacks.

As described earlier, evidence indicates that higher levels of engagement are associated with improved health outcomes, and the literature observes dose-response relationships. However, any clinical outcomes related to the engagement strategies used in this study will need to be tested in future research.

As the study was based on unidentifiable data and no clinical measures were tested, it was deemed exempt from further review by the Evolution Health Institutional Review Board (IRB#000014034, FWA00033737).

Power and Sample Size

The study was designed to have a power of 0.95, indicating a 95% probability of correctly detecting a statistically significant difference in engagement with tips between the 2 treatment groups (arm 2 vs arm 3). On the basis of a 2-tailed t test with a conventional α level (.05) for statistical significance, we required a sample of 1302 members with 651 in each group to detect a small effect size of Cohen $d=0.2$. Refer to Multimedia Appendix 2 for the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [59].

Randomization

During the registration process, new members were assigned to 1 of the 3 arms using a random number generator (Figure 4). Randomization was conducted using simple randomization.

Figure 4. Recruitment process.

Intervention Groups

Arm 1

Members randomized to arm 1 were presented with a dashboard that did not contain behavioral nudges. Figure 5 presents a screenshot of the arm 1 dashboard for a member who chose to engage with the depression course.
Arm 2

Members randomized to arm 2 were presented with a dashboard that contained a tip-of-the-day section containing directive content. The randomization strategy for the 31 directive tips was randomization without replacement.

Figure 6 presents a screenshot of the arm 2 dashboard for a member who chose to engage with the depression course.
Members randomized to arm 3 were presented with 2 sections that contained nudges. The first was a tip-of-the-day section containing social proof and present bias cues. At each log-in, members saw a new tip. The randomization strategy for the tips was randomization without replacement. There were 15 social proof tips and 15 present bias tips.

In addition to the tip of the day, arm 3 featured a to-do checklist that listed 8 course components. When a member clicked on a component, they were brought to the exercise. As mentioned previously, if a member clicks on a to-do checklist item, it is marked complete with a check mark.

Figure 7 presents a screenshot of the arm 2 dashboard for a member who chose to engage with the depression course.
Figure 7. Member home page for arm 3.

Data Collection

A custom data collection interface and reporting mechanism were developed by Evolution Health. Age and gender data were collected at registration or through a secure sign on with various white-label instances. Data were collected for each member who was randomized into the experiment. The course components promoted by tips and to-do checklist items are listed in Textbox 1. Members who participated in both courses were not counted twice.

The following behaviors were tracked in the custom database for each tip and to-do checklist item that was randomly presented to a member:

- If the nudge was shown
- If the nudge was clicked on
- If a member completed the course component described in the tip or to-do checklist item.

It was possible for a single member to participate in both courses. However, the study design was to test the behavioral prompts, not the courses. The participants were randomized to an intervention arm, in which behavioral nudges and prompts were consistent across courses.
Textbox 1. Course components tracked for engagement.

<table>
<thead>
<tr>
<th>Action code and course component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (item in the to-do checklist): Uploading a personal image (avatar) to their profile</td>
</tr>
<tr>
<td>2 (item in the to-do checklist): Completing cognitive behavioral therapy session 1</td>
</tr>
<tr>
<td>3 (item in the to-do checklist): Use of the program diary (mood tracker or symptom tracker)</td>
</tr>
<tr>
<td>4 (item in the to-do checklist): Read a community post</td>
</tr>
<tr>
<td>5 (item in the to-do checklist): Review a worksheet</td>
</tr>
<tr>
<td>6 (item in the to-do checklist): Set personal goals</td>
</tr>
<tr>
<td>7 (item in the to-do checklist): Complete the depression and anxiety test</td>
</tr>
<tr>
<td>8 (item in the to-do checklist): Watch the getting started video</td>
</tr>
<tr>
<td>9: Review another member’s profile</td>
</tr>
<tr>
<td>10: Post in the community</td>
</tr>
<tr>
<td>11: Read other member’s goals</td>
</tr>
<tr>
<td>12: Give a community member a “thumbs up”</td>
</tr>
<tr>
<td>13: Encourage a community member by clicking their “show support” icon</td>
</tr>
<tr>
<td>14: Private message a community moderator</td>
</tr>
</tbody>
</table>

Participants
As there were no barriers to registration and many new members registered with the platform for purposes other than treatment, we removed those who registered but did not return to the intervention (nonparticipants) from the analysis.

Data Analyses
All data were analyzed using mixed effect logistic regression with members as a random variable. Mixed effect logistic regression was conducted using the gImer() function from the lme4 package in R software (R Foundation for Statistical Computing), with the default optimizer Bound Optimization BY Quadratic Approximation, a derivative-free optimization algorithm used for problems with bound constraints.

For the tip-of-the-day section, the outcome variable was whether a user clicked on a tip presented to them (no or yes), and the predictors were arm (2 or 3), gender (female or male), and age group (18 to 30 years to >60 years).

For the to-do checklist items, the outcome variable was whether a user clicked on a checklist item (no or yes), and the predictors were gender and age group.

For the course components, the outcome variable was whether a user completed a component (no or yes), and the predictors were arm (1, 2, or 3), gender, and age group.

Results
Overview
Between November 2021 and May 2022, data were collected from new members who self-registered for Evolution Health’s self-guided treatment program for anxiety and depression. All members were randomized into 1 of the 3 arms.

First, members with test accounts and unauthenticated accounts were removed from the data set, resulting in a population of 13,224 members. Then, members whose accounts were missing demographic data (376/13,224, 2.84%) were removed from the data set, followed by members aged <18 (129/13,224, 0.97%), resulting in a population of 12,719 members.

Finally, of the 13,224 members, 8567 (65.46%) with nonparticipant accounts were removed, resulting in a study population of 4567 (34.53%). Of the 4567 members, 1788 (39.15%) were randomized into arm 1, 865 (18.94%) into arm 2, and 1914 (41.9%) into arm 3 (Figure 8).
**Figure 8.** Study population.

**Demographic Characteristics**

The average age of the members in arms 1, 2, and 3 was 44.4 (SD 10.69), 45.4 (SD 10.62), and 41.8 (SD 11.17), respectively (Table 4). The 51 to 60 years age group was the most populous age category in each of the arms (arm 1: 777/1788, 43.45%; arm 2: 420/865, 48.6%; and arm 3: 586/1914, 30.61%). The members of arm 3 were somewhat younger ($F_{2,4564}=40.97$; $P<.001$) than the other 2 arms.

Regarding gender (Table 5), most members identified as women (arm 1: 1466/1788, 81.99%; arm 2: 751/865, 86.8%; and arm 3: 1770/1914, 92.48%). Few members identified as transgender individuals, nonbinary or third gender individuals, or preferred not to say (arm 1: 5/1788, 0.28%; arm 2: 1/865, 1.2%; and arm 3: 2/1914, 0.1%). Arm 3 had more female members ($\chi^2=92.21$; $P<.001$) than the other 2 arms.

During the study period, there were 13,510 total log-ins, with initial registration considered as 1 log-in (Table 6). The average number of log-ins for male members was 3.89 and female members was 2.83. The average visit duration (AVD) for members who engaged with depression and anxiety course tools ranged from 6 minutes and 45 seconds, with 7.65 pages viewed, to 24 minutes and 21 seconds, with 12.56 pages viewed.

<table>
<thead>
<tr>
<th>Age range (years)</th>
<th>Arm 1 (n=1788), n (%)</th>
<th>Arm 2 (n=865), n (%)</th>
<th>Arm 3 (n=1914), n (%)</th>
<th>Total age category, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 to 30</td>
<td>231 (12.91)</td>
<td>108 (12.5)</td>
<td>368 (19.22)</td>
<td>707 (15.48)</td>
</tr>
<tr>
<td>31 to 40</td>
<td>393 (21.97)</td>
<td>159 (18.4)</td>
<td>492 (25.7)</td>
<td>1044 (22.86)</td>
</tr>
<tr>
<td>41 to 50</td>
<td>329 (18.4)</td>
<td>147 (17.0)</td>
<td>415 (21.68)</td>
<td>891 (19.51)</td>
</tr>
<tr>
<td>51 to 60</td>
<td>777 (43.45)</td>
<td>420 (48.6)</td>
<td>586 (30.61)</td>
<td>1783 (39.04)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>58 (3.24)</td>
<td>32 (3.7)</td>
<td>53 (2.76)</td>
<td>142 (3.11)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Arm 1 (n=1788), n (%)</th>
<th>Arm 2 (n=856), n (%)</th>
<th>Arm 3 (n=1914), n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female members</td>
<td>1466 (81.99)</td>
<td>751 (86.8)</td>
<td>1770 (92.48)</td>
<td>3987 (87.3)</td>
</tr>
<tr>
<td>Male members</td>
<td>317 (17.72)</td>
<td>113 (13.1)</td>
<td>142 (7.42)</td>
<td>572 (12.52)</td>
</tr>
<tr>
<td>Other gender</td>
<td>5 (0.28)</td>
<td>1 (1.2)</td>
<td>2 (0.1)</td>
<td>8 (0.17)</td>
</tr>
</tbody>
</table>
Table 6. Number of log-ins by gender category.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Maximum number of log-ins, n</th>
<th>Minimum number of log-ins, n</th>
<th>Average number of log-ins, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female members</td>
<td>276</td>
<td>1</td>
<td>2.83 (6.04)</td>
</tr>
<tr>
<td>Male members</td>
<td>243</td>
<td>1</td>
<td>3.89 (12.13)</td>
</tr>
<tr>
<td>Other gender identity</td>
<td>6</td>
<td>1</td>
<td>2.6 (1.85)</td>
</tr>
<tr>
<td>Range</td>
<td>276</td>
<td>1</td>
<td>2.38 (7.09)</td>
</tr>
</tbody>
</table>

**Tip of the Day**

In arm 2, there were 31 revolving tips, and in arm 3, there were 30 revolving tips. Therefore, the probability of a member assigned to the arm 2 seeing a specific tip at least once is .088, and the probability of a member assigned to the arm 3 seeing a specific tip at least once is .091.

A total of 11,431 tips were displayed, of which 564 (4.93%) were clicked on (Table 7). In arm 2, 3622 tips were shown, of which 190 (5.24%) were clicked on. In arm 3, 7809 tips were shown, of which 374 (4.79%) were clicked on. Mixed effect logistic regression revealed no statistically significant differences between the 2 arms regarding the number of tips clicked on \( (P=.25) \).

Table 7. Tips shown and clicked on (N=11,431).

<table>
<thead>
<tr>
<th></th>
<th>Arm 2 (n=3622), n (%)</th>
<th>Arm 3 (n=7809), n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tips shown</td>
<td>3622 (100)</td>
<td>7809 (100)</td>
<td>11,431 (100)</td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>190 (5.24)</td>
<td>374 (4.79)</td>
<td>564 (4.93)</td>
</tr>
</tbody>
</table>

In arm 2, female members clicked on 171 (5.82%) tips out of the 2937 tips shown to them and male members clicked on 19 (2.8%) tips out of the 682 tips shown to them, whereas in arm 3, females members clicked on 350 (4.94%) out of the 7081 tips shown to them and male members clicked on 24 (3.3%) out of the 722 tips shown to them (Table 8).

There were no statistically significant differences between number of tips clicked on between age groups (Table 9).

Table 8. Tips shown and clicked on by gender category (N=10,018).

<table>
<thead>
<tr>
<th></th>
<th>Arm 2, n (%)</th>
<th>Arm 3, n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tips shown to female members</td>
<td>2937 (100)</td>
<td>7081 (100)</td>
<td>10,018 (100)</td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>171 (5.82)</td>
<td>350 (4.94)</td>
<td>521 (5.20)</td>
</tr>
<tr>
<td>Tips shown to male members</td>
<td>682 (100)</td>
<td>722 (100)</td>
<td>1404 (100)</td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>19 (2.8)</td>
<td>24 (3.3)</td>
<td>43 (3.06)</td>
</tr>
<tr>
<td>Tips shown to members of other gender identities</td>
<td>3 (100)</td>
<td>6 (100)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Mixed effect logistic regression (Table 10) revealed that female members clicked on more tips than male members \( (z\text{ score}=2.58; \text{ }P=.01) \). Although the role of gender and engagement should be more thoroughly examined in future studies, this finding is consistent with other research on platform components [60].

There were no substantial differences between age groups.
### Table 9. Tip engagement by age group.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Arm 2, n (%)</th>
<th>Arm 3, n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tips shown to members aged 18 to 30 years</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>26 (5.8)</td>
<td>67 (4.81)</td>
<td>93 (5.05)</td>
</tr>
<tr>
<td><strong>Tips shown to members aged 31 to 40 years</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>42 (6.3)</td>
<td>110 (5.16)</td>
<td>152 (5.43)</td>
</tr>
<tr>
<td><strong>Tips shown to members aged 41 to 50 years</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>36 (5.3)</td>
<td>95 (5.23)</td>
<td>131 (5.25)</td>
</tr>
<tr>
<td><strong>Tips shown to members aged 51 to 60 years</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>73 (4.89)</td>
<td>91 (4.16)</td>
<td>164 (4.37)</td>
</tr>
<tr>
<td><strong>Tips shown to members aged &gt;60 years</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tips clicked on</td>
<td>13 (3.9)</td>
<td>11 (5.3)</td>
<td>24 (4.4)</td>
</tr>
</tbody>
</table>

### Table 10. Summary table for tip-of-the-day section.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>ORa (95% CI)</th>
<th>SE</th>
<th>z score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm 3</td>
<td>0.87 (0.69-1.10)</td>
<td>0.1</td>
<td>-1.15</td>
<td>.25</td>
</tr>
<tr>
<td>Gender (male members)</td>
<td>0.59 (0.39-0.88)</td>
<td>0.12</td>
<td>-2.58</td>
<td>.01</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 to 40</td>
<td>1.08 (0.78-1.49)</td>
<td>0.18</td>
<td>0.44</td>
<td>.66</td>
</tr>
<tr>
<td>41 to 50</td>
<td>0.99 (0.71-1.40)</td>
<td>0.17</td>
<td>-0.03</td>
<td>.98</td>
</tr>
<tr>
<td>51 to 60</td>
<td>0.82 (0.59-1.13)</td>
<td>0.13</td>
<td>-1.22</td>
<td>.22</td>
</tr>
<tr>
<td>&gt;60</td>
<td>1.2 (0.66-2.20)</td>
<td>0.37</td>
<td>0.6</td>
<td>.55</td>
</tr>
</tbody>
</table>

aOR: odds ratio.

### To-Do Checklist

As outlined in Textbox 1 and illustrated in Figure 5, the to-do checklist contained 8 items. The members of Arm 3 completed an average of 2.7 (34%) out of 8 course components.

The checklist item with the highest engagement was complete the depression and anxiety test, which was completed by 51.4% (55/107 of the members; Table 11). The second most popular item was watch the getting started video, with 48.2% (923/1914) of the members engaging with the behavioral cue. The checklist item with lowest engagement was upload my image with nearly one-fifth (360/1914, 18.8%) of the members engaging with the behavioral cue.

It should be noted that 2 items read a community post and complete the depression and anxiety test were hidden for some members. This is due to these elements being feature flags and some Evolution Health clients and research partners choosing to hide these features from their membership base.

Although this resulted in a lower number of members seeing these cues and having access to the course components, the percentage of members who viewed the cues and engaged with the components was noteworthy.

Table 12 lists the engagement results for checklist use and gender. Mixed effect logistic regression found that female members clicked on more checklist items than male members (z score=2.07; P=.04). Although the role of gender and engagement should be more thoroughly examined in future studies, this finding is consistent with other research on platform components [60].

Tables 13 and 14 list the engagement with checklist use and age. Members aged 41 to 50 years and 51 to 60 years were significantly less likely to click on a to-do checklist item (z score=2.1; P<.04) than members aged 18 to 30 years (z score=4.35; P<.001). No other differences were found between the age groups.
Table 11. To-do checklist item clicked on (N=1914).

<table>
<thead>
<tr>
<th>To-do checklist item</th>
<th>Clicked, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload my image</td>
<td>360 (18.81)</td>
</tr>
<tr>
<td>Complete CBT a session 1</td>
<td>580 (30.3)</td>
</tr>
<tr>
<td>Use of the mood tracker or symptom tracker</td>
<td>616 (32.18)</td>
</tr>
<tr>
<td>Read a community post</td>
<td>24 (22.4)</td>
</tr>
<tr>
<td>Review a worksheet</td>
<td>606 (31.66)</td>
</tr>
<tr>
<td>Set goals</td>
<td>605 (31.61)</td>
</tr>
<tr>
<td>Complete the depression and anxiety test</td>
<td>55 (51.4)</td>
</tr>
<tr>
<td>Watch the getting started video</td>
<td>923 (48.22)</td>
</tr>
</tbody>
</table>

aCBT: cognitive behavioral therapy.

Table 12. Engagement with the to-do checklist items by gender (N=1914).

<table>
<thead>
<tr>
<th>To-do checklist item</th>
<th>Women (n=1770, 92.48%), n (%)</th>
<th>Men (n=142, 7.42%), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload my image</td>
<td>338 (19.1)</td>
<td>21 (14.79)</td>
</tr>
<tr>
<td>Complete CBT a session 1</td>
<td>549 (31.02)</td>
<td>31 (21.83)</td>
</tr>
<tr>
<td>Use the mood tracker or symptom tracker</td>
<td>572 (32.32)</td>
<td>43 (30.28)</td>
</tr>
<tr>
<td>Read a community post</td>
<td>20 (29.4)</td>
<td>4 (10.8)</td>
</tr>
<tr>
<td>Review a worksheet</td>
<td>572 (32.32)</td>
<td>33 (23.24)</td>
</tr>
<tr>
<td>Set goals</td>
<td>566 (31.98)</td>
<td>38 (26.76)</td>
</tr>
<tr>
<td>Complete the depression and anxiety test</td>
<td>32 (47.0)</td>
<td>21 (56.8)</td>
</tr>
<tr>
<td>Watch the getting started video</td>
<td>862 (48.7)</td>
<td>60 (42.25)</td>
</tr>
</tbody>
</table>

aCBT: cognitive behavioral therapy.

Table 13. Engagement with the to-do checklist items by age (N=1914).

<table>
<thead>
<tr>
<th>To-do checklist item</th>
<th>Age group (years; n=1914, 100%), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18 to 30</td>
</tr>
<tr>
<td>Upload my image</td>
<td>5 (31.25)</td>
</tr>
<tr>
<td>Complete CBT a session 1</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Use the mood tracker or symptom tracker</td>
<td>7 (43.75)</td>
</tr>
<tr>
<td>Read a community post</td>
<td>1 (50)</td>
</tr>
<tr>
<td>Review a worksheet</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Set goals</td>
<td>3 (18.75)</td>
</tr>
<tr>
<td>Complete the depression and anxiety test</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Watch the getting started video</td>
<td>9 (56.25)</td>
</tr>
</tbody>
</table>

aCBT: cognitive behavioral therapy.
Table 14. Summary table for the to-do checklist items.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>ORa (95% CI)</th>
<th>SE</th>
<th>z score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (male members)</td>
<td>0.8 (0.65-0.99)</td>
<td>0.08</td>
<td>−2.07</td>
<td>.04</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 to 40</td>
<td>0.89 (0.76-1.05)</td>
<td>0.07</td>
<td>−1.4</td>
<td>.02</td>
</tr>
<tr>
<td>41 to 50</td>
<td>0.84 (0.71-0.99)</td>
<td>0.07</td>
<td>−2.1</td>
<td>.04</td>
</tr>
<tr>
<td>51 to 60</td>
<td>0.7 (0.59-0.82)</td>
<td>0.06</td>
<td>−4.35</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&gt;60</td>
<td>0.73 (0.52-1.03)</td>
<td>0.13</td>
<td>−1.79</td>
<td>.07</td>
</tr>
</tbody>
</table>

aOR: odds ratio.

Comparison of Course Components Completed by Arm

Tables 15 and 16 outline course component completion rates by arm. Using the set goals example, 17% (304/1788) of the members in the control arm, 19.8% (171/865) in arm 2, and 31.35% (600/1914) in arm 3 completed the goals exercise.

Table 15. Course components completed by arm (N=4567).

<table>
<thead>
<tr>
<th>Course component</th>
<th>Arm 1 (control; n=1788), n (%)</th>
<th>Arm 2 (RCb; n=865), n (%)</th>
<th>Arm 3 (PBc, SP, and TDd; n=1914), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload my image</td>
<td>13 (0.73)</td>
<td>16 (1.85)</td>
<td>357 (18.65)</td>
</tr>
<tr>
<td>Complete CBTe session 1</td>
<td>669 (37.42)</td>
<td>376 (43.47)</td>
<td>806 (42.11)</td>
</tr>
<tr>
<td>Use the mood tracker or symptom tracker</td>
<td>366 (20.47)</td>
<td>224 (25.9)</td>
<td>629 (32.86)</td>
</tr>
<tr>
<td>Read a community post</td>
<td>46 (39.66)</td>
<td>13 (43.33)</td>
<td>51 (47.66)</td>
</tr>
<tr>
<td>Review a worksheet</td>
<td>N/Af</td>
<td>21 (2.4)</td>
<td>606 (31.66)</td>
</tr>
<tr>
<td>Set goals</td>
<td>304 (17)</td>
<td>171 (19.77)</td>
<td>600 (31.35)</td>
</tr>
<tr>
<td>Complete the depression and anxiety test</td>
<td>47 (24.14)</td>
<td>17 (46.67)</td>
<td>60 (51.4)</td>
</tr>
<tr>
<td>Watch the getting started video</td>
<td>N/Af</td>
<td>24 (2.8)</td>
<td>923 (48.22)</td>
</tr>
</tbody>
</table>

aRC: randomized control.
bPB: present bias.
cSP: social proof.
dTD: to-do checklist.
eCBT: cognitive behavioral therapy.
fN/A: not applicable.

Table 16. Summary table for course completion by arm.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>ORa (95% CI)</th>
<th>SE</th>
<th>z score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm 2</td>
<td>1.31 (1.13-1.52)</td>
<td>0.1</td>
<td>3.59</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Arm 3</td>
<td>1.93 (1.71-2.17)</td>
<td>0.12</td>
<td>10.91</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gender (male members)</td>
<td>1.05 (0.90-1.23)</td>
<td>0.09</td>
<td>0.6</td>
<td>.55</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 to 40</td>
<td>0.97 (0.82-1.15)</td>
<td>0.08</td>
<td>−0.32</td>
<td>.75</td>
</tr>
<tr>
<td>41 to 50</td>
<td>1.12 (0.94-1.33)</td>
<td>0.1</td>
<td>1.26</td>
<td>.21</td>
</tr>
<tr>
<td>51 to 60</td>
<td>0.81 (0.69-0.94)</td>
<td>0.06</td>
<td>−2.7</td>
<td>.007</td>
</tr>
<tr>
<td>&gt;60</td>
<td>0.76 (0.54-1.05)</td>
<td>0.13</td>
<td>−1.66</td>
<td>.10</td>
</tr>
</tbody>
</table>

aOR: odds ratio.
The course component with the highest completion rate was *complete the depression and anxiety test*, with 51.4% (60/117) in arm 3, 47% (17/36) in arm 2, and 24.1% (47/195) in arm 1. The course component with the lowest completion rate was *upload my image*, with 18.65% (357/1914) in arm 3, 1.9% (16/865) in arm 2, and 0.73% (13/1788) in arm 1.

As mentioned previously, the 2 items *read a community post* and *complete the depression and anxiety test* were not seen by all members. Some Evolution Health clients do not offer these 2 tools to their membership base. Consequently, there are high engagement percentages but low member use.

For technical reasons, the completion of 2 components *review a worksheet* and *watch the getting started video* were not captured in arm 1. However, about one-third of the members in arm (606/1914, 31.66%) reviewed a worksheet compared to about 2% of those (21/865, 2.4%) in arm 2, and almost half (923/1914, 48.22%) of the members in arm 3 watched the *getting started video* compared to about 3% of those (24/865, 2.8%) in arm 2.

Engagement with 6 of the 8 measurable components was higher in arm 2 than in arm 1 (z score=3.59; P<.001) and higher in arm 3 than in arm 1 (z score=10.91; P<.001). Engagement in 8 of the 8 components was higher in arm 3 than in arm 2 (z score=4.93; P<.001).

Overall, the use of the tips and to-do checklist items resulted in increased engagement. Members in the control arm completed an average of 1.52 course components versus 1.8 course components in arm 2 and 2.11 course components in arm 3.

**CBT Course Completion Rates**

The purpose of CBT is to help individuals deal with overwhelming problems; this is achieved by teaching people how to deal with negative thoughts and beliefs [61]. In traditional in-person therapy, the duration of CBT treatment depends on a variety of factors, including health disorders (eg, depression, anxiety, posttraumatic stress disorder, and sleep disorders), symptom severity, duration of therapy, practitioner availability, cost, willingness of the patient to do homework, convenience, culture, treatment settings, and the therapeutic alliance [62].

The recommended duration of CBT varies. Harvard Medical School recommends 30- to 60-minute sessions over 12 to 20 weeks [63], the United Kingdom’s National Health Services recommends 30- to 60-minute sessions once a week or every 2 weeks for 6 to 20 sessions [64], and the Mayo Clinic states that therapeutic encounters can range from 5 to 20 sessions [65].

According to the US National Institutes of Health’s National Library of Medicine, some people undergoing CBT feel much better after a few sessions, whereas others need treatment for several months [66]. In Canada, the Centre for Addiction and Mental Health notes that some people improve in 4 to 6 sessions, whereas others may need >20 sessions [67].

Structured CBT sessions are 1 of the 8 course components in the Evolution Health platform, and like the 7 other course components examined in this study, members’ engagement is optional. The courses Overcoming Depression and Overcoming Anxiety both comprised 9 sessions. On average, 41% (1851/4567) of the members chose to engage with and complete session 1 (Table 15).

If members completed a course, they received a course completion certificate (Figure 9).

Unlike the platform’s addiction courses, where some members are incentivized by customer sponsors to complete a course (eg, some human resource clients reward their employees for completing the quit smoking behavior change course), to the best of our knowledge, participants in the Overcoming Depression and Overcoming Anxiety courses were not rewarded for course completion or receiving a certificate.

Although the goal of CBT is to decrease symptoms and the duration of treatment varies widely per individual, course completion may be a poor indicator of wellness. However, as engagement is the primary measure of this study and the literature indicates that higher levels of engagement are associated with improved health outcomes, it may be beneficial to form a baseline observation for course completion rates.

Regarding course completion, 31 members in the Overcoming Depression course and 88 members in the Overcoming Anxiety course received course completion certificates. There were no statistically significant differences in the course completion rates between the 3 arms.

In addition, 256 of 4567 members of the N members received a certificate for completing a supplementary 1-session *more help* course, which was less intense and dealt with specific subjects such as overcoming grief, problems in relationships, or role transitions. A greater proportion of members in arm 2 received more certificates for the *more help* course than did the members in arm 1 (P=.02). There were no differences of interaction between gender or age groups.
Discussion

This 3-arm randomized controlled trial tested whether behavioral nudges and prompts could be successfully applied within 2 self-guided treatment digital health programs for anxiety and depression.

Demographic Characteristics

The average age of the members was 43.8 years, and 87.3% (3987/4567) identified as female members. The average number of log-ins for female members was 2.83 versus 3.89 for male members. The average number of log-ins should be interpreted with caution, as in other studies on the Evolution Health population; engagement appears to follow the properties of power laws [53].

H1 Summary: Engagement With Behavioral Nudges and Prompts

Our H1 was to analyze whether the members would engage with the tips and to-do checklist.

Regarding gender, mixed effect logistic regression found that female members clicked on more tips than male members (z score=2.58; P=.01).

Members in arm 3 completed an average of 2.7 (34%) out of the 8 course components featured on the to-do checklist. Engagement ranged from 18.81% (360/1914; upload my image) to 51.4% (55/107; complete the depression and anxiety test).

Mixed effect logistic regression found that female members clicked on more checklist items than male members (z score=2.07; P=.04). Members aged 41 to 50 years and 51 to 60 years were significantly less likely to click on a to-do checklist item (z score=2.1; P=.04) than members aged 18 to 30 years (z score=4.35; P<.001). No other differences were found between the age groups.

Secondary Hypothesis Summary: Members’ Preference of Directive Tips in Arm 2 or Social Proof or Present Bias Tips in Arm 3

There were no statistically significant differences in the number of tips clicked on between directive tips in arm 2 and social proof and present bias tips in arm 3 (P=.25).

Tertiary Hypothesis Summary: Tips and to-Do Checklist

Completion rates of 6 out of the 8 course tools increased in both experimental arms from baseline use in arm 1. Completion rates in arm 2 were greater than those in arm 1 (z score=3.59; P<.001), and completion rates in arm 3 were greater than those in arm 1 (z score=10.91; P<.001).

A total of 2 out of the 8 course tools (review a worksheet and watch the getting started video) did not have arm 1 baseline statistics. However, completion rates for review a worksheet in arm 2 versus arm 3 increased from 2.4% to 31.66%. The completion rate for watch the getting started video increased from 2.8% to 48.22%.
Completion rates for the course component complete CBT session 1 were slightly higher in arm 2 (376/865, 43.5%) than in arm 3 (806/1914, 42.11%). Completion rates of arm 2 were greater than those in arm 1 (z score=2.99; P=.003), and completion rates in arm 3 were greater than those in arm 1 (z score=2.91; P=.004). Completion rates of the both arms 2 and 3 were significantly higher than those in arm 1 (669/1788, 37.42%).

The highest increase in course component use was for the upload my image component, with 0.73% (13/1788) of the members uploading an image in arm 1, 1.9% (16/865) in arm 2, and 18.65% (357/1914) in arm 3. This is interesting as personalization in digital applications is now common; however, the Evolution Health platform is designed to be anonymous, and before uploading, members were advised to only upload nonidentifying images.

However, high z scores require further investigation. For example, high scores may be due to arm 3 members being more likely to complete components than arm 1 members.

**CBT Course Completion Rates**

In the literature, efficacy studies analyzing general CBT completion rates are rare, and this is due to complexities surrounding the delivery of CBT for different indications, symptom severity, duration of therapy, practitioner availability, cost, willingness of the patient to do homework, convenience, culture, treatment settings, and therapeutic alliance.

Owing to the nature of the medium, measuring course completion rates in digital health courses is relatively simple. However, digital health courses should not be held to different standards than traditional in-person therapy, where treatment success is often measured by the therapist by observing decreases in symptoms.

In 2005, Eysenbach [11] published The Law of Attrition, which recognized that a substantial portion of users drop out of eHealth (digital health) trials before completion and that the high dropout rate makes the efficacy of digital health programs less believable. The paper noted that researchers often compare digital health dropout rates with those of clinical drug trials.

Similar to traditional CBT, where treatment can run from a few weekly sessions to >20 sessions, the digital health literature has shown dose-response effects. Users of self-guided digital health courses interact with devices, not trained therapists. Until digital health platforms can unobtrusively and accurately detect symptom severity, measuring course completion rates will continue to be an inaccurate measure of efficacy.

As higher levels of engagement are associated with improved health outcomes [20,21] and due to the complex methodological issues associated with establishing efficacy rates in population-based digital health CBT programs with high reach, research should continue to focus on strategies that increase adherence and engagement [8-10,12].

**Practical Implications**

The arm 3 home page is presently the default setting for the home pages for the Overcoming Anxiety and Overcoming Depression courses. We expect to see increases in course engagement based on the use of behavioral nudges and prompts. Ongoing data collection from all members will contribute to future research.

**Strengths and Limitations**

A strength of this study is that it was conducted in an ad libitum environment. Unlike many digital health studies, a large study population in the thousands was leveraged rather than smaller groups. As participants were not aware of participating in the experiment, this limited participant bias and the Hawthorne effect.

A limitation is that, especially due to the anonymity of members, we have no way of identifying participants or validating their demographic information. Although we have no way of knowing whether the registrants are people with depression or anxiety who are seeking help, the removal of nonparticipants should mitigate the effects on the overall results.

Some Evolution Health clients promote certain course tools, or health care professionals may direct their clients to use certain platform attributes. Therefore, the tips or to-do checklist items may not be a factor in their engagement. Second, many members may simply ignore the behavioral cues and complete certain course tools based on their own preference.

The behavioral cues related to read a community post and complete the depression and anxiety test were not seen by all members. This is due to these elements being feature flags and some Evolution Health clients choosing not to offer these tools to their membership base. Although this resulted in fewer members seeing these cues and having access to the course components, the percentage of members who viewed the cues and engaged with the components was noteworthy.

As this study was designed to form a baseline for future research, there are some methodological issues that can be explored further. For example, unlike the tips presented in arm 2, the use of tips in arm 3 may or may not have been influenced by adding the to-do checklist. Alternatively, using directive tips in arm 3, rather than social proof and present bias tips, may increase the use of the to-do checklist.

Notably, our evidence indicates that nudges and prompts increase engagement in self-guided treatment programs for depression and anxiety. On the basis of these encouraging yet preliminary results, we can proceed with more sophisticated studies that will examine more enhanced strategies designed to increase platform engagement.

**Future Directions**

As mentioned earlier, there has been scant published research regarding the implementation of behavioral economic strategies designed to increase engagement in digital health programs. At minimum, this randomized controlled trial was successful in confirming member engagement through the use of these strategies.

As recommended by Aschbacher et al [68] in a recent study involving digital mental health and dose responses, machine learning models can help enable precision by analyzing
engagement patterns over time. Combined with another recent paper by Forbes et al [24] analyzing digital interventions for depression, which noted that it is important to develop standardized ways of reporting adherence and engagement so that effective comparisons across different interventions could be measured, baseline outcomes need to be established.

The baseline outcomes of this study can serve as guidelines for future studies. With machine learning models in place and with the goal of improving member outcomes and platform efficacy, Evolution Health considers the following research questions:

• Will members of addiction-focused courses (eg, managing drinking and quitting smoking) follow similar engagement patterns if nudges and prompts are made available to them? Which nudges and prompts can be used universally, and which work best for specific mental health or addiction indications?

• The average member logged into the platform 2.38 times. Can specific tips be introduced at onset to promote log-ins?

• There were no statistically significant differences in the number of tips clicked on between directive tips in arm 2 and social proof and present bias tips in arm 3 (P=.25). However, arm 2 featured 31 rotating directive tips and arm 3 featured 15 social proof tips and 15 present bias tips. Which tips were most engaging (eg, directive, social proof, or present bias tips)? Which content areas were most engaging (eg, goal setting, community themed, or specific exercises such as the depression and anxiety tests)? Which tips were the most engaging for specific age groups?

• The average arm 3 member clicked on 2.7 out of 8 to-do checklist items. Which items were those members most likely to click on first? Were there specific patterns of engagement that may influence which tips should be shown to specific members?

• The course components differ in the effort required to complete them. For example, completing a CBT session is more intensive than the few minutes required to complete the depression and anxiety test. Furthermore, one member may take several minutes to contemplate their goals, whereas others may have concrete goals already established. Future research should analyze duration in relation to each course component, as these data may be leveraged to create tailored tips or to-do content for specific user engagement patterns.

• The AVD for members who engaged with depression and anxiety course tools ranged from 6 minutes and 45 seconds, with 7.65 pages viewed, to 24 minutes and 21 seconds, with 12.56 pages viewed. In the future, AVD may be used as a benchmark, as overall engagement may be an important metric to calculate when observing dose-response relationships.

• Members arrived at the platform through the free-to-consumer program and white-label instances licensed by clients that range from employers, insurance companies, employee assistance programs, educational institutions, nonprofit organizations, for-profit health care organizations, and individual therapists. Measuring AVD from these referral sources and patterns of course tool use may assist in creating targeted engagement recommendations.

• The depression and anxiety test has been validated in a separate study, and the algorithm reports whether members qualify for 30 separate mood and anxiety disorders [48]. Symptoms related to these disorders were also collected and reported to the members. Future research should analyze these data to investigate the possible relationships between symptomology, symptom severity, and engagement patterns.

• A recent paper by Forbes et al [24] analyzing patient adherence with digital interventions for depression noted that it is important to standardize reporting adherence and engagement. Our future work will focus on establishing baseline metrics. For example, an acceptable click rate for a tip is x% or an acceptable click rate for a to-do checklist item is y%.

• The platform contains courses for addictive behaviors (eg, smoking cessation and problem drinking), in which members who complete these courses receive a certificate of completion. We are aware that some workforce members received incentives from their human resources department for completing courses and receiving certificates. It may be worthwhile to compare the course completion rates between incentivized addictive behavior courses and nonincentivized mood disorder courses.

• This experiment examined the use of behavioral prompts. Future studies may examine overcoming behavioral barriers such as bounded rationality or choice architecture.

It cannot be assumed that the outcomes observed on the Evolution Health platform will be replicated in other environments. Further research in the combined fields of digital health and behavioral economics is required.

Conclusions

Members of the Evolution Health platform’s self-guided digital health courses engaged with behavioral nudges and prompts. From this preliminary analysis, it appears that both the tips and to-do checklists increased engagement in course components. To the best of our knowledge, this is the first randomized controlled trial designed to test the implementation of behavioral nudges and prompts in web-based self-guided courses for mood disorders. The results of this study may be important because efficacy is related to increased engagement.

Owing to its novel approach, the outcomes of this study should be interpreted with caution but may be used as a guideline for future research.

Data Availability

The data sets generated during and analyzed during this study are not publicly available due to Evolution Health’s data privacy policy, terms of use, and user agreement but are available from the corresponding author on reasonable request.
Research emerging from this and ongoing data collection will be leveraged to train artificial intelligence models to better understand how to increase and encourage healthy behavior changes. Making these data publicly available may have the opposite effect and enable the development of models that can detect and target susceptible populations.

The rich data set from this study, ongoing engagement data that are continually collected, or other platform data sets can be made available to researchers interested in conducting studies for noncommercial purposes. Interested researchers are encouraged to contact Evolution Health.

Conflicts of Interest

TVM is the chief executive officer and founder of Evolution Health, the owner of the Evolution Health digital health care platform. RF is an Evolution Health board member. No funds were received for conducting this study.

Multimedia Appendix 1

Privacy policy and terms of use.

[PDF File (Adobe PDF File), 197 KB - formative_v8i1e52558_app1.pdf ]

Multimedia Appendix 2

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 37389 KB - formative_v8i1e52558_app2.pdf ]

References


33. Black M, Brunet J. A wearable activity tracker intervention with and without weekly behavioral support emails to promote physical activity among women who are overweight or obese: randomized controlled trial. J Med Internet Res 2021 Jun 16;9(12):e28128 [FREE Full text] [doi: 10.2196/28128] [Medline: 34927590]

34. What is positive behavior support? The Association for Positive Behavior Support. URL: https://www.apbs.org/pbs [accessed 2024-02-26]


60. van Mierlo T, Li X, Hyatt D, Ching AT. Demographic and indication-specific characteristics have limited association with social network engagement: evidence from 24,954 members of four health care support groups. J Med Internet Res 2017 Feb 17;19(2):e40 [FREE Full text] [doi: 10.2196/jmir.6330] [Medline: 28213340]


Abbreviations

AVD: average visit duration
CBT: cognitive behavioral therapy
CONSORT-eHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
H1: primary hypothesis

©Trevor van Mierlo, Renante Rondina, Rachel Fournier. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Time Efficiency, Reliability, and User Satisfaction of the Tooth Memo App for Recording Oral Health Information: Cross-Sectional Questionnaire Study

Palinee Detsomboonrat1*, DDS, MSc, PhD; Pagaporn Pantuwadee Pisarnturakit1*, DDS, MSc, DrPH
Department of Community Dentistry, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand
*all authors contributed equally

Corresponding Author:
Pagaporn Pantuwadee Pisarnturakit, DDS, MSc, DrPH
Department of Community Dentistry
Faculty of Dentistry
Chulalongkorn University
34 Henry Dunant Road
Patumwan
Bangkok, 10330
Thailand
Phone: 66 22188545
Fax: 66 22188545
Email: pagaporn.p@chula.ac.th

Abstract

Background: Digitalizing oral health data through an app can help manage the extensive data obtained through oral health surveys. The Tooth Memo app collects data from oral health surveys and personal health information.

Objective: This study aims to evaluate the evaluate the time efficiency, reliability, and user satisfaction of the Tooth Memo app.

Methods: There are 2 sections in the Tooth Memo app: oral health survey and personal oral health record. For the oral health survey section of the Tooth Memo app, different data entry methods were compared and user satisfaction was evaluated. Fifth-year dental students had access to the oral health survey section in the Tooth Memo app during their clinical work. The time required for data entry, analysis, and summary of oral health survey data by 3 methods, that is, pen-and-paper (manual), Tooth Memo app on iOS device, and Tooth Memo app on Android device were compared among 3 data recorders who entered patients’ information on decayed, missing, and filled permanent teeth (DMFT) index and community periodontal index (CPI), which were read aloud from the database of 103 patients by another dental personnel. The interobserver reliability of the 3 different data-entering procedures was evaluated by percent disagreement and kappa statistic values. Laypeople had access to the personal oral health record section of this app, and their satisfaction was evaluated through a Likert scale questionnaire. The satisfaction assessments for both sections of the Tooth Memo app involved the same set of questions on the app design, usage, and overall satisfaction.

Results: Of the 103 dental records on DMFT and CPI, 5.2% (177/3399) data points were missing in the manual data entries, but no data on tooth status were missing in the Android and iOS methods. Transferring data from paper to computer took an average of 55 seconds per case. The manual method required 182 minutes more than the iOS or Android methods to clean the missing data and transfer and analyze the tooth status data of 103 patients. The users, that is, 109 fifth-year dental students and 134 laypeople, expressed high satisfaction with using the Tooth Memo app. The overall satisfaction with the oral health survey ranged between 3 and 10, with an average (SD) of 7.86 (1.46). The overall satisfaction with the personal oral health record ranged between 4 and 10, with an average (SD) of 8.09 (1.28).

Conclusions: The Tooth Memo app was more efficacious than manual data entry for collecting data of oral health surveys. Dental personnel as well as general users reported high satisfaction when using this app.

(JMIR Form Res 2024;8:e56143) doi:10.2196/56143
KEYWORDS

oral health; mobile apps; personal health information; PHI; satisfaction; tooth; teeth; oral; dental; dentist; dentistry; data entry; data collection; mHealth; mobile health; app; apps; applications; periodontal; survey; questionnaire; questionnaires

Introduction

Oral diseases such as tooth decay and gum disease remain prevalent in Thailand. According to the eighth National Oral Health Survey conducted in 2017 in Thailand [1], a significant percentage of the Thai population (31.2%-73.8%) had untreated caries, and less than 20% of the population was free of gum disease. The oral health survey is a crucial epidemiological method in dental health care that helps to understand the extent and prevalence of oral health problems [2]. The survey also provides preliminary data for planning projects to promote oral health locally and nationally [3]. Thailand conducts its national oral health survey every 5 years, recording several complex oral health measures according to the World Health Organization guidelines [4]. However, collecting data on oral health can be a time-consuming and error-prone process, leading to potential inaccuracies in diagnoses and treatment plans. Fortunately, advancements in mobile technology have made it possible to streamline data collection and analysis, providing a more efficient and reliable approach to oral health management.

Collecting data using pen-and-paper can lead to errors when transferring data to an electronic database due to poor legibility, unclear handwriting, smudged or fading ink, etc. Furthermore, manually entering data from a large number of participants into a database can be time-consuming. In today’s modern era, using technological devices in health care is becoming increasingly common [5-7]. Therefore, using an app on smartphones or tablets to input data may be more convenient than using computers and can increase the speed of analyzing and summarizing data [8,9], save time in transferring data from physical documents to electronic forms [10], and minimize paper expenses [10]. In the long term, data collected in mobile apps may help in the development of a database for research and advance the understanding of the state of oral health nationwide. Mobile apps for dental health care can be useful for improving access to oral care information, promoting preventive measures, simplifying appointment scheduling, monitoring the health of children and young people, and potentially offering features for virtual consultations or teledentistry [11,12]. The implementation of mobile health has the potential to enhance the delivery of health services [12]. Unfortunately, no mobile app is currently available for collecting oral health survey data. However, many mobile apps for oral health promotion aim to increase knowledge and promote healthy oral health behaviors [13].

The Oral Health Survey Mobile Application (OHSMA) [14] was created to collect data of oral health surveys. Unfortunately, OHSMA was only available on Android devices. iOS users could only access OHSMA through a web-based platform that required internet connectivity. Dental health professionals found the app inconvenient to use because of its limited availability. Due to these issues, the use of OHSMA was discontinued. A new offline-capable app would be more beneficial for digitizing oral health survey data. It is important to note that dental history plays a significant role in forensic identification. However, obtaining patient records can be challenging because these may be spread across different hospitals and clinics [15]. A dental history record of the general population could be a potential solution to this challenge. Additionally, an individual’s oral health record could provide better insight into their past treatments, which could help dental professionals plan future dental services. A possible solution to address the challenge of oral health care in the general population is to create an individual dental history record. This record could provide valuable information to dental professionals about an individual’s past treatments, enabling them to plan better dental services for the future. Furthermore, people can maintain a personal oral health record to remember past oral health events and share it with their dentists. It is also important to note that the general Thai population does not visit dentists regularly [16], and raising awareness of oral health concerns could help encourage more regular dental checkups.

Introducing Tooth Memo—the revolutionary mobile app designed to digitize oral health survey data and personal oral health information. Unlike OHSMA [14], its predecessor, Tooth Memo is an improved version and is compatible with both Android and iOS devices. Tooth Memo can be used offline, thereby making it very useful for conducting oral health surveys in areas with limited internet connectivity, especially in rural areas. This app offers several upgrades that can help health care workers to interpret the data more efficiently. For example, this app can calculate the mean decayed, missing, and filled permanent teeth (DMFT) index for all participants in each survey instantly and notify the health care worker if the examination is incomplete or complete. Tooth Memo has not only dental health surveys but also other survey forms for dental development and research. The current version of the Tooth Memo app is designed to provide a more constructive experience for health care workers, making it easier for them to conduct oral health surveys and interpret the data more efficiently. This innovative app offers dental health professionals a convenient and user-friendly way to input data, thereby saving time and cutting down on paper expenses. Due to the addition of new design features, enhanced appearance, and increased functions in the Tooth Memo app compared to those in OHSMA, user satisfaction with this app was re-evaluated. By providing dental health professionals with a reliable and efficient tool for data collection, Tooth Memo has the potential to improve the quality of dental care and promote better oral health outcomes for all. This study was conducted to evaluate the time efficiency, reliability, and user satisfaction of the Tooth Memo app.

https://formative.jmir.org/2024/1/e56143

JMIR Form Res 2024 | vol. 8 | e56143 | p.1719

(page number not for citation purposes)
Methods

Study Population and Methodology

This cross-sectional study compares the efficiency of 3 data collection methods for oral health surveys and explores the user satisfaction with the Tooth Memo app. The 3 methods for oral health survey data collection are (1) pen-and-paper (manual), (2) Tooth Memo app in iOS (iOS), and (3) Tooth Memo app in Android (Android). The Tooth Memo app was designed for 2 types of users: dental personnel who record oral health survey data and laypeople who record their own oral health information.

Tooth Memo App

The Tooth Memo app is designed for dental professionals to easily collect and analyze oral health survey data as well as manage personal oral health records. The Tooth Memo app is an improved version of OHSMA [14], in which the pitfalls or weaknesses of OHSMA have been addressed. This app is available for installation from the App Store or Play Store in both iOS and Android devices, respectively, and supports Thai or English language use based on the device setting. iPhone or iPod touch requires iOS 13.0 or later versions. Android phones require Android 6.0 and later versions. There are 2 account types in this app: (1) dental personnel and (2) general user (Figure 1).

Figure 1. Screenshots of the (A) first page, (B) Android sign-in page, (C) iOS sign-in page, and (D) account selection in the Tooth Memo app.

The Tooth Memo app has 2 main sections. The first section, that is, oral health survey (Figure 2), allows dental professionals to record the oral health survey data according to the fourth and fifth editions of World Health Organization Oral Health Surveys-Basic Methods [4,17]. Dental personnel can record dentition status [17]; prosthetic needs; number of posterior occlusal pairs; DMFT index; and the decayed, missing, and filled permanent surfaces (DMFS) index. Gingival health can also be recorded using the community periodontal index (CPI) [4] and simplified oral hygiene index [18]. Tooth Memo provides a function for uploading individual characteristics, including name, gender, and age of each survey participant, for convenient usage. Dental personnel can collect the surveyed data on each device with or without internet connectivity. The recorded data will be retained in each data-entering device, and a summary report can be created and exported as an Excel spreadsheet. The user can also upload the name list of the sample before the survey. There are notices for incomplete data collection.
Figure 2. Screenshots of the oral health survey in the dental personnel feature: (A) first page, (B) gingival status forms, (C) dental status forms, and (D) list of participants.

Figure 3. Screenshots of the personal oral health record in the general user feature: (A) first page, (B) tooth type selection, (C) dental chart for data entry, and (D) status and treatment for each tooth.
Capability Assessment

The study aims to test the interobserver reliability of 3 different data-entering procedures, namely pen-and-paper (manual), iOS app (iOS), and Android app (Android), using the dentition status (DMFT) and gingival status (CPI) of 103 patients from the database of the Department of Community Dentistry, Chulalongkorn University. For each data-entering procedure, 3 data recorders entered each patient’s information on DMFT and CPI, while a dental personnel read out this information aloud simultaneously from the database. The data recording by pen-and-paper was transferred to the computer, and the timing for entering and summarizing the data in the manual method was recorded. The interobserver reliability was assessed using test-retest reliability (κ) and percent agreement among different methods.

Satisfaction Assessment

Oral Health Survey

1. Dental students collected oral health survey data through the Tooth Memo app during their clinical work, and their satisfaction with the app was evaluated using a questionnaire. These students had no experience with other apps for oral health surveys, although they are familiar with mobile apps. Tooth Memo is their first app for oral health surveys. Each dental student examined 4-8 patients during their clinical work.

2. In this study, all fifth-year dental students of the 2022 academic year from the Department of Community Dentistry, Faculty of Dentistry, Chulalongkorn University, were recruited.

Personal Oral Health Record

1. Laypeople who can read Thai and voluntarily participated in this study recorded their dental status and treatments in Tooth Memo, and their satisfaction was evaluated after using the app.

2. The minimum sample size for this study was estimated using the GPower 3.1 Program [19] for a 1-sample case. The 2-sided t test for difference of means from constant (1-sample case) was used for calculating the required sample size by given α (.05), power (.95), and effect size (0.318) [14]. The suggested total sample size was 131.

Overall Satisfaction

User satisfaction was evaluated with a newly developed self-administered questionnaire in Thai via a Google form. The questionnaire had undergone a thorough revision process based on the findings of a previous study [14]. Modifications were made to improve the clarity and relevance of the questions. The revised questionnaire was then pilot-tested in a sample group to evaluate its effectiveness in capturing the intended information. The feedback from the pilot test was used to further refine the questionnaire and ensure that the questions were clear, concise, and easy to understand. The questionnaire had 2 parts, with identical questions for each section of the Tooth Memo app (personal oral health record and oral health survey). Satisfaction with the design and usage of Tooth Memo and the overall satisfaction were evaluated. The satisfaction questions on the design of Tooth Memo were related to the font style, size, and color, appropriate and sufficient content in each page, continuity in content across pages, and the channel for consultation if a problem occurs. The satisfaction questions on the usage of Tooth Memo included registration, recording the data, summary and report, searching the recorded data, and loading speed. Each part of the questionnaire had nine 5-point Likert scale questions concerning users’ satisfaction. The scores ranged from 1 to 5 (1=least appropriate, 2=less appropriate, 3=moderately appropriate, 4=highly appropriate, and 5=most appropriate). Additionally, there was an 11-point rating scale (0-10 points) concerning overall satisfaction.

Data Analysis

Capability Assessment

We compared the errors incurred and time taken for data entry and data summarizing among the 3 methods. Inter-rater reliability was analyzed using Cohen kappa [20], Fleiss kappa, and percent disagreement [21]. This study follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [22] for reporting.

Satisfaction Assessment

User satisfaction was analyzed using SPSS software (version 29.0; IBM Corp) through descriptive statistics, mean, standard deviation, frequency, and percentage. The proportion of each score among the satisfaction questions was analyzed using frequency and percentage.

Ethics Approval

The study protocol was approved by the Human Research Ethics Committee of the Faculty of Dentistry at Chulalongkorn University (HREC-DCU 2023-001) before the study began. Consent was obtained from the Department of Community Dentistry, Chulalongkorn University, to access data for the research. The participants were provided with a clear information sheet outlining the project’s aims, and they were free to choose whether they wanted to participate in this study. The questionnaires were designed to be anonymous. The study participants were informed that they could withdraw from the research at any time and were not obligated to complete the questionnaire. Completing and submitting the questionnaire were considered as participants’ consent to participate in this study. As compensation for their time, a toothbrush was given to the participants.

Results

In 103 dental records, 5.2% (177/3399) data points were missing in the manual entries. However, the Android and iOS methods showed no missing data on tooth status. It is worth noting that complete CPI information was provided by all 3 methods.

Capability Assessment

Our findings showed that analyzing the dmft/DMFT (decayed, missing, and filled primary teeth/decayed, missing, and filled permanent teeth) data on iOS and Android platforms takes less than a minute. The app performed the dmft/DMFT calculations; so, the time required for data analysis is negligible. In the manual method, transferring data from paper records to the
The computer took 95 minutes, averaging approximately 55 seconds for each case. Additionally, cleaning up missing data consumed 63 minutes, while the analysis of DMFT required an additional 25 minutes. As a result, the manual method took 182 more minutes than the iOS or Android method to transfer data to the computer and analyze the data of 103 patients. The summary of the time taken for each step in the data entry methods can be found in Multimedia Appendix 1. Of the 103 patients, 42.7% (44/103) were females and 7-10 years of age, with an average age of 7.40 (SD 0.61) years. The dmft and DMFT of 103 patients were 5.32 and 0.29, respectively. dt/DT (decayed primary teeth/decayed permanent teeth), mt/MT (missing primary teeth/missing permanent teeth), and ft/FT (filled primary teeth/filled permanent teeth) were 4.83/0.23, 0.18/0, and 0.31/0.06, respectively. Moreover, 5.8% (6/103) of the patients had healthy gingival status (CPI score = 0) and 11.6% (12/103) needed scaling (CPI score = 2). The overall Fleiss kappa was 0.93 among the 3 methods, and Table 1 shows the Cohen kappa and percent disagreement among the 3 methods. The Tooth Memo app in both iOS and Android platforms showed superior results compared to those obtained using pen and paper. The iOS and Android versions of the app recorded complete data with no missing information. Additionally, there was a lower percentage of disagreement between the data collected using the app on the iOS and Android platforms.

Table 1. Cohen kappa values and percent disagreement between different methods of oral health survey data recording (n=3399).

<table>
<thead>
<tr>
<th>Methods</th>
<th>Manual versus iOS</th>
<th>Manual versus Android</th>
<th>iOS versus Android</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagreement, n (%)</td>
<td>269 (7.9)</td>
<td>243 (7.2)</td>
<td>59 (1.7)</td>
</tr>
<tr>
<td>Cohen kappa</td>
<td>0.90</td>
<td>0.91</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Satisfaction Assessment

In this study, 109 fifth-year dental students aged 21-23 years collected oral health survey data in Tooth Memo during their clinical work. Of these students, 62.4% (68/109) were males. The overall satisfaction of the users ranged between 3 and 10, with an average (SD) of 7.86 (1.46) (Figure 4A). The users were generally satisfied with the app’s design and usage, with average scores ranging from 4.0 to 4.18. The average design satisfaction scores ranged from 4.0 to 4.17, while the average usage satisfaction scores ranged from 4.06 to 4.18. The medians of the design and usage satisfaction scores were both 4. The design satisfaction scores of each question ranged from 2 to 5 (Table 2). Most fifth-year dental students used tablets rather than mobile phones, and most of them used the iOS platform. The proportions of satisfaction levels for each question are presented in Figure 4B.

Figure 4. (A) Overall satisfaction and (B) percentage of dental personnel who indicated their satisfaction level (less appropriate, moderately appropriate, highly appropriate, and most appropriate) for each question in the oral health survey section.
Table 2. Descriptive statistics of the satisfaction levels of the users who used the Tooth Memo app.

<table>
<thead>
<tr>
<th></th>
<th>Oral health survey (n=109)</th>
<th>Personal oral health record (n=134)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Range</td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Font style, size, and color are appropriate and easy to read</td>
<td>4.17 (0.74)</td>
<td>2-5</td>
</tr>
<tr>
<td>Appropriate content in each page</td>
<td>4.17 (0.69)</td>
<td>3-5</td>
</tr>
<tr>
<td>Continuity in content across pages</td>
<td>4.16 (0.71)</td>
<td>3-5</td>
</tr>
<tr>
<td>Channel for consultation if a problem occurs</td>
<td>4.00 (0.73)</td>
<td>2-5</td>
</tr>
<tr>
<td><strong>Usage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td>4.09 (0.75)</td>
<td>2-5</td>
</tr>
<tr>
<td>Recording the data</td>
<td>4.13 (0.72)</td>
<td>2-5</td>
</tr>
<tr>
<td>Summary and report</td>
<td>4.09 (0.78)</td>
<td>2-5</td>
</tr>
<tr>
<td>Searching the recorded data</td>
<td>4.06 (0.74)</td>
<td>2-5</td>
</tr>
<tr>
<td>Loading speed</td>
<td>4.18 (0.72)</td>
<td>3-5</td>
</tr>
<tr>
<td>Overall satisfaction</td>
<td>7.86 (1.46)</td>
<td>3-10</td>
</tr>
</tbody>
</table>

In our study, 134 laypeople reported their satisfaction after using the personal oral health record section of Tooth Memo. The users were 15-80 years old, and 26.9% (36/134) were males. Equal numbers of laypeople used the iOS and Android platforms to access the personal oral health record section of the Tooth Memo app. The overall satisfaction scores ranged between 4 and 10, with an average (SD) of 8.09 (1.28) (Figure 5A). The laypeople were generally satisfied with the app’s design and usage, with average scores ranging from 4.02 to 4.15. The average design satisfaction scores ranged from 4.02 to 4.16, while the average usage satisfaction scores ranged from 4.06 to 4.15. The medians of the design and usage satisfaction scores were both 4. The design satisfaction scores of each question ranged from 2 to 5, and the proportions of different satisfaction levels for each question are shown in Figure 5B.

**Discussion**

The Tooth Memo app demonstrated good reliability, high time efficiency, and high user satisfaction. Tooth Memo was developed based on the experience of OHSMA [14] and interviews with stakeholders. The Tooth Memo app has various forms for different indexes in the oral health survey. User-friendly functions were developed, such as operating without internet connectivity, a button for patients with no tooth decay, notification of completeness of data collection, and provision of only the necessary active buttons. The recorded data will be stored in each device. All recorded data and the analysis results can be exported into an Excel form.

Tooth Memo provides only the necessary active buttons at each data entry point to minimize data entry mistakes. Screenshots of the active buttons for each data entry point, that is, tooth status, treatment need of each tooth, and gingival status are shown in Multimedia Appendix 2. During the design phase, there was a concern regarding the size of the button. Eventually, it was decided that the button should be large enough to...
accommodate data collection. In mobile devices, the button size is 8x8 millimeters, while in tablets, it is 12x12 millimeters, which makes data entry easier. For process simplicity, registration is only required once for each device during the first use. These points were not included in the satisfaction assessment but were open for comments at the end of the survey. However, no comments were received regarding these points.

This app can benefit researchers, as it has various forms for recording the gingival and dental status. Screenshots of the DMFS, active buttons, and code explanations are given in Multimedia Appendix 3. The dmfs/DMFS (decayed, missing, and filled primary surfaces/DMFS) with the explanation for each code is given in Multimedia Appendix 3. There are functions for reducing the time for data collection, such as the “no tooth decay” button for patients without tooth decay (Multimedia Appendix 3) and the active button’s automatic move after each data entry. The participant name list can be uploaded in the app before going to the site for data collection so that the time spent in filling those data at the site can be saved. Moreover, Tooth Memo can display the data collection status of each participant (Figure 2D). The users who participated in the oral health survey were fifth-year dental students. Although their familiarity with using mobile apps may have influenced their satisfaction results compared to other age groups, they were able to provide valuable insights about the user interface and suggest areas for improvement.

Additionally, Tooth Memo can be an individual’s personal dental history recorder. The tooth status and dental treatment of each tooth can be recorded along with the date of examination and treatment, especially endodontic treatment, which needs further dental procedures. The scatter data of each person can be gathered in their device along with the date and place. The recorded data will benefit future treatment plans or even forensic purposes. The accuracy of recording can be enhanced if each individual can ask the dental personnel about their tooth status and treatment after their dental visit.

This study examines the capability of the “Tooth Memo” for collecting oral health survey data. The time spent on data entering was controlled by simultaneous data entry using the 3 methods. The difference between the time taken in the manual method and the time taken in the iOS or Android methods was attributed to the time spent for transferring data into a computer, cleaning the missing data, and analyzing the data. The manual method takes 182 more minutes than the iOS or Android method because a previous study [14] had already shown higher satisfaction with the mobile app than the pen-and-paper method.

All users indicated high satisfaction with the design and usage of Tooth Memo. Responses to all satisfaction questions indicated high user satisfaction for both sections of Tooth Memo. We did not compare the satisfaction rates between the manual method and the iOS/Android method because a previous study [14] had already shown higher satisfaction with the mobile app than the pen-and-paper method.

As Tooth Memo is a newly developed mobile app, users may not be familiar with the app. However, app unfamiliarity might be present only in the learning period. Nevertheless, some users did indicate low satisfaction with both sections of Tooth Memo. Some comments for improvement were related to enhancing the design by adding new interactive and attractive user-friendly features and increasing the stability of the app.

We do not have any other app to compare with our app. Our app is the first of its kind to collect oral health survey data and provide personalized individual oral health records. We are aware of other data collection apps for health [5,6,12,23-28] and non–health data [29-31], but there is no specific app like ours. There was only 1 app [14] for oral health survey data collection, but it was discontinued due to its inconvenience. Therefore, we were unable to make any comparisons. Interestingly, there are many apps [12,13,32-45] for oral health promotion that aim to promote knowledge and behaviors related to oral health; there is also an artificial intelligence app that can detect dental caries [43]. Our app follows the standard format used by most apps such as banking or other utility apps.

As per the feedback provided by some users, there are some areas for improvement in Tooth Memo’s oral health survey section. Although this app has a user-friendly interface and can record the data of a large number of participants in each survey and can record many surveys, there are some disadvantages that need to be considered. The log-in system is unstable on the new version of iOS, data files cannot be exported on certain devices, and there is no interface for iPad users. Additionally, data cannot be shared among users with the same account, and there were some miscalculations in the summary data. To improve the system, these issues need to be addressed. Despite these drawbacks, the survey provides explanations for each code, collects both dental and gingival status, and provides a summary of each participant’s data, along with the data collection status for each participant. Similar to other mobile apps such as the
electronic medical records app [46] and Ru Tan Ya app [47]. Tooth Memo enhances the effectiveness of self-care, improves continuity of care, simplifies data collection, decreases overhead costs, reduces mortality in various kinds of patients, saves time for professionals, and helps to avoid transcription errors.

The Tooth Memo mobile app is an innovative tool that empowers individuals to take control of their oral health. By providing accurate data on previous treatments, this app helps users make informed decisions about their dental care. It also serves as a helpful reminder for any untreated teeth that require attention, ensuring that users stay on top of their oral hygiene. Additionally, this app offers a convenient mobile record of an individual’s dental health history, which can be a valuable resource for those who do not regularly visit the dentist. Overall, this app is an excellent resource for anyone looking to improve their oral health and increase their oral health literacy. But Tooth Memo is not just a game changer for data collection. This app also has the potential to significantly aid in forensic identification, as a record of an individual’s oral health history can provide valuable insight into their past treatments and dental services. Furthermore, having a record of the dental history of the general population can make it easier for the health care system to obtain patient records, which are often scattered across different hospitals and clinics.

There are several mobile apps for data collection [29-31,48] for various types of research. These apps incur a lower cost and are more effective for data collection than the pen-and-paper method similar to the findings reported in our study. Digital data collection can provide more data security, accountability, and accuracy, save time, and even reduce costs.

One limitation of this study is that the Tooth Memo app is currently in the development stage and is not widely used. Therefore, we cannot conduct a long-term study yet to explore and analyze the results from different groups of participants. However, once this app is launched to the public, we will be able to gather more data and make necessary improvements to provide an effective and user-friendly app that meets the needs of the users. Since this is the first app for oral health survey data collection and personalizing individual dental history, there is no comparable information to compare the satisfaction levels of users. Nonetheless, the feedback provided by the users was useful for further improvement.

The Tooth Memo app significantly reduces the time and effort required for data entry and analysis. This app also had fewer missing data points and lesser disagreement between data entry platforms. Users expressed high levels of satisfaction with the app’s design and functionality.

Acknowledgments
We wish to express our gratitude to the Ratchadaphiseksomphot Endowment Fund, Chulalongkorn University, for funding this project. We also extend our thanks to Mr Kiratijuta Bhumichitr and his team for developing the software. We appreciate the support and assistance provided by all the participants and colleagues in completing this project. Additionally, we would like to thank Professor Lakshman Samaranayake for his invaluable suggestions during the rewriting process.

Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
PPP analyzed the data and interpreted the results. PD and PPP conceived the ideas, designed the software and this study, collected and validated the data, wrote, critically revised, and gave the final approval of this manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Summary of the time taken for each step in the data recording.
[PPTX File, 38 KB - formative_v8i1e56143_app1.pptx]

Multimedia Appendix 2
Screenshots of the active buttons for each data entry point: (A) tooth status, (B) treatment need of each tooth, and (C) gingival status.
[PPTX File, 178 KB - formative_v8i1e56143_app2.pptx]

Multimedia Appendix 3
Screenshots of (A) decayed, missing, and filled surfaces; (B) active buttons; and (C) code explanation.
[PPTX File, 175 KB - formative_v8i1e56143_app3.pptx]
References


29. Powerful and intuitive data collection tools to make an impact. The Kobo Organization. URL: https://www.kobotoolbox.org/ [accessed 2024-03-07]


Abbreviations

CPI: community periodontal index
DMFS: decayed, missing, and filled permanent surfaces
dmfs: decayed, missing, and filled primary surfaces
DMFT: decayed, missing, and filled permanent teeth
dmft: decayed, missing, and filled primary teeth
dt/DT: decayed primary teeth/decayed permanent teeth
ft/FT: filled primary teeth/filled permanent teeth
mt/MT: missing primary teeth/missing permanent teeth
OHSMA: Oral Health Survey Mobile Application
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

©Palinee Detsomboonrat, Pagaporn Pantuwadee Pisarntrurkit. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Race and Socioeconomic Status as Predictors of Willingness to Use Digital Mental Health Interventions or One-On-One Psychotherapy: National Survey Study

Lorenzo Lorenzo-Luaces1, PhD; Akash Wasil2, MA; Corinne N Kacmarek3, PhD; Robert DeRubeis4, PhD

1Indiana University-Bloomington, Bloomington, IN, United States
2Center for AI Safety, San Francisco, CA, United States
3American University, Washington, DC, United States
4University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Lorenzo Lorenzo-Luaces, PhD
Indiana University-Bloomington
1101 E 10th St
Bloomington, IN, 47405
United States
Phone: 1 8128560866
Email: lolorenz@indiana.edu

Abstract

Background: There is an ongoing debate about whether digital mental health interventions (DMHIs) can reduce racial and socioeconomic inequities in access to mental health care. A key factor in this debate involves the extent to which racial and ethnic minoritized individuals and socioeconomically disadvantaged individuals are willing to use, and pay for, DMHIs.

Objective: This study examined racial and ethnic as well as socioeconomic differences in participants’ willingness to pay for DMHIs versus one-on-one therapy (1:1 therapy).

Methods: We conducted a national survey of people in the United States (N=423; women: n=204; mean age 45.15, SD 16.19 years; non-Hispanic White: n=293) through Prolific. After reading descriptions of DMHIs and 1:1 therapy, participants rated their willingness to use each treatment (1) for free, (2) for a small fee, (3) as a maximum dollar amount, and (4) as a percentage of their total monthly income. At the end of the study, there was a decision task to potentially receive more information about DMHIs and 1:1 therapy.

Results: Race and ethnicity was associated with willingness to pay more of one’s income, as a percent or in dollar amounts, and was also associated with information-seeking for DMHIs in the behavioral task. For most outcomes, race and ethnicity was not associated with willingness to try 1:1 therapy. Greater educational attainment was associated to willingness to try DMHIs and 1:1 therapy. Income was inconsistently associated to willingness to try DMHIs or 1:1 therapy.

Conclusions: If they are available for free or at very low costs, DMHIs may reduce inequities by expanding access to mental health care for racial and ethnic minoritized individuals and economically disadvantaged groups.

(JMIR Form Res 2024;8:e49780) doi:10.2196/49780

KEYWORDS
digital mental health; ethnicity; health disparities; internet-based CBT; cognitive behavioral therapy; intervention; mental health; mental health care; race; therapy
psychiatric medication as a first-line treatment, and more likely to be forcibly detained for mental health concerns [2,3]. The existence of these biases may, understandably, lead socially and economically disadvantaged individuals to distrust traditional mental health services [4,5], such as face-to-face, one-on-one therapy (1:1 therapy), in which an individual patient receives assessment and intervention from a specialty provider. Digital mental health interventions (DMHIs) leverage technology (eg, websites and apps) to provide mental health assessment or intervention. DMHIs such as internet-based cognitive behavioral therapy have demonstrated efficacy relative to control conditions such as waiting lists and care as usual and may have roughly similar efficacy to 1:1 therapy [6]. These interventions may have the potential to reduce the public health burden of psychopathology because both the general public and health providers appear to find them acceptable for use and dissemination [7-9].

Despite the promise of DMHIs, there is an ongoing debate about whether DMHIs can reduce racial and ethnic, as well as socioeconomic, inequities in access to mental health care. DMHIs may reduce economic barriers because most popular DMHIs offer a “freemium” model, in which users can access some content for free and pay to receive the full version of the intervention. Notably, even the “premium” versions of a DMHI are generally much less expensive than other forms of treatment. The 2 most popular smartphone apps for depression and anxiety, for example, require users to pay US $13-15 a month [10,11]. Furthermore, given that many popular DMHIs offer unguided self-help [10,12], DMHIs may appeal to individuals whose trust in traditional services has been undermined.

On the other hand, racial and ethnic minoritized individuals have been poorly represented in research on DMHIs [13]. Experts have also raised concerns that lower internet access, digital health literacy, awareness of digital health interventions, availability of culturally sensitive interventions, and ability to pay for digital health interventions may exacerbate existing inequities in access to care [14]. These factors may make DMHIs less appealing rather than more appealing to members of racial and ethnic minoritized groups and individuals from lower socioeconomic status [15-17]. Before touting DMHIs as having the potential to reduce disparities in access to mental health care, there should be evidence that these interventions are equally acceptable, or more acceptable, than traditional mental health services for racial and ethnic minoritized individuals or those from lower socioeconomic statuses [18].

In order to study the acceptability of DMHIs and 1:1 therapy is to measure the self-reported willingness of individuals to engage with these interventions. While querying willingness to use as a measure of potential engagement is not equivalent to the more face valid option of offering DMHIs and 1:1 therapy to large numbers of individuals and capturing racial and ethnic and socioeconomic differences in engagement, it is more feasible. According to the Theory of Planned Behavior [18], attitudes, norms, and perceptions, such as self-reported willingness to use, can be used to predict many different types of behaviors, such as seeking mental health treatment [19,20]. Thus, the relationship between attitudes and willingness to use self-help interventions could predict future use. Understanding the demographics of individuals most likely to use DMHIs and 1:1 therapy can help target engagement efforts and, consequently, broaden the reach of these evidence-based self-help interventions to racially and ethnically marginalized groups.

Objective

We conducted a nationally representative survey on participants’ willingness to use DMHIs and 1:1 therapy. Participants were adults living in the United States and recruited through the web-based survey research tool Prolific. Participants rated their willingness to use each treatment (1) for free, as well as their willingness to pay for the treatments (2) for a small fee, (3) as a maximum dollar amount, and (4) as a percentage of their total income. At the end of the study, we gave participants the option to engage in a behavior that we observed: information-seeking about DMHIs or 1:1 therapy. We also compare the relative willingness of participants to learn more about DMHI and 1:1 therapy.

Methods

Participants

Participants (N=423) were recruited through Prolific [21], a web-based survey platform. Participants were eligible if they were aged 18 years or older and lived in the United States. We obtained a sample of adults meant to be representative of the intersection of age, race and ethnicity, and sex-assigned at birth using US census data.

Measures

Demographics

We collected information on age (in years), gender identity (male, female, and nonbinary), yearly income (in US dollars), highest educational attainment, race, and ethnicity. Race and ethnicity were combined and defined as Asian, Hispanic, non-Hispanic Black, non-Hispanic White, or other (eg, multiracial or Middle Eastern).

Internalizing Distress

We measured internalizing distress with the Kessler Psychological Distress Scale (K6) [22,23]. K6 is a 6-item scale that asks participants the frequency of distress symptoms (eg, depression and nervousness) they have experienced over the past month on a 4-point scale (0=none of the time and 4=all of the time). Scores can range from 0 to 24, with higher scores indicating higher distress. K6 has been demonstrated to have criterion validity [24] and was an internally consistent measure of internalizing distress in this sample (α=.87).

Willingness to Pay

First, participants received descriptions of unguided DMHIs and 1:1 therapy. DMHIs were described as “websites, computer programs, or smartphone apps” that “include information and exercises designed to help people learn skills that improve their...
mental health or well-being.” It was further described that in unguided DMHIs, “individuals learn content from a website or an app on their own. They do not have access to a coach or mentor.” 1:1 Therapy was defined as “counseling in which people receive support from a trained mental health professional who has completed a degree in counseling psychology, clinical psychology, or a related field.” It was further clarified that in “one-on-one therapy from a professional, one individual receives support from one therapist.”

Willingness to pay was evaluated with a series of different outcomes, which were presented in a randomized order by treatment (DMHI questions first vs 1:1 therapy questions first):

1. For free: Participants were asked to rate their agreement with the statement “I would be willing to use an unguided web-based self-help program or smartphone app for free.” For 1:1 therapy, a parallel question was used, replacing “an unguided web-based self-help program or smartphone app” with “weekly one-on-one therapy with a professional.” Responses were recorded on a 7-point Likert scale (1=strongly disagree and 7=strongly agree) with a higher score indicating a greater willingness to use.

2. Low cost: Participants were asked to rate their agreement with the statement “I would be willing to pay US $13 per month for an unguided web-based self-help program or smartphone app.” For 1:1 therapy, a parallel question was used, replacing “an unguided web-based self-help program or smartphone app” with “weekly one-on-one therapy with a professional” and replacing “US $13 per month” with “US $100 per month (US $25/session).” Responses were recorded on a 7-point Likert scale (1=strongly disagree and 7=strongly agree) with a higher score indicating a greater willingness to pay. The value for DMHIs was chosen to reflect the cost of premium versions of unguided DMHIs. For 1:1 therapy, the values were chosen to reflect the cost of therapy with insurance coverage.

3. Maximum dollar amount: Participants were asked to type “the maximum dollar amount” they would be “willing and able to pay” for “an unguided web-based self-help program or smartphone app” and for “one-on-one therapy with a professional.” Responses were recorded as dollar amounts starting from US $0 and with a maximum of US $10,000,000, although they were capped at US $800 for 1:1 therapy and US $60 for DMHIs.

4. Percentage income: Participants were asked to enter “the maximum percentage” of their monthly that they thought they would be willing and able to pay for “an unguided web-based self-help program or smartphone app” or “one-on-one therapy with a professional.” Responses were recorded as percentage amounts in a 0% to 100% range. We opted to ask participants this question, in addition to a maximum exact dollar amount, given evidence that response quality tends to be better with percentage metrics than when asking people to answer in dollar amounts [25].

**Statistical Analyses**

The data and code for all analyses are available in the Open Science Framework website [26]. All analyses were performed in the R programming language [27] using R (version 4.3.1; R Core Team) with the R Studio GUI (version 2023.6.0.421; R Studio Team) [28]. A P<.05 was chosen as the criterion for statistical significance given the exploratory nature of the study. First, we present descriptive statistics to characterize the sample demographics by race and ethnicity. For categorical and ordered variables, we present the number of individuals endorsing each level of the variable. For continuous variables, we present means, SDs, and IQRs. Next, we report simple descriptive statistics (ie, median and IQR) on the various willingness to pay metrics by race and ethnicity.

To address potential differences in willingness to use DMHI and 1:1 therapy, we regressed the various willingness to use outcomes on race and ethnicity, educational attainment, and income, controlling for age and distress. For the ordinal outcomes (ie, agreement with willingness to use for free or for a small fee), the regressions were ordinal logistic regression [29]. For the percentage and raw dollar amount outcomes, we used linear regressions. For the binary outcome (ie, the decision to learn more or not about DMHI and 1:1 therapy), the regression of interest was a binary logistic regression predicting the selection of information (yes vs no). We verified that multicollinearity was low for the variables ultimately included in our model (ie, race and ethnicity, educational attainment, income, distress, and age). We also verified the results were not sensitive to influential cases and different modeling strategies (eg, using beta regression for the bounded percentage outcomes). Additionally, for the ordinal outcomes, we verified the proportional odds assumptions of ordinal regression were met for all the variables.

R packages were used for general programming needs (conflicted [30] and tidyverse [31]), to facilitate data cleaning and analysis (MASS [32], effects [33], broom [34], psych [35], rstantix [36], and effectsize [37]), and for making for tables (gtsummary [38], labelled [39], kableExtra [40], and flextable [41]) and figures (ggpubr [42], ggpubr [43], and scales [44]). When examining willingness to pay as a maximum percentage of financial resources, we addressed outliers through winsorization. Responses that were >3 SDs above the mean were set to the value 3 SDs above the mean (6 values were winsorized for willingness to pay for unguided DMHIs and 9 values were winsorized for one-on-one psychotherapy).

**Ethical Considerations**

The University of Pennsylvania Institutional Review Board approved this study (#843424). Participants provided written informed consent to take part in the study. The study consisted of a web-based survey lasting approximately 15 minutes. The study was described as being about attitudes toward different mental health interventions. Participants were informed that their data would be collected and analyzed in an anonymous fashion. Individuals were paid US $5 for their participation in the study.

**Results**

**Demographics**

The sample appeared fairly representative of the United States in terms of age, gender identity, as well as race and ethnicity.
There were differences between the racial and ethnic groups in age, income, and educational attainment (Table 1). Descriptive statistics for the various willingness to use and pay for DMHIs and 1:1 therapy are presented in Figures 1 and 2, respectively. Willingness to use DMHIs was relatively high when they are described as being “free.” Participants were much less likely to say they would be willing to use DMHIs if they had to incur a small fee to do so. Across these outcomes, racial and ethnic differences suggested that relative to non-Hispanic White individuals, racial and ethnic minoritized individuals were more willing to use or pay for DMHIs. Similarly, willingness to use 1:1 therapy was very high when they are described as being “free.” Willingness dropped when participants were asked to pay a small monthly fee (ie, US $100 or $25 per session) for 1:1 therapy.

Table 1. Racial and ethnic differences in sociodemographic factors for adults in a nationally representative sample in the United States (N=423).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Non-Hispanic White (n=293)</th>
<th>Non-Hispanic Black (n=52)</th>
<th>Hispanic (n=25)</th>
<th>Asian (n=31)</th>
<th>Other (eg, multiracial and Middle Eastern) (n=22)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>47.15 (16.16)</td>
<td>43.19 (15.45)</td>
<td>33.56 (15.28)</td>
<td>37.45 (13.74)</td>
<td>47.18 (14.25)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Income (in US $10,000), mean (SD)</td>
<td>7.45 (4.94)</td>
<td>5.94 (4.10)</td>
<td>5.80 (3.14)</td>
<td>7.35 (5.62)</td>
<td>6.07 (6.55)</td>
<td>.04</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
</tr>
<tr>
<td>Man</td>
<td>49.83 (146)</td>
<td>46.15 (24)</td>
<td>52 (13)</td>
<td>51.61 (16)</td>
<td>50 (11)</td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>47.78 (140)</td>
<td>51.92 (27)</td>
<td>44 (11)</td>
<td>48.39 (15)</td>
<td>50 (11)</td>
<td></td>
</tr>
<tr>
<td>Nonbinary</td>
<td>2.39 (7)</td>
<td>1.92 (1)</td>
<td>4 (1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Educational attainment, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High school or less</td>
<td>10.24 (30)</td>
<td>15.38 (8)</td>
<td>4 (1)</td>
<td>6.45 (2)</td>
<td>4.55 (1)</td>
<td></td>
</tr>
<tr>
<td>Some college (eg, associate’s degree)</td>
<td>26.28 (77)</td>
<td>25 (13)</td>
<td>64 (16)</td>
<td>9.68 (3)</td>
<td>63.64 (14)</td>
<td></td>
</tr>
<tr>
<td>College graduate</td>
<td>38.23 (112)</td>
<td>42.31 (22)</td>
<td>24 (6)</td>
<td>67.74 (21)</td>
<td>18.18 (4)</td>
<td></td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>25.26 (74)</td>
<td>17.31 (9)</td>
<td>8 (2)</td>
<td>16.13 (5)</td>
<td>13.64 (3)</td>
<td></td>
</tr>
<tr>
<td>Sexual orientation, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.56</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>86.01 (252)</td>
<td>86.54 (45)</td>
<td>76 (19)</td>
<td>80.65 (25)</td>
<td>90.91 (20)</td>
<td></td>
</tr>
<tr>
<td>LGBTQ+ab</td>
<td>13.99 (41)</td>
<td>13.46 (7)</td>
<td>24 (6)</td>
<td>19.35 (6)</td>
<td>9.09 (2)</td>
<td></td>
</tr>
</tbody>
</table>

aKruskal-Wallis rank sum test; Fisher exact test for count data with simulated P value (based on 2000 replicates).
bLGBTQ+: lesbian, gay, bisexual, transgender, queer.
Figure 1. Willingness to use digital mental health interventions (DMHIs) or one-on-one therapy (1:1 therapy) for free or a low cost in a nationally representative sample of Prolific users in the United States (N=423). (A) Willing to try DMHIs for free; (B) willing to try DMHIs for a small fee (US $13); (C) willing to try 1:1 therapy for free; and (D) willing to try 1:1 therapy for a small fee (US $100).
Figure 2. Willingness to use digital mental health interventions (DMHIs) or one-on-one therapy (1:1 therapy) as a percentage of income or a raw dollar amount in a nationally representative sample of Prolific users in the United States (N=423). (A) Willingness to pay for DMHIs as a percentage of income; (B) willingness to pay for DMHIs as a raw dollar amount; (C) willingness to pay for 1:1 therapy as a percentage of income; and (D) willingness to pay for 1:1 therapy as a raw dollar amount.

Willingness to Try DMHIs or 1:1 Therapy, if Free or a Small Fee

Table 2 shows the results of ordinal logistic regressions predicting willingness to use DMHIs or 1:1 therapy, when described as for “free” or for a small fee (ie, US $13 for DMHI vs $25 a week for 1:1 therapy), from race and ethnicity, education, and income, controlling for sociodemographic factors and internalizing distress. Race and ethnicity did not predict willingness to try DMHIs or 1:1 therapy for free or for a small fee, although the differences suggested racial and ethnic minoritized individuals were more rather than less willing to try DMHIs. Educational attainment appeared associated with a greater willingness to try DMHIs or 1:1 therapy for free, or 1:1 therapy for a small fee, but the effects of educational attainment on willingness measured crossed the \( P<.05 \) threshold. Income was not associated with willingness to willingness to try DMHIs or 1:1 therapy for free. It was associated with a willingness to try DMHIs or 1:1 therapy for a small fee, although the effect was small (Table 2).
Table 2. Race and ethnicity and socioeconomic factors as predictors of willingness to try digital mental health interventions (DMHIs) versus one-on-one therapy (1:1 therapy) for free or a small fee in a nationally representative sample of US adults (N=423).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>DMHI OR^a (95% CI)</th>
<th>P value</th>
<th>1:1 therapy OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question 1: for free</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race and ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1.76 (0.88-3.52)</td>
<td>.20</td>
<td>1.29 (0.60-2.92)</td>
<td>.70</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.49 (0.69-3.25)</td>
<td></td>
<td>1.05 (0.44-2.59)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>1.58 (0.92-2.72)</td>
<td></td>
<td>0.81 (0.45-1.45)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (eg, multiracial and Middle Eastern)</td>
<td>0.78 (0.36-1.71)</td>
<td>.20</td>
<td>0.63 (0.28-1.45)</td>
<td></td>
</tr>
<tr>
<td>Distress (K6^c)</td>
<td>1.00 (0.97-1.04)</td>
<td>.04</td>
<td>1.02 (0.99-1.06)</td>
<td>.20</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.98 (0.97-0.99)</td>
<td>.002</td>
<td>0.99 (0.97-1.00)</td>
<td>.04</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college (eg, associate’s degree)</td>
<td>2.10 (1.09-4.04)</td>
<td>.08</td>
<td>1.21 (0.59-2.43)</td>
<td>.052</td>
</tr>
<tr>
<td>College graduate</td>
<td>2.26 (1.21-4.23)</td>
<td></td>
<td>1.35 (0.68-2.65)</td>
<td></td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>1.92 (0.96-3.84)</td>
<td></td>
<td>2.36 (1.10-5.06)</td>
<td></td>
</tr>
<tr>
<td>Income (in US $10,000)</td>
<td>0.97 (0.93-1.01)</td>
<td>.10</td>
<td>0.99 (0.95-1.03)</td>
<td>.50</td>
</tr>
<tr>
<td><strong>Question 2: for a small fee</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race and ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1.30 (0.61-2.73)</td>
<td>.30</td>
<td>0.87 (0.44-1.72)</td>
<td>.60</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.44 (0.63-3.25)</td>
<td></td>
<td>0.89 (0.41-1.95)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>1.75 (1.00-3.06)</td>
<td></td>
<td>1.02 (0.59-1.76)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (eg, multiracial and Middle Eastern)</td>
<td>1.37 (0.63-2.93)</td>
<td>.30</td>
<td>0.54 (0.25-1.18)</td>
<td></td>
</tr>
<tr>
<td>Distress (K6^c)</td>
<td>0.97 (0.93-1.00)</td>
<td>.051</td>
<td>1.02 (0.98-1.05)</td>
<td>.30</td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.01 (0.99-1.02)</td>
<td>.40</td>
<td>0.99 (0.98-1.00)</td>
<td>.20</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college (eg, associate’s degree)</td>
<td>0.74 (0.38-1.45)</td>
<td>.15</td>
<td>1.51 (0.79-2.88)</td>
<td>.06</td>
</tr>
<tr>
<td>College graduate</td>
<td>0.90 (0.48-1.72)</td>
<td></td>
<td>1.48 (0.79-2.79)</td>
<td></td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>0.54 (0.26-1.09)</td>
<td></td>
<td>2.43 (1.21-4.89)</td>
<td></td>
</tr>
<tr>
<td>Income (in US $10,000)</td>
<td>1.05 (1.01-1.09)</td>
<td>.02</td>
<td>1.10 (1.06-1.14)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

^aOR: odds ratio.
^bNot available.
^cK6: Kessler Psychological Distress Scale.

**Willingness to Try DMHIs or 1:1 Therapy, as Percent of Income or Raw Dollar Amount**

Tables 3 and 4 show the results of regressions predicting willingness to pay for DMHIs or 1:1 therapy as a percentage of monthly disposable income (Table 3) or as a raw dollar amount (Table 4). Compared to non-Hispanic White adults, racial and ethnic minoritized individuals were willing to pay more for DMHIs as a percentage of their income (Table 3) or as a raw dollar amount (Table 4). Examining the pairwise contrasts revealed that these differences were the largest and statistically significant when comparing non-Hispanic Black adults to non-Hispanic White adults for a percentage of their income (P<.001; Table 3) or a raw dollar amount (P<.001; Table 4), as well as for Hispanic adults for a raw dollar amount (P=.04). The other minoritized groups, Asian and Other (eg, multiracial and Middle Eastern), appeared somewhat more willing to pay but the differences were small and not statistically significant.
Educational attainment was inconsistently associated with willingness to pay. For example, it appeared unrelated to willingness to pay for either DMHIs or 1:1 therapy (Table 3). However, when willingness was assessed on a dollar metric, greater educational attainment was associated with a greater willingness to pay for 1:1 therapy, with effects most pronounced for college graduates and those with a master’s degree or greater educational attainment. By contrast, although educational attainment was associated with greater willingness to pay for DMHIs, the pairwise contrasts comparing the educational groups to those with a high school degree or less were small and not statistically significant. Income was associated with a greater willingness to pay for both DMHIs and 1:1 therapy in raw dollar amounts (Table 4) but not as a percentage of income.

Table 3. Race and ethnicity and socioeconomic factors as predictors of the percent of income willing to pay for face-to-face therapy in a nationally representative sample of US adults (N=423).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Percent income willing to pay for DMHI&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Percent income willing to pay for 1:1 therapy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (unstandardized; SE)</td>
<td>t test (d.f=412)</td>
<td>β (standardized; 95% CI)</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.53 (0.35)</td>
<td>4.41</td>
<td>0.00 (~0.09 to 0.09)</td>
</tr>
<tr>
<td>Race and ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0.32 (0.23)</td>
<td>1.40</td>
<td>0.07 (~0.03 to 0.17)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.32 (0.26)</td>
<td>1.25</td>
<td>0.06 (~0.04 to 0.16)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>0.69 (0.18)</td>
<td>3.82</td>
<td>0.19 (0.09 to 0.28)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>_&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (eg, multiracial and Middle Eastern)</td>
<td>0.01 (0.27)</td>
<td>0.05</td>
<td>0.00 (~0.09 to 0.10)</td>
</tr>
<tr>
<td>Distress (K6&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>0.01 (0.01)</td>
<td>0.66</td>
<td>0.04 (~0.07 to 0.14)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>−0.01 (0.00)</td>
<td>−1.32</td>
<td>−0.07 (~0.18 to 0.04)</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college (eg, associate’s degree)</td>
<td>−0.12 (0.22)</td>
<td>−0.54</td>
<td>−0.05 (~0.21 to 0.12)</td>
</tr>
<tr>
<td>College graduate</td>
<td>−0.16 (0.21)</td>
<td>−0.77</td>
<td>−0.07 (~0.24 to 0.10)</td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>−0.24 (0.23)</td>
<td>−1.05</td>
<td>−0.08 (~0.24 to 0.07)</td>
</tr>
<tr>
<td>Income (in US $10,000)</td>
<td>0.01 (0.01)</td>
<td>1.09</td>
<td>0.06 (~0.04 to 0.15)</td>
</tr>
</tbody>
</table>

<sup>a</sup>DMHI: digital mental health intervention.

<sup>b</sup>Not available.

<sup>c</sup>K6: Kessler Psychological Distress Scale.
Table 4. Race and ethnicity and socioeconomic factors as predictors of the raw dollar amount willing to pay for face-to-face therapy in a nationally representative sample of US adults (N=423).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Dollar amount willing to pay for DMHI</th>
<th>Dollar amount willing to pay for 1:1 therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (unstandardized; SE)</td>
<td>t test (df=412)</td>
</tr>
<tr>
<td>Intercept</td>
<td>3.26 (0.51)</td>
<td>6.37</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0.45 (0.34)</td>
<td>1.32</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.78 (0.38)</td>
<td>2.04</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>0.97 (0.26)</td>
<td>3.68</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>-b</td>
<td></td>
</tr>
<tr>
<td>Other (eg, multiracial and Middle Eastern)</td>
<td>0.30 (0.39)</td>
<td>0.76</td>
</tr>
<tr>
<td>Distress (K6)</td>
<td>-0.02 (0.02)</td>
<td>-0.94</td>
</tr>
<tr>
<td>Age (years)</td>
<td>-0.01 (0.01)</td>
<td>-1.49</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Some college (eg, associate’s degree)</td>
<td>0.20 (0.32)</td>
<td>0.63</td>
</tr>
<tr>
<td>College graduate</td>
<td>0.39 (0.31)</td>
<td>1.26</td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>0.16 (0.34)</td>
<td>0.46</td>
</tr>
<tr>
<td>Income (in US $10,000)</td>
<td>0.04 (0.02)</td>
<td>2.35</td>
</tr>
</tbody>
</table>

\( a\) DMHI: digital mental health intervention.

\( b\) Not available.

\( c\) K6: Kessler Psychological Distress Scale.

**Decision to Learn More About DMHIs and 1:1 Therapy**

Race and ethnicity predicted the decision to learn more about DMHIs \( (P=.02) \) and 1:1 therapy \( (P=.02; \) Table 5). Compared to non-Hispanic White individuals, non-Hispanic Black individuals, Hispanic individuals, and individuals classified as “Other” (eg, multiracial and Middle Eastern) were more likely to choose to learn more about both DMHIs and 1:1 therapy. Asian individuals were somewhat less likely to choose to learn about DMHIs but somewhat more likely to learn about 1:1 therapy. Educational attainment was associated with a decision to learn more about DMHIs \( (P=.02) \) but not 1:1 therapy \( (P=.30) \). Income was not associated with the decision to learn about DMHIs \( (P=.92) \) or 1:1 therapy \( (P=.20) \).
Table 5. Decision to learn more about digital mental health interventions (DMHIs) or one-on-one therapy (1:1 therapy) in a nationally representative sample of respondents (N=423).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>DMHIs OR (95% CI) P value</th>
<th>1:1 therapy OR (95% CI) P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race and ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0.85 (0.29-2.18) .02</td>
<td>1.27 (0.38-3.60) .02</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5.00 (1.85-13.4)</td>
<td>1.87 (0.47-6.23)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>1.85 (0.88-3.78)</td>
<td>4.03 (1.81-8.85)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (eg, multiracial and Middle Eastern)</td>
<td>1.26 (0.33-3.94) &lt;.001</td>
<td>1.79 (0.38-6.33) &lt;.001</td>
</tr>
<tr>
<td>Distress (K6b)</td>
<td>1.12 (1.07-1.17) &lt;.001</td>
<td>1.13 (1.07-1.19) &lt;.001</td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.10 (0.99-1.03) .40</td>
<td>0.98 (0.96-1.01) .14</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Some college (eg, associate’s degree)</td>
<td>0.96 (0.35-2.83) 1.08</td>
<td>0.35-3.58</td>
</tr>
<tr>
<td>College graduate</td>
<td>2.45 (0.97-6.90) 1.63</td>
<td>0.58-5.18</td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>2.13 (0.77-6.42) 2.28</td>
<td>0.73-7.84</td>
</tr>
<tr>
<td>Income (in US $10,000)</td>
<td>1.00 (0.95-1.06) 1.04</td>
<td>0.98-1.11 20</td>
</tr>
</tbody>
</table>

aOR: odds ratio.
bK6: Kessler Psychological Distress Scale.

Discussion

Principal Findings

Across several metrics, we found that members of racial and ethnic minoritized groups were either willing to pay more for DMHIs than non-Hispanic White participants or had similar levels of willingness. These differences were especially large when comparing non-Hispanic Black and Hispanic adults to non-Hispanic White adults. Race and ethnicity was not a consistent predictor of willingness to use 1:1 therapy. The other sociodemographic factors had somewhat more predictable relation to willingness to use DMHIs and 1:1 therapy. More educated individuals were more likely to say they would pay for 1:1 therapy or try DMHIs if the interventions were free. Income was associated with a higher willingness to pay for services, although the associations were less consistent.

Limitations

While we recruited a nationally representative sample of individuals, our sample size (N=423) may have been too low to detect small associations between sociodemographic factors and willingness to use or pay that would nonetheless be of interest. Additionally, we did not measure participants’ experiences with DMHI, which was likely low, or 1:1 therapy, which was likely higher overall [45]. Finally, while our results are informative in terms of capturing attitudes toward DMHIs and 1:1 therapy, the data we collected are self-reports of a prospective, relatively low burden behavior and may not reflect the decision to use mental health services in the real world. A notable bias of self-report data includes a limitation on self-knowledge (eg, individuals may not be aware of how they would act in the event they needed to choose between DMHIs and 1:1 therapy) as well as a sometimes low correspondence between self-report and behavior [46]. Other possible biases include overly positive responding and social desirability (eg, individuals presenting themselves as more willing to seek treatment than they are likely to do in real life). It is worth noting that for these biases to affect our results, which pertain to racial or ethnic differences, the biases would have to operate differentially across the groups we considered.

Several strengths of this study are worth noting. First, we operationalized willingness to use and willingness to pay in a variety of ways that support a similar conclusion: that racial and ethnic minoritized individuals are willing to use and pay for DMHIs. Additionally, we measured a behavioral proxy of treatment seeking—a willingness to learn more about different interventions—to further contextualize our results. Finally, we explored racial and ethnic and socioeconomic differences in a diverse sample of individuals.

Comparison With Previous Work

Our results suggest that racial and ethnic minoritized individuals are roughly equally likely, or perhaps even more likely, to use DMHIs than non-Hispanic White individuals. Previous work has suggested that racial discrimination in health care may lead members of minority groups to lose trust in health care systems [4,5]. Given this history of discrimination and inequitable treatment in health care settings, alternatives to traditional services—such as unguided digital self-help interventions—may be especially appealing to members of socioeconomically disadvantaged groups and racial and ethnic minoritized individuals. An implication of these findings is that when low...
levels of engagement are seen in DMHIs, it is unlikely that these effects are due to racial and ethnic minoritized groups having an overall low willingness to use DMHIs and may instead reflect access issues (eg, DMHI advertisement not reaching racial and ethnic minoritized individuals). Additionally, our results revealed a notable reaction to costs: around two-thirds (65.2%, 276/423) of individuals are unwilling to use DMHIs if there is even a small cost associated with the interventions. These findings reiterate how costs may be a barrier to mental health treatment and support other calls to enhance the accessibility of DMHIs for racial and ethnic minoritized individuals [14,47]. Recent research has compared strategies to increase the adoption of DMHIs in health care [48]. Future research can explore the effectiveness of these strategies on increasing adoption of DMHIS, specifically, among racial and ethnic minoritized individuals.

Conclusions

These findings do not support the concern that DMHIs appeal selectively to racial and ethnic majority members and wealthier individuals. Instead, racial and ethnic minoritized individuals indicated a greater willingness to pay for DMHIs, and income was inconsistent with willingness to pay for DMHIs. The promotion of effective and affordable DMHIS could be an important way to reduce inequities and expand access to mental health care for socially and economically disadvantaged groups. Importantly, our results suggested that the willingness to use interventions when they are delivered for free is quite high for both DMHIs and 1:1 therapy.

Data Availability

The data sets generated during and/or analyzed during this study are available in the Open Science Foundation website [26].

Conflicts of Interest

None declared.

References

8. Peipert A, Krendl AC, Lorenzo-Luaces L. Waiting lists for psychotherapy and provider attitudes toward low-intensity treatments as potential interventions: survey study. JMI Form Res 2022;6(9):e39787 [FREE Full text] [doi: 10.2196/39787] [Medline: 36112400]

15. Fiiss-Healy EA, Nagy GA, Kollins SH. It is time to REACT: opportunities for digital mental health apps to reduce mental health disparities in racially and ethnically minoritized groups. JMIR Ment Health 2021;8(1):e25456 [FREE Full text] [doi: 10.2196/25456] [Medline: 33406050]


Abbreviations

1:1 therapy: one-on-one therapy
DMHI: digital mental health intervention
K6: Kessler Psychological Distress Scale

Niki O'Brien¹, MSc; Roberto Fernandez Crespo¹, PhD; Fiona O’Driscol¹, MBChB; Mabel Prendergast¹, BSc, MBBS; Deeph Chana², MSci, PhD; Ara Darzi¹, PC, KBE; Saira Ghafur¹, MBChB, MSc

¹Institute of Global Health Innovation, Imperial College London, London, United Kingdom
²Institute for Security Science and Technology, Imperial College London, London, United Kingdom

Abstract

Background: Cybersecurity is a growing challenge for health systems worldwide as the rapid adoption of digital technologies has led to increased cyber vulnerabilities with implications for patients and health providers. It is critical to develop workforce awareness and training as part of a safety culture and continuous improvement within health care organizations. However, there are limited open-access, health care–specific resources to help organizations at different levels of maturity develop their cybersecurity practices.

Objective: This study aims to assess the usability and feasibility of the Essentials of Cybersecurity in Health Care Organizations (ECHO) framework resource and evaluate the strengths, weaknesses, opportunities, and threats associated with implementing the resource at the organizational level.

Methods: A mixed methods, cross-sectional study of the acceptability and usability of the ECHO framework resource was undertaken. The research model was developed based on the technology acceptance model. Members of the Imperial College Leading Health Systems Network and other health care organizations identified through the research teams’ networks were invited to participate. Study data were collected through web-based surveys 1 month and 3 months from the date the ECHO framework resource was received by the participants. Quantitative data were analyzed using R software (version 4.2.1). Descriptive statistics were calculated using the mean and 95% CIs. To determine significant differences between the distribution of answers by comparing results from the 2 survey time points, 2-tailed t tests were used. Qualitative data were analyzed using Microsoft Excel. Thematic analysis used deductive and inductive approaches to capture themes and concepts.

Results: A total of 16 health care organizations participated in the study. The ECHO framework resource was well accepted and useful for health care organizations, improving their understanding of cybersecurity as a priority area, reducing threats, and enabling organizational planning. Although not all participants were able to implement the resource as part of information computing technology (ICT) cybersecurity activities, those who did were positive about the process of change. Learnings from the implementation process included the usefulness of the resource for raising awareness and ease of use based on familiarity with other standards, guidelines, and tools. Participants noted that several sections of the framework were difficult to operationalize due to costs or budget constraints, human resource limitations, leadership support, stakeholder engagement, and limited time.

Conclusions: The research identified the acceptability and usability of the ECHO framework resource as a health-focused cybersecurity resource for health care organizations. As cybersecurity in health care organizations is everyone’s responsibility, there is potential for the framework resource to be used by staff with varied job roles. Future research needs to explore how it
Cybersecurity is a growing challenge for health systems worldwide as the rapid adoption of emerging technologies in health care has led to increased vulnerabilities to cyber threats; these threats can significantly erode public trust and compromise patient safety [1]. While the challenge has been increasing for some time, the COVID-19 pandemic put the health sector into the spotlight, leading to increased numbers of cyberattacks during the pandemic period. Such was the scale of activity in this period that the World Health Organization (WHO) reported a 5-fold increase in attempted cyberattacks directed at WHO staff and the wider public and called for increased vigilance globally [2]. Health care systems need to be prepared for threats to continuous operations and develop their resilience in a world with substantial and accelerated technological changes.

The health sector has unique vulnerabilities to cyberattacks, most importantly the threat to patient safety and the potential loss of human life. Other unique vulnerabilities include limited funding available (both for information computing technology [ICT] services and cybersecurity), particularly in public institutions, resulting in a lack of human resources and the ability to address legacy infrastructure across health systems. Cybercriminals have forced health and social care providers to pay large sums in ransom following cyberattacks to regain access to vital technology and systems that are essential for the day-to-day functioning and care of patients [3,4]. Excluding ransom sums demanded by hackers, the average cybersecurity incident costs a health care organization US $10 million [5].

Research into health care cybersecurity has highlighted important vulnerabilities at the global level. While there are a range of technological challenges in securing medical devices and systems, human error is the primary driver of cyber breaches. Estimates from electronic health record breaches in the United States suggest that 73% of incidents were due to poor human security (eg, carelessness or negligence and falling victim to phishing scams) [6]. In May 2021, the Irish health care system was struck by the Conti ransomware attack through a phishing email with a malicious Microsoft Excel (Microsoft Corporation) file attached, which affected more than 80% of its ICT infrastructure [7]. The implications of the attack on patients and services were substantial, as staff were locked out of systems and patient care was disrupted for several months [8]. Despite these challenges, there remains a lack of commonality of language and published global documents from multilateral organizations that provide comprehensive guidance for the health sector in strengthening cybersecurity.

Given the high stakes of maintaining secure health care ICT systems, it is critical to develop workforce awareness and training as part of a safety culture and continuous improvement within health care organizations. While there is limited research, current evidence suggests that providing cybersecurity training to staff is associated with improved cyber hygiene practices, including reduced phishing email click rates [9]. Although most cybersecurity toolkits and frameworks focus on practices across all critical sectors and are not specific to the health care context, more guidance for the health sector has been developed in recent years, including the Health Care and Public Health Sector Cybersecurity Framework Implementation Guide [10]. Nonetheless, such guidance is in its infancy, and there remain limited open-access, health care–specific resources to help organizations at different levels of maturity develop their cybersecurity practices.

In 2020, the Institute of Global Health Innovation, Imperial College London, published the Essentials of Cybersecurity in Health Care Organizations (ECHO) framework, developed through research on capacity and maturity levels across health care organizations worldwide and Delphi research with leading figures in the fields of cybersecurity, ICT, and health policy [11,12]. The ECHO framework, which includes 6 dimensions (Figure 1 [11]), outlines the most important elements for health care organizations to consider and can act as a “minimum standard” or an aspirational checklist, depending on an organization’s resources and its cyber maturity. Following its release, the research team worked with cybersecurity experts and web developers to create the web-based ECHO framework resource. This resource provides key guidance based on each component of the ECHO framework and a checklist that can be used by health care organizations to track progress. As the ECHO framework resource and checklist is designed to be used across high-, middle-, and low-income countries with varied critical infrastructure, it can also be downloaded and used offline.

This study aimed to assess the usability and feasibility of the ECHO framework resource. A secondary aim was to evaluate the strengths, weaknesses, opportunities, and threats associated with implementing the ECHO framework resource at the organizational level.
Methods

Study Design

We conducted a mixed methods, cross-sectional study of the acceptability and usability of the ECHO framework at individual health care organizations. A convergent parallel (also known as simultaneous triangulation) mixed methods design was used, which involved conducting qualitative and quantitative components simultaneously and giving them equal priority. The research team kept both components independent during data collection and analysis and only mixed data during interpretation. This design enabled the collection of different but complementary data on the same topic, which offers the benefits of reducing the limitations of qualitative or quantitative methodologies on their own, triangulating findings, and developing a fuller understanding of the research subject [13].

Ethical Considerations

The study protocol was approved by the Imperial College’s Science, Engineering, and Technology Research Ethics Committee (21IC6775). All participants were emailed information about the study, including a full participant information sheet, and asked if they would be interested in taking part. Participants were made aware that their participation is voluntary, and they were free to withdraw at any stage of the study by contacting the research team. Signed participant consent was received from all participants. The data collected were stored securely on the Imperial College London OneDrive and accessed on Imperial-owned, password-protected computers. Survey responses were pseudonymized, with geographic coordinates, IP addresses, names, and contact details removed before the data analysis, and stored in secure folders only accessible to the research team. Participants in the research were not compensated.

Model

The research model was developed based on the technology acceptance model (TAM) [14]. The TAM measurement scale assesses an individual’s acceptance of technology based on perceived usefulness and perceived ease of use, which are hypothesized to be fundamental determinants of user acceptance. Though not developed for the health sector, the TAM has been used by health researchers across disciplines to measure the acceptance of digital technologies in health care and education settings [15-17].

Survey questions were adapted from previously published scales. The survey instrument was tested for clarity and comprehensiveness with 1 methodological expert and 1 ICT professional before implementation. We included additional elements of the technology acceptance model-2 (TAM2) framework, specifically subjective norm (eg, perception of the organizational factors and individuals’ approval or disapproval), job relevance (eg, the extent of relevance of the resource to job function), and attitude (eg, extent of positive perception of the resource topic). Using a 5-point Likert scale for each construct, the quantitative survey (Multimedia Appendix 1) assessed (1) perceived usefulness, (2) perceived ease of use, (3) attitude, (4) intention to use, (5) job relevance, and (6) organizational factors or external control.

Material

Participants were provided with a URL, website link, and access password to the web-based ECHO framework resource and written instructions on how to navigate the content and checklist. Participants were also informed how they could download the content and checklist for offline use. When designing the web-based resource, care was taken to ensure the website was easy to navigate and clearly identified the 6 core dimensions of the ECHO framework (Figure 2). The website was hosted on Squarespace, with website traffic encrypted by Secure Sockets Layer (SSL).
Participants and Recruitment

Members of the Imperial College Leading Health Systems Network and other health care organizations identified through the research teams’ networks were invited to participate in the research by email. The invitation email outlined the requirement for an individual with ICT or cybersecurity oversight within the organization to participate on the organization’s behalf. Convenience sampling was used, and participants from mixed ICT roles, depending on the organizational structure, were included. The inclusion criteria were health care provider organization that uses an individual or individuals with an ICT- or cybersecurity-focused job role. Participants were excluded if they were not health care providers or did not have an ICT or cybersecurity function.

Data Collection and Analysis

Study data were collected through Qualtrics web-based surveys 1 month and 3 months from the date the ECHO framework resource was received by the participants. The survey instrument (Multimedia Appendix 1) was broken down into 2 parts. Part 1 explored technology acceptance quantitatively (based on the TAM or TAM2 framework). Part 2 explored technical and content acceptability, feasibility, and usability qualitatively (including strengths, weaknesses, opportunities, and threats analysis).

Quantitative Analysis

Quantitative data were analyzed using R software (version 4.2.1; R Foundation for Statistical Computing). Descriptive statistics of the answers for each survey question were calculated by calculating the mean and 95% CIs. To determine significant differences between the distribution of answers by comparing results from the 2 survey time points, 2-tailed t tests were used. The significance level for all statistical tests was set at a P value of <.05, and a 2-sided hypothesis was considered for all tests. Analysis was conducted by comparing all responses from surveys across time points as well as by comparing survey respondents based on whether they were public or private organizations and whether they were from high-income countries (HICs) or low- and middle-income countries (LMICs).

Qualitative Analysis

Following completion of the survey data collection on Qualtrics, we imported and qualitatively compared the long-form survey responses from the baseline, 1-month, and 3-month surveys through directed content analysis in Microsoft Excel.

The thematic analysis relied largely on the use of deductive coding, which uses a top-down approach to making connections and categorizing themes under the TAM framework. As such, the 4 nodes that formed the starting point of the analysis were the process of change, acceptability and feasibility, content appraisal, and experiences with the framework. The research team also used an inductive approach to capture additional concepts, using a line-by-line review of long-form responses to derive additional codes with regard to experiences of implementing the framework. The 2 researchers independently coded the long-form responses from the baseline (NO and FO) and 1-month and 3-month surveys (NO and MP). A third research team member (SG) was used to ensure that codes were clearly defined and being applied consistently.

Results

Participant Characteristics

In total, 16 health care organizations participated in the baseline survey. A total of 14 (87%) health care organizations went on to complete the 1-month survey and 12 (75%) completed the 3-month survey. Table 1 outlines the sector and country each

Figure 2. Essentials of Cybersecurity in Health Care Organizations (ECHO) framework resource website content (home page).
participant organization belonged. Participants represented a variety of country classifications as defined by the World Bank [18], with baseline participation from low-income countries (LICs; n=3, 18%), LMICs (n=4, 25%), middle-income countries (n=3, 18%), and HICs (n=6, 37%). Different types of health care organizations were also represented from across the public (n=9, 56%), private (n=6, 37%), and nongovernmental organization (NGO; n=1, 6%) sectors.

Table 1. Study participant population characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Baseline (N=16), n (%)</th>
<th>1 month (n=14), n (%)</th>
<th>3 months (n=12), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Canada</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Colombia</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Ethiopia</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Iceland</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Nigeria</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Norway</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Pakistan</td>
<td>3 (18)</td>
<td>2 (14)</td>
</tr>
<tr>
<td></td>
<td>Singapore</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>South Africa</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td>1 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>Thailand</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td></td>
<td>United Kingdom</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Country classification</td>
<td>LIC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3 (18)</td>
<td>2 (14)</td>
</tr>
<tr>
<td></td>
<td>LMIC&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4 (25)</td>
<td>3 (21)</td>
</tr>
<tr>
<td></td>
<td>MIC&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3 (18)</td>
<td>3 (21)</td>
</tr>
<tr>
<td></td>
<td>HIC&lt;sup&gt;d&lt;/sup&gt;</td>
<td>6 (37)</td>
<td>6 (42)</td>
</tr>
<tr>
<td>Sector</td>
<td>Public</td>
<td>9 (56)</td>
<td>9 (64)</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>6 (37)</td>
<td>4 (28)</td>
</tr>
<tr>
<td></td>
<td>NGO&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>LIC: low-income country.
<sup>b</sup>LMIC: low- and middle-income country.
<sup>c</sup>MIC: middle-income country.
<sup>d</sup>HIC: high-income country.
<sup>e</sup>NGO: nongovernmental organization.

Baseline Analysis

Participants in the baseline survey reported ECHO framework resource usefulness across 2 domains: framework effects and framework features. The framework effects considered useful were the ability to better improve cybersecurity in health care organizations, reduce threats, and protect data. Participants also noted the framework was useful in prompting the development of cybersecurity policies and protocols and in facilitating organizational evaluation of cybersecurity practices. The framework features participants considered most useful were its health industry focus, its comprehensiveness, its actionability, and the ease with which it could be understood.

Participants reported what they liked most across 3 domains: content, structure or presentation, and perspective taken. The content areas specifically highlighted as most liked were “Dimension 2: Governance” and “Dimension 3: Organizational Strategy.” On structure or presentation, respondents liked the clear, concise, and focused design of the ECHO framework resource. As reported when asked about usefulness, the health sector focus, as well as a focus on “the human side” of cybersecurity and that the ECHO framework was...
complementary to other frameworks, were most liked. Participants reported what they least liked across 2 domains: content and framework features. Several domains and components were mentioned as aspects that individual respondents did not like, along with a lack of methodology and information on how to assess cybersecurity in the resource. Some respondents also noted that there was no comparison with other frameworks within the resource. Regarding framework features, participants noted a high level and lack of detail as the aspect least liked.

**Figure 3** outlines the barriers to implementation described by participants in the baseline survey. Workforce challenges were noted based on an existing organizational commitment to other frameworks, health sector staff shortages, and ICT staff shortages. Competing high-level commitments were outlined as a governance challenge. Time and funding to implement the ECHO framework resource were also reported as challenges by respondents.

**Figure 3.** Concept map of identified barriers to implementing the framework. ICT: information computing technology.

---

**Quantitative Analysis Results**

The TAM explores the attitudes the respondents have regarding the usefulness of the technology. Across the 2 surveys, the lowest mean score on usefulness statements (4.46, SD 1.38) was in response to “Using the ECHO framework makes it easier to do my job,” while the highest mean score (5.23, SD 1.33) was for “Using the ECHO framework for cybersecurity improves the quality of work I do.” In comparing responses to “Using the ECHO framework for cybersecurity improves the quality of the work I do” between participants in HICs and LMICs between 1- and 3-month surveys, there was a trend toward significance in the LMIC subgroup analysis over time ($P = .06$), where the mean Likert score reduced from 6.12 (SD 0.99) to 4.42 (SD 1.90; **Figure 4**), but this was not statistically significant.

Responses to ease-of-use statements ranged from the lowest average Likert score (5.04, SD 1.04) in response to the statement “Interacting with the ECHO cybersecurity framework is often encouraging” to the highest score (5.27, SD 1.15) to “The ECHO cybersecurity framework provides helpful guidance in performing tasks.” Mean scores for attitude questions, including “I think the ECHO framework for cybersecurity scale up is a good/wise idea” and “I am positive towards the ECHO framework for cybersecurity,” were consistently above 5.

External control questions explored the impact of external forces on the usability and acceptability of the ECHO framework resource. While mean scores related to the statement “I have no difficulty accessing and using the ECHO framework for cybersecurity in online and/or PDF format” remained above 5 across the 1- and 3-month surveys, subgroup analysis revealed a trend toward a significant increase in mean score in the 3-month survey among HIC participants ($P = .09$), but this was not statistically significant.

Mean responses to how well the 6 components of the ECHO framework captured the cybersecurity needs of the organization were as follows: context (5.19, SD 1.36); governance (5.00, SD 1.13); organizational strategy (5.07, SD 1.23); risk management (5.08, SD 1.23); awareness, education, and training (5.23, SD 1.36); and technical capabilities (5.08, SD 1.32). While not significant in the HIC group, subgroup analysis of responses on awareness, education, and training showed a significant ($P = .04$) reduction in responses between the 1-month and 3-month survey among LMIC participants (Figure 5). Mean responses on ease of adoption were as follows: context (4.77, SD 1.33); governance (4.61, SD 1.27); organizational strategy (4.50, SD 1.50); risk management (4.42, SD 1.30); awareness, education, and training (5.04, SD 1.25); and technical capabilities (4.54, SD 1.36). In subgroup analysis, changes to the scoring of the adaptability of risk management in public sector facilities trended toward significant ($P = .06$), but this was not statistically significant.
Qualitative Analysis Results

Acceptability and Feasibility

Thematic analysis of responses describing the experience of implementing the resource outlined varied experiences between organizations. Some respondents were unable to implement the resource, while others had only partly implemented it by the time they completed the 1-month survey. Of those who had used the resource, the experience was described as positive, based on it being easy to use and understand, and negative, based on its redundancy based on existing guidelines and tools used by some health care organizations. In responses to the 3-month survey, more participants reported that they had implemented the survey and noted that it had been useful in discussions with organizational leadership but required time to communicate it to all relevant stakeholders.

Participants most liked that the ECHO resource was well structured, user-friendly, comprehensive, current, relevant, easy to understand, teachable, practical, and globally relevant. It was also noted that the checklist was straightforward. They least liked that the resource can only be used as a reference, the time needed to implement, specific dimensions (eg, context, governance, awareness, education, and training dimensions), and the checklist functionality. The need for digital literacy was also noted as a challenge in some LMIC contexts.

Process of Change

Participant responses were positive in response to the process of change questions, although some noted that they were unable to implement the ECHO framework resource in their practice. Thematic analysis of 1-month survey responses to the question “What have you learned from the ECHO framework?” included the usefulness of resources for raising awareness, usefulness as a reference guide, and its ease of use based on familiarity with other standards, guidelines, and tools (eg, NIST, ISO027001, and other frameworks). A non-ECHO–specific learning was the importance of feasible and sustainable cybersecurity planning. The 3-month survey responses reflected the same themes but additionally included that the ECHO framework was useful for developing organizational strategy.
Thematic analysis of responses to the question “Have you noticed any changes in how your organization is approaching cybersecurity?” found participants were taking cybersecurity more seriously or organizations making it more of a priority, putting more importance on the “people” aspect of cybersecurity, and in some health care organizations, implementing incident management processes and other aspects suggested in the ECHO framework.

Content Appraisal
Participants noted the context section was difficult to engage with, particularly in relation to implementation costs, cultural factors, and staff willingness.

It was noted that developing best practice in cybersecurity, as discussed in the governance dimension, sounds straightforward but seems to be the hardest thing to do without proper step-by-step guidance. Specific organizational-level challenges related to the components of the dimension were also discussed, including the challenge of developing work-from-home policies, and clinical safety assessment hindered by lack of awareness from stakeholders.

Challenges raised in operationalizing the components outlined in “Dimension 3: Organizational Strategy” included budget constraints, engaging the board, getting other stakeholders involved, and a lack of time. Specific organizational challenges related to the components of the dimension were also discussed, including developing firewall protocols, communication strategies, etc.

Challenges related to risk management were specific organizational challenges faced by individual participants, including third-party or supply chain risk, asset identification and management, and simulation, due to human resource limitations. The lack of trained personnel, lack of time, and limited budget were challenges raised by several participants.

In responses to the question on “Dimension 5: Awareness, Education, and Training,” participants noted an organizational-level challenge in engaging a large employee base in cybersecurity that is not addressed in the ECHO framework but also recognized that it was a domain often neglected. One participant also noted the practical challenges in ensuring appropriate access to systems based on training and qualifications. Again, capacity challenges and a lack of time were highlighted as barriers to implementing components within the awareness, education, and training domains of the framework.

Challenges raised in operationalizing the technical capabilities components outlined in the ECHO framework included budget constraints, both in funding cybersecurity activities or personnel in the organization and in replacing legacy systems with new technology; lack of time; and difficulty acquiring human capital with appropriate cybersecurity knowledge and skills. Some organizations also reported a challenge around the large scale of the infrastructure that requires attention as part of cybersecurity activities and the regular patching and software updates required to maintain security.

Experiences With the Framework
A limited number of participants expressed concerns about their ability to take part in the research. In baseline data collection, participants were asked, “Now that you know the detailed timeline for this feasibility study, do you think you/your organization will face difficulties in implementing the framework?” A total of 9 (56%) out of 16 participants responded “yes,” with the majority listing time and staffing resources as potential barriers to implementation. The subsequent surveys asked, “Have you had any difficulties to taking part in the study? If the answer is ‘Yes,’ what are they?” A total of 4 (29%) out of 14 participants responded “yes” in the 1-month survey, and 3 (25%) out of 12 participants responded affirmatively in the 3-month survey. Affirmative responses to the questions in the 1- and 3-month surveys came exclusively from participants in HICs, and the reasons for the difficulties in taking part were primary time barriers, followed by a lack of human and financial resources and a lack of prioritization of the study in the organization.

Discussion
Principle Findings
The ECHO framework resource was well accepted and useful for health care organizations, improving understanding of cybersecurity as a priority area in health care organizations, reducing threats, and enabling users to develop organizational planning. Participants particularly liked the ECHO framework’s health sector focus and the resource’s easiness to understand, comprehensiveness, and actionability. The mean score participants gave in the 1- to 3-month surveys to “Using the ECHO framework for cybersecurity improves the quality of the work I do” trended toward significant among LMIC participants, but these results were not statistically significant. This suggests a potential challenge in the long-term usefulness of the ECHO resource in its current format. Based on reported barriers to implementation, it is possible that continued engagement with the framework resource over time further challenges the constraints identified by diverting time and resources, which could be perceived to reduce its potential to improve the quality of work.

Although not all participants were able to implement the ECHO framework resource as part of ICT cybersecurity activities (see challenges noted in Figure 3), those who were able to implement were positive about the process of change. Learnings from the implementation process included the usefulness of the resource for raising awareness as a reference guide and that it was easy to use based on familiarity with other standards, guidelines, and tools. More broadly, it was reported that the resource was useful in driving discussions on the importance of cybersecurity with leadership. Some participants also noted that the introduction of the framework resource encouraged organizations to take cybersecurity more seriously, prioritize it, and put more importance on the “people” aspect of cybersecurity. Subgroup analysis of the statement “I have no difficulty accessing and using the ECHO framework for cybersecurity in online and/or PDF format” also revealed a trend toward a significant increase in mean score over time among HIC participants, suggesting
the resource also became easier to use over time among this cohort, but this was not statistically significant.

Participants had overarching comments on the content appraisal of the 6 dimensions of the ECHO framework. They noted that several of the sections were difficult to operationalize due to costs or budget constraints, human resource limitations, leadership support, stakeholder engagement, and limited time. Select dimension-specific challenges noted were the challenge of operationalizing governance dimension components without step-by-step guidance and the difficulty undertaking activities to secure the large scale of the infrastructure outlined in the technical capabilities dimension in some health care organizations. “Dimension 5: Awareness, Education, and Training” received the highest mean score on how well the 6 components of the ECHO framework captured the cybersecurity needs of the organization. While recognizing this domain is often neglected, the qualitative analysis identified challenges moving forward, including how to engage a large employee base in cybersecurity. It may be for this reason that a significant reduction in mean score among LMIC participants between the 1-month and 3-month surveys was identified through subgroup analysis (Figure 5).

Comparison With Previous Literature

Existing research notes the inadequacy of informatics and cybersecurity education among health care professionals [19,20]. Kamerer and McDermott [19] note that existing education calls for nurses to meet a minimal competency in informatics but does not outline the intersection between security with informatics and patient safety. However, findings from this study suggest nuances in the need for education and training within resource-limited environments with competing priorities. While participants frequently expressed their views on the importance of the awareness, education, and training component of the ECHO framework resource, they also noted major institutional capacity challenges. In quantitative analysis, despite this component scoring the highest on usefulness, participants significantly reduced usefulness scores over time. Taken together, these results highlight the need for increased understanding and research on balancing the need for cybersecurity education and training as a key priority in health care organizations with long-term priorities and capacity. Alami et al [21] note that appropriate cybersecurity can become a value-creation mechanism, suggesting cost-effectiveness and cost-benefit analyses of the resource may generate evidence on its longitudinal impact on financial expenditure and human resource time.

Barriers to implementing the ECHO framework resource described by participants, specifically limited time and funding, have also been echoed as implementation challenges in previous research. Branley-Bell et al [22] noted time pressure and fatigue as a barrier to secure behavior in the health care context, and financial barriers were cited in interviews with cybersecurity experts in Canadian and American health care organizations [23]. Our research findings expand on these more commonly reported barriers, as participants further noted workforce shortages and competing governance priorities, including the mandated use of other frameworks and overlapping compliance and reporting requirements, as barriers to implementing ECHO. Standardizing requirements and using select international standards would help eliminate the high burden of dealing with the many frameworks, regulations, and standards surrounding cybersecurity. This would also be consistent with previous digital health research that has found LMIC settings often have poor physical infrastructure and limited human resource capacity and expertise [24,25]. Further research is required to analyze these barriers in greater detail, including an exploration of where governance may be simplified and where additional human resources would be best placed to enable increased interaction with cybersecurity initiatives.

Strengths and Limitations

The research study provides a comprehensive usability and feasibility analysis of a resource for developing cyber resilience within health care organizations. Developing usable resources on this topic is of particular importance as health care organizations face increasing cyber vulnerabilities as the use of digital technology in health increases and the technology itself becomes more complex. The worldwide diversity of participant organizations in the study is a key strength. Participants were from HICs, middle-income countries, and LICs and represented public, private, and NGO providers to capture diverse perspectives and a wide range of implementation experiences. Finally, the mixed methods approach used by a multidisciplinary research team leads to more nuanced insights into the process of change, acceptability and feasibility, and resource content appraisal.

The limitations of the research must also be discussed. Although there is precedent for using small sample sizes in usability research [26], the small sample size of the study challenged the ability to determine statistical significance in subgroup analyses. Another limitation is the implementation of the framework resource within ICT teams in health care organizations. While the approach enabled homogeneity among participants, the findings are not generalizable if used by other actors within health care organizations, for example, leadership or clinically facing staff. Finally, participant organizations in the research were self-selecting, with implications for the interpretation of findings, as those with more experience and interest in health care cybersecurity may have been more likely to opt into the study and influence the findings.

Implications for Research, Practice, and Policy

Future research should build on the results of this formative research study to improve the design and content of the ECHO cybersecurity framework resource for use in health care organizations. High mean scores and positive qualitative comments by participants highlight that the resource is useful and applicable in health care organizations worldwide. However, lower mean scores on statements such as “Using the ECHO framework makes it easier to do my job,” coupled with the workforce shortages, competing governance priorities, limited time, and funding described by participants, suggest that further iteration of the resource is required to make it more responsive to these challenges. There is the potential for a larger-scale trial of version 2.0 to both gain participation from a larger group of participants internationally and validate the resource across HICs.
and LMIC contexts. However, version 2.0 may benefit from additional qualitative work to determine how ECHO could be made, at least in part, context-specific for different organizations and environments based on the ability to resource for implementation or adoption.

Previous research and academic commentary have highlighted the need for increased cybersecurity education and training [7,27]. As such, a key area of need for further investigation is how the ECHO framework resource can be used as a training tool on cybersecurity for all health care staff or organizational leadership, as well as an informational resource for ICT teams. For example, it may be possible to develop short educational snapshots on aspects of the framework that can be presented to all staff members in health care organizations through email updates; posters; other quick, attention-grabbing, and time-mindful actions; or short training courses that cover the basics of health care cybersecurity as presented in the ECHO framework. Care must be taken to ensure training courses enable staff to engage in cybersecurity topics and risk reduction from the beginning of their employment with the organization and through continuous learning over time while remaining mindful of the workloads and time required to engage in the learning. Such initiatives will require cocreation with intended users and additional testing on acceptability and feasibility.

Beyond the ECHO framework, the findings of the research study have further shown the perceived importance of cybersecurity standards, guidance, and tools for health care organizations. While there are several internationally recognized standards, guidance, and tools, their lack of focus on the health sector specifically neglects the unique aspect of cybersecurity as a critical element of patient safety and its implications for health care organizations in reducing avoidable harm. Existing standards, guidance, and tools are also often complex and technical language-heavy, which impacts their usability in low-resource health settings, often within LMIC health systems, where the level of knowledge of cybersecurity may be less advanced. As such, there is an urgent need for policy makers to fill the gap in providing targeted resources, such as standards to follow, to enable health care organizations to comply with international best practices. Policy makers must also enable an improved understanding of the unique challenges faced by health care organizations in developing more secure systems and how cyber resilience in this critical sector must feature as a priority as part of a national security agenda.

Acknowledgments

The development of the Essentials of Cybersecurity for Health Care Organizations (ECHO) resource was funded by the Qatar Foundation.

Authors’ Contributions

NO and SG conceptualized the manuscript and developed the data collection instruments. NO managed recruitment and oversaw the data collection. NO, RFC, MP, and FO completed the data analysis. NO and RFC wrote the original draft with support from SG. All authors equally contributed to the writing, reviewing, and editing. AD and SG supervised the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instruments.

[DOCX File, 24 KB - formative_v81e50968_app1.docx ]

References


Abbreviations

ECHO: Essentials of Cybersecurity for Health Care Organizations
HIC: high-income country
ICT: information computing technology
LIC: low-income country
LMIC: low- and middle-income country
NGO: nongovernmental organization
SSL: Secure Sockets Layer
TAM: technology acceptance model
TAM2: technology acceptance model-2
WHO: World Health Organization

©Niki O'Brien, Roberto Fernandez Crespo, Fiona O'Driscoll, Mabel Prendergast, Deeph Chana, Ara Darzi, Saira Ghafur. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Mental Distress, Label Avoidance, and Use of a Mental Health Chatbot: Results From a US Survey

Kristin Kosyluk¹, PhD; Tanner Baeder², MSW; Karah Yeona Greene², MSW; Jennifer T Tran¹, MEd, PhD; Cassidy Bolton¹, BS; Nele Loecher¹, MS, PhD; Daniel DiEva², BSW; Jerome T Galea², MSW, PhD

¹Department of Mental Health Law & Policy, University of South Florida, Tampa, FL, United States
²School of Social Work, University of South Florida, Tampa, FL, United States

Corresponding Author:
Kristin Kosyluk, PhD
Department of Mental Health Law & Policy
University of South Florida
13301 Bruce B Downs Boulevard
MHC 2735
Tampa, FL, 33612
United States
Phone: 1 8139746019
Email: kkosyluk@usf.edu

Abstract

Background: For almost two decades, researchers and clinicians have argued that certain aspects of mental health treatment can be removed from clinicians’ responsibilities and allocated to technology, preserving valuable clinician time and alleviating the burden on the behavioral health care system. The service delivery tasks that could arguably be allocated to technology without negatively impacting patient outcomes include screening, triage, and referral.

Objective: We pilot-tested a chatbot for mental health screening and referral to understand the relationship between potential users’ demographics and chatbot use; the completion rate of mental health screening when delivered by a chatbot; and the acceptability of a prototype chatbot designed for mental health screening and referral. This chatbot not only screened participants for psychological distress but also referred them to appropriate resources that matched their level of distress and preferences. The goal of this study was to determine whether a mental health screening and referral chatbot would be feasible and acceptable to users.

Methods: We conducted an internet-based survey among a sample of US-based adults. Our survey collected demographic data along with a battery of measures assessing behavioral health and symptoms, stigma (label avoidance and perceived stigma), attitudes toward treatment-seeking, readiness for change, and technology readiness and acceptance. Participants were then offered to engage with our chatbot. Those who engaged with the chatbot completed a mental health screening, received a distress score based on this screening, were referred to resources appropriate for their current level of distress, and were asked to rate the acceptability of the chatbot.

Results: We found that mental health screening using a chatbot was feasible, with 168 (75.7%) of our 222 participants completing mental health screening within the chatbot sessions. Various demographic characteristics were associated with a willingness to use the chatbot. The participants who used the chatbot found it to be acceptable. Logistic regression produced a significant model with perceived usefulness and symptoms as significant positive predictors of chatbot use for the overall sample, and label avoidance as the only significant predictor of chatbot use for those currently experiencing distress.

Conclusions: Label avoidance, the desire to avoid mental health services to avoid the stigmatized label of mental illness, is a significant negative predictor of care seeking. Therefore, our finding regarding label avoidance and chatbot use has significant public health implications in terms of facilitating access to mental health resources. Those who are high on label avoidance are not likely to seek care in a community mental health clinic, yet they are likely willing to engage with a mental health chatbot, participate in mental health screening, and receive mental health resources within the chatbot session. Chatbot technology may prove to be a way to engage those in care who have previously avoided treatment due to stigma.

(JMIR Form Res 2024;8:e45959) doi:10.2196/45959
Introduction

Over 10 years ago, the Annapolis Coalition for Behavioral Health Workforce Development declared a behavioral health workforce crisis in a report summarizing the key resource challenges in the field [1]. Some of the key challenges noted in this report included the observation that behavioral health care workers are too few and poorly distributed throughout the country, leaving many communities without a trained behavioral health workforce. National projections for 2013-2025 suggest that behavioral health provider supply has not kept pace with demand [2]. Simply training more professionals will not be enough to address these issues of supply and demand [3].

In total, 5 key characteristics have been proposed for service delivery models that might effectively meet the demands placed on the behavioral health care system: (1) the capacity to reach individuals not usually served, (2) scalability (the ability to implement an intervention on a large scale), (3) affordability, (4) expansion of the nonlicensed (ie, peer and paraprofessional workforce via task shifting), and (5) expansion of settings (bringing interventions to locales where those in need are likely to participate) [4]. To a great extent, these 5 key characteristics can be addressed using technology [5-7].

One specific form of technology that holds great promise for addressing behavioral health care workforce demands is chatbot technology. Chatbots are conversational interfaces (eg, Amazon Alexa and customer service chatbots used by banks or insurance companies) that use text or speech in a conversational, human-like manner to deliver information. A major strength of bots is their centralized programming (ie, its “brains”) on secure, cloud-based computing servers, which permits users to interact with the chatbot via multiple, existing platforms like SMS texts, WhatsApp, and Facebook messenger without the need to install special software or apps. Chatbots are heavily used in consumer settings because of their ability to quickly provide tailored information and increase purchasing probability. In the health sector, chatbot use is less prevalent but has been applied to mental health intervention delivery [8,9].

Researchers and clinicians have argued for almost two decades that certain aspects of treatment can be removed from clinicians’ responsibilities and allocated to technology. Research suggests, for example, that some aspects of treatment, such as exposure for panic disorders or cognitive restructuring for depression can be delivered via technology with comparable outcomes to delivery by a clinician [10,11]. Such reallocations of clinical tasks can preserve valuable clinician time and should alleviate the burden on the behavioral health care system by only requiring that services tasks where efficacy is dependent on human interaction are delivered by a human. The most recent comprehensive review of the literature summarizing existing research on blended models of service delivery—models combining some aspects of face-to-face intervention with technology-based intervention [12]—categorized blending according to both the ratio of services delivered via technology versus face-to-face and the order in which the technology or face-to-face components were delivered. One type of sequential model presents internet intervention before face-to-face intervention to engage patients during wait times and uses stepped care to reduce clinician burden. Further, 2 of the studies included in the review showed significant differences in positive outcomes (ie, symptom reduction) between those individuals engaged via technology during waiting times versus those not engaged [13,14]. However, 1 study failed to show any difference between wait-list patients engaged via the internet before face-to-face intervention and patients engaged via a self-help booklet [15]. Unfortunately, none of the studies available at the time of this review examined the cost-effectiveness of this sequential blended service delivery model, though 1 published protocol describes plans to examine the cost savings of a stepped care model [16].

Several technology-based tools exist that deliver interventions using digital therapeutics. Examples of such tools include reSET-O (Digital Therapeutics Alliance) [17], a digital therapeutic for treating opioid use disorder, and Woebot (Woebot Health) [8] and Wysa (Wysa Ltd) [9], cognitive behavioral therapy–based conversational agents for addressing a range of clinical concerns among various populations. To our knowledge, less attention has been given to developing, testing, and using conversational agents (chatbots) for other clinical tasks. In addition to aspects of treatment, other service delivery tasks that could arguably be allocated to technology to alleviate clinician burden without negatively impacting patient outcomes include screening, triage to services of appropriate intensity (as in a stepped care model), and referral. This has been attempted in the field of substance abuse via the Substance Abuse and Mental Health Services Administration’s Screening, Brief Intervention, and Referral (SBIRT) program, which applies the principles of stepped care to substance abuse and aims to funnel people in need into substance use treatment [18]. Woudles and colleagues [18] argue that technology-based delivery of SBIRT may be equally effective to human-delivered SBIRT and that technology-based delivery would increase the accessibility of this intervention and reduce the impact of stigma as a barrier to care for certain populations (eg, pregnant women using substances).

Here, we describe the results of a national survey. This survey was exploratory, designed to understand the relationship between potential mental health chatbot users’ demographics and actual chatbot use and uptake of mental health resources provided by a chatbot; the rate of completion of mental health screening when delivered by a chatbot; and the acceptability of a prototype chatbot designed for mental health screening and referral. This chatbot not only screened participants for psychological distress, but also referred them to appropriate resources (breathing exercises, self-help, peer resources, and crisis lines) that matched their level of distress and preferences. The goal of this study was to determine whether a mental health...
screening and referral chatbot would be feasible and acceptable to users. If indeed this type of technology is feasible and acceptable, it may hold promise to be tested in further research to help ease the strain on the behavioral health care system by assuming the role of screening and referral and delivering low-intensity interventions (eg, breathing exercises and self-help interventions) to those who are experiencing less distress, reserving referral to higher-intensity resources (eg, crisis lines and care from licensed clinicians) for those in greater distress. Additionally, conducting screening via chatbot technology may provide an opportunity to engage those who are currently experiencing significant distress, but are not yet willing to seek the help of a licensed clinician due to environmental barriers such as stigma, with some resources (eg, self-help or stigma reduction interventions) when otherwise they may receive no support.

Methods

Study Overview

During May and June 2021, we conducted a national, hybrid, cross-sectional, internet-based survey among a sample of US-based, English-speaking persons aged 18 years or older registered with the Amazon Mechanical Turk (MTurk) system; a subset of all survey users continued beyond the initial survey to interact with a prototype mental health chatbot and completed a second survey. Though not specifically designed as a research panel, MTurk has been increasingly used to recruit for research studies across social science disciplines [19] and is both more diverse and more attentive than college student samples [20]. Further, MTurk is largely representative of a broader population, skewing toward younger women of minority status [21].

Figure 1 displays this study’s design. MTurk’s recruitment parameters were set to offer study participation only to users meeting the above inclusion criteria. Users interested in participating completed informed consent and then completed a web-based survey collecting demographic data and mental health history (from the PhenX toolkit) [22], along with a battery of measures assessing behavioral health and symptoms, stigma (label avoidance and perceived stigma), attitudes toward treatment-seeking, readiness for change, and technology readiness and acceptance (see Measures).

Next, participants were provided a brief description of chatbots (ie, what they are, common examples of chatbots, and their utility) and invited to either connect to a prototype chatbot (called “Tabatha”) designed to screen users for psychological distress and provide mental health resources and referrals, or end study participation; compensation was not linked to chatbot use. For participants declining to use the chatbot, the reason for their decision was solicited by the choices: “I have no interest in chatbots,” “I do not have the time to use a chatbot,” “I do not know what a chatbot is,” “I do not need mental health services,” “I prefer speaking to a human about my mental health.” An “other” option was also available so participants could provide additional reasons.

Participants agreeing to use the chatbot clicked on a link in this study’s questionnaire that opened a separate internet browser window in which the Tabatha chatbot appeared with text reading, “Hello, my name is Tabatha… What name would you like to go by for our conversation today?” (Figure 2). Users could provide any name by which they wished to be identified. Tabatha then greeted the participant using their name and stated that its purpose was to provide mental health screening and resource navigation, making clear that—despite the human-like nature of the text messages—it was a computer program and not a human responding. Further, Tabatha provided users with phone numbers for emergency services if they were experiencing a crisis. Next, Tabatha administered the Patient Health Questionnaire-9 (PHQ-9), a widely used, 9-question depression screener that scores the severity of depressive symptoms on a score of 0 (none) to 27 (severe) [23]. Participants were then provided with their PHQ-9 score and an explanation of its meaning (see screenshot of PHQ-9 screener and results presentation in Figure 3 [24-27]) and provided mental health resources commensurate to their level of distress based upon mental health professional feedback. Table 1 provides the interpretation of PHQ-9 scores in terms of the level of distress, the chatbot’s response based on the PHQ-9 scores, and resources provided based on distress levels.

Participants who chose to use the chatbot were asked to respond to four acceptability statements adapted from the Acceptability of Intervention Measure [28,29]. The acceptability statements included, “this mental health chatbot is appealing to me,” “this mental health chatbot meets my approval,” “I welcome mental health screening using a chatbot,” “I like this mental health chatbot,” and “I will use a chatbot like this in the future.” Finally, Tabatha asked participants if they would like to receive a single follow-up message 2 weeks later asking whether the referrals provided were used. For participants declining Tabatha’s follow-up message, study participation ended, while participants accepting the follow-up message were asked to provide a mobile telephone number capable of receiving SMS messages (this was necessary because, up until this point, the use of Tabatha was anonymous; a telephone number allowed for direct communication with the participant’s mobile phone). Further, 2 weeks later, Tabatha contacted participants who provided a mobile number, asking whether mental health resources were used.
Figure 1. A diagram depicting procedures completed by all study participants. PHQ-9: Patient Health Questionnaire.
Figure 2. A screenshot of the Tabatha chatbot introduction received by each participant. HIPAA: Health Insurance Portability and Accountability Act.
Figure 3. Screenshots of PHQ-9 chatbot presentation and minimal depression score resources presentation. HIPAA: Health Insurance Portability and Accountability Act; PHQ-9: Patient Health Questionnaire.
Table 1. PHQ-9\textsuperscript{a} scores and corresponding responses and resources utilized in a cross-sectional survey examining the feasibility and acceptability of a mental health screening and referral chatbot among a sample of US adults.

<table>
<thead>
<tr>
<th>PHQ-9 score range</th>
<th>Depression level</th>
<th>Tabatha response</th>
<th>Resources provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>Minimal</td>
<td>Thank you for answering those questions, you did great! Based on your responses, you have a score of _____. Based on this score, you may be experiencing none or minimal depression symptoms.</td>
<td>Our focus should be on prevention methods and techniques. This resource may help you recognize symptoms of depression should you experience them in the future [24]. This video may help and give you more information about how you could be feeling [25].</td>
</tr>
<tr>
<td>5-9</td>
<td>Mild</td>
<td>Thank you for answering those questions, you did great! Based on your responses, you have a score of _____. Based on this score, you may be experiencing mild depression symptoms.</td>
<td>I would like to offer you some resources and tips to help address how you have been feeling. To learn more about what you may be experiencing and to find some helpful tips for reducing depression symptoms, visit [24].</td>
</tr>
<tr>
<td>10-14</td>
<td>Moderate</td>
<td>Thank you for answering those questions, you did great! Based on your responses, you have a score of _____. Based on this score, you may be experiencing moderate depression symptoms.</td>
<td>I would like to offer you some resources and tips to help address how you have been feeling. To learn more about what you may be experiencing and to find some helpful tips for reducing depression symptoms, visit [24].</td>
</tr>
<tr>
<td>15-19</td>
<td>Moderately severe</td>
<td>Thank you for answering those questions, you did great! Based on your responses, you have a score of _____. Based on this score, you may be experiencing moderately severe depression symptoms.</td>
<td>The Crisis Text Line may be able to further assist you with the symptoms you are experiencing. Text HOME to 741741 and a team member will support you and connect you to the appropriate resources. Or if you would prefer to speak with a person on the phone, dial 1-800-662-HELP (4357).</td>
</tr>
<tr>
<td>20-27</td>
<td>Severe</td>
<td>Thank you for answering those questions, you did great! Based on your responses, you have a score of _____. Based on this score, you may be experiencing severe depression symptoms.</td>
<td>I would like to connect you to a Crisis Support Lifeline. Please follow this link for support [26]. You can also text HOME to 741741 if you would prefer. An additional resource for you can be found here [27].</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PHQ-9: Patient Health Questionnaire.

Ethical Considerations

This study was reviewed and approved by the University of South Florida’s Institutional Review Board (IRB; STUDY002142). The IRB determined that this research was exempt from IRB oversight. Before completing the survey, participants were presented with an informed consent document. This study was granted a waiver of signed informed consent and participants were asked to acknowledge consent to participate in this study by clicking a radial box within the survey. All participants were compensated US $5.00 (the standard rate for MTurk users) for study participation regardless of chatbot use per Figure 1. The US $5.00 compensation was paid to the MTurk user through the MTurk platform. Study data were deidentified and stored in a password-protected [30] folder that could only be accessed by members of this study’s team.

Measures

Symptoms

Symptoms were measured within the chatbot session using the PHQ-9, a 9-item depression symptom checklist [23]. The 9 items are scored from 0 to 3 on a Likert scale of “not at all” to “nearly every day.” The PHQ-9 internal reliability and test-retest reliability are high [23]. The PHQ-9 is meant to be self-administered, making the chatbot administration of this screening tool a reasonable adaptation. In the survey completed before initiating the chatbot session, symptoms were assessed using the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) Level 1 Cross-Cutting Symptom Measure [31], a self-rated measure assessing mental health domains that are important across psychiatric diagnoses. This adult version of the measure contains 23 questions assessing 13 psychiatric domains, including depression, anger, mania, anxiety, somatic symptoms, suicidal ideation, psychosis, sleep problems, memory, repetitive thoughts and behaviors, dissociation, personality functioning, and substance use. Each item asks how much or how often the individual has been bothered by the specific symptom during the past 2 weeks. The measure was found to be clinically useful and to have good test-retest reliability in the DSM-5 field trials that were conducted in adult clinical samples across the United States and Canada [31].

Readiness for Change

The URICA (University of Rhode Island Change Assessment) was used to measure the stage of change (precontemplation,
contemplation, action, and maintenance) of a user of Tabatha. This 32-item scale uses a Likert scale from 1 (strongly disagree) to 5 (strongly agree) to answer statements describing how a person might feel toward treatment or approaching problems [32]. The URICA has good reliability with $\alpha$ from .79 to .89 and good construct validity supported through factor analyses [32].

**Attitudes Toward Treatment-Seeking**

Attitudes toward treatment-seeking were measured using an adapted version of the ATMHT (Attitudes Toward Mental Health Treatment) scale [33]. The scale was adapted from Fischer and Turner's Attitudes Toward Seeking Professional Psychological Help scale [34] to update language and culturally meaningful items. This 2-factor scale (beneficial attitudes and pessimistic attitudes toward mental health services) has demonstrated adequate internal consistency (beneficial attitudes toward mental health services $\alpha=.84$; pessimistic attitudes toward mental health services $\alpha=.79$), reliability, and validity. The ATMHT scale contains 20 items assessed on a 4-point Likert scale (4=strongly agree).

**Perceived Stigma**

We measured perceived stigma using the perceived devaluation-discrimination scale, a 12-item instrument asking participants to indicate the extent to which they agree with statements indicating that most people devalue individuals with mental illness (eg, most people feel that entering a psychiatric hospital is a sign of a personal failure) [35]. Participants respond using a 6-point Likert scale (6=strongly disagree). Higher scores on the perceived devaluation-discrimination scale represent greater perceived stigma. The scale has demonstrated good reliability, with $\alpha$ ranging from .86 to .88 [35] and validity [36].

**Label Avoidance**

The Self-Stigma of Seeking Help (SSOSH) scale measured label avoidance [37]. The SSOSH is a 10-item measure asking participants to answer on a 5-point Likert scale (5=strongly agree). The SSOSH has been found to have good internal consistency ($\alpha=.91$), and test-retest reliability ($r=.72$) [37].

**Technology Readiness and Acceptance**

The TRAM (Technology Readiness and Acceptance Model) is a questionnaire that incorporates both the Technology Readiness Index and the Technology Acceptance Model for a deeper understanding of the use of technology and an individual’s readiness and acceptance [38,39]. All items are measured on a 7-point Likert scale (7=strongly agree). The TRAM has good reliability for each of the subscales (optimism, $\alpha=.95$; innovativeness, $\alpha=.95$; discomfort, $\alpha=.90$; insecurity, $\alpha=.92$; perceived usefulness, $\alpha=.95$; perceived ease of use and use intention, $\alpha=.92$) [38]. The TRAM has been found to have adequate model fit.

**Acceptability**

Acceptability of the chatbot following use was measured using four items adapted from the Acceptability of Intervention Measure.

**Analytic Approach**

Feasibility was modeled descriptively based on willingness to use the chatbot. We used $\chi^2$ tests to examine the relationship between categorical demographic variables and the dichotomous outcome variable, willingness to use chatbot (yes or no). Continuous variables were examined as predictors of willingness to use the chatbot via correlations and logistic regression. Acceptability data were examined using descriptive statistics.

**Results**

**Overview**

Guidelines for Transparent Reporting of Evaluations with Nonrandomized Designs were followed. This study enrolled 640 individuals; 329 individuals were included in the analyses. Correlations between all variables for the total sample can be found in Table S1 of Multimedia Appendix 1. Further, 80 participants were excluded from analyses as they answered 1 of the 3 attention check questions dispersed throughout the survey incorrectly. These questions were included to ensure that participants were paying attention to this study and not just selecting random responses and that the participant was not a bot. These 3 items consisted of a question asking the participant to spell the word “horse” backward, a question asking the participant to describe a picture of a picnic scene, and a question asking the participant to respond with “somewhat agree.” In total, 71 participants were excluded because they responded to the survey more than once, even given explicit instructions that they could only respond to the survey 1 time. Further, 160 individuals were excluded from analyses as they had taken less than 10 minutes to complete the survey. Table 2 presents the demographic characteristics of participants.
Table 2. Demographic characteristics of a sample of US adults from a cross-sectional survey examining the feasibility and acceptability of a mental health screening and referral chatbot.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Agreed to chatbot use</th>
<th>Declined chatbot use</th>
<th>Chi-square (df)</th>
<th>P value</th>
<th>Cramér V</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>119 (36.2)</td>
<td>85 (71.4)</td>
<td>34 (28.6)</td>
<td>139.3 (2)</td>
<td>.19</td>
<td>0.10</td>
</tr>
<tr>
<td>Male</td>
<td>209 (63.5)</td>
<td>137 (65.6)</td>
<td>72 (34.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intersex</td>
<td>0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None of these</td>
<td>0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>1 (0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>206 (62.6)</td>
<td>135 (65.5)</td>
<td>71 (34.5)</td>
<td>5.98 (4)</td>
<td>.20</td>
<td>0.14</td>
</tr>
<tr>
<td>Woman</td>
<td>117 (35.6)</td>
<td>84 (71.8)</td>
<td>33 (28.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonbinary</td>
<td>1 (0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transgender</td>
<td>1 (0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None of these</td>
<td>2 (0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>2 (0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian c</td>
<td>12 (3.6)</td>
<td>5 (41.7)</td>
<td>7 (58.3)</td>
<td>11.82 (5)</td>
<td>.04</td>
<td>0.19</td>
</tr>
<tr>
<td>Black or African American d</td>
<td>45 (13.7)</td>
<td>34 (75.6)</td>
<td>11 (24.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White d</td>
<td>262 (79.6)</td>
<td>179 (68.3)</td>
<td>83 (31.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>1 (0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not know</td>
<td>0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>2 (0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>6 (1.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino c</td>
<td>81 (24.6)</td>
<td>67 (82.7)</td>
<td>14 (17.3)</td>
<td>13.16 (2)</td>
<td>&lt;.001</td>
<td>0.20</td>
</tr>
<tr>
<td>Not Hispanic or Latino d</td>
<td>247 (75.1)</td>
<td>155 (62.8)</td>
<td>92 (37.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not know</td>
<td>0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>1 (0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>222 (67.5)</td>
<td>160 (72.6)</td>
<td>62 (27.9)</td>
<td>11.40 (7)</td>
<td>.12</td>
<td>0.19</td>
</tr>
<tr>
<td>Divorced</td>
<td>15 (4.6)</td>
<td>9 (60)</td>
<td>6 (40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>76 (23.1)</td>
<td>44 (58.7)</td>
<td>31 (41.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A member of an unmarried couple</td>
<td>11 (3.3)</td>
<td>6 (54.5)</td>
<td>5 (45.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>2 (0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separated</td>
<td>1 (0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>2 (0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Do you have dependents or children? n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>10.94 (2)</td>
<td>&lt;.01</td>
<td>0.18</td>
</tr>
<tr>
<td>Yes c</td>
<td>201 (61.1)</td>
<td>148 (73.6)</td>
<td>53 (26.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Inferential statistics
Agreed to chatbot use Declined chatbot use Chi-square (df) P value Cramér V

<table>
<thead>
<tr>
<th>Nod</th>
<th>Total</th>
<th>Agreed to chatbot use</th>
<th>Declined chatbot use</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>124 (37.7)</td>
<td>73 (58.9)</td>
<td>51 (41.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>4 (1.2)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

**Level of education**

- High school graduate: 20 (6.1) | 13 (65) | 7 (35) | 11.20 (9) | .26 | 0.19
- Some college, no degree: 28 (8.5) | 16 (57.1) | 12 (42.9) |
- Associate degree: 25 (7.6) | 17 (68) | 8 (32) |
- Master’s degree: 58 (17.6) | 129 (70.1) | 55 (29.9) |
- Bachelor’s degree: 184 (55.9) | 42 (72.4) | 16 (27.6) |
- No high school diploma: 4 (1.2) | — | — |
- GED or equivalent: 4 (1.2) | — | — |
- Professional school degree: 2 (0.6) | — | — |
- Doctoral degree: 2 (0.6) | — | — |
- Do not know: 0 (0) | — | — |
- Prefer not to answer: 2 (0.6) | — | — |

**Employment**

- Employed: 299 (90.9) | 208 (69.6) | 91 (30.4) | 9.82 (1) | .007 | 0.17
- Unemployed: 27 (8.2) | 14 (51.9) | 13 (48.1) |
- Prefer not to answer: 3 (0.9) | — | — |

**Do you have health insurance coverage?**

- Yesc: 273 (83) | 195 (71.4) | 78 (28.6) | 12.49 (3) | <.01 | 0.20
- Nod: 49 (14.9) | 24 (49) | 25 (51) |
- Prefer not to answer: 7 (2.1) | — | — |

---

aCell counts <5 not presented for χ² tests.
bNot available.
cThere was a statistically significant difference in the proportion of individuals agreeing to use the chatbot versus not agreeing between this category and those marked with "d."
dThere was a statistically significant difference in the proportion of individuals agreeing to use the chatbot versus not agreeing between this category and those marked with "c."

GED: General Educational Development.

**Feasibility**

Table 2 provides the proportions of participants willing to use the chatbot by demographic characteristics. Of the 329 individuals included in the analyses, 222 (67.4%) agreed to use the chatbot. In terms of racial and ethnic variables, we found a significantly greater proportion of Black and White versus Asian participants (n=319, χ² =11.82; P<.05) and a greater proportion of Hispanic versus non-Hispanic participants agreed to use the chatbot (n=328, χ² =13.16; P<.001). We found that other demographic variables also contributed, with a significantly larger proportion of participants who had dependents versus no dependents (n=325, χ² =10.94; P<.01), those with insurance versus without insurance (n=322, χ² =12.49; P<.01), and those who were employed versus unemployed (n=326, χ² =9.82; P<.01) agreeing to use the chatbot.

Table 3 presents the mental health characteristics of participants and the proportions of participants willing to use the chatbot by mental health characteristics. Current mental health variables were related to agreeing to use the chatbot with a significantly greater proportion of those who were currently receiving treatment versus not currently receiving treatment (n=325, χ² =15.13; P=.001), those with a diagnosis versus no diagnosis (n=322, χ² =9.89; P<.01), and those currently distressed versus not currently distressed (n=312, χ² =21.69; P<.001) agreeing to use the chatbot. All effect sizes for these findings were small, with Cramér V ranging between 0.17 and 0.26. For individual effect sizes, refer to Table 3.

---

https://formative.jmir.org/2024/1/e45959
Correlations between continuous variables and willingness to use the chatbot for the entire sample and for those individuals who indicated that they were currently experiencing distress can be found in Multimedia Appendix 1. Table 4 displays the results of forward stepwise logistic regressions with willingness to use the chatbot as the outcome variable for the overall sample as well as for those individuals who indicated that they were currently experiencing distress. Significant correlates of chatbot use were included as predictor variables in the regression models. For the overall sample, the perceived usefulness of the chatbot and symptoms significantly predicted chatbot use, with greater perceived usefulness and greater symptoms predicting the likelihood that the participant agreed to use the chatbot. For the distressed sample, higher levels of label avoidance significantly predicted a greater likelihood of agreeing to use the chatbot.

Table 3. Mental health characteristics of a sample of US adults from a cross-sectional survey examining the feasibility and acceptability of a mental health screening and referral chatbot.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Agreed to chatbot use</th>
<th>Declined chatbot use</th>
<th>Inferential statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Are you currently receiving mental health treatment (ie, therapy, medication, and peer support)?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes(^b)</td>
<td>109 (33.1)</td>
<td>89 (81.7)</td>
<td>20 (18.3)</td>
<td>15.13 (2)</td>
</tr>
<tr>
<td>No</td>
<td>216 (65.7)</td>
<td>131 (60.6)</td>
<td>85 (39.4)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>4 (1.2)</td>
<td>_c</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td><strong>Have you ever received a diagnosis of a mental health condition from a doctor or counselor?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes(^b)</td>
<td>116 (35.3)</td>
<td>91 (78.4)</td>
<td>25 (21.6)</td>
<td>9.89 (2)</td>
</tr>
<tr>
<td>No(^d)</td>
<td>206 (62.6)</td>
<td>127 (61.7)</td>
<td>79 (38.3)</td>
<td></td>
</tr>
<tr>
<td>Unsure</td>
<td>7 (2.1)</td>
<td>4 (57.1)</td>
<td>3 (42.9)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>_</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td><strong>Are you currently experiencing any mental distress?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes(^b)</td>
<td>83 (25.3)</td>
<td>73 (88)</td>
<td>10 (12)</td>
<td>21.69 (2)</td>
</tr>
<tr>
<td>No(^d)</td>
<td>229 (69.6)</td>
<td>140 (61.1)</td>
<td>89 (38.9)</td>
<td></td>
</tr>
<tr>
<td>Unsure</td>
<td>17 (5.2)</td>
<td>9 (52.9)</td>
<td>8 (47.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>_</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td><strong>DSM-5 Cross Cutting Symptoms, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30.0 (22.4)</td>
<td>_</td>
<td>_</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Cell counts <5 not presented for \(\chi^2\) tests.

\(^b\)There was a statistically significant difference in the proportion of individuals agreeing to use the chatbot versus not agreeing between this category and those marked with "d.”

\(^c\)not available.

\(^d\)There was a statistically significant difference in the proportion of individuals agreeing to use the chatbot versus not agreeing between this category and those marked with “b.”

\(^e\)DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.
Table 4. Predictors of chatbot use among a sample of US adults from a cross-sectional survey examining the feasibility and acceptability of a mental health screening and referral chatbot.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SE (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full sample (N=329)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.05</td>
<td>0.42</td>
<td>.59</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>0.27</td>
<td>0.09 (1.10-1.56)</td>
<td>.002</td>
</tr>
<tr>
<td>DSM-5 CC(^c) symptoms</td>
<td>0.02</td>
<td>0.01 (1.01-1.03)</td>
<td>.002</td>
</tr>
<tr>
<td><strong>Distressed participants (n=83)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.61</td>
<td>1.72</td>
<td>.59</td>
</tr>
<tr>
<td>Label avoidance</td>
<td>0.13</td>
<td>0.06 (1.01-1.29)</td>
<td>.04</td>
</tr>
</tbody>
</table>

\(^a\chi^2=28.10; P<.001.\)
\(^b\) not available.
\(^c\) DSM-5 CC: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition cross cutting symptoms.
\(^d\chi^2=4.16, P<.01.\)

Acceptability
Of the 222 participants who agreed to use the chatbot, 168 (75.7%) completed the PHQ-9 screening questions in the chatbot, and 164 (73.9%) answered all of the acceptability items. The average depression score was 8.6 (SD 7.2), which constitutes mild depression among the sample of respondents who chose to complete the PHQ-9 screening within the chatbot session. Overall, participants agreed with the acceptability of the chatbot. The mean acceptability score was 19.8 (SD 4.2).

Participants who chose not to use the chatbot (n=107) indicated that they chose not to do so because they did not believe they needed mental health services (n=40), they had no interest in chatbots (n=36), they preferred to speak to a human about their mental health (n=33), they did not have time to use a chatbot (n=20), they did not know what a chatbot is (n=8), or they indicated some other reason (n=3). Frequencies of the reasons participants selected for not using the chatbot can be found in Table S2 of Multimedia Appendix 1. If they agreed to use the chatbot, participants were asked if they would be willing to provide their phone numbers within the chatbot session so that the chatbot could follow up with them in 2 weeks to see if any resources they received were used and whether they were helpful. Further, 63 (28.4% of those participants who agreed to use the chatbot) individuals agreed to provide their phone numbers. Furthermore, 56 (25.2% of those who agreed to use the chatbot) individuals agreed to allow the chatbot to follow up with them in 2 weeks; however, not all of these individuals provided a phone number for follow-up purposes. The reasons participants provided for not wanting to provide their phone numbers can be found in Table S3 of Multimedia Appendix 1, along with frequencies for each reason. Reasons provided for not wanting to provide one’s phone number included “my contact information is private” (n=61), “I do not give my number to strangers” (n=34), “I do not want unsolicited calls” (n=32), “it does not feel confidential” (n=20), “I do not want to give my number to a robot” (n=20), “I do not trust chatbots” (n=3), or “another reason not listed” (n=8).

Of the 56 individuals who received a follow-up text at 2 weeks, 9 (16.1%) individuals clicked on the link in the follow-up text to engage with the follow-up chatbot. Further, 4 (7.1%) individuals stated that they did not use the resources. Reasons included “I didn’t have time to look them over” (N=2) and “I didn’t feel like I needed them” (N=1). Of the 5 people who reported using the resources, all stated they were helpful. Further, 1 participant reported that the resources they received “…really helped me to move on to the next stage.” Another stated, “it was good advice.”

Discussion
Principal Findings
We found in this exploratory study that a chatbot designed to screen for mental distress and refer to appropriate resources matching one’s distress level was feasible and acceptable among a convenience sample of US adults. Moreover, label avoidance was the single significant, positive predictor of chatbot use among distressed participants. Most participants (n=222, 67.4%) were willing to try the chatbot, indicating that using a chatbot to screen for mental distress and connect patients or participants to resources is feasible. Of the 222 individuals who agreed to use the chatbot, 168 (75.7%) completed the PHQ-9 screening within the chatbot session, and 164 (73.9%) persisted in completing the acceptability questionnaire at the end of the chatbot interaction. Again, these findings speak to the feasibility of using chatbot technology to facilitate mental health screening and referral.

A small number of participants agreed to provide their phone number within the chatbot session for follow-up purposes and a small subgroup of these participants clicked on the link within the follow-up text they received 1 month after completing the initial survey and chatbot session. Understanding the reasons for not using the chatbot and not providing one’s phone number are important to modify the chatbot and develop messaging and marketing around the chatbot that will address potential user concerns. The most cited reasons for not using the chatbot included a perception that one did not need mental health screening and referral chatbot.
services, did not have an interest in chatbots, and preferred to speak with a human about one’s mental health. To address some of these points, future iterations of the chatbot might include attention to the marketing of the chatbot, incorporating messaging around the need for mental health screening for preventative care as well as helping potential users to understand that the chatbot is not meant to replace human interaction and screening may result in referral to a human such as a peer or a licensed mental health professional. The top reasons for not providing one’s phone number to enable follow-up included a desire to keep one’s contact information private, a policy that one does not share one’s number with strangers, and not wanting solicited calls. Future iterations of this chatbot may entertain other possible mechanisms for follow-up as an alternative to providing a phone number and more clear messaging about how a phone number would be used (ie, only for a follow-up text as opposed to a call).

Overall, the participants who did agree to try the chatbot found it acceptable. The feasibility and acceptability findings of this study are in line with recent work by Shah and colleagues [40] to develop and iteratively test a chatbot to screen users for eating disorders and refer them to care. Demographic predictors of chatbot use included being White or Black or African American, identifying as Hispanic or Latino, having dependents, having insurance coverage, being employed, having used mental health services in the past, having received a diagnosis of a mental health condition, and reporting current distress. Taking these demographic characteristics into consideration when designing future iterations of the chatbot will be crucial to making sure the chatbot is available to those who need it, is culturally relevant and responsive, and is marketed appropriately. Positive correlates of chatbot use among the full sample included technology discomfort and insecurity, symptoms, beneficial and pessimistic attitudes toward treatment-seeking, perceived usefulness of the chatbot, and reported intentions to use the chatbot. Among the full sample, the only significant predictors of chatbot use were perceived stigma and DSM-5 cross-cutting symptoms. Among those participants who endorsed current psychological distress, the only significant predictor of chatbot use was label avoidance.

Limitations

This feasibility and acceptability study had limitations. First, we recruited participants through Amazon’s MTurk, which limits the generalizability of these findings. Second, we did lose a substantial portion of our sample due to either inattention to or an unreasonably short time taken to complete the survey. On the one hand, this may lead one to question the integrity of the data provided by participants of MTurk. On the other hand, we have greater confidence in the quality of the data that we did include in our analyses due to the data-cleaning procedures that we applied. Finally, the effect sizes for our statistical tests ($\chi^2$ tests in particular) are small. A replication of this study with a larger sample is called for to confirm these findings.

Conclusions

This study found that label avoidance was the single significant, positive predictor of chatbot use among distressed participants. The existing research literature suggests that label avoidance, the desire to avoid mental health services to avoid the stigmatized label of mental illness, is a significant negative predictor of care-seeking [41,42]. Therefore, our finding regarding label avoidance and chatbot use has significant public health implications in terms of facilitating access to mental health resources and stigma reduction programs and messaging. Those who are high on label avoidance are not likely to seek care in a community mental health clinic, yet they are likely willing to engage with a mental health chatbot, participate in mental health screening, and receive mental health resources within the chatbot session.

In 2020, among the 52.9 million adults living in the United States with any mental illness, less than half (46.2%) received mental health services in the past year [43]. This technology-facilitated approach to connecting people with mental health resources holds promise to reach the nearly 54% of US adults living with a mental illness who are currently going without care and presents an opportunity for intervention as well. Research suggests that many adults in the United States living with the most common mental health condition, major depressive disorder, are receiving treatment from their primary care provider [44]. This is an important consideration for future work with this chatbot for a couple of reasons. First, primary care offices may be a deployment site for a chatbot like the one we tested. This could help ensure behavioral health screening and referral and minimize demand on primary care providers. Second, the use of technology for screening and referral in primary care settings may help to eliminate the possible effects of clinician bias on behavioral health screening and referral [45].

Stigma reduction interventions (ie, videos of people sharing their stories of mental health challenges, service usage, and recovery) might be delivered via the chatbot to support those who need it to seek mental health services from a licensed clinician. Research suggests that video-based contact is an effective means of reducing the stigma surrounding mental illness [46]. Additionally, motivational interviewing strategies might be used within the chatbot session to move users through the transtheoretical model of health behavior change [47], toward a decision to seek evidence-based treatments. Others have demonstrated the possibility of deploying motivational interviewing conversational sequencing using chatbot technology for stress management [48] and smoking cessation [49]. Chatbot technology may prove to be a way to engage those in care who have previously avoided treatment due to stigma, which could have a significant public health impact by way of shrinking the mental health treatment gap.

The future vision for Tabatha, given the feasibility and acceptability of the prototype chatbot, is as a conversational agent aimed at assuming the service delivery tasks of screening, referral, and stigma reduction. Tabatha might be deployed in various settings with diverse populations, and refer to a plethora of existing resources, including other conversational agents designed to deliver digital therapy (ie, reSET-O, Woebot, and Wysa). Future research should focus on examining Tabatha’s effectiveness in reducing stigma and navigating users to existing behavioral health resources of various intensities. The current Tabatha prototype is a rule-based chatbot, as opposed to a
machine-learning chatbot [50]. Future research might consider how machine learning might be used to facilitate screening, referral, and stigma reduction through Tabatha.

Acknowledgments

KK and JTG designed this study. TB and DD coordinated this study and managed data collection. KK analyzed the data and drafted the initial paper with participation from KYG, CB, NL, and JTT. All authors reviewed and approved the final paper. This study was funded by the University of South Florida’s Pandemic Research Response Network through COVID-19 Rapid Response Research (100268). No generative artificial intelligence was used in any portion of this paper. The table of contents image for this paper is credited [51].

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Tables with additional statistics.

References


25. I had a black dog, his name was depression. YouTube. URL: https://youtu.be/XiCrniLQGYc

26. 988 Suicide & Crisis Lifeline. URL: https://suicidepreventionlifeline.org [accessed 2024-03-16]


51. Vecteezy. URL: https://www.vecteezy.com/ [accessed 2024-03-16]

Abbreviations

ATMHT: Attitudes Toward Mental Health Treatment  
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition  
IRB: Institutional Review Board  
MTurk: Amazon Mechanical Turk  
PHQ-9: Patient Health Questionnaire-9  
SBIRT: Screening, Brief Intervention, and Referral  
SSOSH: Self-Stigma of Seeking Help  
TRAM: Technology Readiness and Acceptance Model  
URICA: University of Rhode Island Change Assessment

Promoting a Patient-Centered Understanding of Safety in Acute Mental Health Wards: A User-Centered Design Approach to Develop a Real-Time Digital Monitoring Tool

Gemma Louch¹,², PhD; Kathryn Berzins³, PhD; Lauren Walker⁴, BSc; Gemma Wormald⁵, BSc; Kirstin Blackwell⁵, BSc; Michael Stephens⁶, BSc; Mark Brown⁷; John Baker¹, PhD

¹School of Healthcare, University of Leeds, Leeds, United Kingdom
²Yorkshire Quality and Safety Research Group, Bradford Institute for Health Research, Bradford, United Kingdom
³Health Technology Assessment Unit, Applied Health Research Hub, Implementation and Capacity Building Team, NIHR Applied Research Collaboration North West Coast, University of Central Lancashire, Preston, United Kingdom
⁴School of Health and Psychological Sciences, City, University of London, London, United Kingdom
⁵Thrive by Design, Leeds and York Partnership NHS Foundation Trust, Leeds, United Kingdom
⁶Ayup Digital, Leeds, United Kingdom
⁷Social Spider CIC, London, United Kingdom

Corresponding Author:
Gemma Louch, PhD
School of Healthcare
University of Leeds
Baines Wing
Leeds, LS2 9JT
United Kingdom
Phone: 44 01133433 ext 7173
Email: g.e.louch@leeds.ac.uk

Abstract

Background: Acute mental health services report high levels of safety incidents that involve both patients and staff. The potential for patients to be involved in interventions to improve safety within a mental health setting is acknowledged, and there is a need for interventions that proactively seek the patient perspective of safety. Digital technologies may offer opportunities to address this need.

Objective: This research sought to design and develop a digital real-time monitoring tool (WardSonar) to collect and collate daily information from patients in acute mental health wards about their perceptions of safety. We present the design and development process and underpinning logic model and programme theory.

Methods: The first stage involved a synthesis of the findings from a systematic review and evidence scan, interviews with patients (n=8) and health professionals (n=17), and stakeholder engagement. Cycles of design activities and discussion followed with patients, staff, and stakeholder groups, to design and develop the prototype tool.

Results: We drew on patient safety theory and the concepts of contagion and milieu. The data synthesis, design, and development process resulted in three prototype components of the digital monitoring tool (WardSonar): (1) a patient recording interface that asks patients to input their perceptions into a tablet computer, to assess how the ward feels and whether the direction is changing, that is, “getting worse” or “getting better”; (2) a staff dashboard and functionality to interrogate the data at different levels; and (3) a public-facing ward interface. The technology is available as open-source code.

Conclusions: Recent patient safety policy and research priorities encourage innovative approaches to measuring and monitoring safety. We developed a digital real-time monitoring tool to collect information from patients in acute mental health wards about perceived safety, to support staff to respond and intervene to changes in the clinical environment more proactively.

(JMIR Form Res 2024;8:e53726) doi:10.2196/53726
KEYWORDS
patient safety; mental health; patient involvement; qualitative; digital innovation; real time; monitoring; safety; develop; development; design; perception; perceptions; prototype; evidence scan; interview; interviews; logic model; programme theory; dashboard; dashboards; interface

Introduction

Overview

High levels of patient safety incidents are reported within mental health services. Between April 2020 and March 2021, a total of 300,703 incidents were reported in England, often involving self-harming behavior and disruptive, aggressive behavior [1]. Patient safety within mental health services is recognized as a priority within the National Health Service (NHS) Patient Safety Strategy 2021 update [2] and is a key research priority within the extant literature [3–6].

Supporting patients and families to be partners in care safety can be viewed as both a logical and moral imperative [7]. Furthermore, understanding the patient perspective is an important patient safety priority in the mental health field [4]. Recognizing the limitations of patient safety incident reporting systems, research evidence suggests that patients and families are a valuable source of safety intelligence that can inform service improvement in acute [8–10] and primary care settings [11–13]. Indeed, our previous research directly addressed this priority for mental health patient safety research, by evidencing that patients acknowledge the value of their potential involvement in interventions to improve safety and the need to develop interventions that proactively seek the patient perspective of safety [14,15].

In addition to understanding the patient perspective, there is also policy and research impetus to understand safety in real time, to shift from a reliance on retrospective patient safety measures [6,16], and a push toward embracing innovative approaches and digital solutions to safety challenges [6,17].

Considered together, these policy and research priorities provide the rationale for the development of a digital monitoring tool that captures the patient perspective of safety in real time, in an acute mental health ward setting.

Theoretical and Conceptual Underpinning of the Digital Monitoring Tool

The theoretical foundation for the development of a digital real-time monitoring tool is embedded within the measurement and monitoring of safety (MMS) framework [16]. Previous research exploring the framework in practice acknowledged its potential to support a broader and richer approach to organizational safety [18,19]. Specifically, we focus on 1 domain of the framework—sensitivity to operations. This domain describes the need for a collective awareness by staff of the workings of the service and their ability to be sensitive and responsive to subtle changes and disturbances. This domain highlights the crucial but often overlooked activity of “monitoring” the safety of care as it is delivered in real time, and it is a domain where patients and families are recognized as potential key sources of information. Within the wider patient safety literature, the importance of safety “monitoring” as opposed to solely “measurement” and the potential of prospective clinical surveillance as a means of promoting safety within organizations have gained traction [20,21]. Supporting the idea of prospective clinical surveillance as a means of promoting safety is particularly important within the acute mental health care context, where fluctuations in the dynamic of the inpatient group and the interplay between patients, staff, and the environment can occur rapidly, with individual patient needs creating immediate knock-on effects for other patients, their quality of care, and their safety.

Two further concepts that informed our thinking and the development of the real-time digital monitoring tool are milieu and contagion. The concept of milieu is often aligned with the notion of ward atmosphere, which involves the interplay between the physical environment, social structures, and social interactions [22]. In this work, we conceptualize the milieu as being akin to the ward atmosphere. Safety may be related to tensions in the milieu of the ward [23], and improving the milieu may improve safety [24]. There is research to support the notion of a “contagion effect” of safety incidents (violence and self-harm behaviors) within an inpatient mental health ward setting [25], with evidence to suggest that patient aggression and self-harm behaviors do not occur at random intervals, but cluster temporally. Therefore, a key safety issue is the potential for 1 incident to increase the likelihood of further incidents occurring as a result of disturbed ward milieu and contagion.

Considering contagion and milieu alongside the domain sensitivity to operations of the MMS framework [16] underpinned our operationalization of how we might capture perceptions of safety in real time. A digital monitoring tool capturing how the ward is perceived by patients (ie, the milieu) at any given time may provide intelligence that supports staff to respond proactively to emerging safety issues and intervene earlier, which may prevent future incidents from occurring via disturbed ward milieu and contagion.

Aim

The aim of this study is to design and develop a digital real-time monitoring tool that collects and collates information from patients in acute mental health wards about their perceptions of safety, in order to support staff in monitoring and improving the safety of the clinical environment.

Approach

This paper focuses on the design and development of the digital monitoring tool. Two stages are presented: stage 1—a synthesis of exploratory research, existing research, and stakeholder engagement and stage 2—design and development. The methods and results are presented for each stage separately. The findings relating to the implementation and evaluation (qualitative) of the tool are reported elsewhere [26] and include further information about how the tool was operationalized in the implementation phase.
Methods

Stage 1: Exploratory Research, Existing Research, and Stakeholder Engagement Synthesis

Objective

This stage focused on bringing together and synthesizing key findings from our exploratory research, previous research, and early stakeholder engagement discussions. The purpose of this synthesis was to generate themes, preliminary ideas, and needs for the monitoring tool from multiple perspectives.

Activities and Analysis

We conducted (1) a systematic review [27] and evidence scan, (2) semistructured interviews with health professionals [28] and patients, and (3) stakeholder engagement sessions (see Textbox 1).

Textbox 1. Overview of activities contributing to data synthesis.

<table>
<thead>
<tr>
<th>Systematic review and evidence scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Systematic review of patient involvement in safety interventions in an acute mental health setting [27]:</td>
</tr>
<tr>
<td>• Included 52 articles</td>
</tr>
<tr>
<td>• Narrative synthesis</td>
</tr>
<tr>
<td>• Included 13 articles (see Multimedia Appendix 1)</td>
</tr>
<tr>
<td>• Narrative synthesis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Semistructured interviews with health professionals and patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Interviews with patients (n=8; 2 participants identified as female, and 6 as male) older than 18 years of age with current or recent experience (during the past 2 years) of being an inpatient in an acute mental health ward from 2 Mental Health Trusts in the North of England.</td>
</tr>
<tr>
<td>• Interviews with mental health professionals (n=17), methods and findings reported elsewhere [28].</td>
</tr>
<tr>
<td>• The interviews aimed to understand perspectives on safety issues and how patients and health professionals can contribute toward the measurement and monitoring of ward safety in an acute mental health setting.</td>
</tr>
<tr>
<td>• Reflexive thematic analysis [29,30]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stakeholder engagement sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Sessions held with 2 existing patient focused stakeholder groups via video conferencing software, hosted by a Mental Health Trust in the North of England in August 2020, focusing on:</td>
</tr>
<tr>
<td>• How to phrase or ask patients about whether wards feel safe or unsafe</td>
</tr>
<tr>
<td>• Language and how best to talk about patient safety (eg, words or phrases that should or should not be used)</td>
</tr>
<tr>
<td>• What to consider when developing digital interventions to be used in acute mental health wards</td>
</tr>
</tbody>
</table>

Ethical Considerations

The interview study with health professionals received ethical approval from the University of Leeds, School of Healthcare Ethics Committee (HREC 19-028), and the interview study with patients received ethical approval from South Central—Berkshire B Research Ethics Committee (20/SC/0360). Prior to the interviews, researchers explained the purpose and process of the interview to the participant, read the consent form out loud, and recorded consent verbally. Interviews were audio recorded, anonymized, and transcribed professionally. Patients who were interviewed received a £10 (US $12.79) shopping voucher as a thank you for their time.

Data Synthesis

We produced summary documents for the systematic review and evidence scan, ongoing interviews with patients and health professionals, and stakeholder engagement sessions. The findings were organized and reviewed through the lens of implications for the design and development phase and discussed on an ongoing basis between the core research team (including 2 co-investigators with lived experience), the project Steering Group, digital partners (Ayup Digital), and co-design partners (Thrive by Design).

Stage 2: Design and Development

Objective

This stage was informed by stage 1 and focused on the design and development of a digital monitoring tool to enable a real-time understanding of the patient perspective of safety in acute mental health wards.

Activities and Analysis

This stage was significantly affected by the COVID-19 pandemic [31]; pragmatic adjustments were made to the original
co-design plan to produce a feasible alternative in which a single NHS Trust permitted limited visits from a clinical member of the co-design team who was familiar with the local protocols.

At the outset of this stage, members of the core research team, digital partners, and co-design partners drew from the ongoing stage-1 synthesis to inform the basis of the subsequent activities. The reworked co-design activities included face-to-face discussions and opportunistic feedback from patients and staff during ward visits in 2 wards in the North of England. This stage also involved further stakeholder engagement sessions with the same 2 stakeholder groups described in stage 1. We used a collaborative, human-centered, and sprint-based or agile approach [32,33], incorporating techniques such as “personas” (typical users with specific characteristics including experience of disability and different levels of digital skills), storyboards (to elicit user goals), patient journey mapping (the path of an individual patient and health care staff member in a ward), user stories (to elicit specific user requirements), and prototyping (rapid creation of paper prototypes of a digital tool). The first cycle of activities generated requirements for the digital tool; subsequent activities tested and refined the tool as it developed and considered its use within the clinical workflow of the ward environment.

Ward-Based Activities

Visit 1 (December 2020)

Draft ideas for the digital products were discussed, for example, wording, format, color, and how they would work in relation to existing systems. People wrote comments and provided feedback on all aspects of the design. From the staff perspective, we explored whether the visuals looked like anything already in use, to avoid unintended consequences arising from confusion.

- Ward A (adult acute ward caring for male patients aged 18-65 years): opportunistic feedback and discussion with 3 patients and 3 members of staff (clinical and administrative).
- Ward B (adult acute ward caring for female patients aged 18-65 years): opportunistic feedback and discussion with 5 patients and 12 members of staff (clinical and administrative); a total of 3 members of staff and 5 patients participated in a co-design workshop (approximately 40 minutes duration).

Once activities commenced in the wards, members of the core research team, digital partners, and co-design partners continued to meet fortnightly to interrogate and interpret the feedback on the digital product ideas in more detail, focusing on who would benefit, considering the “must haves,” and recording when there was consensus on what not to include. Discussions centered on how many of the “must haves” were realistic to implement, and alongside input from the project Steering Group fed into the development of the second iteration of the digital products.

Visit 2 (January 2021)

Discussions and activities focused on the refined prototype ideas for the digital products. For both Ward A and Ward B, multiple versions of the latest iteration of products were presented and feedback elicited, in addition to revisiting discussions from visit 1. People provided feedback on the wording, icons, colors, and representations, that is, was it what they imagined? From the staff perspective, the requirements of the staff dashboard were revisited, and from the patient and staff perspective, the acceptability of a public-facing ward interface was explored. As there was a high turnover of patients, many new patients were involved, although many of the same staff contributed, which enabled discussions from the first visits to be revisited.

- Ward A: opportunistic feedback and discussion with 5 patients and 3 members of staff (clinical and administrative).
- Ward B: opportunistic feedback and discussion with 6 patients and 5 members of staff (clinical and administrative).

Feedback from the ward-based discussions and activities was collated and prioritized using the Must Have; Should Have; Could Have; Won’t Have this time (MoSCoW) method [34], which then guided conceptual and technical development. Prototype ideas were refined following the prioritization exercise.

Stakeholder Engagement

Sessions followed via video conferencing software with the same 2 stakeholder groups described previously in stage 1 (July 2021). The most recent prototype ideas were presented and discussed, focusing on (1) the design of the patient interface: Would you feel comfortable completing it? (2) Could you say why you would feel more or less comfortable? (3) Do you think completing a report on the ward atmosphere on a digital tablet would make you feel less or more distressed? and (4) public-facing ward interface: How would you feel about this being displayed in a ward?

Results

Stage 1: Exploratory Research, Existing Research, and Stakeholder Engagement Synthesis

The findings from the synthesis are organized around 7 themes with associated implications for the subsequent design and development phase (see Textbox 2). Supporting information for each theme is outlined in Tables S1-S7 in Multimedia Appendix 2.
**Textbox 2.** Themes and design and development implications.

### Conceptualization of safety
- Allow for a multifaceted conceptualization of safety.
- Consider how the word “safety” may be misinterpreted—provide a clear description.

### Anonymity
- Anonymity essential.
- Reinforce feedback is anonymous and that everyone’s experience is important—all feedback valued.
- Give clear information about the use and purpose of collecting information via digital technology.

### Milieu (ward atmosphere), contagion, and incidents
- The technology needs to be sensitive to subtle changes in a negative direction in order to anticipate the potential for an incident occurring.
- Explore whether the technology can be sensitive to context, for example, highlight when a safety incident has recently occurred.
- Location: potential for the technology to be sensitive to location within the ward, for example, increased anxiety or input activity by patients into the technology (prior to and after an incident).
- Night and day: the technology needs to recognize night and day as different contexts.
- The technology may need to account for or recognize ward profiles (eg, all male, all female, and mixed wards).
- The technology may need to account for or recognize differences between individual patients.

### Digital technology in the ward
- Ensure the purpose of the monitoring tool is clearly described to patients and staff.
- The technology needs to be inclusive across a wide range of patients and digital abilities.
- Staff may need support with using the technology and with data interpretation.
- Consider multiple mechanisms for data collection other than a single device or reporting “point.”
- Consider other relevant systems in place for staff.

### Involving patients in understanding safety
- Technology needs to incorporate the input of quantitative and qualitative data to give patients choice about the level of detail they provide.
- Explore using language or imagery that is universal and include a free text option to expand.
- Consider different levels of providing feedback, for example, level 1 reporting—a color or smiley face and level 2—more detailed information.
- Explore whether thresholds need to be built into the system to trigger alerts or action.
- Consider how the phrases used to describe safety converge with those used by other relevant organizations.
- Consider the location of data collection and timings.
- Consider the implications of the type of ward.

### Feeding data back
- The technology and mechanism of feeding data back needs to work within existing trust and ward infrastructure.
- Explore frequency or timing of monitoring the data to ensure staff can be responsive.
- Consider approaches to displaying the data.
- Consider who can see the data and when.
- Staff may need support with data interpretation and action.

### Unintended consequences
- Important to provide choice and freedom on when to provide feedback.
- May need to offer different levels of providing feedback.
- Consider how the technology sits alongside verbal information and feedback.
Stage 2: Design and Development

Intervention Design and Description

Overview
The latest version of the digital monitoring tool, WardSonar (prototype 3), includes 3 components—a patient recording interface, a health professional dashboard, and a public-facing ward interface. Further information about the prioritization process and conceptual and technical development is provided in Multimedia Appendix 3.

Patient Recording Interface
The recording interface is an app accessed on a tablet device (see Figure 1). The interface provides brief background information and prompts patients to answer a series of questions. No patient data are collected, and all data collected on the interface are anonymized at the point of entry. The interface uses a weather analogy with questions, such as “How does the ward atmosphere feel right now?” (very calm to very stormy), as the co-design phase identified the need to provide different ways for people to be able to express how the ward was feeling, such as pictures and text. Therefore, free-text options were also included. The design and development phase highlighted that more options for how the ward was feeling would be beneficial to reflect that this is often more than “good, bad and okay” and that the feeling is more of a spectrum. Therefore, the interface includes 5 options to describe how the ward is feeling, and an additional question to indicate the direction things are moving in (ie, getting better, getting worse, or staying the same). Attention was paid to understanding the contextual constraints and pressures that patients may be experiencing (eg, location when entering data and current mental state at the time of data entry), and feedback from the design and development phase suggested that a reporting “point” could create negativity toward the area it is placed and prevent people from using it.
The idea of patients using their own mobile phones to provide feedback was also considered in the design and development phase. This was not progressed at this initial stage as not all patients have a smartphone or access to data, and from a technological development perspective, in the subsequent implementation phase, each ward required a secure login to access the app. Organizing this for each patient on either their own or a study-bought device would have required technological support in the wards, which would not have been feasible during the COVID-19 pandemic, and we were aiming for the approach to be implemented with minimal support. Therefore, in the subsequent implementation phase, it was decided that staff would supply patients with the tablet device (provided by the research study) optimally 3 times per day so patients could enter real-time safety perspectives. Staff supplying the tablet device to patients also addressed staff concerns about the security of the tablet devices and who would have responsibility for them.

**Staff Dashboard**

The staff dashboard is the main interface used by staff to view data submitted by patients on a tablet device or desktop computer (see Figure 2). It is accessed in specific, authorized locations, such as the ward office, and provides real-time snapshot data and greater informational insights through the use of data visualizations. The co-design phase with staff,
statisticians, and researchers resulted in the dashboard including barometer-style visualization of “How is the ward feeling,” statistical process control charts, graphs, and statistical metrics, and the functionality for monthly, weekly, daily, or shift aggregate data to be comparable to historic time periods.

**Figure 2.** Prototype-3 staff dashboard.

**Public-Facing Ward Interface**

The public-facing ward interface displays the average ward atmosphere rating for the current shift on, for example, a television screen, desktop computer monitor, or tablet device (see Figure 3). The outcome measure is displayed in the form of a barometer.
Stakeholder Engagement

The 2 groups were positive about the prototype-3 patient recording interface, in particular, the simplicity of the interface and the anonymity it provides. Overall, people felt the patient interface would be a safe space to report concerns. However, it was emphasized that this system should not replace reporting in person to staff. Regarding the question of factors affecting how the ward feels, there was consensus that it was important to be able to select more than 1 factor and this change was subsequently implemented. A number of potential unintended consequences were emphasized such as people being concerned that others were making reports about them; that the technology might trigger people’s symptoms; and that the ward interface could be confusing, as patients may think it referred to the weather outside. Of particular note were discussions around the potential for a poor ward atmosphere when a ward is understaffed, meaning there may be limited support from staff to circulate the device at these times. There were concerns raised around the accessibility and inclusivity of the patient recording interface which would need to be addressed in future iterations, for example, people who may experience barriers related to language or literacy, people with learning disabilities, and people with visual impairments. As potential unintended consequences were highlighted regarding the public-facing ward interface, this component of the tool did not progress to the subsequent implementation phase [26].

Technical Specifications

The digital products were built using open web standard technologies (eg, HTML5, CSS3, JavaScript, and PHP) in a componentized and scalable way, to allow modifications and future developments, and were based on the Government Digital Service Agile Delivery methodology framework and NHS standards. All digital products are hosted in a custom cloud environment built on top of the Amazon Web Services infrastructure. Data are recorded in real time and sent to an application programming interface. The interface was built to Web Content Accessibility Guidelines (WCAG) 2.1 AA [35] standards and is device agnostic.

Patients are able to input data via a device supplied by ward staff. Patient data are monitored through the staff dashboard by authorized members of staff with access via a tablet device or desktop computer. Any data input by patients are recorded via the app and is automatically sent back in real time to the central database. If Wi-Fi is temporarily unavailable in the ward, then the data that are input are stored locally on the device in offline mode. When the device comes back into the Wi-Fi range the data are automatically submitted to the central database. Further information about the technologies used and infrastructure is provided in Multimedia Appendix 4.

Logic Model and Program Theory

A logic model was generated to articulate how the various components of WardSonar related to each other (see Figure 4). In conjunction with the logic model, Multimedia Appendix 5 provides a narrative description of the program theory explaining how WardSonar might facilitate change in proximal and distal outcomes.
**Figure 4.** The logic model for the WardSonar monitoring tool. *For example, gender, ethnicity, disorder, and digital literacy; **for example, qualification and gender; ***for example, skill mix and staffing levels.

**Discussion**

**Principal Findings**

We successfully developed WardSonar, a theory-informed, patient-centered, real-time digital safety monitoring tool for acute mental health wards. The WardSonar monitoring tool aims to collect daily information from patients in acute mental health wards about perceived safety to support health professionals in monitoring and improving the safety of the clinical environment.

The process of design and development included synthesizing findings from a systematic review, evidence scan, interviews with patients and health professionals, and stakeholder engagement, to inform a cycle of design activities. The synthesis highlighted broad themes and potential unintended consequences to be explored further within the subsequent design phase. For instance, the importance of the tool allowing for a multifaceted conceptualization of safety, ensuring patient anonymity, being sensitive to subtle changes, and being inclusive across a wide range of patients and digital abilities.

The design and development process resulted in 3 components of the monitoring tool. First, a patient recording interface to assess how the ward feels. Patients were keen to have different ways to be able to express how the ward was feeling, including pictures, text, and being able to write in free text. The need to capture whether the direction is changing, that is, “getting worse” or “getting better” was felt to be really important from a health professional perspective. The second component was a staff dashboard with functionality to interrogate the data at different levels, and the third component was a public-facing ward interface. Potential unintended consequences and challenges were highlighted, for example, the public-facing ward interface being potentially triggering if displaying a negative ward atmosphere. Embedding the principles of WardSonar within the MMS framework [16], and within the domain sensitivity to operations specifically, has produced a safety monitoring tool that recognizes patients as a key source of safety information and facilitates health professionals being aware of subtle changes and disturbances.

**Implications for Research and Practice**

The next step is for the WardSonar monitoring tool and its components to be implemented in practice, and for this to be accompanied by a robust evaluation. This will generate evidence around key questions pertaining to feasibility and acceptability, for example, is it feasible and acceptable to collect safety data from patients in mental health wards? and, how do staff use the data collected from patients? Such an evaluation may support further refinement of WardSonar, and over the longer term, would explore the assumptions articulated in the WardSonar logic model, for instance, our hypothesized distal outcomes, leading to a refined logic model and program theory in line with Medical Research Council guidance on developing and evaluating complex interventions [36].
A key feature of the design and development phase was the need for WardSonar to be accessible for people with different language and literacy needs and for it to be accessible for a range of digital abilities. Our most recent stakeholder engagement identified concerns around accessibility and inclusivity of the patient interface. Therefore, it will be essential for these concerns to be examined further in subsequent evaluation work to ensure future iterations of the WardSonar tool can be accessed equitably.

**Strengths and Limitations**

To our knowledge, the WardSonar monitoring tool is the first system that actively seeks the perspective of safety from patients in real time for use within acute mental health wards. A key strength of this work is our theory-based approach and the multiple data sources brought into the design and development phase. A limitation of our work is the level of co-design and stakeholder engagement activity we were able to undertake, as our original plans had to be amended considerably due to the impact of the COVID-19 pandemic on the project.

**Conclusions**

Recent patient safety policy and research priorities encourage innovative approaches to measuring and monitoring safety, and there is a need for a patient-centered understanding of safety in an acute mental health setting. At present, no interventions or tools exist to address this need. We developed a digital real-time monitoring tool to collect information from patients in acute mental health wards about perceived safety, to support health professionals to respond, and intervene to changes in the clinical environment more proactively. Further research is required to evaluate the implementation of WardSonar to further refine and improve this innovative approach.

**Acknowledgments**

This project was funded by the National Institute for Health and Care Research (NIHR) Health and Social Care Delivery Research program (HSDR, NIHR 12/80/70) and was supported by the NIHR Yorkshire and Humber Patient Safety Translational Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. We would like to express our thanks to everyone who participated in interviews and discussions, which informed the development of WardSonar, and to the stakeholder groups and research staff involved in the WardSonar project. We would also like to thank Steven Taylor from Ayup Digital and the WardSonar Steering Group for their contributions, support, and steer. The project was supported by 2 coinvestigators with lived experience of mental health services and mental health services and digital expertise. Stakeholder engagement fed into the development of the monitoring tool.

**Data Availability**

The data sets generated and analyzed during this study data are available from the Principal Investigator (JB) on reasonable request.

**Authors’ Contributions**

GL led the study draft. K Berzins led the patient interviews and contributed to the study draft. LW led the engagement work and contributed to the study draft. GW led the workshops and discussions for tool development and contributed to the study draft. K Blackwell contributed to tool development and contributed to the study draft. MS led technological development and contributed to the study draft. MB contributed to the study draft. JB contributed at all stages.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Evidence scan included articles.

[PDF File (Adobe PDF File), 615 KB - formative_v8i1e53726_app1.pdf ]

**Multimedia Appendix 2**

Supporting information for each theme.

[PDF File (Adobe PDF File), 731 KB - formative_v8i1e53726_app2.pdf ]

**Multimedia Appendix 3**

Feedback following the ward visits and prioritization.

[PDF File (Adobe PDF File), 643 KB - formative_v8i1e53726_app3.pdf ]

**Multimedia Appendix 4**

Technologies used and infrastructure.
Multimedia Appendix 5
WardSonar programme theory narrative description.

References


Abbreviations

- MMS: measurement and monitoring of safety
- MoSCoW: Must Have; Should Have; Could Have; Won’t Have this time
- NHS: National Health Service
- WCAG: Web Content Accessibility Guidelines

Please cite as:
Promoting a Patient-Centered Understanding of Safety in Acute Mental Health Wards: A User-Centered Design Approach to Develop a Real-Time Digital Monitoring Tool
JMIR Form Res 2024;8:e53726
URL: https://formative.jmir.org/2024/1/e53726
doi: 10.2196/53726
PMID:38607663
The Effects of a Single-Session Virtual Rumination Intervention to Enhance Cognitive Functioning in Veterans With Subjective Cognitive Symptoms: Multimethod Pilot Study

Tara Austin1,2,3, PhD; Jennifer Smith3, AAS; Borsika Rabin1,4,5, PhD; Laurie Lindamer1,2, PhD; James Pittman1,2, PhD; Staley Justice3, MSc, LCSW; Elizabeth W Twamley1,2,6, PhD; Crystal Lantrip3,7, PhD

1Research Service, VA San Diego Healthcare System, San Diego, United States
2Center of Excellence for Stress and Mental Health, VA San Diego Healthcare System, San Diego, CA, United States
3The VA Center of Excellence for Research on Returning War Veterans, Waco, TX, United States
4Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, CA, United States
5UC San Diego Altman Clinical and Translational Research Center Dissemination and Implementation Science Center, University of California San Diego, San Diego, CA, United States
6Department of Psychiatry, University of California San Diego, San Diego, CA, United States
7Department of Psychology and Neuroscience, Baylor University, Waco, TX, United States

Corresponding Author:
Tara Austin, PhD
Research Service
VA San Diego Healthcare System
3350 La Jolla Village Drive
San Diego, 92161
United States
Phone: 1 8585528585
Email: tara.austin@va.gov

Abstract

Background: Subjective cognitive concerns (SCCs) entail perceived difficulties in thinking or memory, often reported without substantial objective evidence of cognitive impairment. These concerns are prevalent among individuals with a history of brain injuries, neurological conditions, or chronic illnesses, contributing to both psychological distress and functional limitations. They are increasingly considered to be a risk factor for future objective decline. A considerable number of individuals reporting SCCs also exhibit mental health symptoms, such as a history of trauma, depression, or anxiety. Interventions that address modifiable emotional and cognitive factors related to SCC could improve functioning and quality of life. Therefore, the use of emotion regulation strategies, especially those directed at minimizing rumination, could serve as a promising focus for interventions aimed at mitigating subjective cognitive concerns in veteran populations.

Objective: This pilot study explored the feasibility, acceptability, and preliminary efficacy of a brief, 1-session emotion regulation intervention called “Worry Less, Remember More.” The Worry Less, Remember More intervention was designed to reduce rumination and improve subjective cognitive functioning in veterans with subjective cognitive changes (N=15).

Methods: We randomized 15 veterans to either the active telehealth condition or waitlist control and completed the intervention. Participants were aged between 31 and 67 (mean 49.5, SD 10.1) years, and the sample was primarily male (12/15, 83%) and White (10/15, 67%). The most common diagnoses were posttraumatic stress disorder and depression. Following the intervention, veteran input was sought through semistructured interviews with a subset of 12 participants, examining feasibility, acceptability, and perceived efficacy. Preliminary efficacy was also measured using pre- and postintervention self-report measures.

Results: Veterans reported that this intervention was acceptable, with 92% (11/12) of the sample reporting that they benefited from the intervention and would recommend the intervention to others with similar difficulties. Semistructured interviews revealed difficulties with feasibility, including problems with the remote consenting process, forgetting appointments, and needing additional strategies to remember to consistently use the interventions. The intervention improved self-reported cognitive symptoms on quantitative measures but did not improve self-reported rumination.
Conclusions: This pilot study establishes the preliminary feasibility, acceptability, and efficacy of the Worry Less, Remember More intervention for veterans with subjective cognitive symptoms. Future iterations of the intervention may benefit from simplifying the electronic consent process, providing reminders for appointments, and incorporating compensatory cognitive strategies to assist with using the telehealth system, as well as applying the strategies learned in the intervention. While future research is needed with larger samples, including nonveteran populations, the intervention may also be a useful clinical tool to bridge care between neuropsychology clinics and mental health treatment.

KEYWORDS
army; cognition; cognitive; emotion regulation; memory symptoms; memory; military; rumination; subjective cognitive decline; telehealth; telemedicine; veteran; worry

Introduction
Subjective cognitive concerns (SCCs) refer to perceived challenges in thinking or memory reported with either no evidence or only minimal objective evidence of cognitive impairment [1]. SCCs are common in people with histories of brain injuries, neurological illnesses, and other chronic illnesses and contribute to psychological distress and functional impairment [2-5]. More than 10% of US veterans aged 45 years or older report subjective cognitive symptoms [6]. Interest in SCCs has increased due to cognitive symptoms, such as “brain fog,” following the COVID-19 infection [7]. Research has found that SCCs are often related to potentially modifiable factors that influence cognition, such as mood, and many individuals with SCCs have mental health symptoms, including a history of trauma, depression, or anxiety [8-11]. A recent study in a neuropsychology clinic found that childhood trauma predicted SCCs and that this relationship was mediated by a ruminative thought style [12]. Emotion dysregulation, including rumination, contributes to and maintains psychopathology as well as cognitive dysfunction, particularly in those with affective illnesses [13] and has been implicated in the maintenance of neuropsychiatric illness [14,15]. Rumination, or the tendency to passively and persistently dwell on negative or problematic aspects of life, and the ruminative thought process, which includes the act of focusing on the potential causes and outcomes of the negative or problematic aspects of life, are forms of emotional dysregulation, which are patterns of emotional experience or expression that interfere with goal-directed activity [14]. Ineffective coping strategies perpetuate emotion dysregulation and, thereby, depression and other psychological disorders [16,17].

Emotion regulation approaches vary in their impact on both emotional health (eg, depression relapse and severity) and cognitive functioning. Less effective regulation strategies, such as rumination (eg, perseverative thought processes focusing on negative content) [18] and suppression (inhibiting the outward signs of inner feelings), are suggested to be cognitively taxing, thus diminishing cognitive resources [19]. Rumination and suppression have been associated with poor health and negative psychological outcomes in the general population as well as in veteran populations [20]. In veterans, rumination has been found to moderate the association between posttraumatic stress disorder (PTSD) or depression and risky behavior [21] and between moral injury and negative mental health symptoms [22]. Rumination is also associated with sleep problems in veterans with PTSD and depression [22]. There is also preliminary evidence that rumination may moderate the relationship between attentional difficulties and PTSD symptoms [23].

As such, emotion regulation strategy use, particularly strategies aimed at reducing rumination, may be a good target for intervention to reduce SCCs in veteran populations. In order to address these difficulties, we created a 1-session treatment called “Worry Less, Remember More,” integrating elements from Watkins’s [24] rumination-focused cognitive-behavioral therapy for depression and Gilbert’s [25] compassion-focused therapy (Multimedia Appendix 1 [24,26,27]). The psychoeducation portion consisted of concepts from evolutionary psychology as described by Gilbert [25], including an evolutionarily adapted attentional bias toward negative information, information about emotional regulation systems and their responses to trauma and stress, and attention as a limited resource that can be redirected. We used Watkins’s [24] 12-session rumination-focused cognitive-behavioral therapy for depression to provide specific examples of rumination, purposes of rumination, and 3 short intervention exercises. The session was designed to be delivered through telehealth, as there is evidence that telehealth and web-based interventions are feasible, acceptable, and efficacious in bridging care and supplementing existing mental health treatment [28,29], including in populations with cognitive difficulties [30].

In addition to establishing the feasibility and efficacy of emotion regulation interventions to improve subjective cognitive functioning, it is also important to establish the acceptability of the intervention to individuals with SCCs to increase treatment engagement, compliance, and completion [31]. For this study, we operationalized feasibility (including demand, implementation, practicality, and integration into existing systems) and acceptability according to Bowen et al [32], Pearson et al [33], and Sekhon et al [34]. The goal of this study was to establish the feasibility, acceptability, and preliminary efficacy of a 1-session, rumination-focused intervention for veterans with SCCs compared to a waitlist control condition.

Methods

Ethical Considerations
This study was approved by the Central Texas Veterans Healthcare System Institutional Review Board (IRB number...
00697). All participants were informed of the purpose of the study, as well as the possibility of dropping out at any time, and signed written informed consent before participation. All included study data are deidentified, and all quotes are anonymous and have been carefully reviewed to have all potentially identifiable data removed. Participants were paid US $40 per visit (US $120 in total).

**Recruitment**

Veterans were recruited from the Veterans Administration (VA) neuropsychology specialty clinics, primary care clinics, and mental health clinics. The majority of veterans who participated were referred by mental health providers, though some veterans self-referred through flyers placed on the medical campus.

Inclusion criteria included verbal endorsement of cognitive difficulties and veteran status. Exclusion criteria included (1) a score of <23 on the Montreal Cognitive Assessment (MoCA); (2) current substance abuse; or (3) a diagnosis of serious mental illness, such as schizophrenia or psychotic disorders. We screened 49 veterans for eligibility, of whom 9 were self-referred and 40 were referred by mental health providers. Of the 49 veterans referred, 23 did not return to complete informed consent, 4 declined to participate, 2 had MoCA scores <23, another 2 were ineligible due to substance use, and 1 was ineligible due to a diagnosis of severe mental illness. Thus, a total of 17 veterans were enrolled and randomized to either the intervention or waitlist control condition before baseline testing. Of the veterans randomized, 15 attended the first visit (baseline testing and intervention for the intervention group; baseline testing only for the waitlist control group), and 12 attended both the initial assessment and follow-up testing. The 12 participants who completed the intervention also completed quantitative assessments and qualitative interviews at the 8-week follow-up. All study procedures (including screening, consenting, assessment, and intervention) were delivered remotely using VA video telehealth software.

**Sample Characteristics**

Participants' were aged between 31 and 67 (mean 49.5, SD 10.1) years (Table 1), and the sample was primarily male (12/15, 83%) and White (10/15, 67%). The most common diagnoses were PTSD and depression. There were no significant differences between the full sample and the qualitative interview participant subset on age, gender, race or ethnicity, rates of PTSD or sleep difficulties, or the percentage of participants who completed all study visits. The qualitative interview group had a higher percentage of brain injuries and lower rates of depression.
<table>
<thead>
<tr>
<th>Demographic</th>
<th>Intervention group (n=7)</th>
<th>Waitlist control group (n=8)</th>
<th>Qualitative sample subset (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants who completed the study, n (%)</td>
<td>6 (86)</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>10 (83)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>51.7 (11.5)</td>
<td>49.6 (8.7)</td>
<td>49.4 (10.1)</td>
</tr>
<tr>
<td>Gender (male), n (%)</td>
<td>5 (71)</td>
<td>7 (88)</td>
<td>10 (83)</td>
</tr>
<tr>
<td><strong>Self-reported race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0)</td>
<td>1 (13)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>White</td>
<td>5 (71)</td>
<td>6 (75)</td>
<td>8 (67)</td>
</tr>
<tr>
<td>Black</td>
<td>2 (29)</td>
<td>1 (13)</td>
<td>4 (33)</td>
</tr>
<tr>
<td><strong>Self-reported ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 (14)</td>
<td>1 (13)</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>CPRS diagnoses of interest, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3 (42)</td>
<td>3 (38)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Depression</td>
<td>4 (57)</td>
<td>3 (38)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Sleep</td>
<td>1 (14)</td>
<td>2 (25)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>ADHD&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Traumatic brain injury</td>
<td>2 (28)</td>
<td>2 (25)</td>
<td>4 (33)</td>
</tr>
<tr>
<td><strong>Baseline self-report measures, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI-II&lt;sup&gt;d&lt;/sup&gt;</td>
<td>31.55 (10.54)</td>
<td>34 (8.83)</td>
<td>N/A</td>
</tr>
<tr>
<td>BAI&lt;sup&gt;e&lt;/sup&gt;</td>
<td>20.75 (5.11)</td>
<td>17.36 (5.33)</td>
<td>N/A</td>
</tr>
<tr>
<td>RRS-SF&lt;sup&gt;f&lt;/sup&gt;</td>
<td>19.6 (3.67)</td>
<td>23.33 (6.15)</td>
<td>N/A</td>
</tr>
<tr>
<td>BRIEF-A&lt;sup&gt;g&lt;/sup&gt;</td>
<td>96 (26.7)</td>
<td>82.75 (30.96)</td>
<td>N/A</td>
</tr>
<tr>
<td>NSI&lt;sup&gt;h&lt;/sup&gt;</td>
<td>44 (12.51)</td>
<td>37.38 (9.02)</td>
<td>N/A</td>
</tr>
<tr>
<td>Rivermead&lt;sup&gt;i&lt;/sup&gt;</td>
<td>42.33 (8.96)</td>
<td>29.44 (5.71)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.
<sup>b</sup>PTSD: posttraumatic stress disorder.
<sup>c</sup>ADHD: Attention deficit hyperactivity disorder.
<sup>d</sup>BDI-II: Beck Depression Inventory-Second Edition.
<sup>e</sup>BAI: Beck Anxiety Inventory.
<sup>f</sup>RRS-SF: Ruminative Response Scale and Perseverative Thinking Questionnaire-Short Form.
<sup>g</sup>BRIEF-A: Behavior Rating Inventory of Executive Functioning-Adult.
<sup>h</sup>NSI: Neurobehavioral Symptom Inventory.
<sup>i</sup>Indicates statistically significant difference between groups (P<.05).

**Study Procedure**

The pilot study was a 1-visit randomized controlled trial with 2 follow-up visits to gather outcome data. All study visits were conducted remotely, using the VA telehealth system. Before the start of the study, participants were randomized to either an intervention or waitlist control condition. At the first visit, participants in both conditions completed a comprehensive preintervention self-report battery. Immediately following the initial baseline measures, those randomized to the intervention condition then participated in a 30-minute psychoeducation and rumination-focused intervention, whereas those in the waitlist control condition were excused. Then, 8 weeks later, both groups completed follow-up behavioral measures, and the waitlist control group received the intervention. Both groups completed follow-up behavioral measures at the third visit, approximately 8 weeks after the second visit.

**Measures**

Measures administered included the Beck Depression Inventory-Second Edition (BDI-II) [35], the Beck Anxiety Inventory (BAI) [36], the Ruminative Response Scale and Perseverative Thinking Questionnaire-Short Form (RRS-SF) to assess rumination [37], the Behavior Rating Inventory of Executive Functioning-Adult (BRIEF-A) self-report [38], the Neurobehavioral Symptom Inventory (NSI) [39], and the
Rivermead Post-Concussion Symptoms Questionnaire [40]. Raw scores from the BRIEF-A and cognitive items from the Rivermead and NSI were summed into a cognitive index score for each participant.

**Statistical Analysis**

Quantitative analyses were performed in R (R Core Team) using statistical packages and the original code [41]. The data were checked for normality and outlier values; no corrections were needed. When <10% of data were missing at random, missing values were estimated and imputed using a random forest–based imputation, which has been shown to be appropriate for imputing mixed continuous or categorical data even in the presence of potential interactions and nonlinearity [42]. Differences in pre-post outcome measures (rumination and cognitive composite score) were assessed using repeated measures ANOVA with group as a between subject factor and time point (pre or post) and test as within-subject factors. Significant ANOVAs were followed up with 2-tailed t tests. Post hoc power analysis showed that with our sample size, we were only 15% powered to find an effect similar to previous single-session studies, which found small effects on emotional functioning (d=0.10-0.30). Based on the effect sizes in these studies, 139 participants per group would be needed for an α=.05 and a power of 0.80. As such, the purpose of this study is only to determine preliminary efficacy.

**Qualitative Data Collection and Analysis**

We created an interview guide based on acceptability and feasibility concepts from Sekhon et al [34] and Bowen et al [32]. One portion of the interview guide consisted of 2 questions asked of all participants (“would you recommend this intervention to others with thinking difficulties?” and “did you feel you benefited from this intervention?”). The second portion of the interview guide contained flexible questions and prompts related to intervention cohesiveness, perceived effectiveness, usability, feasibility, and preferences. Example prompts from the interview guides are included in Table 2.

Interviews were conducted by 2 researchers at least 1 month after the second visit, and each interview was conducted by a researcher not involved in intervention delivery for each participant. Interviews lasted 20–45 minutes and were recorded and transcribed verbatim by the primary researcher (TA). Following each interview, a memo was written describing the conversation, themes observed, and experience of the interviewer. During the interview and transcription, responses were put under the construct addressed. If a response did not appear to address a construct, it was written at the bottom of the interview form. Following transcription, these responses were read and coded according to pre-established codes. Pre-established codes and themes were identified through discussions with participants before initiating the formal qualitative portion of the project. Additionally, themes that appeared in both the data memos before the transcription process and after all interviews were completed were coded. All interviews were read and coded for evidence of emerging codes. A secondary coder met to review and discuss the assigned codes. Discrepancies between coders were resolved by discussion.
Table 2. Quotes about veterans’ experiences of the intervention.

<table>
<thead>
<tr>
<th>Constructs and examples of interview questions</th>
<th>Veterans’ responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant satisfaction</strong></td>
<td></td>
</tr>
<tr>
<td>• Would you recommend this program to other Veterans?</td>
<td>92% (11/12) said “yes”</td>
</tr>
<tr>
<td>• Do you feel you benefited from this intervention?</td>
<td>92% (11/12) said “yes”</td>
</tr>
<tr>
<td><strong>Perceived difficulties: preintervention</strong></td>
<td></td>
</tr>
<tr>
<td>• What lead you to sign up for this intervention? Did you have any specific goals going into the intervention or areas you wanted help with?</td>
<td>“It takes me longer to complete tasks,…. Longer and longer to complete tasks”</td>
</tr>
<tr>
<td>• What were some areas of difficulty before the intervention?</td>
<td>“I’m having [a] lot of issues with concentration staying on task, forgetfulness, when I was younger I had ADD [attention deficit disorder]. I didn’t know about adult ADD and it is something you don’t grow out of?”</td>
</tr>
<tr>
<td>• What were some areas of difficulty before the intervention?</td>
<td>“To improve memory- that was the hope, but no expectations”</td>
</tr>
<tr>
<td>• What were some areas of difficulty before the intervention?</td>
<td>“[I] got out of the military, can’t concentrate. I quit things, can’t wrap my head around them”</td>
</tr>
<tr>
<td>• What were some areas of difficulty before the intervention?</td>
<td>“Hard time coming up with my own thoughts”</td>
</tr>
<tr>
<td>• What were some areas of difficulty before the intervention?</td>
<td>“Don’t remember things”</td>
</tr>
<tr>
<td>• What were some areas of difficulty before the intervention?</td>
<td>“Walk from one room to another room, go in the room and forget why I’m there”</td>
</tr>
<tr>
<td><strong>Initial expectations</strong></td>
<td></td>
</tr>
<tr>
<td>• What were your initial expectations of the intervention?</td>
<td>“It was supposed to help with my memory, correct? It sounds like a cliché but I didn’t really have any expectations.”</td>
</tr>
<tr>
<td>• How did your referring provider explain the intervention to you?</td>
<td>“It was about memory loss, wasn’t it? I guess thinking, maybe some ideas on how to improve my memory:”</td>
</tr>
<tr>
<td>• How did your referring provider explain the intervention to you?</td>
<td>“Well, they were doing a study for Veterans who were having cognitive issues.”</td>
</tr>
<tr>
<td>• How did your referring provider explain the intervention to you?</td>
<td>“I didn’t really know what to expect.”</td>
</tr>
<tr>
<td>• How did your referring provider explain the intervention to you?</td>
<td>“I originally from what I heard and read into, I thought it would be giving me some skills to help combat the cognitive decline I have had over the past of the few years, not mentally a decline, but a slowing of my thought process.”</td>
</tr>
<tr>
<td>• How did your referring provider explain the intervention to you?</td>
<td>“Not really sure, I didn’t really, didn’t really have much. I guess to help others.”</td>
</tr>
<tr>
<td><strong>Perceived effectiveness</strong></td>
<td></td>
</tr>
<tr>
<td>• Have any specific things in your life changed?</td>
<td>“The wife has noticed my attitude and like that I have helped and she thought my participation was positive”</td>
</tr>
<tr>
<td>• Do you feel you benefited from the intervention?</td>
<td>“My wife has picked up the fact I am a bit more in tune, especially if we are in a public situation.”</td>
</tr>
<tr>
<td>• Has another (family members, coworkers, or medical team) noticed changes in either your mood or cognitive functioning?</td>
<td>“My wife has mentioned [me] being in a better mood. Getting up and doing more things. Now being a bit more detail focused”</td>
</tr>
<tr>
<td>• Has another (family members, coworkers, or medical team) noticed changes in either your mood or cognitive functioning?</td>
<td>“I do. It’s not like taking a pill and automatically fixed. It takes work on my part. Still can’t get over that hump in real life.”</td>
</tr>
<tr>
<td><strong>Barriers to participation or using strategies</strong></td>
<td></td>
</tr>
<tr>
<td>• Were there any barriers to participation?</td>
<td>“If they could email appointment reminders, like a hard copy, just as a physical reminder. And y’all may have done that, I don’t know”</td>
</tr>
<tr>
<td>• Were there any barriers to participation?</td>
<td>“I just… I just hardly ever have my ringer on on my phone. I have problems following through with stuff like that”</td>
</tr>
<tr>
<td>• Were there any barriers to participation?</td>
<td>“I have never actually been tested, but I think I have a learning disability, but I think I have dyslexia. Yes, that’s one of the barriers. Had to read the questions a couple different times.”</td>
</tr>
<tr>
<td>• Were there any barriers to participation?</td>
<td>“Technical difficulties with VVC [VA video connect system], took about an hour or so for the questions.”</td>
</tr>
<tr>
<td><strong>Target population</strong></td>
<td></td>
</tr>
<tr>
<td>• Who do you think would benefit from a program like this?</td>
<td>“Combat veterans for sure”</td>
</tr>
<tr>
<td>• Who do you think would benefit from a program like this?</td>
<td>“Would be helpful to do post-deployment”</td>
</tr>
</tbody>
</table>
Results

Quantitative Findings
There were no significant differences between groups on preintervention measures of rumination, perseverative thinking, or mood. There was a significant difference on one measure of neuropsychiatric symptoms, with higher reported cognitive difficulties on the NSI in the treatment group (Table 1). Both groups reported high levels of neuropsychiatric symptoms, rumination, and perseverative thinking, as well as severe symptoms of depression and moderate symptoms of anxiety.

Acceptability
A total of 92% (11/12) of participants reported they benefited from the intervention and would recommend the intervention to other veterans with similar difficulties. The 1 veteran who reported they did not benefit directly from the program reported benefiting somewhat from the intervention, as it was able to provide a referral to more intensive cognitive rehabilitation treatment.

Efficacy
When examining the postintervention outcome measures (rumination and the cognitive composite index), there was a significant interaction between time and group for the cognitive composite index ($F_{1,13}=5.97; P=.03$), with significant improvement in the intervention group and not the control group (small effect $d=0.10$). There was not a significant interaction between time and group on measures of self-reported rumination.

Qualitative Findings
Preintervention Expectations and the Motivation to Participate
Veterans reported few, if any, expectations for the intervention, with comments like “I didn’t really know what to expect” or reported general expectations about improving cognition, such as “It was supposed to help with my memory, correct?” (Table 2 contains participant responses). A total of 92% (11/12) veterans reported that their primary motivation to start the intervention was to have tools to help with their daily difficulties with cognition. One veteran reported that his primary motivation to participate was to improve future treatments for other veterans experiencing cognitive problems.

Acceptability
When asked about recommending the intervention to others with similar difficulties, several veterans (n=4 with similar comments) reported that the intervention “would be helpful to do postdeployment” and should be given to all “combat veterans for sure” before the development of cognitive difficulties. The feedback on perceived effectiveness was generally positive, with most veterans reporting improvement in daily functioning. Notably, more veterans reported their spouse or partner had noticed improvement (n=6) compared to those who stated they had noticed improvement themselves (n=4).

Feasibility
A barrier to participation in the study was the required remote consenting procedures. A total of 23 veterans expressed interest but did not return the consent form. Those who participated in the intervention reported that the remote consenting process was more difficult compared to previous in-person consenting for research participation. Another barrier to participation was forgetting appointments, and many participants had to be rescheduled multiple times due to forgetting appointments. Veterans were called the week of their appointment to remind them of their appointment; however, they requested additional reminders and reminders in digital modalities, such as “email appointment reminders, like a hard copy, just as a physical reminder.” Retention rates improved when the telehealth system started to send out SMS text message reminders of the appointment along with the email reminder.

Even though 92% (11/12) of participants said they recommended the intervention, only 2 out of the 12 veterans were able to describe and recall an exercise from the study, and several indicated that they forgot what the exercises were. Of the 10 veterans who did not remember the specifics of the intervention, 2 reported it was helpful to learn more about additional resources for both cognitive skills and treatments for mood, which they were able to pursue after the intervention. A veteran reported it was helpful as “this helped me put things into words and helps me understand. Sometimes you don’t know how to put things into words.” He reported that being able to explain both his cognitive difficulties and mood symptoms had improved communication with his family and members of his health care team. In contrast to the intervention exercises, the psychoeducation portion was noted as helpful in interviews with 10 veterans.

Discussion
Overview
We found the preliminary efficacy of a 1-session rumination-focused intervention to improve cognitive symptoms; however, there was no change in rumination, which was the proposed treatment target and mechanism. In qualitative interviews, veterans reported this intervention was acceptable and beneficial, as evidenced by over 92% (11/12) stating they benefited from the intervention and would recommend the intervention to others with similar difficulties. In addition to being acceptable to veterans, there was a high level of perceived effectiveness and intervention cohesiveness. Due to the discrepancy between self-reported improvement and the scores on the RRS-SF, further research may benefit from exploring the use of other rumination or perseverative thought measures.

The psychoeducation and brief intervention modules have potential utility in bridging care after neuropsychological or neurological evaluations and subsequent referrals to mental health treatment. Brief, internet-administered interventions have been shown to be effective in bridging care for mental health symptoms [28]. The brief interventions described in the module and handouts can then be used to build awareness while waiting to be seen by mental health clinicians or for existing mental health patients to understand how their treatment may improve...
their cognitive functioning. In these cases, the intervention can be incorporated into the feedback session and may yield additional “buy-in” from veterans to fully participate in mental health treatment. Due to the large number of veterans reporting SCC, telehealth-based intervention is an efficient and cost-effective way to meet these needs [43].

There are several limitations in this study, including the small sample size, which limited statistical analyses. Our pilot work suggests that veterans with subjective cognitive changes are amenable to psychological treatment and perceive benefit from a short, 1-session intervention. Further work with a larger sample size is needed to fully evaluate efficacy and whether the intervention described here should be implemented more broadly. Additionally, a larger sample size will allow further exploration of how changes in rumination may or may not mediate the relationship between SCCs and cognition. Further studies can then evaluate whether the psychoeducation about rumination is sufficient to fulfill the long-term goals of the intervention, namely, to bridge care following referral to mental health services and increase veteran buy-in for participation in psychological treatments to improve cognition. Further research in this area could also explore the use of a longer intervention as a standalone treatment.

**Conclusion**

In conclusion, we found preliminary evidence for the feasibility, acceptability, and efficacy of a 1-session rumination-focused intervention for veterans with SCCs, which will benefit from continued evaluation of this intervention as well as comparison to routine clinical practice.

**Data Availability**

The data sets generated during or analyzed during this study are available from the corresponding author on reasonable request.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

"Worry less, remember more" intervention and resources.

[DOCX File, 34 KB - formative_v81e48525_app1.docx ]

**References**


Abbreviations

BAI: Beck Anxiety Inventory
BDI-II: Beck Depression Inventory-Second Edition
BRIEF-A: Behavior Rating Inventory of Executive Functioning-Adult
MoCA: Montreal Cognitive Assessment
NSI: Neurobehavioral Symptom Inventory
PTSD: posttraumatic stress disorder
RRS-SF: Ruminative Response Scale and Perseverative Thinking Questionnaire-Short Form
SCC: subjective cognitive concern
VA: Veterans Administration

Edited by A Mavragani; submitted 26.04.23; peer-reviewed by A Ecker, L Yu; comments to author 01.11.23; revised version received 19.12.23; accepted 21.12.23; published 12.04.24.

Please cite as:
Austin T, Smith J, Rabin B, Lindamer L, Pittman J, Justice S, Twamley EW, Lantrip C
The Effects of a Single-Session Virtual Rumination Intervention to Enhance Cognitive Functioning in Veterans With Subjective Cognitive Symptoms: Multimethod Pilot Study
JMIR Form Res 2024;8:e48525
URL: https://formative.jmir.org/2024/1/e48525
doi:10.2196/48525
PMID:38608264

©Tara Austin, Jennifer Smith, Borsika Rabin, Laurie Lindamer, James Pittman, Staley Justice, Elizabeth W Twamley, Crystal Lantrip. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Novel Approach for Detecting Respiratory Syncytial Virus in Pediatric Patients Using Machine Learning Models Based on Patient-Reported Symptoms: Model Development and Validation Study

Shota Kawamoto, MMG; Yoshihiko Morikawa, MD; Naohisa Yahagi, MD, PhD
Graduate School of Media and Governance, Keio University, Fujisawa, Japan

Corresponding Author:
Naohisa Yahagi, MD, PhD
Graduate School of Media and Governance
Keio University
5322 Endo
Fujisawa, 252-0882
Japan
Phone: 81 466 49 3404
Email: yahagin@sfc.keio.ac.jp

Abstract

Background: Respiratory syncytial virus (RSV) affects children, causing serious infections, particularly in high-risk groups. Given the seasonality of RSV and the importance of rapid isolation of infected individuals, there is an urgent need for more efficient diagnostic methods to expedite this process.

Objective: This study aimed to investigate the performance of a machine learning model that leverages the temporal diversity of symptom onset for detecting RSV infections and elucidate its discriminatory ability.

Methods: The study was conducted in pediatric and emergency outpatient settings in Japan. We developed a detection model that remotely confirms RSV infection based on patient-reported symptom information obtained using a structured electronic template incorporating the differential points of skilled pediatricians. An extreme gradient boosting–based machine learning model was developed using the data of 4174 patients aged ≤24 months who underwent RSV rapid antigen testing. These patients visited either the pediatric or emergency department of Yokohama City Municipal Hospital between January 1, 2009, and December 31, 2015. The primary outcome was the diagnostic accuracy of the machine learning model for RSV infection, as determined by rapid antigen testing, measured using the area under the receiver operating characteristic curve. The clinical efficacy was evaluated by calculating the discriminative performance based on the number of days elapsed since the onset of the first symptom and exclusion rates based on thresholds of reasonable sensitivity and specificity.

Results: Our model demonstrated an area under the receiver operating characteristic curve of 0.811 (95% CI 0.784-0.833) with good calibration and 0.746 (95% CI 0.694-0.794) for patients within 3 days of onset. It accurately captured the temporal evolution of symptoms; based on adjusted thresholds equivalent to those of a rapid antigen test, our model predicted that 6.9% (95% CI 5.4%-8.5%) of patients in the entire cohort would be positive and 68.7% (95% CI 65.4%-71.9%) would be negative. Our model could eliminate the need for additional testing in approximately three-quarters of all patients.

Conclusions: Our model may facilitate the immediate detection of RSV infection in outpatient settings and, potentially, in home environments. This approach could streamline the diagnostic process, reduce discomfort caused by invasive tests in children, and allow rapid implementation of appropriate treatments and isolation at home. The findings underscore the potential of machine learning in augmenting clinical decision-making in the early detection of RSV infection.

(JMIR Form Res 2024;8:e52412) doi:10.2196/52412

KEYWORDS
respiratory syncytial virus; machine learning; self-reported information; clinical decision support system; decision support; decision-making; artificial intelligence; model development; evaluation study; detection; respiratory; respiratory virus; virus; machine learning model; pediatric; Japan; detection model
Introduction

Every winter, respiratory syncytial virus (RSV) causes acute lower respiratory tract infections in approximately 33.8 million children younger than 5 years worldwide [1]. Approximately all children are infected at least once, and half are infected twice or more by the age of 24 months [2]. Newborns and children with underlying medical conditions are particularly susceptible to severe infection [3-5]. Therefore, reducing the number of RSV-infected patients is paramount for reducing the number of associated deaths. Consequently, there is an urgent need to develop a quick and accurate detection system for RSV infection [6].

Laboratory testing of RSV, including rapid antigen testing and polymerase chain reaction tests, provides reasonably accurate infection-related information [7]. However, the collection of nasopharyngeal secretions, a necessary step for these tests, can cause discomfort in children. Furthermore, given that RSV test results rarely influence treatment decisions, these tests are not routinely conducted [8,9]. On a positive note, RSV infections are not serious in most cases, and there is no active treatment, indicating that most patients can be treated at home under the constant supervision of parents or caregivers while taking precautions [10,11]. Thus, remote identification of RSV infection allows patients to be cared for at home, preventing infection spread [12]. This approach could also ease the burden on health care workers during epidemics by providing remotely procured information. Nonetheless, a home-based detection method that matches the accuracy of a rapid antigen test has yet to be recognized.

The symptoms and signs of RSV infection may help establish remote detection strategies. Studies that focused on the detection of RSV infection based on symptoms either lacked discriminatory accuracy or highlighted difficulties because of the diverse clinical manifestations of RSV infection [13-18]. However, symptom onset of RSV infection may not appear as a cross-sectionally typical pattern at a specific time point but rather as a pattern that is diverse in characteristics, including the longitudinal aspects of symptoms. Particularly, the symptoms of RSV infection peak 4-5 days after infection and change with the increase or decrease in viral load [19,20]. Dyspnea and other lower respiratory symptoms, including wheezing, moaning, and tachypnea, which are typical symptoms of severe RSV infection, occur when infected ciliated bronchial epithelial cells drop into the lower respiratory tract, thereby delaying the manifestation of upper respiratory symptoms [21-23]. Contrarily, most studies linking RSV infection to overt symptoms used cross-sectional data based on symptoms at specific time points and did not consider the time course of symptom onset in the longitudinal profile of individual children.

Therefore, we propose that cross-sectional studies based on specific time points of signs and symptoms expressed with RSV infection appear to be unrelated to RSV infection, but this may not represent the unique disease trajectory of RSV infection in individual patients infected with RSV. Therefore, machine learning models based on symptom data structured to include longitudinal characteristics may enable highly accurate identification of viral infection. As the progression of symptoms can exhibit multiple patterns in each individual, considering aspects such as the size of patient bronchi and machine learning algorithms, which are already widely used to diagnose and classify diseases based on symptom characteristics, are equally suitable for identifying RSV infection [24-27]. Here, we sought to leverage the longitudinal diversity of symptoms using machine learning based on patient-reported information, aiming to confirm the presence of RSV infection remotely at home with sensitivity and specificity comparable to those of rapid antigen testing [28]. If this strategy is recognized, it will contribute to reducing the physical burden on children, saving medical costs, and preventing nosocomial infections. The purpose of this study is to develop and validate a machine learning–based RSV infection identification model using patient self-reported symptom information for outpatients. This study is also valuable as it is one of the few studies conducted in a cohort of outpatients with mild infections.

Methods

Overview

The results are presented according to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) statement [29].

Ethical Considerations

The study was conducted according to the Declaration of Helsinki and Japan’s ethical guidelines. The institutional review board of Yokohama City Municipal Hospital approved the protocol (18-05-04), and we obtained informed consent from the patients’ parents in the form of an opt-out clause. In this study, analysis was conducted using data that had been anonymized to ensure the privacy and confidentiality of participants. No compensation was provided to the participants.

Data Collection Setting

This observational retrospective cohort study involved patients aged ≤24 months who visited the pediatric or emergency department of Yokohama City Municipal Hospital between January 2009 and December 2015 and had RSV rapid antigen test results. According to the facility policy, all outpatients were required to fill an electronic template, and those who exhibited test results. According to the facility policy, all outpatients were required to fill an electronic template, and those who exhibited cold symptoms were subjected to RSV rapid antigen testing. These patients were extracted from a prospectively curated database and enrolled in this study. In this study, we used the immunochromatographic Quick-Navi RSV test (Denka Seiken Co Ltd) for nasopharyngeal swab fluid testing as the gold standard. Patients who presented weakly positive results in the rapid antigen test were excluded.

Data Preparation

An automated medical interview system with an electronic template was introduced to standardize the entry of clinical symptoms, and the patient’s parents completed the form before the hospital visit. A group of highly trained general pediatric attending physicians with over 15 years of clinical experience created the template for entering symptoms and signs, allowing parents to select up to 3 symptoms per entry. Once a symptom
was selected, the system presented additional questions based on the selection, with all responses except temperature being optional. Therefore, the status of each symptom, along with the number of elapsed days since symptom onset, was recorded as a categorical variable. Additional questions for each symptom were presented differently for each age group; all questions and options used are listed in Multimedia Appendix 1. The feature set was solely based on information from the electronic template and did not include any additional data from the examination or treatment. Based on medical insights, the symptoms introduced to the models were limited to cough, runny nose or nasal congestion, and wheezing, using only baseline characteristics and overall health status features. Statistical feature selection was not performed.

**Experimental Process**

We developed and evaluated a machine learning model that outputs binary information on RSV infection based on symptom and sign information. Users can obtain information about RSV infection status by entering symptoms to determine whether to seek medical attention.

**Models**

Random forest, extreme gradient boosting (XGBoost), and support vector machine models were used to determine appropriate machine learning algorithms. We used grid search in a hyperparameter space for all classifiers, optimizing the hyperparameters based on 10-fold cross-validation. The area under the receiver operating characteristic curve (AUC-ROC) was calculated for each model using the optimized hyperparameters. Finally, we selected the machine learning algorithm and its corresponding hyperparameters that performed best.

**Model Performance Evaluation**

Model performance was assessed through calibration and discrimination. Calibration was evaluated graphically using a calibration plot and Hosmer-Lemeshow test with 10 groups, where \( P < 0.05 \) indicated a poor model fit. The AUC-ROC sensitivity, and specificity were used to evaluate model discrimination ability. Sensitivity and specificity were calculated using the Youden index, and performance was calculated under conditions, where 1 parameter was fixed to be equivalent to that of the rapid antigen test. Additionally, we calculated model discriminatory power based on the number of elapsed days since the onset of illness to assess the use of rapid patient isolation. The performance of a valid model should not be considerably worse even in short periods after disease onset. The discrimination metrics of the final model were evaluated to estimate stability using the 1000 times bootstrap method. This resampling technique, which involves generating multiple bootstrap samples and using them to train and test the model, offers a robust estimate of evaluation indices even with small samples by correcting for optimism in the model’s performance [30].

To further assess the effect of additional symptom information on detection accuracy, we constructed a baseline model that classified patients based solely on the presence or absence of symptoms. As RSV infection is seasonal and its prevalence varies by season, a model excluding only the month of hospital visit was also created for comparison.

**Interpretability Evaluation**

We evaluated the interpretability of the final model using Shapley additive explanations (SHAP), calculated using an algorithm that mimics the Shapley value used in game theory to evaluate the relative importance of each feature on discrimination performance while considering interactions. Therefore, it was used to corroborate the presence of essential variables and interactions. R (version 4.2.0; R Foundation for Statistical Computing) was used for all analyses.

**Results**

**Patient Characteristics**

Between January 2, 2009, and December 31, 2015, a total of 7362 patients underwent rapid antigen tests, and their parents provided the necessary information through an electronic template. Of these, 4182 patients who were aged 24 months or younger were included in the analysis. One patient with weak positive test results and 7 with inaccurate age information were excluded. Figure 1 depicts the process of patient exclusion, data selection, and missing value completion. Of the remaining 4174 patients, 619 (14.8%) were positive and 3555 (85.2%) were negative for RSV infection. Table 1 presents the demographic and clinical characteristics of patients. The details of the features used in the model are provided in Multimedia Appendix 1.
Figure 1. Sequence of steps in patient selection and data preprocessing.

Table 1. Demographics and some important variables of the selected data (n=4174).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (days), median (IQR)</td>
<td>340.0 (174.0-501.0)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>2348 (56.3)</td>
</tr>
<tr>
<td>Respiratory syncytial virus-positive, n (%)</td>
<td>619 (14.8)</td>
</tr>
<tr>
<td>Current body temperature (°C) median (IQR)</td>
<td>38.0 (37.0-39.0)</td>
</tr>
<tr>
<td>Maximum body temperature (°C) median (IQR)</td>
<td>39.0 (38.0-40.0)</td>
</tr>
<tr>
<td>Cough, n (%)</td>
<td>2323 (55.7)</td>
</tr>
<tr>
<td>Wheezing, n (%)</td>
<td>1388 (33.3)</td>
</tr>
<tr>
<td>Nasal discharge or nasal distraction, n (%)</td>
<td>1944 (46.7)</td>
</tr>
</tbody>
</table>

aRSV: respiratory syncytial virus.
Machine Learning Algorithm Selection

During the algorithm selection process, XGBoost demonstrated the highest AUC-ROC at 0.825 (95% CI 0.772-0.875), outperforming other algorithms such as support vector machine and random forest, which showed AUC-ROCs of 0.704 (95% CI 0.665-0.740) and 0.798 (95% CI 0.756-0.857), respectively. Therefore, we adopted XGBoost for the proposed and baseline models.

Calibration and Discrimination Ability of the Differential Models

The model with the XGBoost algorithm fit well visually, with good calibration in the Hosmer-Lemeshow goodness-of-fit test ($P=.27$). For discrimination ability, the AUC-ROC of the estimated model calculated with 1000 times bootstrap was 0.811 (95% CI 0.784-0.833). The sensitivity and specificity were 73.5% (95% CI 66.8%-79.4%) and 73.9% (95% CI 70.3%-77.2%), respectively. To validate the exclusion performance of the proposed model, the threshold was adjusted according to the discrimination performance of the rapid antigen test [28]. When the sensitivity was set to 71.6% (95% CI 64.8%-77.8%), approximately 68.7% (95% CI 65.4%-71.9%) of the total patients were predicted to be negative for RSV infection. Patient samples were predicted to be positive (6.9%, 95% CI 5.4%-8.5%) when the specificity was set to 96.6% (95% CI 95.3%-97.7%). For the baseline model, which was considered to determine the effect of additional symptom information on performance improvement, the AUC-ROC was 0.766 (95% CI 0.739-0.792) for the model that excluded the month of visit (model 2), 0.703 (95% CI 0.677-0.731) for the model that considered only the presence of symptoms (model 3), and 0.521 (95% CI 0.484-0.556) for the null model that relied solely on age (model 4). The discrimination performance of the proposed model is summarized in Table 2.

Model prediction performance is calculated by adjusting thresholds under different conditions. The scores are based on 1000 times bootstrap, with means and 95% CIs of each sample.

Table 2. Discrimination performance under different conditions.

<table>
<thead>
<tr>
<th>Condition and metric</th>
<th>Values, mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (AUC-ROC)</td>
<td>0.811 (0.784-0.833)</td>
</tr>
<tr>
<td><strong>Youden index maximized (with threshold: 0.142)</strong></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>73.5 (68.8-79.4)</td>
</tr>
<tr>
<td>Specificity</td>
<td>73.9 (70.3-77.2)</td>
</tr>
<tr>
<td><strong>Sensitivity equivalent to that of the rapid antigen test (with threshold: 0.152)</strong></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>71.6 (64.8-77.8)</td>
</tr>
<tr>
<td>Specificity</td>
<td>75.7 (72.0-78.9)</td>
</tr>
<tr>
<td>Percentage predicted negative</td>
<td>68.7 (65.4-71.9)</td>
</tr>
<tr>
<td><strong>Specificity equivalent to that of the rapid antigen test (with threshold: 0.463)</strong></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>27 (21.0-33.6)</td>
</tr>
<tr>
<td>Specificity</td>
<td>96.6 (95.3-97.7)</td>
</tr>
<tr>
<td>Percentage predicted positive</td>
<td>6.9 (5.4-8.5)</td>
</tr>
</tbody>
</table>

*AUC-ROC: area under the receiver operating characteristic curve.

Discrimination Based on Days Since Onset

Elapsed time was measured by counting the number of days since the initial onset of symptoms. A total of 46.8% (n=1952) of patients reported that symptoms such as cough, nasal discharge, or wheezing began within 3 days, and 77.4% (n=3231) of RSV-positive patients developed these symptoms within 8 days. The AUC-ROC was 0.721 (95% CI 0.628-0.815) for patients on the day of symptom onset and 0.746 (95% CI 0.694-0.794) and 0.779 (95% CI 0.749-0.808) for patients within 3 and 8 days of onset, respectively. To assess the robustness of these AUC-ROC values, 1000 bootstrap samples were used for computation, as depicted in the box and whisker plot in Figure 2. This figure illustrates the distribution of AUC-ROC values across the elapsed days since symptom onset, highlighting that the variance in AUC-ROC values diminishes as the sample size increases with more elapsed days.
Interpretability of the Final Model

Variables contributing to the prediction were examined using SHAP. The variable that primarily contributed to the relative prediction performance was the month of visit, followed by the number of days from the onset of cough and maximum body temperature. Based on the Beeswarm plot, the number of days from the onset of the cough variable exhibited bifurcation along the x-axis, indicating variability in its impact on the model’s output. However, age or current body temperature did not show a clear trend between the feature value and SHAP (Figure 3).

Figure 2. Area under the receiver operating characteristic curve (AUC-ROC) for elapsed days in the proposed model.
Figure 3. Average Shapley additive explanations (SHAP) value for each feature from the top in order of importance.

Discussion

Principal Results

We developed a machine learning tool that leverages longitudinal diversity of symptoms for the remote detection of RSV infection at home. The model that incorporated the time course of the onset of symptoms and their evolution (model 2) exhibited enhanced discriminative performance. However, our final model, including information on onset days and month of visit (model 1), performed effectively in terms of sensitivity and specificity. By adjusting the threshold based on 2 exclusion criteria set at the same level as the rapid antigen test, this model applied thresholds of 0.152 or 0.463, thereby performing an exclusion accuracy comparable to the standard RSV detection method and successfully identifying 75.6% of infected patients for exclusion. This approach could potentially reduce the need for further confirmatory testing—most positive cases comprised samples with values above the threshold based on high sensitivity. Samples with values below the threshold had almost no positives and could be considered negatives. Samples with values below the threshold based on high specificity comprised mostly of negatives, indicating that samples with values above the threshold could be considered as positives. Nevertheless, rapid testing should be continued for patients who fall within the defined thresholds. The consistency between the estimated positive probability and the actual percentage was confirmed using calibration. Hence, there is no reason to perform additional testing on samples labeled based on the 2 thresholds.

When considering the number of days since disease onset, our proposed model demonstrated an AUC-ROC of approximately 0.721 for patients who began to experience symptoms on day 1, which is the visiting day, showing particularly high accuracy when including patients with symptoms that emerged within 5 days. The results indicate that discriminatory ability improved within approximately 6 days; however, it was high even on the day of onset.

Comparison With Previous Studies

The model used here showed a higher discriminative performance than those used in other studies on symptom-based RSV infection detection. For children, a model with 80% sensitivity, 68% specificity, and an AUC-ROC of 0.66 has been reported [15]. A model with 72.8% sensitivity and 73.2% specificity has also been reported; however, this model included x-ray and laboratory test results as features, which differed from the variables we used, where patients even outside the hospital could ascertain RSV infection by themselves [16]. Therefore, the results are considered noteworthy in terms of identification accuracy. To the best of our knowledge, this is the first study to include outpatients and obtain symptom information through nonmedical personnel, which is in contrast to previous studies that involved inpatients or data obtained by health care professionals [13-18].
Our results indicate that adequately accurate predictions can be acquired using machine learning and symptom information. The detection of RSV infection based solely on patient-reported symptoms is still in its early stages; however, our tool shows robust capabilities in distinguishing positive and negative results. The tool estimates RSV infection based on symptom data and progression entered by the patient’s caregivers at home, indicating that the intervention of medical personnel is not required. This remote detection strategy could potentially reduce the risk of nosocomial infections and physical burden on children. Moreover, by applying this tool, isolation measures can be implemented before visiting a medical facility. This will allow RSV infections to be detected at home, reducing the need to visit hospitals, thereby preventing secondary and nosocomial infections and reducing the burden on health care providers, especially during an epidemic, and protect them from coinfection with RSV and severe acute respiratory syndrome coronavirus 2 [32]. In addition, it may limit the spread of the virus in the community.

Another unique feature of this study is that the model was developed based on 3 primary symptoms: cough, nasal discharge, and wheezing, which were selected and extracted from the system based on the clinical manifestations of RSV infection. Therefore, although other symptoms were recorded in the electronic template, they were not used here based on clinical rationale.

**Limitations**

There are several limitations to this study. First, participant demographics were limited as this study was conducted at a single institution. However, it is worth noting that Japan’s insurance system, which allows for free access, may mitigate potential economic-related biases in our findings. Additionally, the rapid antigen test may yield false positives when there is a low disease incidence [33,34]. Thus, the final diagnostic result, used as the gold standard here, may vary. The RSV prevalence pattern may already be changing and combining this with a surveillance system would further improve accuracy. Furthermore, model calibration must be confirmed based on race to assess performance differences between races when introduced to populations with widely varying demographics. In this study, we used a single cohort for internal validation and did not perform external validation, necessitating further testing on additional data sets to confirm generalizability.

**Future Research Directions**

Future prospective studies are required to assess the generalizability of this algorithm to all patients because they may differ from the retrospective cohort used in this study. The model developed in this study is specific for RSV infection; however, similar methods may be used to construct models to detect infections with other viruses using their respective symptoms.

**Conclusions**

Our detection tool was based on patient-reported symptoms and basic attribute information; nevertheless, it effectively detected RSV infection. Furthermore, our findings highlight the necessity to develop machine learning models and support the use of structural data for capturing complex patterns for symptom-based detection of RSV infection. The presented model leverages the distinct temporal patterns of RSV symptoms, allowing accurate identification of the infection even at early stages and with symptom evolution. Health care providers can perform model analysis before an outpatient visit to direct infected patients to home treatment or an appropriate isolation cohort. Applying this model to other settings can validate a standardized and comprehensive approach to improve RSV infection detection at home, and it could then be applied to other viruses.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

**Authors' Contributions**

SK analyzed the patient-reported data. YM interpreted the patient data and results and substantially contributed to manuscript writing. NY led and supervised the study. All authors read and approved the final manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

All questions and options for model development.

[DOCX File, 28 KB - formative_v8i1e52412_app1.docx ]

**References**


Abbreviations

- **AUC-ROC**: area under the receiver operating characteristic curve
- **RSV**: respiratory syncytial virus
- **SHAP**: Shapley additive explanations
- **TRIPOD**: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
- **XGBoost**: extreme gradient boosting

Please cite as:

Kawamoto S, Morikawa Y, Yahagi N

**Novel Approach for Detecting Respiratory Syncytial Virus in Pediatric Patients Using Machine Learning Models Based on Patient-Reported Symptoms: Model Development and Validation Study**

JMIR Form Res 2024;8:e52412
URL: https://formative.jmir.org/2024/1/e52412
doi:10.2196/52412
PMID:38608268

©Shota Kawamoto, Yoshihiko Morikawa, Naohisa Yahagi. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e52412 Jmir Form Res 2024 | vol. 8 | e52412 | p.1805 (page number not for citation purposes)
Developing a Multiprofessional Mobile App to Enhance Health Habits in Older Adults: User-Centered Approach

Andressa Crystine da Silva Sobrinho¹; Grace Angelica de Oliveira Gomes²; Carlos Roberto Bueno Júnior¹

¹Faculty of Medicine of the University of São Paulo, Ribeirão Preto, Brazil
²Department of Gerontology at the Federal University of São Carlos, São Carlos, Brazil

Corresponding Author:
Andressa Crystine da Silva Sobrinho
Faculty of Medicine of the University of São Paulo
Bandeirantes Avenue 9000
Ribeirão Preto, 14055-230
Brazil
Phone: 55 16988155152
Email: andressa.sobrinho@usp.br

Abstract

Background: Although comprehensive lifestyle habits are crucial for healthy aging, their adherence tends to decline as individuals grow older. Sustaining a healthy life over time poses a motivational challenge. Some digital tools, such as smartphone apps aimed at promoting healthy habits, have been used to counteract this decline. However, a more profound investigation is necessary into the diverse experiences of users, particularly when it concerns older adults or those who are unfamiliar with information and communications technologies.

Objective: We aimed to develop a mobile app focused on promoting the health of older adults based on the principles of software engineering and a user-centered design. The project respected all ethical guidelines and involved the participation of older adults at various stages of the development of the app.

Methods: This study used a mixed methods approach, combining both quantitative and qualitative methodologies for data collection. The study was conducted in Ribeirão Prêto, São Paulo, Brazil, and involved 20 older adults of both genders who were aged ≥60 years and enrolled in the Physical Education Program for the Elderly at the University of São Paulo. The research unfolded in multiple phases, encompassing the development and refinement of the app with active engagement from the participants.

Results: A total of 20 participants used a mobile health app with an average age of 64.8 (SD 2.7) years. Most participants had a high school education, middle-class status, and varying health literacy (mean score 73.55, SD 26.70). Overall, 90% (18/20) of the participants owned smartphones. However, 20% (4/20) of the participants faced installation challenges and 30% (6/20) struggled with web-based searches. The focus groups assessed app usability and satisfaction. Adjustments increased satisfaction scores significantly (Suitability Assessment of Materials: 34.89% to 70.65%; System Usability Scale: 71.23 to 87.14). Participant feedback emphasized font size, navigation, visual feedback, and personalization, and suggestions included health device integration, social interaction, and in-app communication support.

Conclusions: This study contributes to the development of health care technologies tailored to the older adult population, considering their specific needs. It is anticipated that the resulting app will serve as a valuable tool for promoting healthy habits and enhancing the quality of life for older adults.

(JMIR Form Res 2024;8:e54214) doi:10.2196/54214

KEYWORDS
information and communications technologies; ICTs; health care; digital inclusion; focus groups; health promotion; user; usability; health literacy; digital competencies; digital skills; mobile phone

Introduction

The older adult population in Brazil is continuously growing [1]. As this demographic expands, critical questions arise regarding how to support older adults in their quest for independence and vitality during the aging process. Furthermore, it becomes essential to identify effective strategies to promote...
In the contemporary scenario, marked by constant technological changes, digital products have become an integral part of the context and social life [3,4]. Therefore, there is an urgent need to develop mechanisms that not only enable older adults to comprehend these tools and their potential but also enable older adults to have access to digital products that cater to their specific needs. As Mansell and Tremblay [5] pointed out, knowledge plays a fundamental role in achieving social, economic, and cultural goals, being essential for cultural integration, political participation, and market inclusion.

In this context, information and communications technologies have emerged as vital instruments for developing digital tools that offer numerous benefits to society while contributing to the autonomy and well-being of older adults [6]. These benefits span from the realm of entertainment to the professional sphere and, critically, the field of health, where technology can enhance the well-being of older adults, improve their quality of life, and keep them healthier [7,8].

Tidd and Bessant [8] emphasized that innovation not only is limited to opening new markets but also involves creating meaning and developing new ways of serving. To achieve this goal, it is fundamental to engage all stakeholders in the knowledge-building process, promoting the dissemination of information through interactions such as initiatives coordinated by coalitions of interested professionals and research projects that consider local opinions and choices [5].

Involving the end user is a crucial goal during the development of innovative solutions, not only for evaluation but also in co-design, following a user-centered strategy. Indeed, it is a significant asset of research to base the work on a user-centered approach because it allows building a platform that will address the real needs of the users [3,4].

Digital inclusion emerges as the primary gateway to the use of information and communications technology, enabling individuals to develop the capacity to seek, understand, and apply information according to their needs, acquiring information literacy and digital literacy [9,10]. However, older adults often face obstacles when dealing with these tools as many of them are developed without considering the limitations of users with limited or no digital knowledge and without addressing age-specific issues [3,4]. This mismatch between the condition of older adults and their effective inclusion in the digital environment demonstrates the need for tools that facilitate learning and interaction.

Therefore, this study aims to address the urgent need for research and the search for solutions that can mitigate existing gaps and effectively bring older adults closer to the technological environment. In this context, the central question to be addressed in this research is how to develop a multiprofessional mobile app aimed at promoting health habits in older adults, integrating the perspective of older adults themselves through focus groups to optimize the design and ensure the effective usability of this technological tool for an older adult population.

**Methods**

**Ethical Considerations and Trial Design**

This research was conducted by the lead researcher CRBJ (associate professor), GAAO (associate professor and research assistant), and ACSS with a degree in health sciences (Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo [USP]). This project followed the ethical guidelines established by Resolution 466/2012 of the National Health Council and the basic principles of bioethics. All individuals involved in the research received the informed consent form and were informed of confidentiality, with all information disclosed only in scientific events or academic publications. The protocols for this randomized crossover, triple-blind, and placebo-controlled trial were approved by the Ribeirão Preto School of Physical Education and Sport Research Ethics Committee (Process: 58433922.0.0000.5659, on October 26, 2022).

The study details were registered on The Brazilian Clinical Trials Registry (RBR-6wgkszS8: development of an application to improve health in older people). All participants read and signed an informed consent form with detailed information about the experimental protocols.

**Study Type**

This study is part of a larger project and describes one of the initial steps. This study consists of technological production research for the development of a mobile app. The approach involves both quantitative and qualitative aspects, with a cross-sectional design. Data were collected from older adults in 3 focus group meetings. In this case, the specific topic was the design of apps for older adults. Data collected in focus group studies are typically used to gain insight into people’s perspectives and experiences. In this study, the data can be used to inform the design of apps that are easier to use and accessible for older adults.

**Study Location and Period**

The study was conducted in the city of Ribeirão Preto, São Paulo, Brazil, with the participation of older adults recruited from the convenience sample of participants in the Physical Education Program for the Elderly (PEFI) at Escola de Educação Física e Esporte de Ribeirão Preto (EEFERP; EEFERP-USP). This study was conducted between January and July 2023.

**Participants**

The study included older adults aged ≥60 years who had participated at some point in the physical training program offered at the EEFERP-USP through the culture and extension project. This PEFI program includes a WhatsApp group that facilitates recruitment through media and technology for all older adults who have already participated in the program. For a focus group testing the app design, it is essential to have a variety of participants with different demographic profiles, interests, and needs [11]. Therefore, for this focus group, we included 10 men and 10 women. A flowchart regarding the recruitment of participants is described in Figure 1.
The inclusion criteria included individuals aged \( \geq 60 \) years who possessed a mobile device capable of supporting the use of the app and its features—the criterion of having a mobile device is fundamental as the 3 stages of the focus group use this equipment. The exclusion criteria included limitations in performing the assessments outlined in the research, such as visual impairments, hearing impairments, and severe cognitive impairments. The Montreal Cognitive Assessment (MoCA) [12] instrument was used to assess cognitive function and screen for dementia. It was used to analyze whether the difficulty in using the app was related to cognitive performance or the design of the app.

Research Development

The project was divided into phases to meet the proposed objectives (Figure 2).

Figure 1. Flowchart for the recruitment of participants.
Development and Adjustment of the App With the Involvement of Older Adults

In this phase, the goal was to provide information for the development of a mobile app prototype that tracks and promotes health habits in older adults. Initially, the objectives were established in accordance with the documents and processes recommended by Serviço Brasileiro de Apoio às Micro e Pequenas Empresas for business and technology development. The following documents were created to align project understanding and initiate system modeling:

- **Phase 1: business plan**: describes the business objectives and the necessary steps to achieve them and identifies the market, product, and existing tools.
- **Phase 2: scope**: defines the initial project idea and goals and serves as a guide for project production and control.
- **Phase 3: Supplier, Input, Process, Output and Customer**: a tool from the Six Sigma methodology that manages and optimizes the flow of inputs and products.
- **Phase 4: use case diagram**: documents the main system functionalities from the user’s perspective.
- **Phase 5: flowchart**: graphically represents the sequence of screens and their access paths, providing an overall view of the product.
- **Phase 6: wireframing**: an Agile technique to create a first impression of the project and visualize the layout and flow. We used Figma (Figma, Inc), a cloud-based graphic design and user interface prototyping software. After the initial documents were prepared, wireframing was performed using Figma. The Agile methodology was adopted for communication with the lead researcher. Usability testing was conducted with 20 participants aged ≥59 years who were part of the PEFI at the EEFERP-USP. The testing was conducted using the design thinking method, allowing for feedback collection and app adjustments based on user needs.
- The app was developed with Vue and TypeScript for the frontend as a progressive web app and distributed through a Kubernetes cluster with backend communication via an application programming interface developed in Node.js, also using TypeScript. MongoDB was used as the database, hosted within Atlas to facilitate data cross-referencing and subsequent analysis (Figure 3).

These technologies were chosen for their performance and because they are native web technologies, allowing for broader distribution across various scenarios compared with a more complex framework such as Django, for example. Development time is also crucial, as changes can be made with a short build time afterward.

- **Phase 7: execution of focus groups**
Focus Group Objectives
The meetings were conducted at the EEFERP-USP, in a reserved room, a quiet location free of distractions, with audio recording for a duration of 60 to 80 minutes. In addition to the participants, there was a mediator, represented by the lead researcher, and an observer who took notes during the focus groups.

Open-ended questions about the user experience with the system were used during the meetings. In the first and last meetings, the study questionnaires were administered. Closed-ended questions were used in the meetings with caregivers to focus on the system evaluation. Video and audio recordings were used for recording and transcribing the participants’ comments on data security [11]. The layout of the room where the focus group was conducted is illustrated in Figure 4, showing the positions of participants, mediators, and recording equipment.

Focus group 1 comprised a total of 20 participants, equally divided between men and women, with 10 men and 10 women. The primary aim of the first focus group was to comprehend the perceptions of older adults regarding the purpose of the app. The tasks undertaken during this focus group included exploring the health and well-being needs of the older adults, identifying health habits deemed significant by them, gathering suggestions and tips from the older adults for the functionality and practicality of the app, and understanding the expectations of older adults regarding the design and usability of the app.

During the execution of focus group 2, some participants could not attend owing to unforeseen circumstances, medical appointments, and other events, resulting in a total of 13 participants, including 6 men and 7 women. The objective of the second meeting was to assess the early-stage mobile app and gather feedback for adjustments. The tasks carried out during the second focus group meeting included collecting opinions on the appearance, navigability, and features of the app; recording suggestions for improvements and specific adjustments desired by the participants; and exploring the willingness and readiness of older adults to use the app in the future.

In the third focus group, it was possible to bring together all 20 participants with the goal of conducting practical tests with the app, collecting feedback on its usability, and refining the ideal script for future focus group sessions. To achieve this objective, the conduct of the focus group revolved around tasks such as allowing participants to use the app in a controlled environment, documenting observations on how the older adults interacted with the app, gathering feedback on any difficulties or issues encountered during use, and identifying specific aspects of the app that needed improvement.

Questionnaires Applied
Before the development phase, various questionnaires were applied to gather comprehensive data crucial for the study.
objectives. These questionnaires covered a range of aspects including participant characterization, cognitive functioning, usability assessment, material suitability, socioeconomic classification, digital competencies, and health literacy.

1. Anamnesis: an initial questionnaire was developed to gather participant characterization for the study. This questionnaire included questions on gender, age, occupation, marital status, profession, and mobile phone brand and model.

2. MoCA: the MoCA is a cognitive screening questionnaire designed to assess cognitive functioning in older adults. The questionnaire consists of 11 items that evaluate attention, memory, reasoning, language, calculation, and orientation. Each MoCA item is scored from 0 to 3, with a maximum score of 30 [12]. A score of ≥26 is considered normal, whereas a score of ≤25 may indicate cognitive decline. However, it is important to note that the MoCA is only a screening test, and a more precise diagnosis should be obtained by consulting a health care professional [12].

3. System Usability Scale (SUS): this is a 10-item questionnaire that measures the usability of a system. It was developed by John Brooke in 1986 [13] and is one of the most widely used usability assessment tools [14]. SUS is a self-administered questionnaire used to evaluate the usability of various systems, including mobile apps, websites, software, and hardware [14]. The questionnaire comprises 10 items that assess system usability on a scale of 1 to 5, with 1 representing “strongly disagree” and 5 representing “strongly agree.” The SUS items evaluate usability in dimensions such as ease of learning, ease of use, efficiency, satisfaction, and memorability [14].

4. Suitability Assessment of Materials (SAM) questionnaire: SAM is a 10-item questionnaire that assesses the suitability of materials for specific applications. It was translated and adapted into Portuguese by Souza et al [15] and is a widely used material assessment tool. SAM is a self-administered questionnaire used to evaluate material suitability for a variety of applications, including medical products, consumer goods, and industrial products. The questionnaire comprises 10 items that assess material suitability on a scale of 1 to 5, with 1 representing “inadequate” and 5 representing “adequate.” SAM items evaluate material suitability based on physical properties, compliance with application requirements, and material safety [15].

5. Brazilian Socioeconomic Classification (CSEB): CSEB is a socioeconomic classification system developed by the Brazilian Institute of Geography and Statistics in 2010 [16]. It is based on a combination of factors, including income, education, and occupation. CSEB is used for research and social planning purposes and is an essential tool for understanding socioeconomic inequalities in Brazil. It can also be used to develop more equitable public policies [16].

6. Digital Competencies and Skills Questionnaire: digital competencies and skills were assessed through a questionnaire created by the researchers based on the Mobile Learning Competence Model for Seniors, specifically designed for older adults [17]. This questionnaire used a 3-point assessment scale: “Yes,” indicating full competence or skill in the evaluated area; “Yes, but with some difficulty,” indicating competence or skill with some challenges; and “No,” indicating the absence of competence or skill in the individual being assessed [18]. The questionnaire covered 6 fundamental categories of digital competencies and skills, including basic technology knowledge, internet navigation skills, mobile app use, digital communication, and digital resource use. The Mobile Learning Competence Model for Seniors provides a solid conceptual framework for measuring the digital competencies of older adults, allowing for a precise and contextualized analysis of their digital skills [17,18].

7. Health Literacy Test (TLS): TLS, validated from the Test of Functional Health Literacy in Adults, was developed by the Federal University of Rio Grande do Sul in 2019 [19]. It is a widely used health literacy assessment tool in Brazil. TLS for Seniors is a 20-item questionnaire that evaluates health literacy in individuals aged ≥60 years [19]. The questionnaire consists of 20 items that assess health literacy on a scale of 0 to 4, with 0 indicating “doesn’t know” and 4 indicating “knows very well.” TLS for Seniors assesses health literacy in dimensions such as understanding medical information, the ability to make health decisions, and the capacity to use health resources effectively [19]. This process will ensure the development of a mobile app that meets the needs and expectations of older adults, promoting greater adherence and effectiveness in health promotion within this population.

Data Analysis

Data analysis plays a fundamental role in understanding the perceptions, expectations, and needs of older adult participants regarding the mobile app developed in this study. The information collected through specific questionnaires was recorded in a Microsoft Excel software database. Descriptive analyses, such as mean, SD, and percentages, were conducted to characterize the sample in terms of age, gender, education, MoCA, occupational status, marital status, and economic classification. These variables were subjected to a detailed analysis using SPSS statistical software (IBM Corp), with a 2-tailed t test to check for potential differences between men and women.

Subsequently, the results of the focus groups were subjected to a qualitative content analysis. The transcriptions of the discussions were coded to identify emerging themes and patterns related to older adult participants’ perception of the app. This qualitative analysis allowed for an in-depth understanding of the participants’ opinions and suggestions presented during the focus groups. Content analysis was carried out according to Bardin [20]. A set of communication analysis techniques was used to obtain, through systematic and objective procedures for describing the content of messages, indicators that allow inferences about the production or reception conditions (inferred variables) of these message points. This analysis was divided into (1) preanalysis, (2) material exploration, and (3) data processing and interpretation.

To validate the questionnaires as a whole, the content validity index was calculated, with the result needing to be >0.90 (>90%) to be considered validated. The purpose was to measure the percentage of older adults who agreed on certain aspects of the
instrument and its items. A Likert scale was used with scores ranging from 1 to 4, where 4=entirely appropriate, 3=appropriate, 2=partially appropriate, and 1=inappropriate. The content validity index score was calculated by summing the agreement of the items marked “4” or “3” by 20 older adults.

Results

Participant Characterization

In the Results section, we provide a description of our sample based on various key variables. These data are essential for understanding the profile of the study participants. Table 1 displays the means and SDs of age and MoCA scores for both men and women.

Table 1. Sample characterization through mean, SD, and percentage of study participants (N=20).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Men (n=10)</th>
<th>Women (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>64.8 (2.7)</td>
<td>65.1 (4.3)</td>
</tr>
<tr>
<td>MoCA(^a) (points), mean (SD)</td>
<td>23.3 (4.4)</td>
<td>23.5 (4.8)</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>2 (20)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>High school</td>
<td>6 (60)</td>
<td>8 (80)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>2 (20)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>6 (60)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Single</td>
<td>2 (20)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Divorced</td>
<td>2 (20)</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Widowed</td>
<td>0 (0)</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retire</td>
<td>4 (40)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Salesperson</td>
<td>1 (10)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Engineer</td>
<td>1 (10)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Mason</td>
<td>1 (10)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Electrician</td>
<td>1 (10)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gardener</td>
<td>2 (20)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Homemaker</td>
<td>0 (0)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Nurse</td>
<td>0 (0)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Teacher</td>
<td>0 (0)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Secretary</td>
<td>0 (0)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Cook</td>
<td>0 (0)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Psychologist</td>
<td>0 (0)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Receptionist</td>
<td>0 (0)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Socioeconomic class, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle class</td>
<td>3 (30)</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Upper class</td>
<td>4 (40)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Lower class</td>
<td>3 (30)</td>
<td>2 (20)</td>
</tr>
</tbody>
</table>

\(^a\)MoCA: Montreal Cognitive Assessment.

We observed that, men had an average age of 64.8 (SD 2.7) years, whereas women had an average age of 65.1 (SD 4.3) years. The mean scores on the MoCA test were similar for men (mean 23.3, SD 4.4) and women (mean 23.5, SD 4.8). These variables did not show any statistical differences. Furthermore, we analyzed the distribution of participants according to other important variables. Regarding education, we noticed that most participants (6/10, 60% of men and 8/10, 80% of women) had completed high school. In contrast, a college degree was more common among men (2/10, 20%) than among women (1/10, 10%).

Regarding marital status, 60% (6/10) of the men were married, whereas none of the women in the sample were married. Furthermore, both groups had a similar number of single individuals (2/10, 20%), whereas the number of divorced individuals was higher among women (3/10, 30%) compared with men (2/10, 20%). A significant proportion of women were widowed (4/10, 40%), whereas none of the men in the sample fell into this category.

The occupation of the participants varied, with some groups represented in only one of the genders. For example, all retirees were men, whereas the categories “Homemaker” and the professions such as nurse, teacher, secretary, cook, psychologist, and receptionist were exclusive to women in the sample.

Finally, we explored the socioeconomic class of the participants. The middle-class status was more prevalent among women (6/10, 60%), whereas the upper-class status was more common among men (4/10, 40%). The lower class was relatively uniform among both sexes, with 30% (3/10) of men and 20% (2/10) of women belonging to this category.

These pieces of information provide a comprehensive overview of the demographic and socioeconomic profile of our participants, offering a solid foundation for the analysis and interpretation of the subsequent study results.

Characterization of Mobile Devices

Participants were also asked to provide the brand and model of their mobile device. There were 4 Samsung smartphones, 2 Motorola smartphones, 2 Redmi smartphones, and 2 iPhones. Of the 20 participants, 18 (90%) reported that the app worked on their mobile models, resulting in a functioning rate of 90% (equivalent to 9 mobile devices). The mobile model that did not work with the app was the Samsung Galaxy A02s model.

Health Literacy

A variation in scores was observed, ranging from 37 to 98. The average score of the sample was 73.55 (SD 26.70), indicating that most individuals had a good understanding of health information. Some scores stood out, such as 45% (9/20) of the participants who missed only 1 question, achieving 98 points.

- Low health literacy (20-40): Four participants scored in this range, indicating a low level of health understanding, which may affect their ability to make informed decisions about medical care.
- Moderate health literacy (40-60): Four participants scored in this range, suggesting a moderate level of health understanding. They can probably understand basic information, but may have difficulty with more complex material.
- High health literacy (60-100): Twelve participants scored in this range, indicating a high level of health understanding. They are likely to be able to understand and use health information effectively to make informed decisions about their care.

Participant classification was based on the results of the TLS. Of the 20 participants, 12 (60%) were classified as having “adequate” health literacy, with scores above 71; a total of 4 (20%) had “inadequate” health literacy, scoring below 40; and 4 (20%) had “marginal” health literacy, with scores ranging from 40 to 70. This suggests that most participants exhibited an adequate level of health literacy, although there was variation in the results.

Digital Competencies and Skills

This study addressed various categories related to participants’ knowledge and skills concerning technology and the use of mobile devices. The results provide a comprehensive overview of the technological competencies of the sample, as well as areas where improvements may be considered.

In category 1, which assessed basic technology knowledge, it was observed that all participants owned mobile devices, which is a positive indicator of the penetration of these technologies. However, 10% (2/20) of the participants reported difficulties in turning on and off these devices, suggesting the need for additional guidance in this aspect. In addition, 20% (4/20) of the participants encountered challenges in installing apps, highlighting an area for potential improvement.

In category 2, which investigated internet navigation skills, it was noted that all participants knew how to open a web browser. However, 30% (6/20) of the participants reported difficulties in typing website addresses or conducting web-based searches. Nevertheless, most participants possessed this fundamental knowledge, which was encouraging.

Category 3 addressed the use of mobile apps for learning. Most participants had already used mobile apps to learn something new, demonstrating an interest in leveraging these educational tools. However, 20% (4/20) of the participants had never used them, suggesting an opportunity to promote the use of these resources. Moreover, 40% (8/20) of the participants reported difficulties in searching for and downloading relevant apps, indicating a need for assistance in this area.

In category 4, which analyzed the ability to conduct web-based research, the vast majority of participants had already conducted web-based research to obtain information. However, 30% (6/20) of the participants faced challenges in selecting appropriate keywords, and some struggled to assess the quality of the information found. This underscores the importance of promoting effective search skills and critical content evaluation.

In category 5, which explored digital communication, most participants used emails and instant messaging apps. However, 30% (6/20) of the participants encountered difficulties with messaging apps, indicating an area where training could be beneficial. In addition, most participants demonstrated an awareness of proper digital communication etiquette, which is essential for effective web-based interaction.

Finally, in category 6, which assessed the use of digital resources, most participants exhibited satisfactory skills in handling cameras and media sharing features. Nevertheless, some researchers have encountered challenges in attaching images, photos, and videos in digital communications.

The results of this study indicate that participants have a basic understanding of technology and a general willingness to adopt mobile devices and apps in their lives. However, areas were identified in which some participants faced challenges and
difficulties. These insights are valuable in guiding the development of digital empowerment programs and ongoing education to enhance technological skills and promote digital inclusion.

**Satisfaction and App Usability**

Table 2 presents the results obtained at 2 points in the focus group, the first interaction with the app (focus group 1) and after adjustments with the users (focus group 3), regarding 2 evaluation metrics: the SUS Validation Index and the SAM Validation Index. These focus groups were conducted to assess user satisfaction and the usability of specific systems or products.

**Table 2.** App satisfaction and usability evaluation at 2 focus group sessions (n=20).a.

<table>
<thead>
<tr>
<th></th>
<th>SAMb score, mean (SD)</th>
<th>CVIb SAM (%)</th>
<th>SUSd score, mean (SD)</th>
<th>CVI SUS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus group 1</td>
<td>34.89 (5.7)</td>
<td>66.11</td>
<td>71.23 (4.7)</td>
<td>70</td>
</tr>
<tr>
<td>Focus group 3</td>
<td>70.65 (13)</td>
<td>89.12</td>
<td>87.14 (5.9)</td>
<td>91.13</td>
</tr>
</tbody>
</table>

aThe values in the table represent the scores of the participants in 2 different focus group sessions, with higher scores indicating higher satisfaction and usability.
bSAM: Suitability Assessment of Materials.
cCVI: Content Validity Index.
dSUS: System Usability Scale.

In focus group 1, participants had an average of 34.89 (SD 5.7) on the SAM questionnaire, indicating moderate satisfaction with the interface of the systems or products under analysis. Regarding the CVI of the SAM questionnaire, the average was 66.11%, suggesting a reasonable assessment of satisfaction. After adjustment, there was a 35.76 increase in the SAM questionnaire results. Therefore, in focus group 3, the results were substantially superior. The participants achieved an average of 70.65 (SD 13) on the SAM questionnaire, reflecting significant satisfaction with the interface. The CVI of the SAM questionnaire, with an average of 89.12%, indicated high satisfaction.

For the SUS questionnaire, participants obtained an average score of 71.23 (SD 4.7), which is a positive evaluation of usability. The CVI of the SUS questionnaire, with an average of 70%, indicated acceptable usability. These results suggest that participants had a more positive user experience in terms of usability than satisfaction with the interface. In the third focus group, there was a 15.91 increase in the SUS questionnaire, with an average score of 87.14 (SD 5.9), indicating excellent usability, whereas the CVI of the SUS questionnaire, with an average of 91.13%, demonstrated that usability was very satisfactory. These results reflect an improvement in the overall user experience, with positive outcomes in both usability and participant satisfaction in focus group 3.

**Results From the Focus Group: Improvements in the Design of the Health App for Older Adults**

During the conduct of the focus groups, with a user-centered approach, older adults were highly engaged in discussing and evaluating the current design of the health app. The aim of these meetings was to gather valuable feedback that allowed us to enhance the user experience and make the app more user-friendly for the older adult population. We present the main findings and adjustment points commented on and documented through audio and video recordings by the focus group participants.

Regarding font size and readability, the following comments were made: “The text is too small. I need my glasses to read it”; “There should be an option to increase the font size”; and “It would be great if the text could be read out loud.”

The need for simplified navigation was addressed with comments such as, “Sometimes, I get lost on the home screen. It should be more straightforward”; “The icons are not self-explanatory. We need hints, maybe without images and just text like WhatsApp”; and “There should be a ‘Back’ button on all screens.”

In terms of visual feedback, the following comments were made: “When we complete a task, it’s difficult to know if it was done correctly; there could be a confirmation message”; “Adding a visual confirmation signal would be helpful”; and “Having a contact with a professional within the app encouraging us to use the system would be a great motivator.”

The personalization of notifications was addressed: “I receive too many notifications. It should be possible to choose which ones to receive” and “It would be nice to set up custom reminders for medications.”

Accessibility was emphasized with comments such as, “Some colors are hard to distinguish for people with weak vision”; “The app should be compatible with screen readers”; and “Bright colors make us more excited to use the app.”

Health measurement integrations touched on points such as “Could it connect automatically to our medical devices?”; “A blood pressure recording feature would be wonderful”; “Is it possible to share health tips on social networks?”; and “Can we present this record to our private doctors?”

Concerning in-app communication support, it was suggested that: “Having a customer support chat for questions would be great”; “Could we share our progress with other users of the same age?”, and “Can we create conversation groups on specific topics, such as cooking and recipes?”
In terms of the simplicity of registration, participants raised the following questions: “The registration process is too long. It should be easier”; “The password is too complicated. It should accept simple passwords”; and “Colors could be more distinct to serve as a guide in the instructions.”

Participants indicated the need for an interactive tutorial with the following remarks: “A step-by-step tutorial would help us navigate better”; “The app should explain its features when we start using it”; and “There could be a video section where I can revisit and see how to use it.” They also discussed points related to social experience, including “Can we share our achievements with our friends?”; “Can we invite our friends to view a map of sports facilities in the city?”; “A rewards system to encourage physical activity would be motivating”; and “Can we create competition groups?”

Figure 5 shows the final design of the screens based on the adjustment points mentioned by the focus groups. These points were crucial in improving the health app and making it more accessible and enjoyable for older adults.

Discussion

Principal Findings

The main findings of this study include the characterization of its participants, revealing similarities in age and cognitive performance between men and women, but significant differences in education, marital status, and occupation. Furthermore, most participants demonstrated an adequate level of health literacy, albeit with some variability in the results. Regarding digital skills, most older individuals showed basic knowledge of technology, internet navigation skills, and mobile app use, but with some difficulties in certain areas. Finally, the satisfaction and usability of the app significantly improved after adjustments, with older adults highlighting areas for improvement such as font size, simplified navigation, and communication support. These results provide valuable insights for future digital empowerment programs and the development of apps tailored to the older adult population.

Participant Characteristics

The discussion of the results obtained in our study concerning literature data is essential for contextualizing our findings and assessing how they align or diverge from existing evidence. Regarding age and scores on the MoCA test, our results indicated that there were no significant differences between men and women. This is in line with several previous studies that did not find consistent differences in cognitive function between the genders [21,22]. However, it is important to note that some studies have reported small differences in certain aspects of cognition, such as verbal memory [23]; however, these discrepancies are not uniform in all studies, as demonstrated by this meta-analysis.

The analysis of education revealed that women in our sample had higher educational levels compared with men, a finding consistent with the literature [24,25]. This may suggest that women in our study had a potential cognitive advantage because of higher education, which should be considered in subsequent analyses.

The discrepancy in marital status is also noteworthy, with only 11% (1/10) married women in our sample, while more than half of the men (6/10, 60%) were married. This may reflect trends broader social contexts, indicating that women are less likely to marry compared to men in some cultures [26,27]. This difference in marital status can influence factors such as social support and support networks, which may have implications for cognitive health.

The occupational differences identified in our sample were also consistent with gender patterns observed in previous studies [28-30]. For example, the exclusivity of certain occupations among women, such as “Homemaker” and some specific professions, reflects gender occupational segregation that persists in many societies. These differences may be related to specific environmental exposures that influence cognitive health.

Regarding the socioeconomic class, the predominance of the upper class among men and the middle class among women is a noteworthy finding that deserves attention. Previous studies have suggested that socioeconomic status can influence cognitive health, with cognitive advantages associated with higher levels of income and education [31]. Therefore, this disparity in socioeconomic class may be an important factor to consider when interpreting our results.

Characterization of Mobile Devices

The analysis of mobile devices used by the participants revealed a significant presence of mobile phone models from Samsung, Motorola, Redmi, and iPhone, reflecting the diversity of choices available in the current mobile device market [31-33]. This finding is consistent with the current technological reality in which consumers have a wide range of device options to meet their individual needs.

The most notable result was that 90% (18/20) of the participants reported that the app worked smoothly on their devices. This
is a positive indicator of the effectiveness of app development for multiple platforms and operating systems, ensuring a satisfactory experience for most users. Such a high level of compatibility is crucial for maximizing the reach and utility of mobile health apps, which often play a significant role in promoting well-being and disease management [34].

The incompatibility detected with Samsung Galaxy A2S models represents a significant issue that deserves attention. Although the overall high rate of app functionality is encouraging, the identification of specific devices that faced problems highlights the complexity of mobile app development. These findings underscore the crucial importance of considering device compatibility during the development process, which is often underestimated but has a substantial impact on user experience [32,33]. Mobile device research and app development have increasingly focused on creating accessible and effective experiences for a wide variety of devices [34]. Variations in hardware, operating systems, and individual settings can create challenges for developers, requiring extensive testing and ongoing optimization to ensure that the app functions reliably in all scenarios. The incompatibility with Samsung Galaxy A2S models highlights the need to consider device and operating system diversity from the outset of the development process [33].

**Health Literacy**

The analysis of the HLS results provided an important insight into health literacy among the older adult population. The scores varied considerably, with most participants demonstrating an adequate level of understanding of health information, as evidenced by the average score of 73.55. Furthermore, approximately 45% (9/20) of the participants achieved scores very close to the maximum score, reaching 98 points. These results are in line with a series of previous studies that have also highlighted the relevance of health literacy in the older adult population [35,36]. For example, a systematic review conducted by Berkman et al [35] consistently highlighted the relationship between low health literacy and adverse health outcomes, including lower treatment adherence, inadequate understanding of medical information, and increased health care costs.

Furthermore, classifying participants into health literacy categories (“adequate,” “inadequate,” and “borderline”) underscores the diversity within the older adult population. This variability in health literacy levels has been consistently observed in previous studies [36,37] and reinforces the need for public health approaches that consider the different needs of specific groups based on their health literacy. However, it is important to emphasize that identifying individuals with “inadequate” and “borderline” health literacy underscores the importance of targeted educational strategies and specific interventions for these groups.

Health literacy promotion programs have been shown to be effective in previous studies [38], highlighting the ability to improve the understanding of health information and, consequently, informed decision-making about health among those with low health literacy.

**Digital Competencies and Skills**

The assessment of digital competencies and skills of the participants in this study revealed a comprehensive picture of how the investigated population deals with digital technology in various categories. In general, most participants demonstrated a reasonable level of digital knowledge and skills, especially when it comes to basic tasks such as turning on and off mobile devices, opening internet browsers, and sending emails. These findings are in line with the increasing penetration of digital technology into contemporary society [39].

However, areas where some participants faced challenges were also identified, such as selecting appropriate keywords for web-based searches, evaluating the quality of information found on the internet, and the effective use of instant messaging apps. These findings highlight the importance of personalized educational approaches to improve digital skills for specific groups, focusing on the areas where they face difficulties [40].

The results of this study will have significant implications for the development of digital empowerment programs and lifelong education. The ability to use digital technology is essential in an increasingly digitized world, especially for accessing health information, education, and communication. Therefore, investing in initiatives that promote digital inclusion and improve digital skills is crucial to ensure that all individuals, regardless of their age or background, can effectively harness the benefits of digital technology [41].

In summary, this study provides a comprehensive insight into the digital competencies and skills of a group of participants and emphasizes the importance of targeted educational strategies to improve these skills, thus promoting digital inclusion and full participation in the digital society.

**Satisfaction and Usability Assessment**

The evaluation of app satisfaction and usability through focus groups provided significant insights into the user experience, and comparing these results with previous studies in the literature enriches our understanding.

Initially, in focus group 1, participants showed moderate satisfaction with the app interface, reflected in an average SAM questionnaire score of 34.89. This initial result is similar to those of studies that have highlighted that satisfaction with app interfaces can vary significantly depending on usability and alignment with user preferences [42].

However, the process of adjusting and refining the app, reflected in focus group 3, had a substantial positive impact on user experience. The participants reported considerable satisfaction, with an average SAM questionnaire score of 70.65, indicating that the implemented improvements aligned the interface with user expectations. This improvement in satisfaction is consistent with the findings of studies that emphasized the importance of adaptability and user-centered design in promoting user satisfaction [43].

Regarding usability, the initial results in focus group 1 already indicated a positive evaluation, with an average SUS questionnaire score of 71.23. This initial result suggests that the usability of the app was in line with previous studies that
emphasized the importance of usability in effectiveness and user satisfaction [44]. However, after the app’s improvements and the subsequent assessment in focus group 3, there was a significant improvement in usability. The average SUS questionnaire score increased to 87.14, reflecting excellent usability. This increase in usability is supported by research emphasizing that improvements in the interface, task simplification, and problem-solving can lead to enhanced usability [45].

In summary, the results of this study and comparisons with the literature emphasize the importance of involving users in the development and refinement of digital systems. This not only improves user satisfaction but also enhances usability, promoting a more positive and effective user experience.

Results From the Focus Groups: Improvements in the Design of the Older Adults Health App

Collecting feedback from the older adults through focus groups played an essential role in identifying areas for improvement in the design of the health app, thereby demonstrating the value of a user-centered approach. The participants highlighted several critical issues that directly affect usability and the user experience.

Regarding readability, concerns about font size and the need for an option to increase it reflect the importance of accessibility for older adults, especially those with vision problems. This is in line with studies that emphasized the need for designs adapted to the specific needs of older adults and bodily changes that occur with aging, such as visual decline [46].

Simplifying navigation was another significant concern, with participants emphasizing the need for a more intuitive interface and self-explanatory icons. This issue is consistent with the literature, which emphasizes the importance of clarity and simplicity in designing apps for older adults—a more straightforward design had better acceptance and adherence in this population [47].

The need for visual feedback to confirm task completion suggests a concern with the effectiveness of actions taken within the app, which can boost user confidence. This aspect is supported by studies highlighting the importance of feedback in user motivation and engagement, showing that it enhances the user’s bond and credibility with the design’s goal [48]. Customization of notifications and integration with medical devices indicate the desire of older adults for a more adaptable and relevant experience. This customization is an important element for improving usability and app acceptance [49].

Accessibility, including compatibility with screen readers and the use of vivid colors, is crucial to meet the needs of older adults with varying levels of visual abilities. This aligns with inclusive design guidelines [50]. The suggestion of communication support and social features reflects the desire of older adults to connect and share experiences with their peers. This social dimension can significantly enhance user engagement and satisfaction [51]. Simplicity in the registration process and the availability of an interactive tutorial are also essential elements to ensure that older adults can make the most of the app. Simplifying the registration process is in line with the literature that emphasizes the importance of minimizing entry barriers and developing fewer back-and-forth paths within screens [52]. Finally, the emphasis on rewards and competitions suggests that gamification elements can be motivating for older adults. This aligns with research demonstrating the benefits of gamification in user motivation and engagement [53].

In summary, the integration of feedback from older adults into the health app design process resulted in a series of improvements that met the needs and preferences of this population. These findings underscore the importance of involving end users, especially the older adults, in the creation of user-centered apps, ensuring a more accessible, user-friendly, and satisfying experience.

Implications for Public Health

The results of this study have significant implications for public health, especially in the context of an aging population and the growing importance of digital technology in delivering health information and services. First, the analysis of participant characteristics highlighted the importance of considering factors such as gender, education, marital status, occupation, and socioeconomic status when developing health promotion strategies. These variables directly influence cognitive health and the ability to use digital technology.

The assessment of digital skills revealed areas where participants can benefit from additional training and empowerment, especially in more advanced tasks such as evaluating web-based information. Therefore, digital literacy programs targeted at older adults can play a crucial role in empowering these individuals to make the most of digital technologies in health and communication.

Furthermore, the assessment of satisfaction and usability emphasized the importance of designing health apps in a user-centered manner based on user feedback. Continuous improvement of the user interface and usability is essential to ensure that apps effectively meet the needs of older adults.

Finally, the broader implications of this study extend to digital health promotion and the development of digital communication strategies targeting older adults. The increasing digitalization of health information and services requires approaches that ensure equitable access to and proper understanding of health information by older adults, considering their diversity in terms of digital skills and health literacy.

Limitations and Suggestions for Future Research

Despite providing valuable insights into the interaction of older adults with health apps, this study has limitations that require consideration. First, the sample used in this study was relatively small, which may compromise the generalizability of the results. Efforts to include a more substantial number of participants could improve the representativeness of this research. Another limitation of this study is related to the inherent selection bias in the sample, which was composed of participants already predisposed to using health apps. This may limit the applicability of the conclusions to a broader population of older individuals who may not have the same degree of familiarity or willingness to adopt such technologies.
In addition, it is worth noting that much of the data collected relied on self-reports from participants, which may have introduced memory bias and inaccuracies in responses. Obtaining objective data, supplemented by external assessments, could strengthen the validity of the findings.

In terms of suggestions for future research, we consider some promising directions. First, the inclusion of more diverse samples, covering different ethnic groups, geographic locations, and socioeconomic characteristics, would allow for a more comprehensive understanding of the interaction dynamics between older adults and health apps.

Comparatively, another promising international research suggestion would be the evaluation of the health literacy questionnaire and digital skills of older adults in different countries and cultural contexts, thus providing insights into the contextual determinants of these variables. Furthermore, future studies can explore the health impacts associated with the use of health apps by older adults, including improvements in the self-management of chronic conditions and quality of life. Specifically, the analysis of specific health apps and their features can reveal detailed information about the design and functionality factors that best meet the needs and preferences of older adults.

Conclusions
In this comprehensive study, we thoroughly examined participant characteristics, mobile device compatibility, health literacy, digital skills, and user satisfaction with a health app designed for older adults. The findings highlight the complexity of the interactions among these factors and underscore the importance of a user-centered approach in the development of digital health apps.

The results suggest that considering the individual differences among older adults, such as gender, education, marital status, and socioeconomic class, is essential to meet their health care needs and ensure an effective digital experience. Furthermore, investing in digital empowerment programs and promoting health literacy can enhance the ability of older adults to use digital technology effectively for accessing health information and services.

The evaluation of user satisfaction and usability of the app emphasized the importance of listening to users and implementing continuous improvements in the user interface. This is crucial to ensure that digital health apps are well received and effective in promoting health and supporting older adults in their health care needs.

Ultimately, this study contributes to the understanding of the intersection between health, technology, and aging, providing valuable insights to inform digital health promotion strategies targeted at this increasingly important population.

Acknowledgments
The authors would like to thank the study participants.

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (2017/21361-2), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (485045/2015-3 and 141720/2017-4), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (finance code 001).

Authors’ Contributions
ACSS conducted extensive literature research and compiled a comprehensive background section for the research paper and designed and executed experiments, collected data, and performed statistical analysis for the study. GAOG contributed to the theoretical framework of the research, providing critical insights into the conceptual foundation of the study. He played a crucial role in drafting and revising the manuscript, ensuring clarity and coherence in the written content. CRBJ secured the necessary research funding and managed the financial aspects of the project, ensuring the smooth execution of the study. He provided guidance and oversight throughout the research project, contributing to the overall project management and research direction.

Conflicts of Interest
None declared.

References


42. Costa KN, dos Santos FA. Usabilidade aplicada aos sistemas m-learning: análise de dispositivos móveis como ferramenta de ensino. Conex Ciênc Tecnol 2022 Nov 08;16:e022027 [FREE Full text] [doi: 10.21439/conexoes.v16i0.2317]


Abbreviations

- **CSEB**: Brazilian Socioeconomic Classification
- **EEFERP**: Escola de Educação Física e Esporte de Ribeirão Preto
- **MoCA**: Montreal Cognitive Assessment
- **PEFI**: Physical Education Program for the Elderly
- **SAM**: Suitability Assessment of Materials
- **SUS**: System Usability Scale
- **TLS**: Health Literacy Test
- **USP**: Universidade de São Paulo

Edited by A Mavragani; submitted 01.11.23; peer-reviewed by KP Rodrigues; comments to author 25.01.24; revised version received 14.02.24; accepted 19.02.24; published 15.04.24.

Please cite as:

Sobrinho ACDS, Gomes GADO, Bueno Júnior CR
Developing a Multiprofessional Mobile App to Enhance Health Habits in Older Adults: User-Centered Approach
JMIR Form Res 2024;8:e54214
URL: https://formative.jmir.org/2024/1/e54214
doi: 10.2196/54214
PMID: 38619865

©Andressa Crystine da Silva Sobrinho, Grace Angelica de Oliveira Gomes, Carlos Roberto Bueno Júnior. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Clinical Decision Support System for Guidelines-Based Treatment of Gonococcal Infections, Screening for HIV, and Prescription of Pre-Exposure Prophylaxis: Design and Implementation Study

Saugat Karki, MD, MS; Sarah Shaw, MPH; Michael Lieberman, MD, MS; Alejandro Pérez, MPH; Jonathan Pincus, MD; Priya Jakhmola, MBA, MS; Amrita Tailor, MPH, PhD; Oyinkansola Bukky Ogunrinde, MHSA; Danielle Sill, MSPH; Shane Morgan, MS; Miguel Alvarez, MSEE; Jonathan Todd, PhD; Dawn Smith, MD, MPH; Ninad Mishra, MD, MSHI

Division of STD Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States
Public Health Informatics Institute, Decatur, GA, United States
OCHIN, Portland, OR, United States
Oregon Health & Sciences University, Portland, OR, United States
Codman Square Health Center, Boston, MA, United States
Division of HIV Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States
†deceased

Corresponding Author:
Saugat Karki, MD, MS
Division of STD Prevention
Centers for Disease Control and Prevention
1600 Clifton Rd NE
Atlanta, GA, 30333
United States
Phone: 1 4047187483
Email: skarki@cdc.gov

Abstract

Background: The syndemic nature of gonococcal infections and HIV provides an opportunity to develop a synergistic intervention tool that could address the need for adequate treatment for gonorrhea, screen for HIV infections, and offer pre-exposure prophylaxis (PrEP) for persons who meet the criteria. By leveraging information available on electronic health records, a clinical decision support (CDS) system tool could fulfill this need and improve adherence to Centers for Disease Control and Prevention (CDC) treatment and screening guidelines for gonorrhea, HIV, and PrEP.

Objective: The goal of this study was to translate portions of CDC treatment guidelines for gonorrhea and relevant portions of HIV screening and prescribing PrEP that stem from a diagnosis of gonorrhea as an electronic health record–based CDS intervention. We also assessed whether this CDS solution worked in real-world clinic.

Methods: We developed 4 tools for this CDS intervention: a form for capturing sexual history information (SmartForm), rule-based alerts (best practice advisory), an enhanced sexually transmitted infection (STI) order set (SmartSet), and a documentation template (SmartText). A mixed methods pre-post design was used to measure the feasibility, use, and usability of the CDS solution. The study period was 12 weeks with a baseline patient sample of 12 weeks immediately prior to the intervention period for comparison. While the entire clinic had access to the CDS solution, we focused on a subset of clinicians who frequently engage in the screening and treatment of STIs within the clinical site under the name “X-Clinic.” We measured the use of the CDS solution within the population of patients who had either a confirmed gonococcal infection or an STI-related chief complaint. We conducted 4 midpoint surveys and 3 key informant interviews to quantify perception and impact of the CDS solution and solicit suggestions for potential future enhancements. The findings from qualitative data were determined using a combination of explorative and comparative analysis. Statistical analysis was conducted to compare the differences between patient populations in the baseline and intervention periods.

Results: Within the X-Clinic, the CDS alerted clinicians (as a best practice advisory) in one-tenth (348/3451, 10.08%) of clinical encounters. These 348 encounters represented 300 patients; SmartForms were opened for half of these patients (157/300, 52.33%)
and was completed for most for them (147/300, 89.81%). STI test orders (SmartSet) were initiated by clinical providers in half of those patients (162/300, 54%). HIV screening was performed during about half of those patient encounters (191/348, 54.89%).

**Conclusions:** We successfully built and implemented multiple CDC treatment and screening guidelines into a single cohesive CDS solution. The CDS solution was integrated into the clinical workflow and had a high rate of use.

**(JMIR Form Res 2024;8:e53000)** doi:10.2196/53000

**KEYWORDS**
clinical decision support systems; CDS; gonorrhea; pre-exposure prophylaxis; PrEP; HIV; sexually transmitted infections; electronic health records; guideline adherence

**Introduction**

**Background**

**Overview**

Research has shown that a simultaneous, integrated approach to testing and other services targeting multiple diseases on the same population can be feasible and effective in improving accessibility and health outcomes among patients [1]. This integrated approach can be supported by point-of-care resources such as clinical decision support (CDS) system tools that can combine electronic patient health data and up-to-date guidelines and clinical protocols. This paper describes the usability phase of a study on the integration of a CDS tool for managing gonorrhea, screening for HIV, and identifying patients for HIV pre-exposure prophylaxis (PrEP) at Codman Square Clinic.

**Gonorrhea, HIV, and PrEP**

Gonorrhea was the second most commonly reported sexually transmitted infection (STI) in the United States in 2020, with a total of 677,769 cases reported to the Centers for Disease Control and Prevention (CDC) and has shown a 71.49% (395,216/677,769) increase since 2015 [2]. Among both men and women, untreated gonorrhea can cause serious and painful health problems, infertility, and in rare cases, even life-threatening conditions [3-7]. To mitigate these risks, annual screening is recommended for all sexually active women younger than 25 years and those at increased risk for infection. Timely diagnosis through routine screening and prompt, effective treatment—adhering to up-to-date CDC treatment guidelines—is paramount to both reducing gonorrhea and slowing down the threat of antimicrobial resistance [3,8-10].

HIV causes an infection that attacks the body’s immune system, specifically the cluster of differentiation antigen 4 T lymphocytes [11,12]. If left untreated, it can lead to AIDS, opportunistic infections, malignancies, and death [11]. HIV PrEP has been shown to reduce the risk of acquiring HIV infection [13,14]. As a result, the CDC has developed guidance for prescribing PrEP to individuals considered at high risk, and gonorrhea infection is identified as one of those risk factors [15,16].

**Syndemic of STIs: Gonorrhea and HIV Infection**

Patients with STIs have a 2- to 3-fold increased risk of having concomitant HIV infection compared to those without any STI [17,18]. Gonococcal infections, specifically, have also been associated with increased risk of HIV [18-23]. Gonococcal infections cause immune reactions that, among others, recruit cluster of differentiation antigen 4 cells, which can potentially enhance HIV in vivo replication and facilitate acquisition and transmission of HIV infection [20,24-26]. Previous studies have further advanced our understanding of this syndemic and describe the biological, behavioral, social, and structural determinants and their synergisms [20,22,25]. As a result, it has been a long-standing recommendation to screen patients diagnosed with STIs, including gonorrhea, for HIV [8,27].

This synergistic and potentially concomitant nature of gonorrhea and HIV and the overlap between gonorrhea treatment, HIV screening, and PrEP prescription present us with the opportunity to develop a CDC guidelines-based synergistic intervention tool that could target adequate treatment for gonococcal infections, screening for HIV based on risk factors associated with the diagnosis of gonorrhea, and PrEP prescription to relevant populations.

**CDS Systems**

CDS systems are computational tools that leverage information available in electronic health records (EHRs) to provide person-specific evidence-based information, intelligently filtered, and presented at appropriate times to help inform decisions regarding a patient’s care to improve patient outcomes and lead to higher quality care [28-31]. A systematic review of 160 articles representing 148 unique studies describing CDS implemented across diverse settings found that CDS interventions were efficacious on health care process outcomes, but data on clinical and economic outcomes were sparse [32]. The Community Preventive Services Task Force conducted a systematic review of 23 studies and reported that CDS increased HIV screening for both the general population and people at higher risk for HIV infections [33]. Based on the strong evidence of effectiveness, the Community Preventive Services Task Force has now recommended CDS to increase HIV screening [33]. The Sexually Transmitted Infections National Strategic Plan has also recommended increasing the implementation of CDS to support high-quality sexual health assessments, screen for STIs, integrated care models, and reduce adverse outcomes [34]. Gonorrhea treatment recommendations are updated regularly to represent the latest evidence to provide adequate treatment and curb the rise of antimicrobial resistance [3,8]. Keeping up to date with the latest guidelines could potentially pose a challenge to clinical providers, as evidenced with varying degrees of adherence to gonorrhea treatment guidelines [9,10]. Coupled with the synergistic nature of the infection with HIV [18,20-26], this presents an opportunity to develop a
multifaceted CDS intervention to address the needs of the same patient population and improve clinical provider adherence to CDC guidelines on gonorrhea, HIV screening, and providing PrEP for HIV infection, when indicated.

**Objectives**

The primary goal of this study was to translate portions of CDC treatment guidelines for gonorrhea and relevant portions of guidelines for HIV screening and PrEP prescribing, into EHR-based CDS interventions to aid clinicians’ adherence to these guidelines and improve respective patients’ health outcomes. CDC guidelines–based CDS tools for STIs have not been reported in published literature; similarly, 2 or more guidelines have not previously been cohesively built into a single CDS tool. A secondary goal was to assess whether the translated CDS solution worked in a real-world clinical workflow initiated by patients with a gonorrhea diagnosis.

**Methods**

**Ethical Considerations**

This project was designed and executed as a quality improvement project at OCHIN, the implementing partner, and therefore was not considered human subjects research, and institutional review board input was not obtained. The impact of the project was assessed using only aggregated deidentified data. All OCHIN members include language in their patient privacy notices stating that deidentified data may be used for research purposes. There was no compensation for human subjects’ research.

**EHR Platform and Clinical Partner**

**Overview**

We partnered with OCHIN, Inc, a nonprofit community–based health center–controlled network as the EHR provider [35,36]. We selected OCHIN’s Epic EHR [37] as it serves as the primary EHR to almost 1000 health care sites with 21,000 clinicians in 45 states who serve over 6 million patients [38]. OCHIN provides 1 instance of Epic consisting of enterprise-wide master patient index; patients have a single medical record across all clinics in the network, and all data are managed centrally.

Within the OCHIN network, we partnered with Codman Square Health Center, a Federally Qualified Health Center in Dorchester, Massachusetts, to pilot the CDS intervention [39]. We selected Codman Square Health Center as both patients with gonorrhea and HIV are regularly managed in their clinics and they have a well-established STI-reporting practice in place. Typically, they have over 115,000 client contacts per year.

**Intervention Clinic**

While all of Codman Square Health Center had access to the CDS solution, a subset of clinicians within the Internal Medicine clinic, specifically, were selected as the focus of the intervention. This subset of clinicians frequently engaged in the screening and treatment of STIs and provide care under the name “X-Clinic” to allow a level of discretion to the patients they serve. The X-Clinic clinicians were selected for the focus of the intervention due to the high rates of gonorrhea infection and the potential for use of the CDS solution during the pilot period. A clinical champion was identified to provide insight and encourage participation among staff along with 2 project leads, 1 representing the providers and 1 for support staff. While the CDS tool was available to the broader Codman Square Health Center, the X-clinic received targeted training and support to use the tool. This training and support were not extended to the rest of Codman Square Health Center.

**CDS Solution Design**

**Study Design**

An internal review by OCHIN, the implementing partner, determined this work as quality improvement and deemed that an institutional review board review was not necessary. A 12-week study period was planned, beginning August 31, 2021, and ending on November 23, 2021. A mixed methods pre-post design [40] was used to measure the feasibility, use, and usability of the CDS solution. During the time of the planned intervention, COVID-19 altered the typical pattern of patients seeking STI-related care [41,42]. As a result, we determined that the 12-week period immediately prior to the intervention period would be the most representative sample of baseline patient data.

**Practice Coach Engagement**

In addition to the direct training provided, the X-Clinic clinicians received targeted and continual support on the use of the CDS solution using the practice coaching methodology. This included supporting adaptive skills and learning to build capacity and capability for the care team to effectively use the CDS solution, explore change ideas, provide feedback to support the team’s progress as the team becomes more self-sufficient, and engage with iterative learning. Practice coaching (also known as practice facilitation) has been demonstrated to be effective in successfully implementing tools and new evidence into clinical practice [43-46]. A key part of the intervention design included a dedicated practice coach who regularly engaged with the clinical champion and project leads at the X-Clinic. The coach was part of the project team’s kick-off meeting, which included a warm welcome from the CDS solution developer, project orientation from the project manager, training by the trainer to demonstrate the solution, and an overview of the coaching engagement from the coach. The facilitation meetings were held biweekly for the duration of the project, and they typically included discussions about general assessment and feedback, use, perceived effectiveness, and clinical integration of the CDS solution. When available, CDS solution–related data were reviewed for discussion of key insights, opportunities for improvement, and when a need was identified, the practice coach engaged subject matter experts to provide added support and facilitated referrals for technical and application support.

**CDS Solution Development**

OCHIN Epic EHR [37] has native tools to develop CDS solutions in a variety of formats, each performing specific functions. For this CDS solution, we developed four tools: 1) a sexual history data capture form using “SmartForm” (Epic’s name for a method for capturing responses to questions in a structured format) and a patient questionnaire with the same
questions that can be made available to the patient prior to the visit, (2) a “best practice advisory” (BPA; Epic’s name for rule-based alerts), (3) an enhanced STI order set—that included diagnoses as well as links to documentation templates—using “SmartSet” (Epic’s name for order sets), and (4) a documentation template using “SmartText” (Epic’s name for a method of notetaking that prompts clinician with standard information collected and presented during a visit).

We developed a CDS solution by integrating translated information from 3 separate CDC guidelines, specifically to guide clinicians to choose appropriate therapy for gonorrhea [3,8] and—as per the CDC’s guidelines to prevent or diagnose new HIV infections [47-49] and link those individuals considered at risk to relevant prevention and medical services [15,16]—prompt them to screen for HIV and consider PrEP, if patients met those criteria (Figure 1). Health data standards were used wherever possible; ICD-10-CM (International Classification of Diseases, Tenth Revision, Clinical Modification) was used to evaluate the diagnosis of gonococcal infections.

The Sexual History SmartForm and corresponding patient questionnaire were developed to collect sexual history and STI-related symptoms from patients. The form and questionnaire were structured as point and click questions and answers for ease of use; patients had access to the questionnaire via MyChart [50] (Epic’s patient portal) prior to the visit (Figure 1, step 1; Table 1). At the beginning of the clinic visit, nurse navigators reviewed the information collected using the SmartForm and completed the questionnaire, if incomplete (Figure 1, step 2). Based on CDC’s guidelines [15,16,48], patients were considered to be at risk for HIV infection for the following criteria: (1) positive result or diagnosis for gonococcal infection in the last 6 months; (2) reason for visiting the clinic was related to STI; and (3) sexual risk behaviors, known partner with any STI, and if patient or partner using or sharing needles. If the patient’s response suggested a high risk for HIV infection, the SmartForm algorithm presented further questions to gauge awareness and interest for PrEP (Table 1).

Figure 1. Workflow and details of the clinical decision support solution. CDS solution workflow; represented in numerical order in the diagram: 1. "SmartForm" (a native Epic tool for capturing responses to questions) captured detailed sexual history from the patient. It was made available to patients in advance via MyChart (Epic’s patient portal) on their phone app or web browser. 2. At the beginning of the visit, the nurse navigator entered a reason for the visit and chief complaint, reviewed information entered via SmartForm, and completed the SmartForm, if incomplete. 3. When the clinical provider accessed the patient’s chart, a "best practice advisory" (BPA; Epic’s name for rule-based alerts) directed the clinical provider to a "SmartSet" (Epic’s name for order sets) that represented the Centers for Disease Control and Prevention treatment and prevention guidelines for gonococcal and HIV infections. 4. If the clinician opted to open the SmartSet, a detailed list of orders, treatment, and diagnosis options was provided. 5. The clinician documented the visit using a "SmartText" (Epic’s name for a method of taking clinical notes) documentation template specifically designed for sexually transmitted infections (STIs) to standardize the information collected and presented during the visit. Gonorrhea treatment, HIV screening, and PrEP (pre-exposure prophylaxis) prescription guidelines, represented in alphabetical order in the diagram: a. If the patient was considered at risk for HIV infection, additional questions related to PrEP prescriptions appeared in the SmartForm. b. The BPA was displayed to the clinician if the patient met the criteria to treat for gonococcal infection. c. The clinical provider was offered to use the SmartSet which included guidance for screening and treating STIs, and screening and monitoring for PrEP. CDS: clinical decision support.
Table 1. Information collected, criteria, and type of decision support provided by each clinical decision support tool.

<table>
<thead>
<tr>
<th>Clinical decision support tool</th>
<th>Type of information collected</th>
<th>Criteria for providing clinical decision support</th>
<th>Clinical decision support provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>SmartText</td>
<td>Prefilled template to document clinical encounter for STI-related visit</td>
<td>SmartSet recommendation to use SmartText</td>
<td>SmartForm was embedded in SmartText</td>
</tr>
</tbody>
</table>
| SmartForm                    | • Sexual history  
• STI symptoms | • Positive laboratory result for gonococcal infection in the last 6 months  
• Diagnosed with gonococcal infection in the last 6 months  
• Reason for visiting the clinic was related to STIs  
• Current partner has HIV or AIDS  
• Patient or partner using or sharing needles in the last 6 months  
• Syphilis or gonorrhea in the last 6 months  
• Infrequent use of condoms with partners at risk of HIV (persons who inject drugs or men who have sex with men),  
• Male who has sex with males who has had anal sex without condoms in the last 6 months (outside of a monogamous relationship with an HIV-negative partner),  
• Male who has sex with males who has had any STI (including chlamydia) | Additional PrEP-related questions prompted:  
• “Have you heard of HIV PrEP?”  
• “Have you ever taken HIV PrEP?”  
• “Are you interested in HIV PrEP?” |
| Best practice advisory       | • SmartForm:  
• Sexual history  
• STI symptoms | • Confirmed positive result for gonococcal infection in the last 14 days  
• Chief complaint related to gonorrhea  
• STI symptoms or concern  
• Penile, urethral, rectal, or vaginal discharge  
• STI and HIV-related screening, testing, follow-up, or visit | Alerted clinical provider to use SmartSet to aid clinical decisions |
| SmartSet                     | • SmartForm:  
• Sexual history  
• STI symptoms | • Best practice advisory recommendation to use SmartSet | Recommended laboratory tests to screen for STI and HIV  
• Recommend PrEP to at-risk patients  
• Provide recommendations for medication regimens to treat gonococcal infections  
• Provide a link to SmartText |
|                             | • EHR:  
• Prior testing history  
• Chief complaint  
• Treatment information | | |

aSTI: sexually transmitted infection.  
cEHR: electronic health record.

For patients who (1) had a confirmed positive result for gonococcal infection in the last 14 days, (2) had a chief complaint related to gonorrhea (STI symptoms or concern; penile, urethral, rectal, or vaginal discharge and STI-related screening, testing, follow-up, or visit), or (3) were treated presumptively due to a partner with known gonococcal infection, a BPA would alert the clinical provider (Figure 1, step 3) to use the SmartSet to assist in providing standardized clinical care (Figure 1, step 4; Table 1). The SmartSet was designed to (1) recommend laboratory tests to screen for STI, including HIV, (2) provide PrEP prescription options, (3) present CDC guidelines–based recommendations for medication regimens as preferred options to treat gonococcal infections, and (4) provide a link to gonorrhea treatment guidelines for anything beyond uncomplicated gonorrhea and a link to the SmartText (Figure 1, step 5; Table 1). Opening the BPA did not necessarily mean that clinical providers selected diagnoses or ordered laboratory tests or medication directly from the SmartSet but merely that those options were made available for selection and action. Diagnoses, laboratory tests, and medication orders could also be entered for the patient outside of the SmartSet. The SmartText template was developed to prompt the clinical provider to document in a standardized manner (Table 1). The SmartText included common heading for note taking, prepopulated STI-related laboratory results from the patient’s chart, and risk reduction suggestions.

Use of CDS Solution and Outcomes of Intervention

We evaluated CDS solution use and outcome metrics specifically for any person with confirmed case of gonorrhea or with an STI-related chief complaint. Data were collected on the demographics of the patient population to ensure appropriate
comparison between the baseline and intervention populations, \textit{use} metrics of CDS solutions by X-Clinic, and outcomes following the implementation of the CDS solution. To measure patient and disease characteristics for the baseline period (immediate 12-week period prior to the intervention period), a 1-time data extraction was performed (Table 2). During the intervention period, quantitative data were extracted on a nightly basis.

Table 2. Goal of assessing the clinical decision support solution and their respective data sources.

<table>
<thead>
<tr>
<th>Assessment area and data source</th>
<th>Assessment question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
</tr>
<tr>
<td>Baseline period: 1-time data extraction</td>
<td>What are demographic characteristics of the patient population in the baseline period?</td>
</tr>
<tr>
<td>Intervention period: nightly data extraction</td>
<td>What are demographic characteristics of the patient population in the intervention period and how do they compare with the patient population in the baseline period?</td>
</tr>
<tr>
<td><strong>Use</strong></td>
<td></td>
</tr>
<tr>
<td>Intervention period: nightly data extraction</td>
<td>Are clinical providers using the CDS\textsuperscript{a} solution and how are they using it?</td>
</tr>
<tr>
<td>Midpoint survey: questionnaire for X-Clinic clinical providers</td>
<td>Identify challenges for course correction, if necessary</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Intervention period: nightly data extraction</td>
<td>Are there any changes in outcomes of gonorrhea treatment, screening for HIV and prescription of HIV PrEP\textsuperscript{b}?</td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td></td>
</tr>
<tr>
<td>Midpoint survey: questionnaire for X-Clinic clinical providers</td>
<td>What are the provider perspectives on the characteristics of the CDS?</td>
</tr>
<tr>
<td>End of intervention period: key informant interviews</td>
<td>Which features do the clinicians like or do not like? How did these factors influence use?</td>
</tr>
</tbody>
</table>

\textsuperscript{a}CDS: clinical decision support.  
\textsuperscript{b}PrEP: pre-exposure prophylaxis.

**Clinical Provider Perspectives and Usability of the CDS Solution**

We conducted 1 midpoint survey to confirm the \textit{use} of CDS solution and identify any challenges faced by the clinical providers using the solution (Table 2; Multimedia Appendix 1). Four end users—or clinical providers who used the CDS solution—completed the survey. We also conducted key informant interviews with 3 end users at the end of the pilot period to assess the \textit{usability} of the CDS solution (Multimedia Appendix 2). The CDS solution design encompassed multiple end user roles such as the SmartForm, which was anticipated to be completed by nurse navigators, and SmartSet, which was anticipated to be completed by clinical providers. We interviewed individuals who represented the various roles such as leadership, clinical provider, and nurse navigator perspectives. The purpose of the key informant interviews was to understand treatment and screening workflows prior to CDS solution implantation, perception, and impact of the CDS solution and solicit suggestions for potential future enhancements (Table 2). We measured the \textit{use} and \textit{usability} of the CDS solution based on the clinical provider role to reveal how perception and familiarity with the CDS solution may have impacted use.

**Data Analysis**

We used quantitative and qualitative means to examine the \textit{feasibility} of this guidelines-based CDS solution in a real-world clinical workflow, where 2 separate guidelines have been cohesively built into a single CDS tool solution, initiated by patients with a gonorrhea diagnosis. Quantitative data were used to describe the \textit{use} metrics and outcomes post implementation of CDS solution and were summarized by descriptive statistics using counts and percentages. To assess \textit{usability} of the CDS tool, qualitative data were systematically coded from interview transcripts, and findings were determined using a combination of explorative and comparative analysis to examine end user (clinical provider) perspectives.

Statistical analysis was conducted to compare the differences between patient populations in the baseline and intervention periods. For continuous variables, to test for a statistically significant difference among populations, a 2-sample \textit{t} test was used. To test for a statistically significant association among outcomes, all categorical variables were tested using Pearson chi-square test, and if an expected count of 5 was not observed for a cell, Fisher exact test was used. If a statistically significant association was observed, odds ratios were calculated. All statistical analyses were conducted using SAS (version 9.4; SAS Institute).

**Results**

**CDS Solution Design**

**Description of Patient Populations: Pre- and Post-CDS Intervention**

During the intervention period, 12,048 patients were provided with clinical care in Codman Square Health Center and 37 were diagnosed with gonorrhea; similarly, 40 patients out of 11,269...
were diagnosed with gonorrhea in the baseline period. The mean age of patients seen in the baseline and intervention periods was 37 (SD 21.46) years and 36 (SD 21.98) years, respectively ($P<.001$). A 2-sample $t$ test analysis of the age groups revealed that the intervention period included younger individuals compared to the baseline period and cannot be considered similar; however, the largest age group of 25-44 years old remained the same with 30.78% (3469/11,269) in the baseline period and 29.58% (3564/12,048) in the intervention period. Using Fisher exact test, it was determined that the patients in the baseline and intervention periods were comparable regarding gender ($P=.25$). While ethnicity distribution was found to be similar ($P=.56$) using Pearson chi-square test, baseline and intervention groups differed significantly by race ($\chi^2=4.36; P=.04$). Nevertheless, the largest patient population served by Codman Square Health Center was Black or African American in both baseline (9040/11,269, 80.22%) and intervention (9565/12,048, 79.39%) periods.

**Use of CDS Solution and Outcomes of Intervention**

**Use**

Throughout Codman Square Health Center, the BPA was triggered 950 (4.07%) times. Within the X-Clinic specifically, the BPA was triggered in one-tenth of all the clinical encounters (348/3451, 10.08%); these 348 encounters represented 300 patients (Figure 2). The SmartForm was opened for about half of the patients (157/300, 52.33%), and in instances when the SmartForm was opened, it was completed 89.81% (141/157) of the time (Figure 2). Of note, while the BPA did not prompt the user to open the SmartForm, the patient cohort where the BPA presented was used for analysis as this represents the target population for this study. For those patients whose responses determined them as high-risk for HIV, further PrEP-related questions were asked (Table 1). Some information about PrEP was known to 63.31% (88/139) of the patients who were asked that question; 6.62% (9/136) of patients reported having taken PrEP and 16.06% (22/137) patients showed interest in PrEP (Figure 2).

Clinical providers could open SmartSet from the BPA (Figure 1, step 4), and this action was taken for about half of the patients (162/300, 54%) for whom BPA was presented. In the target population, the most common diagnosis was “Vulvovaginitis” (ICD-10-CM: N76.0); it was assigned to 34 patients, 21 were assigned to patients for whom BPA was presented, and 13 were assigned where no BPA was presented (Table 3). Diagnoses were assigned, and laboratories were ordered both from BPA-prompted SmartSet and outside of BPA without using the CDS solution (Table 3). No patients were diagnosed with cervicitis but were diagnosed with urethritis, pharyngitis, and proctitis. The most ordered laboratory test was for HIV. More than half (191/348, 54.89%) of the encounters where a BPA was presented received these orders for HIV testing.

In the X-Clinic, SmartText note template was used in most of the clinical encounters (313/348, 89.9%), where the BPA was presented. Clinical providers used SmartText extensively (2343/3103, 75.5%) even during encounters where the BPA did not trigger. The SmartText was accessible either through the SmartSet or a standard “quick button” in the documentation area. The clinical site continued to use the CDS tool beyond the study period.

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Description of patient interaction with SmartForm, percent completed, and answers selected. BPA: best practice advisory; PrEP: pre-exposure prophylaxis.
Table 3. Use of best practice advisory by diagnosis in the X-Clinic.

<table>
<thead>
<tr>
<th>Diagnoses selected and tests ordered from SmartSet</th>
<th>Best practice advisory presented (n=302), n (%)</th>
<th>Best practice advisory not presented (n=3103), n (%)</th>
<th>Total (n=3451), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Urethritis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urethritis, unspecified [N34.2]</td>
<td>5 (62.5)</td>
<td>3 (37.5)</td>
<td>8 (0.23)</td>
</tr>
<tr>
<td>Gonococcal urethritis [A54.01]</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Chlamydial urethritis [A56.01]</td>
<td>0 (0)</td>
<td>1 (100)</td>
<td>1 (0.03)</td>
</tr>
<tr>
<td><strong>Cervicitis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervicitis [N72]</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gonococcal cervicitis [A54.03]</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Chlamydial cervicitis [A56.09]</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Vulvovaginitis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vulvovaginitis [N76.0]</td>
<td>21 (61.76)</td>
<td>13 (38.24)</td>
<td>34 (0.99)</td>
</tr>
<tr>
<td>Gonococcal vulvovaginitis, unspecified [A54.02]</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Chlamydial vulvovaginitis [A56.02]</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Pharyngitis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharyngitis, unspecified etiology [J02.9]</td>
<td>4 (28.57)</td>
<td>10 (71.43)</td>
<td>14 (0.41)</td>
</tr>
<tr>
<td>Pharyngitis, gonococcal [A54.5]</td>
<td>3 (75)</td>
<td>1 (25)</td>
<td>4 (0.12)</td>
</tr>
<tr>
<td>Pharyngitis, chlamydial [A56.4]</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Proctitis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proctitis [K62.89]</td>
<td>1 (50)</td>
<td>1 (50)</td>
<td>2 (0.06)</td>
</tr>
<tr>
<td>Proctitis, chlamydial [A56.3]</td>
<td>1 (50)</td>
<td>1 (50)</td>
<td>2 (0.06)</td>
</tr>
<tr>
<td>Proctitis, gonococcal [A54.6]</td>
<td>1 (100)</td>
<td>0 (0)</td>
<td>1 (0.03)</td>
</tr>
<tr>
<td><strong>Sexually transmitted infection–related laboratory test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia or Gonorrhoeae RNA, TMA(^a), and throat</td>
<td>36 (72)</td>
<td>14 (28)</td>
<td>50 (1.45)</td>
</tr>
<tr>
<td>Chlamydia or Gonorrhoea, RNA, TMA and rectal</td>
<td>19 (55.88)</td>
<td>15 (44.12)</td>
<td>34 (0.99)</td>
</tr>
<tr>
<td>DNA Probe, chlamydia or gonorrhea, swab</td>
<td>1 (50)</td>
<td>1 (50)</td>
<td>2 (0.06)</td>
</tr>
<tr>
<td>Chlamydia/Gonorrhoea BV(^b) (BHN COD)</td>
<td>4 (80)</td>
<td>1 (20)</td>
<td>5 (0.14)</td>
</tr>
<tr>
<td>Chlamydia/Gonorrhoea, DNA, SDA(^c), urine</td>
<td>12 (32.43)</td>
<td>25 (67.57)</td>
<td>37 (1.07)</td>
</tr>
<tr>
<td>Syphilis IgG(^d)/IgM(^e) screen w/reflex to RPR(^f)</td>
<td>30 (29.7)</td>
<td>71 (70.3)</td>
<td>101 (2.93)</td>
</tr>
<tr>
<td>HIV fourth-generation ELISA(^g)</td>
<td>191 (38.74)</td>
<td>302 (61.26)</td>
<td>493 (14.29)</td>
</tr>
</tbody>
</table>

\(^a\)TMA: transcription-mediated amplification.  
\(^b\)BVS: blind vaginal swab.  
\(^c\)SDA: strand displacement amplification.  
\(^d\)IgG: immunoglobulin G.  
\(^e\)IgM: immunoglobulin M.  
\(^f\)RPR: rapid plasma reagin.  
\(^g\)ELISA: enzyme-linked immunosorbent assay.

**Outcomes of Intervention**

In the intervention period, among patients where a BPA was triggered, there were fewer patients with a diagnosis of gonococcal infections (33/750, 4.4%) as compared to the baseline period (41/649, 6.32%; Table 4). While roughly four-fifths (34/41, 82.93%) of these patients diagnosed with gonorrhea were prescribed antimicrobials for treating the infection, the difference in the number of patients treated in the intervention (26/33, 78.79%) and baseline periods were not statistically significant (P=.65). Individuals in the intervention period had a 26% decrease in odds (odds ratio 0.74, CI: 0.60-0.91) of being screened for HIV (310/750, 41.33%) compared to the baseline period (317/649, 48.84%). PrEP was
prescribed 2 times more to patients in the intervention period (12/750, 1.6%) compared to the baseline period (6/649, 0.92%); however, analysis revealed that these numbers not to be statistically significant ($P=.26$).

### Table 4. Outcomes of intervention, pre- and postclinical decision support solution intervention.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Baseline period (n=11,269), n (%)</th>
<th>Intervention period (n=12,048), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target population</strong></td>
<td>649 (5.76)</td>
<td>750 (6.23)</td>
</tr>
<tr>
<td><strong>Diagnosed with gonorrhea</strong></td>
<td>41 (6.32)</td>
<td>33 (4.4)</td>
</tr>
<tr>
<td>Prescribed with antimicrobial treatment</td>
<td>34 (82.93)</td>
<td>26 (78.49)</td>
</tr>
<tr>
<td><strong>Screened for HIV</strong></td>
<td>317 (48.84)</td>
<td>310 (41.33)</td>
</tr>
<tr>
<td>Diagnosed with HIV</td>
<td>3 (0.95)</td>
<td>3 (0.97)</td>
</tr>
<tr>
<td><strong>Prescribed PrEP</strong></td>
<td>6 (0.92)</td>
<td>12 (1.6)</td>
</tr>
<tr>
<td>Truvada</td>
<td>5 (83.33)</td>
<td>11 (91.67)</td>
</tr>
<tr>
<td>Descovy</td>
<td>1 (16.67)</td>
<td>1 (8.33)</td>
</tr>
</tbody>
</table>

*a*Any person with confirmed case of gonorrhea or with a sexually transmitted infection–related chief complaint.

*PrEP: pre-exposure prophylaxis.*

### Clinical Provider Perspectives and Usability of the CDS Solution

#### Midpoint Survey

Most of the survey responders reported that the CDS solution was not intrusive to their clinical practice workflow (2/4, 50% strongly agreed; 1/4, 25% agreed; and 1/4, 25% was neutral). They also agreed (0/4, 0% strongly agreed; 3/4, 75% agreed; and 1/4, 25% was neutral) that the CDS solution was easy to navigate and provided sufficient time to use it during the patient consultation. They had neutral responses (0/4, 0% strongly agreed; 1/4, 25% agreed; and 3/4, 75% was neutral) about the CDS solution’s ability to present clear recommendations for treatment.

### Key Informant Interviews

The interview revealed 3 major emerging themes. First, the interviewees opined that the CDS solution provided a faster and more standardized approach for capturing the sexual history and treating the patient and planned to keep using it after the pilot was complete. Second, while the CDS solution improved the efficiency and streamlined their existing processes, the interviewees did not change their treatment or screening recommendations based on the solution. Third, they also revealed that the full scope of the CDS solution was not clear to them and reported discovering components of the solution on their own that they were not aware were part of the CDS solution.

There were 3 key themes related to the features of the CDS solution. First, the BPAs alone would not prompt behavior change and use. Interviewees emphasized the importance of education of new guidelines and changes in the EHR tools from someone they know and trust and believed this would promote adherence from clinicians. Second, there were mixed reviews regarding the use of the logic that determined to present PrEP-related questions to some patients considered at high-risk for HIV. While 1 interviewee noted that the unique focus of X-Clinic to specifically address sexual health–related questions and conditions, all patients should be asked PrEP-related questions and noted confusion over the questions being presented to only some patients and not to other patients. Another interviewee indicated that this logic was helpful for colleagues who are less familiar with PrEP to highlight relevance in populations not typically targeted for PrEP, such as heterosexual patients. Third, there was a general sense that SmartSets were hard to navigate. Multiple interviewees commented that they were unable to find SmartSets without BPAs, and once in the SmartSet, it was not easy to leave and come back if an issue outside the scope of the SmartSet needed to be addressed, such as discussions regarding intrauterine devices in relation to gonococcal infections. One interviewee noted handwriting details to remember after completing a task within SmartSet.

### Role-Based Use of CDS Solution

A patient may be seen by multiple providers from more than one provider type category. In the target population (750/12,048, 6.23%) described in Table 4, 32 Doctors of Medicine and Doctors of Osteopathy (MD/DO) provided care for 281 patients and minimally used SmartForms (20/281, 7.12%) but completed most of them once opened (18/20, 90%); 26 advanced practice providers (APPs) provided care to 362 patients, used SmartForms for 20.99% (76/362) of patients, and 81.58% (62/76) completed most of them. In contrast, the “other” category, which included all other clinical staff such as case manager, dentist, laboratory technician, licensed practical nurse, medical assistants, midwives, and social workers, had the highest use, opening SmartForms for half of the patients they served (99/192, 51.56%) and completed most of the forms (83/99, 83.84%). Similarly, “other” category staff administered the highest number of PrEP-related questions, followed by APPs and MD/DOs administered very few. Providers classified as MD/DOs had 309 clinical encounters with 281 patients and opened SmartSets in half of the encounters (157/309, 50.81%) and those classified as APPs had 398 clinical encounters with 362 patients and opened SmartSets slightly lower than half of the encounters (180/398, 45.23%). “Other” staff opened...
SmartSets the least (82/126, 39.42%), with 10 staff who interacted with 192 patients in 208 encounters.

**Discussion**

**Principal Findings**

We integrated translated information from 3 CDC guidelines into a single cohesive CDS solution. This included CDC’s treatment recommendations for gonococcal infections (for appropriate diagnosis, testing, and treatment for gonorrhea) and HIV screening recommendations and PrEP prescription recommendations that stem from the diagnosis of gonorrhea. This CDS solution was implemented at a federally qualified health center where it continued to be used after the intervention period. The CDS solution successfully collected relevant information about the patient, evaluated the patient information against the CDC guidelines to prompt adequate treatment for gonorrhea, and identified at-risk patients for further HIV screening and PrEP prescription. The quantitative data revealed high rates of use of the various tools developed for the CDS solution and demonstrated the feasibility of incorporating such multifaceted guidelines-based CDS solutions in real-world clinical settings.

During our usability testing, clinical providers used all components of the CDS solution, with varying degrees, based on their respective provider type or roles, and familiarity with the CDS solution. Providers classified as MD/DO used the SmartSet more than “other” staff such as medical assistants and social workers and vice versa for the SmartForm that performed the initial screen for patients’ sexual history and PrEP awareness. In the X-Clinic, specifically, the BPA was presented in 10.08% (348/3451) of the encounters; clinicians reported that this was not intrusive and could be used as a benchmark to potentially reduce alert fatigue. The SmartSet was opened in about half of the times alerted by BPA, but fewer orders were directly entered from the SmartSet itself. This does not mean that patients did not receive care. Some insights offered by providers as plausible explanations include (1) clinicians reported benefiting from the reminder and served as a knowledge source even if they performed those actions outside of the SmartSet; (2) clinicians were not accustomed to using SmartSets and reported experiencing difficulties in locating it once they moved away from it during the clinical visit; (3) the BPA alert at the time when the patient record was initially opened may be too early to determine diagnosis, order laboratory tests, and prescribe treatment, delaying the timing of the BPA could be explored; and (4) clinicians had neutral responses when asked about CDS solution’s ability to present clear recommendations since SmartSet presented all treatment options with those aligned to the guidelines marked as “preferred” instead of targeted recommendations for a particular patient.

There was no significant difference between the baseline and intervention patient groups with respect to the treatment for gonococcal infections. It is important to note that X-Clinic is primarily focused on providing sexual health and STI-related care, and the providers were well versed with treatment recommendations, which could be an explanation for the lack of difference. The results also show that one-fifth of the patients in both baseline and intervention periods were not treated with antimicrobials during the visit. A likely explanation is that these patients were empirically treated elsewhere and were referred to X-Clinic for further follow-up, given their specialty in providing STI-related care. Fewer patients were screened for HIV during the intervention period compared to the baseline period. However, key informant interview participants anecdotaly reported an increased interest regarding PrEP in less traditional populations, such as heterosexual women. While PrEP was prescribed to twice as many patients in the intervention period compared to the baseline period, this was not deemed statistically significant.

The widespread adoption of EHRs has provided us with the unique opportunity to leverage CDS approaches to automate and align clinical decisions such as identifying-at-risk patients for further screening and providing adequate treatment with the latest guidelines. This becomes especially important in situations where clinical providers may not be familiar with the latest screening and treatment guidelines. Given the rise in resistance to antimicrobial treatment for gonorrhea, adherence to treatment guidelines is even more important. Further development of standards and tools and examination of workflows to implement useful guidelines-based CDS solutions with the ultimate goal of improving patient care are needed.

**Limitations**

The original design was based upon a 6-month intervention period. A longer pilot period would have allowed the opportunity for a larger data set of patients diagnosed with gonorrhea and potentially more compelling statistical correlations, along with additional time for the clinical staff to acclimate to the new solution.

For this phase of our work, it was important to test the feasibility of the guidelines-based CDS logic, measure the use of the CDS tools, and solicit feedback from end users regarding their experience using the CDS tool. Prioritizing these factors, we decided to leverage EHR-native but vendor-specific tools such as SmartForm, BPA, SmartSet, and SmartText. We factored that OCHIN, our partner organization has many clinics using their instance of Epic EHR across the United States [37,38], and any lessons learned could potentially be replicated and scaled to other interested clinical sites.

Since the focus of this work was to translate multiple CDC guidelines into a cohesive CDS solution, we selected a clinical site that was well versed in STI diagnosis and treatment with a track record of serving a large number of patients seeking sexual health and STI-related care. Due to the same reasons, this clinical setting may not be the best target of clinical providers to measure any potential outcomes of the CDS solution. The CDS solution would better serve clinical providers and settings, where there is less familiarity of treating patients with STIs.

In this study, we report the feasibility, use, and usability of the CDS solution; we did not measure the effectiveness of the intervention. We did not examine any change in clinical practice and its impact on clinical care or the long-term prevention outcomes in the patient population. As a result, we did not randomize the study and did not examine any potential
confounding. Similarly, while we obtained quantitative and qualitative information from end users of various roles, the small sample size poses challenges to draw firm conclusion of usability at scale. Further examination of the CDS solution could be performed by manual chart review with an annotated gold standard measure.

Comparison With Prior Work
In the past, CDC guidelines–based CDS interventions have been designed to recommend appropriate immunizations [51], screening for alcohol use disorder [52], and screening for cervical cancer [53]. A United States Indian Health Services clinic implemented a chlamydia and HIV screening tool and observed increase in screening frequencies of both chlamydia and HIV; however, this was not strictly based on CDC guidelines [54]. Another study described the implementation of CDS tool to encourage appropriate prescription of azithromycin in primary care clinics with the aim to curb rise of antimicrobial resistance, but this study was not targeted toward STIs and was focused on bronchitis and upper respiratory tract infections [55]. To the best of our knowledge, no CDS tools have been implemented with the goal of improving adherence to guidelines for gonorrhea treatment, HIV screening, and HIV PrEP.

Conclusions
It is feasible to integrate multiple CDC guidelines into a single cohesive CDS solution. The CDS solution showed high rates of use, but given the short study period, we could not adequately measure realistic patient outcomes. The clinical site has opted to continue using the full scope of the CDS solution, and perhaps that decision is a measure of success.

Learning from this experience, we will be developing a standards-based EHR-agnostic CDS solution with a longer study period.

Acknowledgments
This project was supported by the Centers for Disease Control and Prevention (CDC) funding to the Task Force for Global Health under a cooperative agreement (NU38OT000316). The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the US CDC. Mention of company names or products does not imply endorsement by the CDC. The authors have not leveraged generative artificial intelligence in writing any portion of this manuscript.

“Profile” symbol on Figure 1 is by Gregor Cresnar from The Noun Project “Platforms” symbol on Figure 1 is by IconPai from The Noun Project.

Data Availability
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Midpoint survey instrument.
[DOCX File, 16 KB - formative_v8i1e53000_app1.docx]

Multimedia Appendix 2
Key informant interview questions.
[DOCX File, 18 KB - formative_v8i1e53000_app2.docx]

References


12. HIV/AIDS. World Health Organization. URL: https://www.who.int/health-topics/hiv-aids#tab=1 [accessed 2022-09-19]


30. Clinical decision support. The Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/topic/safety/clinical-decision-support [accessed 2022-09-19]

38. OCHIN. URL: https://ochin.org/ [accessed 2022-08-25]


### Abbreviations

- **APP**: advanced practice provider
- **BPA**: best practice advisory
- **CDC**: Centers for Disease Control and Prevention
- **CDS**: clinical decision support
- **DO**: Doctor of Osteopathy
- **EHR**: electronic health record
- **ICD-10-CM**: International Classification of Diseases, Tenth Revision, Clinical Modification
- **MD**: Doctor of Medicine
- **PrEP**: pre-exposure prophylaxis
- **STI**: sexually transmitted infection
Assessing the Accuracy of Generative Conversational Artificial Intelligence in Debunking Sleep Health Myths: Mixed Methods Comparative Study With Expert Analysis

Nicola Luigi Bragazzi1,2,3, MPH, MD, PhD; Sergio Garbarino2,4, MD, PhD

1Human Nutrition Unit, Department of Food and Drugs, University of Parma, Parma, Italy
2Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal/Child Sciences, University of Genoa, Genoa, Italy
3Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada
4Post-Graduate School of Occupational Health, Università Cattolica del Sacro Cuore, Rome, Italy

Corresponding Author:
Nicola Luigi Bragazzi, MPH, MD, PhD
Human Nutrition Unit
Department of Food and Drugs
University of Parma
Via Volturno 39
Parma, 43125
Italy
Phone: 39 0521 903121
Email: nicolaluigi.bragazzi@unipr.it

Abstract

Background: Adequate sleep is essential for maintaining individual and public health, positively affecting cognition and well-being, and reducing chronic disease risks. It plays a significant role in driving the economy, public safety, and managing health care costs. Digital tools, including websites, sleep trackers, and apps, are key in promoting sleep health education. Conversational artificial intelligence (AI) such as ChatGPT (OpenAI, Microsoft Corp) offers accessible, personalized advice on sleep health but raises concerns about potential misinformation. This underscores the importance of ensuring that AI-driven sleep health information is accurate, given its significant impact on individual and public health, and the spread of sleep-related myths.

Objective: This study aims to examine ChatGPT’s capability to debunk sleep-related disbeliefs.

Methods: A mixed methods design was leveraged. ChatGPT categorized 20 sleep-related myths identified by 10 sleep experts and rated them in terms of falseness and public health significance, on a 5-point Likert scale. Sensitivity, positive predictive value, and interrater agreement were also calculated. A qualitative comparative analysis was also conducted. A qualitative comparative analysis was also conducted. A qualitative comparative analysis was also conducted.

Results: ChatGPT labeled a significant portion (n=17, 85%) of the statements as “false” (n=9, 45%) or “generally false” (n=8, 40%), with varying accuracy across different domains. For instance, it correctly identified most myths about “sleep timing,” “sleep duration,” and “behaviors during sleep,” while it had varying degrees of success with other categories such as “pre-sleep behaviors” and “brain function and sleep.” ChatGPT’s assessment of the degree of falseness and public health significance, on the 5-point Likert scale, revealed an average score of 3.45 (SD 0.87) and 3.15 (SD 0.99), respectively, indicating a good level of accuracy in identifying the falseness of statements and a good understanding of their impact on public health. The AI-based tool showed a sensitivity of 85% and a positive predictive value of 100%. Overall, this indicates that when ChatGPT labels a statement as false, it is highly reliable, but it may miss identifying some false statements. When comparing with expert ratings, high intraclass correlation coefficients (ICCs) between ChatGPT’s appraisals and expert opinions could be found, suggesting that the AI’s ratings were generally aligned with expert views on falseness (ICC=0.83, P<.001) and public health significance (ICC=0.79, P=.001) of sleep-related myths. Qualitatively, both ChatGPT and sleep experts refuted sleep-related misconceptions. However, ChatGPT adopted a more accessible style and provided a more generalized view, focusing on broad concepts, while experts sometimes used technical jargon, providing evidence-based explanations.

Conclusions: ChatGPT-4 can accurately address sleep-related queries and debunk sleep-related myths, with a performance comparable to sleep experts, even if, given its limitations, the AI cannot completely replace expert opinions, especially in nuanced and complex fields such as sleep health, but can be a valuable complement in the dissemination of updated information and promotion of healthy behaviors.
KEYWORDS

sleep; sleep health; sleep-related disbeliefs; generative conversational artificial intelligence; chatbot; ChatGPT; misinformation; artificial intelligence; comparative study; expert analysis; adequate sleep; well-being; sleep trackers; sleep health education; sleep-related; chronic disease; healthcare cost; sleep timing; sleep duration; presleep behaviors; sleep experts; healthy behavior; public health; conversational agents

Introduction

An adequate amount of good, restorative sleep is of paramount importance for both individual and public health [1,2]; from an individual standpoint, it helps maintain optimal physical and mental health, facilitating cognitive function, ensuring well-being, and mitigating the risks associated with chronic diseases [3]. In the context of public health, sleep’s impact is profound and multifaceted as well, being a pivotal element in driving the economy, ensuring public safety, and managing health care expenditures. The strategic addressing of sleep-related issues not only alleviates the global burden of disease but also ameliorates the economic strain associated with it [4,5].

The promotion of healthy sleep patterns and the intervention in sleep-related disorders emerge as vital strategies, paving the way for the enhancement of overall societal well-being, boosting productivity, and fostering social cohesion [6]. Such initiatives can yield substantial benefits, at both the individual and community levels, thereby underscoring the role of innovative tools, including digital ones—spanning from dynamic websites to sleep trackers and mobile apps—in promoting and providing education on sleep health [7].

The internet offers a vast, versatile, easily accessible, and cost-effective platform for disseminating up-to-date information about sleep, reaching diverse populations, raising public awareness about the importance of sleep, and providing personalized guidance on sleep health and related topics. People can access the latest findings and recommendations to make informed decisions about their sleep habits, with telemedicine and web-based consultations with sleep experts becoming increasingly popular. The digital realm can enable individuals to monitor their sleep patterns, engaging them in continuous learning about sleep health, and facilitating self-awareness and behavioral changes to improve sleep quality [8].

In the era of generative conversational artificial intelligence (AI) [9,10], characterized by disruptive technological transformation, the importance of sleep health promotion and education becomes even more relevant [11]: conversational AI-based platforms and agents, such as chatbots, can provide instant responses to sleep-related queries, making information readily available at any time. This real-time accessibility can help individuals seeking answers about sleep health, who can receive personalized advice and recommendations based on an individual’s specific sleep patterns and concerns. However, besides being accessible and tailored, this information should also be accurate [12].

There are only a few studies that have assessed sleep-related knowledge of conversational AI-based chatbots, such as ChatGPT-4, which was found very recently to successfully pass the sleep medicine certification board examinations [13] and be conversant in sleep disorders, such as obstructive sleep apnea syndrome [14-16].

On the other hand, conversational AI may contribute to disseminating “factual errors, nonsense, fabricated sources, and dangerous advice” and, thus, spreading biomedical misinformation, including sleep-related misinformation [17]. Therefore, our study was conducted to verify the accuracy of a popular prototype of conversational AI, ChatGPT, in addressing queries concerning sleep health and, in particular, sleep-related myths. These can be defined as widely held “false beliefs about sleep” that “lack an evidence base” and “can persist despite contradicting scientific evidence, potentially impairing” and even degrading population health, by promoting the adoption of unhealthy behaviors and lifestyles, the identification of which “can inform efforts to promote population sleep health” [18].

Methods

Procedure

A list of 20 sleep-related myths, as defined above, was taken from a previously published study [18]. This list was compiled using internet searches of popular press and scientific literature and leveraging a Delphi process that involved 10 sleep experts from the fields of sleep medicine and research. Experts were recruited by convenience sampling, after being identified through literature searches (using PubMed). To be considered an expert, they were required to have published 20 papers that were cited by 20 or more different peer-reviewed sources, and at least one of these publications had to be tagged with the “Medical Subject Headings” “sleep” along with either “circadian rhythms,” “neuroscience,” or “psychiatry.” A total of 20 individuals who fulfilled these requirements were reached out to and out of these 20 experts 10 took part in this study. The Delphi process consisted of selecting and refining myths and was conducted in 3 stages: initially, focus groups were held (phase 1); this was followed by a period of email-based feedback for editing, adding, or removing myths (phase 2); finally, closed-ended surveys were used (phase 3), during which experts assessed the myths. The 20 myths were, then, categorized along six domains: namely, (1) “sleep duration” (n=6), (2) “sleep timing” (n=1), (3) “behaviors during sleep” (n=4), (4) “daytime behaviors that relate to sleep” (n=2), (5) “pre-sleep behaviors” (n=5), and (6) “brain function and sleep” (n=2). Besides providing feedback, experts had to rate myths on 2 dimensions: falseness and public health significance using a 5-point Likert scale from 1 (“not at all false” or “not at all significant”) to 5 (“extremely false” or “extremely significant”) [18].
It should be noted that, while some of these myths are patently false (such as the statement “during sleep, the brain is not active,” which belongs to the “brain function and sleep” domain), other statements such as “lying in bed with your eyes closed is almost as good as sleeping” (belonging to the “behaviors during sleep” domain) contain some elements of truth and other sleep experts may disagree in labeling them as complete misinformation or myths. Indeed, whether “waking rest” and other “resting states” may confer benefits almost as good as deep rest is debated [19,20]: these concepts challenge the conventional dichotomy of sleep and wakefulness conceived as binary and distinct states and suggest that periods of quiet reflection during wakefulness, characterized by a lack of effortful, focused thought and the absence of distracting stimuli, can also contribute significantly to mental rejuvenation, memory consolidation, hormonal regulation, cellular repair, and emotion regulation [21-23]. Altogether, these resting states, defined also as “offline states,” including eyes-closed rest, daydreaming, mind wandering, or inattentive states, represent approximately half of our waking hours [24,25]. However, the existing scholarly literature reports scarce, contrasting, or even negative findings [26], that warrant further research and suggest that these concepts are not yet well established and are continuously evolving [19,20].

Sleep is, indeed, a complex, nonlinear process, and sometimes, our perception of how well we sleep or even whether we are asleep or awake can be incorrect. For instance, “sleep state misperception,” also known as “paradoxical insomnia,” is a condition where individuals believe they are awake for most of the night, despite actually sleeping for a normal duration. This disorder is characterized by a significant discrepancy between perceived and actual sleep time, often associated with personality traits like neuroticism and altered brain activity during sleep, though its causes and prevalence remain under investigation [27-30].

**Ethical Considerations**

Full ethical clearance was waived for this study, as this study is a purely observational study with responses generated by AI (ChatGPT) and secondary analysis of research data, consisting of anonymous or deidentified study data [18].

**Statistical Analysis**

ChatGPT (version 4) was asked both to determine if these 20 sleep-related disbeliefs were true or false and to appraise them using the 5-point Likert scale. To determine if there is a statistical difference between the 6 abovementioned domains in terms of the distribution of true, false, and other categories of responses, a chi-square test of independence was used. This test helped us determine if the differences in proportions across the different domains were statistically significant. Means were also reported for the overall score (along with their SDs) and broken down according to each domain.

In terms of accuracy, the sensitivity and the positive predictive value of ChatGPT in categorizing the sleep-related statements as false were computed.

Finally, ChatGPT’s ratings of falseness and public health significance of sleep-related myths were compared with those provided by sleep experts. The degree of agreement was measured, in terms of consistency, using the interrater reliability analysis, computing the intraclass correlation coefficients (ICCs) [31].

All statistical analyses were carried out using SPSS (version 28 for Windows; IBM Corp). P values less than .05 were considered statistically significant.

**Qualitative Analytical Approach**

A qualitative comparative analysis was also conducted. Initially, responses from ChatGPT and summary responses from the sleep experts [18] were subjected to a line-by-line comparison to identify similarities and differences in content, style, and complexity of information provided. Then, responses were scrutinized to identify themes, concepts, or categories that were entered in a matrix to have a clear snapshot of where ChatGPT and the experts aligned or diverged in their discussions and to make emerging patterns of alignment and divergence between them. This phase was crucial for understanding how ChatGPT’s training data correlated or not with the current consensus among experts and enabled the identification of gaps in ChatGPT’s knowledge base.

**Results**

**ChatGPT’s Falseness Quantitative Analysis of Sleep-Related Myths**

Overall, ChatGPT labeled 45% (n=9) of the statements as “false,” while a further 40% (n=8) of the items were deemed as “generally false.” Of note, concerning the remaining statements, 5% (n=1) and 10% (n=2) of them were considered “true” and “not (entirely or necessarily) true or false,” respectively. In terms of domain, half of the items related to “sleep duration” were considered “false” (n=3), with the remaining half percent being deemed “generally false” (n=3). The statement concerning “sleep timing” was labeled as “generally false” (n=1). Further, 75% (n=3) of the items related to “behaviors during sleep” were correctly identified as “false,” while the remaining 25% (n=1) were classified as “generally false.” All the statements concerning “daytime behaviors related to sleep” were considered “generally false” (n=2). When assessing the accuracy of items concerning “pre-sleep behaviors,” half of them were properly labeled as “false” (n=2), whereas 16.67% (n=1) of the statements were considered “generally false,” with a further 16.67% (n=1) being “not entirely true or false” and the remaining 16.67% (n=1) being even considered “true.” Finally, concerning “brain function and sleep,” half of the statements were correctly appraised as “false” (n=1), with the remaining half being labeled as “not necessarily true or false” (n=1). Further details are reported in Table 1 and Multimedia Appendix 1.
Table 1. ChatGPT’s appraisals of the falseness of sleep-related myths, scored both qualitatively (true or false) and quantitatively (on a 5-point Likert scale, from 1 or “not at all false” to 5 or “extremely false”), encompassing a range of 6 different topics (sleep duration, sleep timing, behaviors during sleep, daytime behaviors that relate to sleep, presleep behaviors, and brain function and sleep).

<table>
<thead>
<tr>
<th>Sleep-related myths</th>
<th>ChatGPT, mean</th>
<th>True or false</th>
<th>On the 5-point Likert scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sleep duration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Being able to fall asleep ‘anytime, anywhere’ is a sign of a healthy sleep system”</td>
<td>Generally false</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>“Many adults need only 5 or less hours of sleep for general health”</td>
<td>False</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>“Your brain and body can learn to function just as well with less sleep”</td>
<td>False</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>“Adults sleep more as they get older”</td>
<td>Generally false</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>“If you can get it, more sleep is always better”</td>
<td>False</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>“One night of sleep deprivation will have lasting negative health consequences”</td>
<td>Generally false</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>Sleep timing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“In terms of your health, it does not matter what time of day you sleep”</td>
<td>Generally false</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>Behaviors during sleep</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Lying in bed with your eyes closed is almost as good as sleeping”</td>
<td>False</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>“If you have difficulty falling asleep, it is best to stay in bed and try to fall back to sleep”</td>
<td>Generally false</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>“Although annoying for bed partners, loud snoring is mostly harmless”</td>
<td>False</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>“A sound sleeper rarely moves at night”</td>
<td>False</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>Daytime behaviors that relate to sleep</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Hitting the snooze when you wake up is better than getting up when the alarm first goes off”</td>
<td>Generally false</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>“If you are having difficulties sleeping, taking a nap in the afternoon is a good way to get adequate sleep”</td>
<td>Generally false</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>Presleep behaviors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Alcohol before bed will improve your sleep”</td>
<td>False</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>“For sleeping, it is better to have a warmer bedroom than a cooler bedroom”</td>
<td>False</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>“Boredom can make you sleepy even if you got adequate sleep before”</td>
<td>True</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>“Watching television in bed is a good way to relax before sleep”</td>
<td>Generally false</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>“Exercising within 4 hours of bedtime will disturb your sleep”</td>
<td>Not entirely true or false</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>Brain function and sleep</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“During sleep, the brain is not active”</td>
<td>False</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>“Remembering your dreams is a sign of a good night’s sleep”</td>
<td>Not necessarily true or false</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

The various response categories did not vary depending on the domain of sleep-related myths ($\chi^2_{15}=14.60$, $P=.48$).

On the 5-point Likert scale, the degree of falseness was computed at 3.45 (SD 0.87), according to ChatGPT’s estimates. The highest scores were recorded for “brain function and sleep” (4.00, SD 1.41), “sleep timing” (4.00, single item), and “sleep duration” (3.67, SD 0.98), while the “behavioral domains” scored the lowest. More in detail, “behaviors during sleep” yielded a value of 3.25 (SD 0.50), followed by “pre-sleep behaviors” (3.20, SD 0.84) and “daytime behaviors that relate to sleep” (3.00, SD 1.41). Further details are presented in Table 1 and Multimedia Appendix 2.

Based on these data, ChatGPT demonstrated an overall sensitivity of 85% and a positive predictive value of 100% in categorizing the statements as “false.”

**Quantitative Comparison of ChatGPT’s and Expert Ratings on the Falseness of Sleep-Related Myths**

When comparing with sleep experts, a good interrater agreement could be found between ChatGPT’s categorization of statements...
and expert rating on their falseness. Statements categorized by ChatGPT as “false” and “generally false” were those that received the highest scores by the experts (4.25 and 3.97, respectively), whereas those judged by the AI as “true” and “not true or false” received the lowest scores by the experts (3.75 and 3.44, respectively), as shown in Figure 1. From a more quantitative standpoint, the association yielded an ICC value of 0.83 ($P < .001$), when ChatGPT was asked to rate the degree of falseness of the statement on the 5-point Likert scale (Figure 2).

**Figure 1.** Comparison of ChatGPT’s and expert qualitative ratings on the falseness of sleep-related myths, showing a general good agreement and alignment between experts and artificial intelligence.

![Falseness](image)

**Figure 2.** Scatterplot of the comparison of ChatGPT’s and expert quantitative ratings (on a 5-point Likert scale, from 1 or “not at all false” to 5 or “extremely false”) on the falseness of sleep-related myths, showing a general good agreement and alignment between experts and artificial intelligence.

**ChatGPT’s Public Health Significance Quantitative Analysis of Sleep-Related Myths**

The overall score was 3.15 (SD 0.99). “Sleep timing” was the domain scoring the highest (4.00, single item), followed by “sleep duration” (3.33, SD 1.37) and “behaviors during sleep” (3.25, SD 0.96). “Daytime behaviors that relate to sleep” scored the lowest (2.50, SD 0.71), while both “brain function and sleep” and “pre-sleep behaviors” yielded a mean score of 3.00 (SD 1.41, and SD 0.71, respectively). Further details are presented in Table 2.
Table 2. ChatGPT’s appraisals of the public health significance of sleep-related myths, scored quantitatively (on a 5-point Likert scale, from 1 or “not at all significant” to 5 or “extremely significant”), encompassing a range of 6 different topics (sleep duration, sleep timing, behaviors during sleep, daytime behaviors that relate to sleep, presleep behaviors, and brain function and sleep).

<table>
<thead>
<tr>
<th>Sleep-related myths</th>
<th>ChatGPT (on the 5-point Likert scale), mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sleep duration</strong></td>
<td></td>
</tr>
<tr>
<td>“Being able to fall asleep ‘anytime, anywhere’ is a sign of a healthy sleep system”</td>
<td>2</td>
</tr>
<tr>
<td>“Many adults need only 5 or less hours of sleep for general health”</td>
<td>5</td>
</tr>
<tr>
<td>“Your brain and body can learn to function just as well with less sleep”</td>
<td>5</td>
</tr>
<tr>
<td>“Adults sleep more as they get older”</td>
<td>3</td>
</tr>
<tr>
<td>“If you can get it, more sleep is always better”</td>
<td>3</td>
</tr>
<tr>
<td>“One night of sleep deprivation will have lasting negative health consequences”</td>
<td>2</td>
</tr>
<tr>
<td><strong>Sleep timing</strong></td>
<td></td>
</tr>
<tr>
<td>“In terms of your health, it does not matter what time of day you sleep”</td>
<td>4</td>
</tr>
<tr>
<td><strong>Behaviors during sleep</strong></td>
<td></td>
</tr>
<tr>
<td>“Lying in bed with your eyes closed is almost as good as sleeping”</td>
<td>4</td>
</tr>
<tr>
<td>“If you have difficulty falling asleep, it is best to stay in bed and try to fall back to sleep”</td>
<td>3</td>
</tr>
<tr>
<td>“Although annoying for bed partners, loud snoring is mostly harmless”</td>
<td>4</td>
</tr>
<tr>
<td>“A sound sleeper rarely moves at night”</td>
<td>2</td>
</tr>
<tr>
<td><strong>Daytime behaviors that relate to sleep</strong></td>
<td></td>
</tr>
<tr>
<td>“Hitting the snooze when you wake up is better than getting up when the alarm first goes off”</td>
<td>2</td>
</tr>
<tr>
<td>“If you are having difficulties sleeping, taking a nap in the afternoon is a good way to get adequate sleep”</td>
<td>3</td>
</tr>
<tr>
<td><strong>Presleep behaviors</strong></td>
<td></td>
</tr>
<tr>
<td>“Alcohol before bed will improve your sleep”</td>
<td>4</td>
</tr>
<tr>
<td>“For sleeping, it is better to have a warmer bedroom than a cooler bedroom”</td>
<td>3</td>
</tr>
<tr>
<td>“Boredom can make you sleepy even if you got adequate sleep before”</td>
<td>2</td>
</tr>
<tr>
<td>“Watching television in bed is a good way to relax before sleep”</td>
<td>3</td>
</tr>
<tr>
<td>“Exercising within 4 hours of bedtime will disturb your sleep”</td>
<td>3</td>
</tr>
<tr>
<td><strong>Brain function and sleep</strong></td>
<td></td>
</tr>
<tr>
<td>“During sleep, the brain is not active”</td>
<td>4</td>
</tr>
<tr>
<td>“Remembering your dreams is a sign of a good night’s sleep”</td>
<td>2</td>
</tr>
</tbody>
</table>

Quantitative Comparison of ChatGPT’s and Expert Ratings on the Public Health Significance of Sleep-Related Myths

Similar trends to those observed for the expert appraisals of the falseness of sleep-related myths could be reported for the expert rating on their public health significance. Items labeled by ChatGPT as “false” and “generally false” corresponded to a score of 3.37 and 3.04, respectively, while statements appraised by the AI as “true” and “not true or false” scored the lowest (2.71 and 2.07, respectively), as pictorially represented in Figure 3.

When comparing the ratings on public health significance provided by ChatGPT with those by the experts, an ICC of 0.79 could be computed (P=.001), as shown in Figure 4.
Qualitative Comparison and Thematic Analysis of ChatGPT’s and Sleep Experts’ Responses

The qualitative comparison between ChatGPT’s and sleep experts’ appraisals of sleep-related myths revealed both similarities and differences across the various misconceptions.

In general, both ChatGPT and sleep experts clarified and, in some instances, strongly refuted these misconceptions, highlighting the importance of adequate sleep for overall well-being, while acknowledging individual differences in sleep needs and patterns, and emphasizing that sleep requirements can vary from person to person. However, differences in responses could be noted: ChatGPT adopted a more conversational and accessible style and provided a more generalized view, focusing on broad concepts, while experts sometimes used technical jargon, providing evidence-based explanations to debunk false beliefs about sleep. Sleep experts tended to provide a more detailed assessment, often including specific medical and physiological contexts, referencing studies, and focusing both on individual (clinical) and public health implications, while a population health perspective was generally missing in ChatGPT’s responses.

Specifically concerning “sleep duration myths,” both ChatGPT and sleep experts clarified the misconceptions, but experts provided a more medically oriented focus, delving into the biological and physiological underpinnings more deeply, while ChatGPT emphasized sleep-related flexibility and individual variability. Further, sleep experts provided a more nuanced...
appraisal of risks and the evidence (or lack thereof) supporting them, focusing not only on temporary or short-term effects but offering detailed perspectives on long-term consequences and recovery too. Similarly, concerning “sleep-related timing myths,” both ChatGPT and sleep experts emphasized the role of circadian rhythms and the importance of sleep timing, but experts gave a more technical assessment focusing on specific physiological processes. Regarding “behaviors during sleep” myths and “daytime sleep-related behaviors” myths, sleep experts delved into physiological, cultural, and habitual aspects and potential adverse outcomes. Concerning “pre-sleep behaviors myths,” sleep experts focused more on the neurophysiological impacts and provided more evidence-based assessments, with more nuanced explanations of the causes and effects underlying sleep phenomena and disturbances. Finally, regarding “brain function and sleep myths,” sleep experts gave a more detailed explanation of brain functions, highlighting the complexity of sleep research and the presence of controversial topics and conflicting results. Further details are reported in Table S1 in Multimedia Appendix 3.

Discussion

Generative Conversational AI and Sleep-Related Myths

AI-driven platforms and agents, including chatbots, can provide sleep-related information and education, ensuring that a diverse global audience can access valuable and customized resources on sleep health. This tailored guidance can be particularly beneficial for improving sleep quality, in that AI-driven chatbots and digital assistants can provide ongoing support and reminders for healthy sleep habits, promoting consistent behavior change over time. However, information disseminated using digital tools must be accurate and reliable.

According to a few studies, the internet can be a useful platform for enhancing sleep-related health literacy, but can also contribute to spreading misinformation, often including commercial biases and incorrect and misleading content. Robbins et al [32] evaluated the understandability, quality of information, and presence of misinformation in popular YouTube videos about sleep compared to those featuring credible experts. The top YouTube videos on sleep or insomnia and 5 expert-led videos were analyzed for clarity and understanding using established tools. Sleep medicine experts agreed on instances of misinformation and commercial bias, with about 67% (n=14/21) of popular videos showing evidence of commercial bias, unlike the expert videos. Misinformation was more prevalent in popular videos that averaged 8.2 million (SD 2.2) views, significantly higher than the expert videos’ 0.3 million (SD 0.2) million views. Most YouTube videos were found to have clickbait and be appealing to shorter attention spans, having engaging content, good visual quality, and being highly relatable to viewers. All this highlights the issue of misinformation and bias in widely viewed sleep or insomnia videos on YouTube and other web-based platforms, suggesting the urgent need for combating digital sleep-related misinformation [33].

ChatGPT is anticipated to play a key role in sleep health promotion and education, enhancing public perceptions of the importance of sleep in daily life and its impact on human health. This analysis demonstrates the potential for AI tools like ChatGPT to provide health information, in particular in the arena of sleep medicine. Considering the overall distribution of responses provided by ChatGPT, a high proportion of sleep-related myths (n=17/20, 85% of the statements) was correctly identified as either false or generally false, suggesting that ChatGPT is aligned with scientific evidence. However, the categorization of some statements as “true” or “not necessarily true or false” indicates ChatGPT’s ability to recognize and label scientific items as accurate can be still improved.

In general, ChatGPT has a good, scholarly understanding of several crucial aspects of sleep health, spanning from sleep duration and timing to behaviors during sleep, while it demonstrates some limitations in the field of sleep hygiene, and in the understanding of sleep-related occupational and public health implications.

Moreover, addressing sleep myths involves a nuanced exploration of sleep-related topics: our qualitative analysis on how common misconceptions are clarified by both AI platforms such as ChatGPT and sleep experts shows a good alignment, though some statements are approached from different angles. From a qualitative comparative perspective, ChatGPT tends to provide more pragmatic advice and tips, emphasizing the importance of regular sleep schedules and practices, even if in the context of a certain degree of flexibility in sleep systems, and the impact of individual behaviors on sleep quality. This approach often includes general recommendations based on a broad understanding of sleep science, aiming to correct misunderstandings such as the notion that less sleep can be habitually sufficient or that lying in bed with eyes closed substitutes for genuine sleep. In contrast, sleep experts delve deeper into the medical and physiological specifics, offering a more detailed assessment that considers individual health conditions, genetic predispositions, and the long-term health risks associated with disrupted sleep patterns. They might focus on the precise effects of sleep deprivation on cognitive function, the specific dangers of certain presleep behaviors, or the complex, nonlinear relationship between sleep stages and overall health. A major difference between ChatGPT and sleep experts is that only the latter have mentioned the public and occupational aspects of sleep, while the former has focused more on the individual level. The dialogue between these perspectives can enrich our understanding of sleep, blending practical guidance with in-depth scientific insights to debunk myths and promote healthier sleep practices across diverse populations.

However, as previously mentioned, some errors by ChatGPT in correctly classifying myths as false statements underscore the current limitations of AI: users should be aware of the shortcomings of AI-based tools in interpreting complex, evolving fields like sleep science and sleep health. ChatGPT’s classifications are not definitive statements of truth but rather reflections of current knowledge and interpretations, which are constantly evolving. In summary, the categorizations by ChatGPT provide an interesting insight into how AI tools process and present information on complex health topics such
as sleep, emphasizing the importance of contextual understanding and the ongoing development of AI capabilities in health education.

**Implications and Future Directions**

ChatGPT’s ability to debunk sleep-related myths has several important implications, both in the field of sleep health and in the context of AI in health care and information dissemination. The ability of ChatGPT to accurately debunk sleep-related myths can significantly contribute to enhancing public health education, including sleep health education. Providing reliable information can help correct widespread misconceptions about sleep, which is vital given the importance of sleep for overall health. ChatGPT can also serve as a tool for supporting health care professionals, helping them to stay abreast of the latest advancements, quickly verify information, and provide evidence-based advice to their patients, potentially improving the quality of sleep health advice given.

Moreover, AI-based platforms such as ChatGPT can make sleep health information more accessible to a broader audience and can offer personalized advice based on individual queries, which is difficult to achieve through traditional health education methods.

This study indicates that AI can be a reliable source of health information. However, it also highlights the need for ongoing evaluation to ensure accuracy, especially in areas with nuanced and complex information, such as sleep health. More in detail, this study suggests that while AI tools such as ChatGPT can be highly effective, they should not replace expert opinion but rather complement it. This is particularly important in complex fields where contextual understanding and professional judgment are crucial. There is a need for continuous learning and updating: AI systems must continuously learn and upgrade their knowledge base to ensure the information they provide stays current with the latest scientific findings and expert consensus.

Moreover, ChatGPT’s ability to identify and correct false information is particularly relevant in an era where misinformation can spread rapidly on the web [34-36]. This capability can play a significant role in public health initiatives. On an individual level, accurate AI-driven advice on sleep health can directly contribute to the prevention of diseases, including sleep-related disorders, which are often linked to chronic diseases such as obesity, diabetes, and cardiovascular issues. By debunking myths and offering personalized guidance on healthy sleep practices, these tools can play a pivotal role in enhancing individual wellness, mental health, and overall quality of life. In the broader context of occupational and public health, provided that the above-mentioned shortcomings of ChatGPT in these fields are properly addressed, the dissemination of reliable sleep-related information via AI platforms can aid in the formulation of more informed public health policies and initiatives. By increasing the general population’s understanding of the importance of sleep, these tools can contribute to a reduction in health care costs associated with sleep disorders and their comorbidities. Furthermore, the implications for public safety are significant. Improved sleep health, guided by AI-based tools, can lead to decreased incidences of accidents and errors attributed to sleep deprivation, such as those in high-risk professions (eg, transportation, health care, etc). This would not only enhance the safety of the individuals in these roles but also safeguard the broader community. Thus, the integration of AI in sleep health education and promotion aligns with broader public health and safety goals, offering a proactive approach to mitigating risks associated with poor sleep and promoting a healthier, safer society.

Finally, this study opens the door for similar applications of AI in other areas of health and wellness, suggesting a potential for AI tools to become more integrated into various aspects of health care delivery, provided that ethical and practical considerations in addressing misinformation and biases are taken into full account. As previously mentioned, there is a need to constantly monitor and improve AI systems to prevent the spread of misinformation and reduce biases in the information provided. Further, in the context of digital health tools, ensuring the privacy and security of user data is paramount, especially when personal health information is involved, underscoring the need for regulatory and ethical oversight in the use of AI in health care to ensure that these tools are used responsibly and for the benefit of individual, occupational, and public health.

**Strengths and Limitations**

This study has some strengths, including its novelty, methodological rigor, and reproducibility. On the other hand, it suffers from several limitations that should be properly acknowledged: future studies should investigate other AI-based tools, such as Google Bard. Not all digital assistants and chatbots have demonstrated efficacy in improving health- and sleep-related behaviors [37,38]. It should be, indeed, considered that each AI-enhanced platform, being trained on different knowledge bases, has specific technical features and capabilities, and, therefore, some AI-based tools may exhibit lower sleep-related knowledge and literacy, demonstrating less capability of correctly identifying the sleep-related statements as false. As such, this implies that monitoring of the AI system should be tool-specific.

**Conclusions**

In the present digital era, the synergy of generative conversational AI and sleep health promotion has the potential to positively impact individual, occupational, and public health by providing easy access to evidence-based information and support. This study’s findings demonstrate the potential of AI tools such as ChatGPT in enhancing public health education, particularly in debunking myths and disseminating accurate information related to sleep health. While promising, it is important to use these tools as supplements to, rather than replacements for, sleep expert opinion and to maintain strict standards of accuracy, privacy, and ethical use.
Acknowledgments

Only the text included in Multimedia Appendices 1 and 2 was generated by generative AI.

Data Availability

All data are available within this paper's text and in Multimedia Appendices 1-3.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ChatGPT full replies to sleep-related false myths in terms of true/false.
[DOCX File, 26 KB - formative_v8i1e55762_app1.docx]

Multimedia Appendix 2
ChatGPT full replies to sleep-related false myths in terms of rating on the 5-point Likert scale.
[DOCX File, 23 KB - formative_v8i1e55762_app2.docx]

Multimedia Appendix 3
Qualitative comparison between ChatGPT’s and sleep experts’ appraisals of the falseness and public health significance of sleep-related false.
[DOCX File, 15 KB - formative_v8i1e55762_app3.docx]

Multimedia Appendix 4
Verbatim transcription from the interaction with ChatGPT about the truthfulness or falseness of twenty sleep-related false myths.
[DOCX File, 34 KB - formative_v8i1e55762_app4.docx]

References


Abbreviations

- AI: artificial intelligence
- ICC: intraclass correlation coefficient

©Nicola Luigi Bragazzi, Sergio Garbarino. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Experiences, Lessons, and Challenges With Adapting REDCap for COVID-19 Laboratory Data Management in a Resource-Limited Country: Descriptive Study

Kagiso Ndlovu1, BSc, MSc, PhD; Kabelo Leonard Mauco2, BSc, MSc, PhD; Onalenna Makhura1, BSc, MSc, PhD; Robin Hu3, BA; Nkwebi Peace Motlogelwa1, BSc, MSc; Audrey Masizana1, BSc, MSc, PhD; Emily Lo3, BA; Thongotho Mphoyakgosi1, BSc; Sikhulile Moyo1, BSc, MSc, MPH, PhD

1Department of Computer Science, University of Botswana, Gaborone, Botswana
2Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden
3College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, United States
4National Health Laboratory, Ministry of Health, Gaborone, Botswana
5Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana

Corresponding Author:
Kagiso Ndlovu, BSc, MSc, PhD
Department of Computer Science
University of Botswana
Private Bag UB 0022
Gaborone, 00267
Botswana
Phone: 267 71786953
Email: ndlovuk@ub.ac.bw

Abstract

Background: The COVID-19 pandemic brought challenges requiring timely health data sharing to inform accurate decision-making at national levels. In Botswana, we adapted and integrated the Research Electronic Data Capture (REDCap) and the District Health Information System version 2 (DHIS2) platforms to support timely collection and reporting of COVID-19 cases. We focused on establishing an effective COVID-19 data flow at the national public health laboratory, being guided by the needs of healthcare professionals at the National Health Laboratory (NHL). This integration contributed to automated centralized reporting of COVID-19 results at the Ministry of Health (MOH).

Objective: This paper reports the experiences, challenges, and lessons learned while designing, adapting, and implementing the REDCap and DHIS2 platforms to support COVID-19 data management at the NHL in Botswana.

Methods: A participatory design approach was adopted to guide the design, customization, and implementation of the REDCap platform in support of COVID-19 data management at the NHL. Study participants included 29 NHL and 4 MOH personnel, and the study was conducted from March 2, 2020, to June 30, 2020. Participants’ requirements for an ideal COVID-19 data management system were established. NVivo 11 software supported thematic analysis of the challenges and resolutions identified during this study. These were categorized according to the 4 themes of infrastructure, capacity development, platform constraints, and interoperability.

Results: Overall, REDCap supported the majority of perceived technical and nontechnical requirements for an ideal COVID-19 data management system at the NHL. Although some implementation challenges were identified, each had mitigation strategies such as procurement of mobile Internet routers, engagement of senior management to resolve conflicting policies, continuous REDCap training, and the development of a third-party web application to enhance REDCap’s capabilities. Lessons learned informed next steps and further refinement of the REDCap platform.

Conclusions: Implementation of REDCap at the NHL to streamline COVID-19 data collection and integration with the DHIS2 platform was feasible despite the urgency of implementation during the pandemic. By implementing the REDCap platform at the NHL, we demonstrated the possibility of achieving a centralized reporting system of COVID-19 cases, hence enabling timely and informed decision-making at a national level. Challenges faced presented lessons learned to inform sustainable implementation of digital health innovations in Botswana and similar resource-limited countries.
Introduction

The onset of the COVID-19 pandemic resulted in a global public health crisis [1]. The pandemic stretched almost all health care systems to their limits and exposed their weaknesses [2]. Previously documented COVID-19 challenges for the health sector include a lack of the health care services needed for the pandemic, inadequate resources, limited testing ability and capacity for a COVID-19 response, as well as overall poor data management within existing health care systems [1]. It was previously projected that countries in sub-Saharan Africa could see a sharp rise in COVID-19 infection rates and deaths, as such challenges are prominent in resource-limited countries [3]. This projected disproportionate impact of COVID-19 in resource-limited countries comes as no surprise. Currently, the health sector in high-income countries is deemed as underfunded, while in most resource-limited countries, it is reportedly heavily underfunded [4]. It is not a surprise that, in 2001, African leaders through the African Union’s Abuja Declaration agreed to “allocate 15% of the state’s annual budget to the improvement of the health sector” [5]. However, in 2013, only 5 African countries had achieved this target, while in 2018, only 2 countries achieved the target [6].

Access to accurate and current information has been recognized globally as a critical requirement for timely COVID-19 pandemic responses [7]. As such, the pandemic presented the need for robust data management systems in response to the evolving nature of COVID-19 [8]. Ideally, eHealth—“cost-effective and secure use of information and communications technologies (ICT) in support of health and health-related fields, including health-care services, health surveillance, health literature, and health education, knowledge and research” [9]—could ensure reliable information reporting and timely decision-making by governments and relevant stakeholders. However, requirements for such eHealth solutions present a greater challenge for resource-limited countries with documented weak ICT infrastructure, limited maintenance budgets, a lack of health human resource capacity to utilize eHealth systems, and nonuniform unique patient identifiers [10]. According to Archer et al [11], other factors affecting successful implementation of eHealth solutions in resource-limited countries include the absence of eHealth agendas, ethical and legal considerations, common system interoperability standards, and reliable power supplies.

Similar to other nations responding to the pandemic, the government of Botswana, through the Ministry of Health (MOH), identified eHealth as a means to improve COVID-19 data management and address complex data capture and transfer processes at various port of entries and the National Health Laboratory (NHL). Botswana’s eHealth infrastructure was previously reported as generally adequate, with functioning computers and some communication systems such as telephones, email, and Internet services [12]. According to Seifio-Kgokgwe et al [12], almost all public health facilities in Botswana are now connected to the government data network (GDN) with an average Internet connectivity of about 2 Mbps. The authors further highlighted that Botswana’s eHealth initiatives have always operated within a very weak policy and regulatory framework characterized by inadequate health information legislation, national policy, and strategic plan. Fragmentation and inefficient eHealth initiatives were noted as other contributing factors to poor utilization of information in Botswana’s health sector, as well as the lack of appreciation of the important role played by health information in managing health services. Currently, over 52 health laboratories operate in Botswana, 8 of which are accredited with national certifications like the South African National Accreditation System (SANAS), Clinical Laboratory Improvement Amendments (CLIA), International Organization for Standardization (ISO) 15189, or the Southern African Development Community Accreditation Services (SADCAS) [13,14].

The Botswana NHL is one of the accredited laboratories tasked with the management of national COVID-19 testing. The NHL has decentralized its COVID-19 testing services from the capital city Gaborone, by operating satellite testing centers in selected districts with sizable populations and key ports of entry into the country. The NHL data flow is such that specimen data go through each of the 4 laboratory stages of (1) Reception Lab (all incoming lab specimens are captured using a barcode scanner), (2) Extraction Lab (specimen processing for nucleic acid isolation and purification), (3) Detection Lab (sample amplification and detection), and (4) Resulting and Verification Lab (lab results are captured and verified for release to clients). The need to scan laboratory specimen samples at each of the laboratory phases is an important quality assurance step. Accession numbers in laboratory processes are critical to link the specimen with a participant. Despite having an already existing eHealth system at the NHL, electronic data transfers within and across the 4 laboratories were considered tedious, time consuming, and a risk to both data quality and timely COVID-19 results reporting. These limitations required immediate attention.

The authors volunteered their technical support toward implementation of a customizable COVID-19 data management system—Research Electronic Data Capture (REDCap)—at the NHL. REDCap was suggested by the authors following its documented benefits including its utility within a resource-limited country context [15,16], as well as its availability locally through the University of Botswana (UB). REDCap is a secure, web-based platform designed to support electronic data capture. It was developed at Vanderbilt University in the United States in 2004 and can be set up to support a variety of health care environments and scenarios. REDCap is compliant with international standards such as the...
Methods

A participatory design approach was adopted to guide the design, customization, and implementation of REDCap for COVID-19 data management at the NHL.

Study Population, Setting, and Design

Participants

All NHL and MOH personnel responsible for processing COVID-19 specimen samples and data were invited to participate. All potential participants were sent an introductory email describing the background and objectives of the study. A consent form was subsequently shared with all those who showed interest. All invited participants agreed to participate in the study, of which 29 were NHL personnel and 4 were based at the MOH. Of the 29 NHL personnel, 12 were based at the Reception Lab, and 17 were based at the Extraction, Detection, and Resulting and Verification Labs. The study spanned from March 2, 2020, to June 30, 2020.

Data collection was conducted in 3 phases.

Phase 1: Participant Engagement

The authors facilitated a 1-day consultative physical meeting or workshop with the study participants to solicit specific requirements for the COVID-19 data management system. Study participants were requested to define requirements for an ideal COVID-19 data management system. The authors recorded all participants’ responses during the session which lasted for 1 hour. Based on the insights gathered from this exercise, iterative design and testing approaches were adopted for each key deliverable from design, customization, and implementation of the REDCap platform to support COVID-19 data management. A minimum of 2 iterations and a maximum of 4 iterations were incurred per deliverable, and this was influenced by the complexity or noncomplexity of the tasks.

Phase 2: Implementation and Assessment of the Platform to Meet User Requirements

Design and customization of the REDCap platform were followed by implementation of the solution as well as evaluating its feasibility to address the previously noted requirements and specifications. This involved participants testing the REDCap platform and providing feedback on any issues or challenges they encountered.

Phase 3: Human Resource Capacity Development

At each phase of the design, customization, and implementation of the REDCap platform, participants were trained on the various system components and had the opportunity to share their feedback to inform next steps.

Ethical Considerations

This study was approved by the Ethics Committee of the University of Botswana (Reference: UBR/RES/IRB/BIO/GRAD/244). All data experiments were performed in accordance with relevant guidelines and regulations such as the Declaration of Helsinki. All participants were informed of the objective of the exercise as well as their voluntary participation, and all gave informed consent for this study. Those who consented to participate were immediately sensitized and granted access to the REDCap instance at the UB. No compensation was provided for participating in the study.

Data Analysis

NVivo 11 software was used for thematic analysis [21] of data to determine participants’ requirements for an ideal COVID-19 data management system and their postimplementation experiences.

Results

Study participants consisted of all personnel involved with handling COVID-19 specimens and managing all relevant laboratory data at both the NHL and MOH (Table 1).
Table 1. Study participants and their roles in COVID-19 data management.

<table>
<thead>
<tr>
<th>Participant type</th>
<th>Participants, n</th>
<th>Roles in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Health Laboratory (NHL) Reception Lab</td>
<td>12</td>
<td>Scanning of COVID-19 specimen samples at point of reception</td>
</tr>
<tr>
<td>NHL Extraction Lab</td>
<td>5</td>
<td>Scanning of COVID-19 specimen samples to create maps/batch lists and for nucleic acid extraction</td>
</tr>
<tr>
<td>NHL Detection Lab</td>
<td>7</td>
<td>Scanning of COVID-19 specimen samples into the analyzers for the detection of SARS-CoV-2</td>
</tr>
<tr>
<td>NHL Resulting and Verification Lab</td>
<td>5</td>
<td>Resulting and verification of COVID-19 specimen results for access in Research Electronic Data Capture (REDCap) and producing lab reports</td>
</tr>
<tr>
<td>Health informatics personnel</td>
<td>4</td>
<td>Served as intermediaries between clinical lab personnel and IT personnel</td>
</tr>
</tbody>
</table>

Consultative meetings with study participants led to the identification of user requirements for an ideal COVID-19 data management system at the NHL. These were categorized under the following 2 themes: functional and nonfunctional requirements (Table 2).

Table 2. Thematic presentation of participants’ requirements for the COVID-19 data management system.

<table>
<thead>
<tr>
<th>Requirement type</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional requirements</td>
<td>• &quot;Authentication by unique username and password&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Automated sign out in case of user inactivity&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Data validation checks&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Temporary data staging between the two systems (REDCap(^a) and DHIS2(^b))&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Automated field data pull from the DHIS2 system&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Automated COVID-19 results sync with the DHIS2 system&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Data summary dashboard&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Barcode data export for data capture, resulting, and verification&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Automated data backups&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Data encryption during transfer and storage&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Automated alert message in case of a positive COVID-19 result&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Reordering of barcodes to be displayed on the detection plate to resemble the 8x12 matrix and preferably printed on two pages&quot;</td>
</tr>
<tr>
<td>Nonfunctional requirements</td>
<td>• &quot;High performance system&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Reliable data management system&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;User-friendly system&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Scalable&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Cross-platform independent&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Secure system&quot;</td>
</tr>
<tr>
<td></td>
<td>• &quot;Interoperable system&quot;</td>
</tr>
</tbody>
</table>

\(^a\)REDCap: Research Electronic Data Capture.

\(^b\)DHIS2: District Health Information System version 2.

Although this paper reports on challenges and resolutions while implementing REDCap at the NHL (Table 3), Figures 1-7 are intended to illustrate the design, customization, and implementation of the REDCap platform at the NHL.
<table>
<thead>
<tr>
<th>Challenges</th>
<th>Resolutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network firewall constraints to access REDCap within the government data</td>
<td>Ministry of Health (MOH) leadership engaged relevant departments to resolve firewall constraints.</td>
</tr>
<tr>
<td>network (GDN)</td>
<td></td>
</tr>
<tr>
<td>Misalignment of barcode scanner configuration</td>
<td>The barcode scanner was reconfigured to allow acceptance of the “tab” key versus “carriage return” key upon capturing a specimen.</td>
</tr>
<tr>
<td></td>
<td>Computers were set up specifically for REDCap use.</td>
</tr>
<tr>
<td></td>
<td>2 scanners were used: one dedicated for REDCap use and another for the pre-existing eHealth system.</td>
</tr>
<tr>
<td>Lack of trained NHL personnel</td>
<td>REDCap users were continuously sensitized and trained.</td>
</tr>
<tr>
<td></td>
<td>User manuals and training slides were developed and shared with the rest of the implementation team.</td>
</tr>
<tr>
<td>Slow Internet connection speed at the NHL (download and upload speeds of</td>
<td>Relevant government departments were informed of the slow internet connectivity, and attempts were made to improve the Internet bandwidth.</td>
</tr>
<tr>
<td>12.59 Mbps and 16.84 Mbps, respectively)</td>
<td>Mobile WiFi routers from private Internet service providers (ISPs) were used to assist during the training sessions.</td>
</tr>
<tr>
<td>Failure of REDCap to automatically pull data from the District Health</td>
<td>A web application programming interface (API) was developed by the authors and supported pulling data from the DHIS2 system, subsequently imported into the REDCap platform.</td>
</tr>
<tr>
<td>Information Software version 2 (DHIS2) platform</td>
<td>A web portal/application was developed by the authors and linked to the REDCap platform. The application supported generating unique barcodes for each of the COVID-19 specimens.</td>
</tr>
<tr>
<td>REDCap unable to generate the barcodes, posing a big issue at the extraction lab</td>
<td>Although not a 2-page 8x12 matrix, the third-party web application developed by the authors supported the ordering of barcodes and reformatting them into a printer-friendly format.</td>
</tr>
<tr>
<td>Reordering of barcodes to be displayed on the detection plate to resemble</td>
<td>The authors added a comment field on the REDCap result form.</td>
</tr>
<tr>
<td>the 8x12 matrix and getting those printed on 2 pages</td>
<td></td>
</tr>
<tr>
<td>Need for a comment field in the REDCap result form</td>
<td></td>
</tr>
</tbody>
</table>

The authors used participants’ feedback to guide the design of an ideal COVID-19 data flow architecture for this study (Figure 1).
Figure 1. Data flow architecture design for COVID-19 data management at the National Health Laboratory (NHL). CRF: case report form; DHIS2: District Health Information System version 2; DHMT: District Health Management Team; DIT: Department of Information Technology; DMC: Data Management Committee; MOH: Ministry of Health; REDCap: Research Electronic Data Capture; UB: University of Botswana.

The REDCap data security architecture at UB separates the web server from the database server (Figure 2).

Figure 2. Research Electronic Data Capture (REDCap) security architecture supporting COVID-19 data management at the National Health Laboratory (NHL). LDAP: Lightweight Directory Access Protocol; SMTP: Simple Mail Transfer Protocol; SSL: Secure Sockets Layer; VLAN: virtual local area network.

In order to facilitate timely decision-making and address a key technical requirement (“Automated COVID-19 results sync with the DHIS2 system” in Table 2), the REDCap system was linked with the DHIS2 system at the MOH for aggregate data reporting.

Interfacing between DHIS2 and REDCap was controlled by the DHIS2 link application programming interface (API; Figure 3).

The DHIS2 API link supported 3 functionalities: (1) data format conversions between DHIS2 and REDCap, (2) avoiding duplicate synchronizing of data between REDCap and DHIS2, and (3) enabling data access for the DHIS2 front end, which relied on the API to make consistent data available for the 2 systems.
In order to access the DHIS2 API, a RestTemplate object short code was implemented with basic authentication secured by username and password (Figure 4).

Study participants accessed REDCap using their personal account details (unique usernames and passwords) before proceeding to collect, manage, and report on COVID-19 at the MOH (Figure 5).
Of important note, the REDCap platform triggered automated email alerts whenever a COVID-19 positive result was recorded. The email alerts were sent to study coordinators at the MOH.

A third-party web application simplified the process of pulling records from the DHIS2 system into REDCap (Figure 6) and enabled the viewing of specimen barcodes through a web browser (Figure 7).

Figure 5. Specimen Receiving Lab personnel capturing data using Research Electronic Data Capture (REDCap).

Figure 6. Angular web application for pulling records from District Health Information System version 2 (DHIS2) and storage in Research Electronic Data Capture (REDCap).
Figure 7. Sample barcodes generated from the Angular web application.

Challenges and resolutions while implementing the REDCap platform to support COVID-19 data management at the NHL are summarized in Table 3.

Discussion

This paper describes the experiences, challenges, and lessons learned while designing, customizing, and implementing the REDCap platform for COVID-19 data management at the NHL in Botswana. REDCap implementation challenges and lessons identified during the study were categorized under the following 4 themes: (1) infrastructure, (2) capacity development, (3) platform constraints, and (4) interoperability.

Infrastructure

The literature emphasizes the critical need for robust ICT infrastructure to serve as the backbone for successful and sustainable digital health implementation [22-25]. The national eHealth strategy tool kit by the World Health Organization (WHO) and International Telecommunication Union (ITU) identifies essential eHealth infrastructure components as high-speed data connectivity, computing infrastructure, identification and authentication services, directory services, health care provider systems, electronic health record repositories, and health information data sets, all of which underpin a national eHealth environment [26]. The tool kit further urges countries to secure long-term funding for investment in national eHealth infrastructure and services. It is against this backdrop that the WHO recognizes a pressing need for countries to invest in infrastructure to support digital transformation; Internet connectivity; and issues related to legacy infrastructure, technology ownership, privacy, and security while adapting and implementing globally recognized standards and technologies [27].

In this study, infrastructural challenges experienced include weak Internet speed and misalignment of computer network firewalls between the GDN and the UB network. These issues resonate with the identified ICT infrastructural challenges within the Botswana National eHealth strategy [28]. During implementation of the REDCap platform at the NHL, initial Internet download and upload speeds were 12.59 Mbps and 16.84 Mbps, respectively (Table 3). The Federal Communications Commission (FCC) considers minimum Internet speeds of 5 Mbps to 25 Mbps as ideal to support online tasks such as file download and telecommuting [29]. However, despite meeting the FCC requirement, study participants reported the need for the Internet speed to be upgraded to adequately support frequent data transactions and data sharing between REDCap and the DHIS2 platform. This could be due to multiple factors. The following are some example quotes from study participants pertaining to their dissatisfaction with the Internet bandwidth:

Internet connectivity is still slow. [participant, NHL Reception Lab]

Internet bandwidth needs to be upgraded. [participant, NHL Resulting and Verification Lab]

As a mitigation for the slow Internet connectivity, relevant government departments were engaged, and measures were put in place to increase the Internet speed. This includes procurement of mobile Internet routers from private Internet service providers (ISPs) for use during the study. The engagement of private sector stakeholders to support implementation of eHealth initiatives is also highlighted as essential within the “Strategy and Investment” pillar of the Botswana National eHealth Strategy, which emphasizes “eHealth planning, with involvement of major stakeholders and sectors” [28].

Another challenge was the restriction introduced by computer network firewalls resulting in data flow constraints between the GDN and UB networks. A similar issue was encountered in another study, in which implementation of the REDCap platform incurred network firewall challenges hindering participants from using computers outside the Veterans Health Administration network to complete a survey on REDCap [30]. According to Nagpure et al [31], firewall policy conflicts can be complex to eliminate but could be addressed through practical resolution methods such as the “first-match resolution” involving “identifying which firewall policy rule involved in a conflict situation should take precedence when multiple conflicting rules (with different actions) filter a particular network packet simultaneously.” At the NHL, the authors resolved the network firewall challenges by engaging relevant ICT authorities at both governmental and UB IT departments to facilitate reconfiguration of the firewall settings to allow data flow between the 2 networks. This necessitated effective measures for data security, privacy, and confidentiality throughout the study. To achieve this, the authors ensured that the REDcap and DHIS2 servers communicate through an encrypted Internet connection using a 128-bit Secure Sockets Layer (SSL) certificate. The use of SSL has been previously considered a standard technology for securing electronic commerce and
Electronic banking transactions over the Internet [32]. However, recent advances in cybersecurity attacks call for using technologies to detect compromised SSL network traffic [33]. This consideration was brought to the attention of both governmental and UB IT departments. Compliance with the Botswana-specific Data Protection Act [34] could improve the necessary safeguards for the right to privacy of individuals and the collection and transfer of their personal data.

**Capacity Development**

Some study participants lacked the requisite skills to effectively use the REDCap platform, which was another challenge encountered during the study. The lack of competent and experienced REDCap users in some low and middle-income countries (LMICs) has been linked to poor “REDCap penetration” in those countries [35]. For example, in Botswana, only 3 institutions have REDCap instances [36], and unless affiliated with any of these institutions, most health care workers are unfamiliar with the platform. Low ICT literacy is common among health care workers in most LMICs [25] and could affect the ability of some participants to competently use the REDCap platform. Recognizing this challenge, the Botswana National eHealth strategy included “Workforce Development” as one of its pillars dedicated to eHealth capacity building among health care workers [28].

A reported benefit of having a health workforce trained in ICT prior to engaging with a health information system is minimizing avoidable errors, which, in this case, may be influenced by a lack of familiarity with the REDCap platform or poor ICT competency. To ensure the effective use of REDCap at the NHL, the authors continuously trained participants at all stages in the study. The ever-evolving nature of the REDCap platform [16,36] also necessitates that those working with the system must continuously familiarize themselves with its core functionalities and data workflows through support from the REDCap Consortium. Almost all training needs for this study were identified during weekly virtual progress report meetings with all study participants. Training content was also informed by the identified system requirements for the REDCap platform (Table 2). According to Nsaghurwe et al [37], “functional requirements describe what the system should do—such as its ability to exchange client-level data in a single repository, search for records with data quality issues; while non-functional requirements describe how the system should perform,” including system performance, reliability, and user-friendliness.

In essence, a central aspect to the successful implementation of any health information management system is the need to train users on how to appropriately utilize the system to capture and manage data [25]. Recommended approaches from the existing literature include ensuring continuous engagement and availing training materials for trainees’ access and reference at a later stage. In this study, the authors made all training material and manuals available via the REDCap “File Repository” feature for the participants to access at their convenience posttraining.

Informed by the level of participants’ familiarity with the REDCap platform, other measures were put in place such as denying participants “data deletion” privileges on the REDCap system. Instead, participants could add, edit, and view data they required to complete their respective tasks. This minimized accidental data deletion and resonated with the famous “principle of least privilege” by Saltzer and Schroeder [38], which states that “every program and every user of the system should operate using the least set of privileges necessary to complete the job.”

**Platform Constraints**

An important lesson learned was that, although REDCap is often regarded as a complete solution for data management [16], in some instances, support from other software applications could augment its limitations. In this study, the REDCap platform was not able to generate barcodes for the NHL personnel to scan to retrieve previously captured information. This limitation was resolved by developing a web-based application supported by the Angular framework [39] for quick production of scannable barcodes required during the Results and Verification Lab stage at the NHL. Some documented benefits of the Angular framework include support for lightweight web applications; faster software development capabilities; and easily readable, testable, and interoperable software solutions [39]. For this study, the authors leveraged the ability to easily link the Angular web application with the REDCap platform. Moreover, only one username and password combination was used to access the data capture forms between the Angular-based web application and REDCap.

Another constraint encountered was the inability of the REDCap platform to automatically store COVID-19 specimen barcodes on the data collection form. Further, the REDCap system requires that the barcode scanner accept a “tab key” to move to the next field, but the scanner configuration used at the NHL accepted a “carriage return key” instead to move to the next field. This meant that each time a barcode was scanned by the study participants, REDCap would not capture it into the appropriate field unless the user pressed the “tab key” on the keyboard to trigger a move to the next field. This became tedious to participants considering the high number of barcodes required scanning at the NHL during the COVID-19 pandemic. To resolve this, the authors reconfigured the barcode scanners to align with REDCap requirements, that is, use of the “tab key” to capture barcodes versus the initial configuration of the “carriage return key.” Moreover, the MOH procured dedicated barcode scanners for use with the REDCap platform as the previous ones accompanied a different system at the NHL. Overall, REDCap ensured auto-saving of the barcodes as they are being scanned, minimizing data loss in case of network or power failure.

**Interoperability**

The Healthcare Information and Management Systems Society (HIMSS) defines interoperability as “the ability of different information systems, devices, and applications (‘systems’) to access, exchange, integrate and cooperatively use data in a coordinated manner, within and across organizational, regional and national boundaries, to provide timely and seamless portability of information and optimize the health of individuals and populations globally” [40]. Some documented benefits of interoperable health care information systems include improved patient management, quality of care, and decision-making, as
well as reduced health care costs [41]. However, several challenges have been noted that hinder the interoperability of health care systems, affecting their successful and sustainable implementation, especially in resource-limited countries [10].

Despite the documented interoperability challenges, REDCap has been successfully linked to other systems previously. For example, a study in the United Kingdom automated the process by which COVID-19 clinical trial registration records were exported from the WHO International Clinical Trials Registry Platform into external software [42]. A linking script subsequently pulled relevant data, aligned it with the data dictionary in REDCap, and directly imported it, removing any duplicates in the process. Other researchers utilized REDCap as a data harmonizer, managing and combining cancer research data from multiple registries and allowing for the reconciliation of disparate data sets as well as their conversion [43,44]. In another previous study, REDCap served as the platform housing all global data collected with modified survey questions to capture nuances and allowed for individualization to study public health interventions for COVID-19 [45]. Consequently, these applications reinforced REDCap’s utility for cross-continental collaborations, extending beyond any singular institution or setting.

In this study, although linking the REDCap platform with the DHIS2 platform was achieved, the failure of the REDCap system to automatically pull data from the DHIS2 platform was noteworthy. This limitation was a result of constraints within the Dynamic Data Pull (DDP) feature in REDCap. The REDCap DDP is a special feature for importing data into REDCap from an external source system [46]. The DDP feature provides an adjudication process whereby REDCap users can approve all incoming data from the source system before the data are officially saved in their REDCap project. Because the REDCap DDP assumes that all incoming data from the source system may not be trusted as valid or that only a subset of the data coming from the source system needs to be imported, it utilizes an adjudication web page inside REDCap’s interface to allow manual review of the data obtained from the source system before confirming that it be imported into the REDCap repository [46]. It is precisely the adjudication process that hindered the automatic data import from the DHIS2 system into REDCap. To address this challenge, a custom API was developed by the authors to support accessing health records from the DHIS2 platform, staging them for quality assurance checks when interacting with the system.

The datasets used and/or analyzed during the study are available from the corresponding author on reasonable request. Further, this approach helped minimize network traffic whenever data synchronization between REDCap and the DHIS2 systems occurred.

Limitations

The REDCap system was implemented at the NHL in response to the emergency COVID-19 pandemic. Consequently, a systematic approach to its implementation could have been compromised, and essential considerations could have been overlooked. Most notably, multiple training and feedback sessions had to be conducted virtually using the Zoom platform due to restrictions on in-person gatherings. Hence, the authors were unable to frequently visit the NHL in person to provide the necessary technical support and training. Participation in the study was limited and often disrupted, as some NHL personnel were reassigned to efforts such as vaccine distribution or other scenarios not applicable for REDCap use. Last, increased workloads for participants due to COVID-19 surges in Botswana likely contributed to the lack of thorough quality checks when interacting with the system.

Conclusion

Implementation of the REDCap platform to support COVID-19 data management at the NHL in Botswana was successful, albeit with challenges. It is worth noting that, like any software, REDCap as a system possesses limitations that were addressed with other applications to meet requirements for this study. Most challenges encountered were exacerbated by the lack of pandemic response preparedness, as was the case in Botswana and around the world. As such, most of these challenges will not be unique to this specific case of REDCap implementation but will continue to affect the sustainable implementation of eHealth innovations until conscious efforts using key digital health strategies are made that create an enabling environment to support implementation of digital health innovations. Another lesson learned is the essential need for collaborations with key stakeholders to minimize technological barriers while implementing eHealth solutions. Despite the challenges encountered, the REDCap and DHIS2 platforms served as readily available and customizable platforms to address COVID-19 data management at the NHL. To this end, effective planning is essential, including the engagement and training of key personnel to optimize the use of eHealth systems beyond the COVID-19 pandemic.

Acknowledgments

The authors express their great appreciation to the National Health Laboratory (NHL) and Ministry of Health (MOH) management for supporting the implementation of Research Electronic Data Capture (REDCap) at their facilities as well as the supportive study participants for their valued contributions. No generative artificial intelligence (AI) such as ChatGPT was used in any portion of this manuscript.

Data Availability

The datasets used and/or analyzed during the study are available from the corresponding author on reasonable request.
Authors' Contributions

All authors jointly conceived the study, and jointly contributed to its design. KN, OM, AM, and NPM contributed to the customization of the Research Electronic Data Capture (REDCap) platform for use at the National Health Laboratory (NHL) and provided REDCap training sessions to the study participants. KN, KLM, and OM completed the initial data analysis and wrote the first draft of the manuscript. SM, TM, RH, and EL provided substantial editorial and intellectual input and contributed to subsequent revisions. All authors approved the final manuscript.

Conflicts of Interest

TM is an employee of the National Health Laboratory (NHL). KN, AM, OM, and NPM form part of the REDCap technical team in Botswana. The remaining authors have no competing interests.

References


9. eHealth. World Health Organization. URL: https://www.emro.who.int/health-topics/ehealth/ [accessed 2024-03-26]


39. Angular. URL: https://angular.io/ [accessed 2024-03-26]


Abbreviations

API: application programming interface  
CLIA: Clinical Laboratory Improvement Amendments  
DDP: Dynamic Data Pull  
DHIS2: District Health Information System version 2  
FCC: Federal Communications Commission  
GDN: government data network  
GDPR: General Data Protection Regulation  
HIMSS: Healthcare Information and Management Systems Society  
HIPAA: Health Insurance Portability and Accountability Act  
HISP: Health Information System Programme  
ICT: information and communication technology  
ISO: International Organization for Standardization  
ISP: private internet service provider  
ITU: International Telecommunication Union  
LMICs: low and middle-income countries  
MOH: Ministry of Health  
NHL: National Health Laboratory  
REDCap: Research Electronic Data Capture  
SADCAS: Southern African Development Community Accreditation Services  
SANAS: South African National Accreditation System  
SSA: sub-Saharan Africa  
SSL: Secure Sockets Layer  
UB: University of Botswana  
WHO: World Health Organization

©Kagiso Ndlovu, Kabelo Leonard Mauco, Onalenna Makhura, Robin Hu, Nkwebi Peace Motlogelwa, Audrey Masizana, Emily Lo, Thongbotho Mphoyakgosi, Sikhulile Moyo. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,
provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Landscape of Digital Technologies Used in the National Health Service in England: Content Analysis

Jake Alan Allcock¹, MA; Mengdie Zhuang², PhD; Shuyang Li³, PhD; Xin Zhao², PhD

¹Department of Sociological Studies, University of Sheffield, Sheffield, United Kingdom
²Information School, University of Sheffield, Sheffield, United Kingdom
³Business School, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Mengdie Zhuang, PhD
Information School
University of Sheffield
The Wave
2 Whitham Road
Sheffield, S10 2AH
United Kingdom
Phone: 44 114 222 6339
Email: m.zhuang@sheffield.ac.uk

Abstract

Background: In England, digital technologies are exploited to transform the way health and social care is provided and encompass a wide range of hardware devices and software that are used in all aspects of health care. However, little is known about the extent to which health care providers differ in digital health technology capabilities and how this relates to geographical and regional differences in health care capacities and resources.

Objective: This paper aims to identify the set of digital technologies that have been deployed by the National Health Services clinical commissioning groups (NHS CCGs) in England. In doing this, we respond to calls to shed light on the internal dynamics and variation in the form of digital capability in England in terms of health service regional differences and health diversity, equity, and inclusion.

Methods: We collected 135 annual reports that belong to 106 NHS CCGs in England, comprising more than 18,000 pages in total, released from 2020 to 2021. Using this data set, we identified 2163 pages related to digital technologies and labeled them using content analysis. We follow the construct taxonomy used by digital options theory, a theory from the management information systems field analyzing organizational resource investment choices, in classifying observed technologies according to digital themes—inherent design patterns that we identified and explained. We then used a hierarchical clustering method to extract groups of NHS CCGs that implement similar technology themes.

Results: We found 31 technologies from the reports and grouped them into 9 digital themes. The 9 themes were further assigned to 1 of the 3 constructs of digital options theory, the identification of patients’ requirements (we identified information portals [76/106], digital health engagement [67/106], and digital inclusion support [45/106]), the development of new work patterns (we identified telehealth [87/106], teledermatology [35/106], and care home technologies [40/106]), the realization of improvements in efficiency and public accessibility (we identified online booking [26/106], online triage [104/106], and digital mental health services [74/106]). The 3 clusters of CCGs are identified based on the 8 themes (Hopkins=0.9914, silhouette=0.186), namely (1) digitally disengaged, (2) digitally engaged, and (3) digital torchbearer.

Conclusions: Our findings show prominent digital themes within each construct group, namely information portals, telehealth, and online triage, covering people’s fundamental health information needs. Almost half of CCGs fell into the digitally disengaged group, and all London CCGs (5/106) belonged to this group. We propose that practitioners should offer specialized assistance to regions with limited digital engagement, emphasizing digital health literacy, inclusion support, and ongoing evaluation, rather than concentrating solely on technical advancements.

(JMIR Form Res 2024;8:e51859) doi:10.2196/51859
KEYWORDS
digital health; healthcare service; regional difference; National Health Service; NHS; digital technology; health equity

Introduction

The health care services in England are in a transformational phase due to the increasing pressure to propose and review digital strategies, use continually emerging new digital technologies, and place varied emphasis on digitization according to regional needs [1,2]. Digital technologies are exploited to transform the way health and social care is provided and encompass a wide range of hardware devices and software that are used in all aspects of health care. These technologies are often grouped and examined by their use cases [3,4], such as medical consultations and treatment, patient management, and information campaigns for access to care. Some researchers emphasize the change these technologies may bring (also referred to as “innovation”) and therefore propose a different type of taxonomy. Zweifel [5] categorized technologies serving the purpose of health care innovation into 3 types, including technologies for product innovation, process innovation, and organizational innovation.

Despite these developments, however, health care practitioners and managers often find it difficult to choose and leverage different forms of digital technologies and innovation in their institutions to overcome challenges or inefficiencies [6]. Furthermore, little is known about the extent to which health care providers differ in technology capabilities regarding improving public access to health resources, and how this relates to geographical and regional differences in health care capacities and resources. In the United Kingdom, the availability and usage of medical treatments (measured by, eg, travel distance to attend treatment, waiting time to receive a treatment, and funding for a primary-care practice) vary across regional health services in England [7-10]. We suspect this variation may also apply to the availability of digital technologies as a result of a local strategy priority, prior digitalization level in the area, and the demographics of the local populations.

Drawing from this emerging stream of research, this study identified the digital technologies adopted in 106 National Health Service clinical commissioning groups (NHS CCGs) in England, using 135 annual reports for the years 2020-2021. The CCGs were clinically led statutory NHS bodies responsible for the planning and commissioning of health care services for their local area, which were dissolved in 2022 to be replaced by the new integrated care boards (ICBs). The CCG annual reports were selected for 2 main reasons. First, the CCGs’ structure (106 CCGs in England as of April 2021) provided a finer spatial granularity than the ICBs’ structure (42 ICBs in England as of July 2022). Therefore, they offered crucial insights and statistics for all involved NHS services across different regions, which could offer valuable information regarding differences in digital capabilities and diversity, equity, and inclusion issues across various fine-grained geographical sites. Second, the years 2020-2021 marked an important step in health care digital transformation as the CCGs afterward focused on structural change and were replaced by ICBs nationally. In addition, in response to the COVID-19 pandemic, CCGs were incentivized to make services accessible through digital platforms, aiming to ensure the continual provision of care while mitigating transmission risks. Therefore, the reports in 2020-2021 highlighted important aspects of digital technologies and themes adopted by the CCGs before the structural shift and in response to the COVID-19 pandemic, which were fed into the newly formed ICBs.

We aim to answer the following research questions: What are the digital technologies adopted by CCGs to improve their services as well as widen public access and engagement? How do they vary across regions?

We use the construct taxonomy used by digital options theory [11] to understand the different types of technologies that are used in health care institutions, and how these options support health care providers’ ability to translate their resources into performance. In the management information systems literature, digital options represent an organization’s investment in and adoption of information technologies [12]. Such adoption, as Sambamurthy and colleagues argued [11], together with changes to the organization’s technological environment, impacts an organization’s information technology capabilities. Therefore, digital options are primarily used to examine the evaluation of information technologies [12,13] and to assess how the technologies and digital capabilities in an organization can be transformed into performance improvement [11,14].

Following this set of ideas, in the context of health care, digital options represent opportunities to use new technical tools and features that will increase the service quality, efficiency, and public accessibility. Previous work involving digital options theory and health care focuses on performance improvement, for example, how cost-effective information technology solutions can enhance the financial performance of resource-constrained hospitals [15]. Little effort has been invested to map out the existing set of digital technologies in health care using the taxonomy of digital options.

Drawing on the concepts of digital options theory [16,17], the technologies serving the purpose of health care efficiency can be grouped using three constructs: (1) identifying patients’ requirements that involve recognizing new technical features for the digital health service (such as online community engagement and digital champion events); (2) developing new work patterns that improve internal coordination and working process (such as artificial intelligence [AI] diagnostics and digital prescribing); and (3) realizing improvements in efficiency and public accessibility (such as virtual consultation and direct online booking). We use these constructs as the theoretical lens to categorize the digital tools used in various CCGs in NHS England.

This paper aims to identify the set of digital technologies that have been deployed by the NHS CCGs in England. This study has several contributions. First, it contributes to the health care literature by providing insights into how regional differences in technology, under an NHS system in England, can vary in
their capabilities and efforts in implementing digital technologies. We empirically draw the landscape of digital transformation in health care by unpacking the role of NHS CCGs in adopting technologies and by presenting the regional differences in England. In doing this, we respond to calls to shed light on the internal dynamics and variation in the form of digital capability and patient engagement in England in terms of health service regional differences and health diversity, equity, and inclusion [18,19]. Second, we extend digital options theory to understand its application of constructs within a health care setting. We applied and examined the 3 key approaches, proposed by Rolland et al [12], for technological options to engage patients as technology users and to improve health care service quality. The study demonstrates the complex choices faced by health care providers: while they need to address large numbers of patient queries and appointments through actionable digital options, they are constrained by regional resources and digital capabilities. Similarly, while different types of digital options could offer health care providers with new opportunities to understand patients' needs and coordinate workflows, the actual adoption of options varies significantly across different regions.

**Methods**

**Overview**

In this study, we followed a text-mining approach and content analysis method to analyze the research data. Specifically, we used a text mining approach to extract key paragraphs and texts from the 135 NHS CCG annual reports released for 2020-2021 identifying digital technologies. We further followed content analysis methods to manually code and categorize the extracted paragraphs and texts through the lens of the digital options theory [16,17]. We then clustered the CCGs based on the corresponding technologies to reveal the similarities between CCGs and regional differences in digital technologies’ availability.

**Data Selection**

Secondary data sources, namely annual reports of each NHS CCG, were used in this paper. We decided to use annual reports as they cover each CCG’s performance and accountability in the period, ranging from performance analysis, progress on key initiatives, and public and patient involvement to actual spending, and are the most recent reports before the start of the ICB-forming stage. The reports were downloaded from individual CCG websites directly. In total, there were 106 CCGs across England as of April 1, 2021, which was reduced from 135 CCGs in 2020 with the merger of 38 CCGs into 9. This paper investigates the 135 annual reports, 18,667 pages in total, released from 2020 to 2021. We further group the results according to the 106-CCG structure as it reflected the most recent structure of the CCG systems for 2021-2022.

**Data Preprocessing**

Given that each report contains 44-222 pages and is labor-consuming to go through manually, highlighting the relevant texts visually can support the researchers in locating useful information for content analysis more quickly. The reports were preprocessed using the semantic matching method [20], a text mining technique, to identify and highlight texts that are relevant to author-selected keywords. The processing was implemented in Python (Python Software Foundation).

First, reports from 10 CCGs were manually screened by the first 3 researchers independently, to gather the initial set of keywords. The selected 10 CCGs covered the main geographic regions in England. The whole research team had regular meetings during and after screening to discuss the expansion of or trimming the keyword set. Frequent and relevant keywords were excluded from the set if they might cover a much larger field, such as “digital tool.” To validate the set, the second author applied semantic matching methods using the proposed keyword set on the 10 reports and manually screened the reports again to make sure the relevant contents were highlighted appropriately. The final set contained the following words: “digital,” “technology,” “AI,” “Machine Learning,” “e-,” and “online,” together with their derivations. Using this method, 2163 pages containing selected keywords were highlighted in the 135 reports.

**Data Analysis**

**Content Analysis**

After the preprocessing, we analyzed the extracted 2163 pages through content analysis. Content analysis [21] was conducted by the first author. According to digital options theory, there are three types of options when engaging with technologies, including: (1) identifying patients’ requirements that involve recognizing new technical features for the digital health service (eg, digital survey to gather patient feedback on services), (2) developing new work patterns that improve internal coordination and working process (eg, AI diagnostic tools), and (3) realizing improvements in efficiency and public accessibility (eg, appointment booking system). We first conducted the initial coding by going through the research data and grouping them into categories according to the purpose of the adopted technologies. Second, we went through the categories in detail and mapped them to the 3 options as themes. Regarding reliability, the research team had weekly meetings during and after screening to discuss the expansion of or trimming the keyword set. Frequent and relevant keywords were excluded from the set if they might cover a much larger field, such as “digital tool.” To validate the set, the second author applied semantic matching methods using the proposed keyword set on the 10 reports and manually screened the reports again to make sure the relevant contents were highlighted appropriately. The final set contained the following words: “digital,” “technology,” “AI,” “Machine Learning,” “e-,” and “online,” together with their derivations. Using this method, 2163 pages containing selected keywords were highlighted in the 135 reports.
Table 1. Digital technologies used in NHS\textsuperscript{a} CCG\textsuperscript{b} and associated themes.

<table>
<thead>
<tr>
<th>Digital options theory construct and theme</th>
<th>Example of technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifying patients’ requirements that involve recognizing new technical features for the digital health service</td>
<td></td>
</tr>
<tr>
<td>Information portal</td>
<td>• Webinar, health care promotion campaigns (eg, through social media), service direction (eg, via a web portal collating quick access to available digital health services), advice and guidance (eg, for GP\textsuperscript{c} to use to contact health specialists for advice), patient access to individual health care records, and fast information sharing portal (eg, eHealthscope, the development of which was led by Dr Michael O’Neil at the Saxon Cross Surgery, in partnership with local practices and the data management team of the 6 Nottingham CCGs)</td>
</tr>
<tr>
<td>Digital health engagement</td>
<td>• Digital champions (eg, to improve digital literacy), podcasts, and online community engagement about health care (eg, online meetings and surveys)</td>
</tr>
<tr>
<td>Digital inclusion support</td>
<td>• Technology for the digitally excluded (eg, providing devices for care homes) and alternative communication methods (eg, text messages)</td>
</tr>
<tr>
<td>Developing new work patterns that improve internal coordination and working process</td>
<td></td>
</tr>
<tr>
<td>Telehealth</td>
<td>• Virtual assistants (eg, Alexa; Amazon Inc \cite{22}), virtual wards, remote monitoring, and self-management software</td>
</tr>
<tr>
<td>Telemedicine</td>
<td>• AI\textsuperscript{d} diagnostics, teledermatology, digital prescribing (eg, OptimiseRx \cite{23}), and applications using VR\textsuperscript{e} headsets</td>
</tr>
<tr>
<td>Care home technologies</td>
<td>• Digital technologies designed specifically for care homes and digital training tools for using these technologies for care home residents or staff</td>
</tr>
<tr>
<td>Realizing improvements in efficiency and public accessibility</td>
<td></td>
</tr>
<tr>
<td>Booking system</td>
<td>• Indirect online booking (eg, the patient will need to submit a request form first and wait for an appointment)</td>
</tr>
<tr>
<td></td>
<td>• Direct online booking (eg, the patient can directly check available appointments and book one online)</td>
</tr>
<tr>
<td>Online triage</td>
<td>• Live online consultation and eConsult (eg, offline consultation system using a patient-filled form)</td>
</tr>
<tr>
<td>Digital mental health services</td>
<td>• IAPT\textsuperscript{f} (it is undergoing a national rebranding and will be called the NHS Talking Therapies for Anxiety and Depression), region-specific mental health service (eg, Mind-BLMK\textsuperscript{g} \cite{24} for people in Bedfordshire, Luton, and Milton Keynes, With Me In Mind \cite{25} for people in Doncaster, Rotherham, and North Lincolnshire), and independent mental health care (eg, Kooth; Kooth PLC)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}NHS: National Health Service.  
\textsuperscript{b}CCG: clinical commissioning group.  
\textsuperscript{c}GP: general practitioner.  
\textsuperscript{d}AI: artificial intelligence.  
\textsuperscript{e}VR: virtual reality.  
\textsuperscript{f}IAPT: Improving Access to Psychological Therapies.  
\textsuperscript{g}BLMK: Bedfordshire, Luton, and Milton Keynes.

\textbf{Cluster Analysis}

Hierarchical clustering \cite{26} with Euclidean distance was used to extract common patterns from the code; by doing so the CCGs implementing similar types of technologies were grouped together. We used the identified 9 digital technology themes from the previous step to cluster the CCGs, in which the number of technology types mentioned for each theme was assigned to each CCG. For example, the Sheffield CCG reported they only had 1 technology (remote monitoring) under the telehealth theme, therefore the value of telehealth for Sheffield CCG was 1. In total, for each CCG, there were 9 values describing how engaged the CCG was with each technology theme. These values were further standardized using the \textit{z} score to achieve balanced similarity weights across all 9 themes. The silhouette and elbow method were used to select the optimal numbers for clusters \cite{27}, resulting in a structure with 3 clusters.

\textbf{Ethical Considerations}

This study is based on the secondary analysis of NHS CCG annual reports, which are public information. The ethics approval for the secondary analysis of all data presented in this study was obtained from the University of Sheffield Research Ethics Committee (045790).
Results

Digital Technologies in NHS CCGs

Overview

In total, 18,667 pages from 135 annual reports were preprocessed and 2163 pages containing selected keywords were highlighted in the 135 reports, representing 106 CCGs. Further, 31 types of technologies were then extracted, covering a wide range of tools and purposes. These technologies were then grouped into 9 themes. The full codebook can be found in Multimedia Appendix 1.

The purposes of the digital technologies are typically associated with the patients’ or health care practitioners’ needs across various use cases and regions. Generalizing across all the observed items, these technologies were first grouped under particle themes, specific to either a location (eg, care home), disease (eg, mental health), or a type of service (eg, booking appointment). We then further assigned the identified 9 themes to 1 of the 3 digital options theory constructs. Table 1 describes the digital technologies observed from the NHS CCG annual reports and their associated themes.

Information Portals

The information portal theme consisted of 6 technologies that all involved access to or communication specific health care information via an online format. In total, 72% (76/106) of CCGs contained a mention of this theme via either 1 or more of the 6 technologies. Health care promotion for raising the public’s awareness of available services (35/106, 33%) was the most popular method, followed by online service catalogues (30/106, 28.3%) and webinars designed to briefly or formally announce digital tools (30/106, 28.3%). Providing patients with their own health care records accounted for 6% (7/106). These 4 above all involve having health care information readily available online for patients to locate and use. Other tools were typically more oriented toward the sharing of health care information between health care professionals or to the patients. For example, advice and guidance (16/106, 14.8%) was a tool that allowed general practitioners (GPs) to contact specialists quickly to obtain information for a patient so they did not have to refer the patient to the specialist. In doing so they saved time and could provide appropriate treatment for the patient on the same day. The health practitioners could also share electronic patient records quickly with relevant members through an online system (47/106, 43.5%) to facilitate discussion.

Digital Health Engagement

Digital health engagement was about engaging with the public through digital means (67/106, 63.2%). Differing from the information portal theme aiming to support patients or health practitioners dealing with immediate health needs, this theme focuses on obtaining feedback on service delivery and raising awareness of a disease or healthy lifestyle in the community in a less formal manner. The most popular technologies within this category involved patients in the operations of the health care service within their area through virtual meetings or surveys (63/106, 59.4%). These activities kept patients informed and created and allowed them to give their input on the operations of the health care service. Around 12% (13/106) of the reports discussed the development of a podcast that patients could access anytime that kept them up to date with any health care developments or gave them advice for better self-care management. Around 17.9% (19/106) of reports mentioned the use of digital champions, who are individuals who help staff and patients struggling with the integration of digital technologies and help them develop their digital skills and confidence.

Digital Inclusion Support

The implementation of many digital technologies was vital to keeping the NHS operating during the COVID-19 pandemic, and there was a conscious effort to keep providing adequate health care to those without a means of accessing these services. In total, 42.5% (45/106) of CCGs included recognition of the problem of digital exclusion and how they would tackle it via an alternative digital approach. The Newcastle Gateshead CCG report reflected on the downside of the increase in digital technology, suggesting that it could impact the NHS’ free at point of care policy (eg, using a phone or the internet is not free). Methods such as providing targeted alternative communication methods (eg, text message or physical newsletters; 14/106, 13.2%) or providing essential technology (typically providing hardware to care homes; 38/106, 35.5%) were mentioned.

Telehealth

Telehealth is a theme that aims to incorporate digital technologies to monitor patients’ health information through real time data and provide long-distance health care. At least one form of telehealth was mentioned by 82% (87/106) of the CCGs and was seen by many of them as vital to providing safe health care during the COVID-19 pandemic. Remote monitoring (69/106, 65.1%) typically involved the use of software or hardware technologies such as pulse oximetry and digital blood pressure monitors. The information from these devices would automatically be sent to the patients’ health care record, which allowed the GP to continuously check the patient’s vitals. This allowed the GP to take quick action as needed as they would be notified if any major problems arose. Further, 26.4% (28/106) of the reports mentioned that telehealth tools also helped patients with self-management, particularly for patients with chronic conditions. In addition, this theme involved some innovation and creative use of technology by some CCGs. For example, 6.5% (7/106) of the reports discussed providing Alexa (Amazon Inc) devices to patients who were not able to use traditional computer devices (due to conditions such as vision impairments). The speech recognition software would allow the patient to keep in contact with their GPs, where they could also be monitored remotely. We note that some reports mentioned the use of telehealth tools in general terms rather than naming the specific technology implemented (such as the name or provider of the tool).

Telemedicine

Telemedicine was one of the least mentioned themes among the CCG groups, with it only being mentioned by 33% (35/106) of CCGs. Telemedicine differs from telehealth in that it seeks...
to provide treatment or make a diagnosis for patients using digital technologies. The more innovative approach to digital technologies would lie within this category. This includes health care practices using virtual reality headsets (2/106, 1.9%) to treat patients for their mental health by having them experience scenarios in a digital space. AI diagnostics, mentioned by 9% (10/106) of reports, makes predictions about a patient’s health (eg, heart-related issues or developing cancer), allowing the health care provider to advise a patient so they may mitigate any future health care concerns. We also found that 10.3% (11/106) of the reports mentioned technologies for online prescribing, often for a repeat prescription. Teledermatology, mentioned by 7.5% of the reports (8/106), refers to the use of static digital images to triage, diagnose, monitor, or assess skin conditions without the patient physically meeting the dermatologist. This technology required the users to submit a clear photo of their skin condition before the appointment (either in the form of online meetings or chat). This dependency on trust in technology (so that the users submit personal information through the tool and believe in the results) as well as effort in the form of hardware (such as a virtual reality headset, laptop, or mobile camera for taking pictures and making video calls) from the users, is consistent for all tools in this theme.

**Care Home Technologies**

We refer to care home technologies as digital technologies designed specifically for the care home or training tools for care home residents and staff to use digital technologies. The technologies under this theme contained accessibility considerations particularly designed to support people in the later stage of life and living in care homes; therefore we separated it as a single theme. Care home technologies were only mentioned by about 37.7% (40/106) of the reports. By 2021, it was predicted that more than 400,000 people would live in care homes in the United Kingdom and would likely need close monitoring and more delicate health care due to age or health conditions than older people staying at home. In total, 35.8% (38/106) of the CCGs mentioned the involvement of training staff within these settings so they could adequately use the new digital technologies to care for their residents.

**Booking System**

In total, 24.5% (26/106) of CCGs had started to adopt various forms of online booking systems, especially within GP practices. This is the alternative to the traditional method of a patient telephoning a GP asking for an appointment through these means. The reasons cited for the uptake of this theme by some CCGs can be attributed to three factors: (1) the integration of NHS 111 (a free-to-call single nonemergency number medical helpline) services with GP and emergency departments (EDs), (2) the desire to ease the burden on telephone lines, and (3) the development of the NHS app. Further, 2 types of technologies, namely indirect booking and direct booking, were found within this theme. Indirect online booking was mentioned by 13.2% (14/106) of the CCGs, and this was the process of NHS 111 having the ability to book patients directly into GP or ED appointments. Multiple reports mentioned that the sharing of data between services (such as GPs, NHS 111, and ED) allowed NHS 111 to filter patients through their lines and book patients into appointments that they urgently needed. For example, the Manchester CCG used the Adastra Digital solution to send information between EDs and NHS 111. This improved patient flow and eased the burden on GPs and EDs. Direct online booking is when the patient themselves can book, manage, and also cancel their own appointment online. This was mentioned by 14% (15/106) of the CCGs and many of them cited the integration of the NHS app as a convenient means of allowing patients access to book appointments themselves.

**Online Triage**

Virtually every CCG (104/106, 98.1%) mentioned some form of online triage that was integrated into their effort to adopt digital technologies. This mostly came from the reports covering the use of virtual consultations by their health care professionals. Many reports discussed how the COVID-19 pandemic forced them to accelerate their plans to integrate digital solutions into their health care plans, and virtual consultations became a must for many health care settings so they could continue operating safely (both for the practitioner and for the patients). Typically, this involved practitioners contacting patients via telephone or some form of video consultation services. The video consultation would usually be operated by private health care platforms such as Doctorlink (HealthHero), askmyGP (Evergreen Health Solutions Ltd), or Attend Anywhere (Induction Healthcare Group PLC), in which case the private companies created a process to help practitioners carry out this health care service (either through a telephone call, a smartphone app, or a web app). eConsult is a form of online triage mentioned by 19.6% (21/106) of CCGs and differs from the virtual consultation format as it involves patients filling out an online form and sending it to their health care practice, where it is reviewed and next steps for treatment are provided by a health care professional without directly speaking to the patient. This service was cited as being useful for prescribing repeat prescriptions and removing pressure on phone lines.

**Digital Mental Health Services**

Digital mental health services were popular among CCGs, being mentioned by 70.3% (74/106) of them. We decided to put these technologies into a separate theme due to the increased mental health needs during the pandemic and the innovations used to carry out traditional treatments (such as psychotherapy or counseling). Some mental health treatments involved using digital tools, such as video calls, to offer counseling services before the pandemic. These mental health services often incorporated other digital technologies such as information portals (eg, Instagram accounts) to share advice and tips for self-management. Around 38.7% (41/106) of CCGs developed their own region-specific digital mental health care services that they were able to refer patients to. For example, Mind-BLMK is available for people in Bedfordshire and Luton as well as Milton Keynes, and With Me in Mind is available for people in Doncaster, Rotherham, and North Lincolnshire. However, around 33.9% (36/106) of CCGs discussed how they had integrated private mental health services that they had partnered with to refer their patients to. Kooth was one of the more popular services as it focused on treatment for children and younger people, although other services such as Qwell specialized...
in treatment for adults. Both sites are similar in layout, and they tailor their services to the targeted user group. For example, Kooth offers the opportunity for young people to engage in mini activities to manage their mental health, while Qwell offers more traditional long-form articles that explore different mental health issues and solutions. Both sites also offer online forums to talk to other patients of a similar age using the services, potentially sharing experiences and offering peer support to build a community between each other. Both sites offer access to therapists who specialize in the targeted age groups, as well.

Clusters

Clustering was used to identify groups of CCGs implementing similar types of technologies. We first calculated Hopkins statistics for the data. The Hopkins statistic was 0.9914 (>0.5), indicating the data were highly clustered. A structure with 3 clusters was the optimal cluster structure (average silhouette width of 0.186, SD 0.163), determined using the silhouette and elbow method [27].

Resulting from hierarchical clustering, the first cluster group contained 51 CCGs, the second cluster contained 35, and the third contained 20. The CCGs differed between clusters in their digital themes. Figure 1 helps establish the differences between each group by showing the weighted proportions of technologies in each theme aggregated across CCGs in each group, which we refer to as “scores” in the following text and Figure 1. There were 9 scores assigned to each CCG, each corresponding to a digital theme. Group 1 (orange) will be named “disengaged” (of digital technologies) due to low scores in all of the digital themes within the CCG reports within this cluster (8 out of 9 themes had scores less than 25%). Cluster 2 (blue) will be named “engaged” (with digital technologies) due to the CCG reports within this cluster having a general interest in many of the digital themes (6 out of 9 themes had scores between 25% and 50%). This is especially true for telemedicine, where it excels compared to the other two clusters; however, there is a low score for the digital engagement and telemedicine themes. Group 3 (green) did not have this problem as the CCG reports within this cluster had a high score for technologies in the digital engagement category, as well as a very high score for those in the information portal category. Therefore, this group will be named “torchbearer” (of digital technologies).

Table 2 presents the number of CCGs in each cluster group represented by the region they belong to according to the UK Office for National Statistics. For example, in the southeast region, there were 5 CCGs that belonged to the digitally disengaged cluster and 6 CCGs that belonged to the digitally engaged cluster. Figure 2 displays the geographical distribution of cluster groups in England. The graph does not present a particular spatial pattern, but there are some notable observations. For one, all of the London CCGs fell into the digitally disengaged group. The northwest and Midlands regions have a proportionally very low presence in the digitally engaged cluster group. This may be due to them having a very high presence in the digital torchbearer group.

Figure 1. Three clusters and their scores for each digital technology theme.
Table 2. Regional distribution of cluster groups (n=106) (number of boroughs in each region).

<table>
<thead>
<tr>
<th>Cluster groups</th>
<th>London, n</th>
<th>Southeast, n</th>
<th>Southwest, n</th>
<th>East of England, n</th>
<th>Midlands, n</th>
<th>Northeast and Yorkshire, n</th>
<th>Northwest, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disengaged</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Engaged</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Torchbearer</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 2. The distribution of cluster groups in England.

Discussion

Principal Results

The pandemic has disrupted health care support services in the United Kingdom, resulting in the rapid adoption of digital technologies [29]. Nevertheless, the digital technology themes guiding the mass adoption of technology and the potential regional disparities remain to be explored. Based on 135 NHS CCG annual reports, this paper identified digital technology themes and associated technologies adopted by the NHS nationally in 2020 and 2021 and examined their regional differences from a digital divide perspective [30].

Informed by digital options theory [16,17], we identified 9 digital technology themes, which were categorized into 3 main groups, namely, the identification of patients’ requirements, the development of new work patterns that improve internal coordination and working process, and realizing improvements in efficiency and public accessibility. First, the identification of patients’ requirements includes the use of information portals (eg, retrieving and sharing health-related information), digital means (eg, using podcasts for promoting health-related information), and digital support (eg, providing hardware or alternative access for the digitally excluded) to achieve effective communication channels between health care providers and patients. Second, the development of new work patterns that improve internal coordination and working process includes digitizing existing health care services, such as remote monitoring of patient health (ie, Alexa), e-prescriptions (eg, AI diagnostics and online prescriptions), and digital training (eg, staff training on using digital tools in care homes). Finally, the realization of improvements in efficiency and public accessibility results in the use of digital methods to widen public access to health care, including creating online booking systems shared by health care providers (eg, NHS 111 and NHS apps), online triage (eg, online consultation), and online mental health services (eg, Kooth). Our data suggest that there are prominent digital themes within each group. Information portals are mostly adopted by CCGs to achieve effective health care communication among patients and health care providers. Telehealth is primarily adopted by CCGs for digitizing health care services, particularly to monitor patient health remotely. Online triage has been widely implemented by CCGs to provide patients with access to health care during the pandemic. Therefore, these top 3 themes, which cover people’s fundamental health needs, could serve as a starting point for future CCGs and other health care providers when adopting and implementing digital solutions.

In addition, our findings contribute to the new research theme of “digital health citizenship” by highlighting that digital tools
and technologies extend beyond operation efficiency to wider patient engagement, and therefore reshape social relations and interactions among patients as health service users [2]. Regional differences in such social relations and interactions could be linked to the equality and inclusion issues in health service provision. Based on the 9 identified digital themes, 3 main clusters were identified: the digitally engaged, the digitally disengaged, and the digital torchbearer. It is concerning that almost half of the CCGs fell into the digitally disengaged group, showing a low uptake of the aforementioned digital themes. Interestingly, most digitally disengaged CCGs belonged to London areas. This seems to be aligned with the data released by the Office for National Statistics [31] in 2019, suggesting that London has the lowest percentage of internet nonusers in the United Kingdom by population. However, Watson et al [32] also pointed out that a lack of digital devices and private spaces for accessing online health care could also act as barriers to digital health care themes in London. The use of such digital devices and online health care services opens up a new set of digital rights, opportunities, and responsibilities for patients [2], and this needs to be balanced across regions to ensure the equality and inclusion of health care. Our data suggest that CCGs are still not sufficiently efficient when it comes to adopting digital themes.

Furthermore, our findings highlight the impact of COVID-19 on the development of a CCG’s digital technology themes. Many of the reports cited COVID-19 as accelerating the need to digitize health care. For example, the use of telehealth to increase remote monitoring of patients or encourage them to manage their own health increased during the pandemic due to patients not being able to reach their own health care practices as freely. Most of the reports were generally very positive about increasing the inclusion of digital technology. Reasons for the positives statements included making it easier to access health care, providing early treatment, providing more data to improve services, and making the health care practice more agile and efficient. However, some reports cited reasons for concern about increasing digital health care when discussing the possibility of digital exclusion for some patients. Further, one report from the Newcastle Gateshead CCG raised concerns about how the integration of digital technology may impact the NHS policy of being free at point of use. This issue is consistent with concerns expressed by other researchers. For example, Clare [33] and Eruchalu et al [34] have noted the potential limitations of telehealth and telemedicine due to broadband connectivity issues resulting from socioeconomic disparities among regions, particularly among the underprivileged, the medically underserved, and in communities of color. With services becoming increasingly online (eg, many mental health services were all online) and not everyone having access to technology, there is a justifiable concern. To realize equitable benefits from health-related technologies across all populations, it is imperative to thoroughly examine and address complex issues, such as social, cultural, and economic factors that hinder accessibility and adoption among different communities. Otherwise, as pointed out by Ramsetty and Adams [35], despite technological advancements, disparities in health care access and outcomes will inevitably persist, particularly among the most vulnerable during times of crisis. Future research could further investigate this area and how these issues could be addressed.

It is essential to highlight that the absence of digital technology mentioned in a CCG report does not necessarily indicate a failure in its adoption. For example, eConsult is an online triage form that patients can fill out and send to their health care practitioner where it is reviewed and the next steps for treatment are accessed by a health care professional for the patients to take. The Kent and Medway CCG did not make any reference to this digital technology within their annual report [36], however, when investigating individual GP practices that operate under them, the technology is being used. This suggests that the lack of mention of a particular technology by a CCG group does not mean it is not being used at all but indicates the action of using such technology is not applied at the CCG level.

Conclusion
This research mapped out the current digital technology themes adopted by CCGs in the United Kingdom when providing health care services during the pandemic. These identified themes can be used by future health care providers to adopt digital solutions to address different health care issues, such as improving communication, digitizing their existing services, and increasing public access to health care. Furthermore, the research highlights the existence of a digital divide within CCGs in terms of adopting digital technology themes, particularly when it comes to regional disparities. The possible solutions could include providing support to the “digitally disengaged” CCGs in using various identified digital technology themes and becoming more digitally “active” or “engaged.” Caution needs to be taken when offering support as well. For example, we need to fully explore and understand the reasons behind their slow uptake of digital technologies in order to offer tailored solutions. To build upon our findings and promote equal access to digital health benefits for all communities, future research could examine the factors that contribute to regional disparities in digital technology adoption and access within and across CCGs. A comprehensive understanding of the underlying causes of these disparities could help policy makers and health care professionals focus their efforts more effectively toward bridging the digital divide and improving access to health resources and support for the public.

Although our study offers great insights into the digital technology themes adopted by CCGs, there are a few limitations that need to be addressed. Our findings are mainly based on CCG reports. The length, focus, and description details may differ slightly between these reports. In addition, there may be underreporting when it comes to digital technology themes by individual CCGs, resulting in misrepresentation when analyzing reports. Finally, the research focus was on the digital technology themes and the adoption of associated technologies rather than the impact these technologies had on patient satisfaction or patient outcomes. Future research could look at the impact of these identified digital technology themes or individual technologies and their effectiveness through in-depth interviews or questionnaire surveys with patients and health care professionals.
Acknowledgments
This study is supported by the Faculty of Social Science Social Research Internship fund, University of Sheffield.

Data Availability
All data analyzed during this study are included in this published article and its supplementary information files.

Authors’ Contributions
The project was conceived and designed by MZ and SL. The analysis was carried out by JAA, MZ, and SL. All authors read and approved the final paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Codebook.
[XLSX File (Microsoft Excel File), 93 KB - formative_v81le51859_app1.xlsx ]

References


23. Patient-Specific Medicines Optimisation | OptimiseRx. FDB (First Databank). First Databank. URL: https://www.fdbhealth.co.uk/solutions/optimiserx-medicines-optimisation [accessed 2023-08-14]

24. Mind BLMK - For better mental health. URL: https://www.mind-blmk.org.uk/ [accessed 2023-08-14]

25. With Me In Mind – Mental Health Support Teams. URL: https://www.withmeinmind.co.uk/ [accessed 2023-08-14]


28. Qwell. URL: https://www.qwell.io/ [accessed 2023-08-14]


Abbreviations

CCG: clinical commissioning group
ED: emergency department
GP: general practitioner
ICB: integrated care board
NHS: National Health Service
Allcock et al

© Jake Alan Allcock, Mengdie Zhuang, Shuyang Li, Xin Zhao. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
AI-Led Mental Health Support (Wysa) for Health Care Workers During COVID-19: Service Evaluation

Christel Lynne Chang¹,², MA; Chaitali Sinha³, MA; Madhavi Roy³, MA; John Chee Meng Wong¹,², MBBS, MSc

¹Yeo Boon Khim Mind Science Centre, National University of Singapore, Singapore, Singapore
²Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
³Wysa Inc, Boston, MA, United States

Corresponding Author:
John Chee Meng Wong, MBBS, MSc
Department of Psychological Medicine
Yong Loo Lin School of Medicine
National University of Singapore
Level 9, NUHS Tower Block
1E Kent Ridge Road
Singapore, 119228
Singapore
Phone: 65 6772 3481
Email: pcmwcmj@nus.edu.sg

Abstract

Background: The impact that the COVID-19 pandemic has had on health care workers’ mental health, in particular, cannot be ignored. Not only did the pandemic exacerbate mental health challenges through elevated stress, anxiety, risk of infection, and social isolation, but regulations to minimize infection additionally hindered the conduct of traditional in-person mental health care.

Objective: This study explores the feasibility of using Wysa, an artificial intelligence–led mental health app, among health care workers.

Methods: A national tertiary health care cluster in Singapore piloted the use of Wysa among its own health care workers to support the management of their mental well-being during the pandemic (July 2020-June 2022). The adoption of this digital mental health intervention circumvented the limitations of in-person contact and enabled large-scale access to evidence-based care. Rates and patterns of user engagement were evaluated.

Results: Overall, the opportunity to use Wysa was well-received. Out of the 527 staff who were onboarded in the app, 80.1% (422/527) completed a minimum of 2 sessions. On average, users completed 10.9 sessions over 3.80 weeks. The interventions most used were for sleep and anxiety, with a strong repeat-use rate. In this sample, 46.2% (73/158) of health care workers reported symptoms of anxiety (Generalized Anxiety Disorder Assessment-7 [GAD-7]), and 15.2% (24/158) were likely to have symptoms of depression (Patient Health Questionnaire-2 [PHQ-2]).

Conclusions: Based on the present findings, Wysa appears to strongly engage those with none to moderate symptoms of anxiety. This evaluation demonstrates the viability of implementing Wysa as a standard practice among this sample of health care workers, which may support the use of similar digital interventions across other communities.

(JMIR Form Res 2024;8:e51858) doi:10.2196/51858

KEYWORDS
AI; app; application; artificial intelligence; COVID-19; digital; health care workers; mental health; pandemic; Wysa

Introduction

During the COVID-19 pandemic, health care systems and workers were under immense pressure as the rates of infection surged. Health care workers experienced elevated levels of stress and psychological burden [1], due to factors such as excessive work hours and an increased risk of infection [2,3]. In a study of Wuhan’s frontline health care workers’ mental health during this pandemic, respondents reported an increase in symptoms of depression and anxiety, and 59% reported moderate to severe...
perceived stress [4]. Chew et al [5] examined the mental health of health care workers in Singapore and India and found a lower prevalence of psychological symptoms than reported elsewhere but a high prevalence of physical symptoms. These physical symptoms could reflect somatization, as psychological outcomes and the presence of physical symptoms were significantly associated. In a separate study, Tan et al [6] reported that nonmedical health care workers such as allied health professionals, technicians, or administrators had a higher prevalence of anxiety compared to medical staff such as physicians and nurses. This highlights the importance of attending to the mental health of not just frontline medical workers but health care workers as a whole.

Meanwhile, efforts to reduce the spread of COVID-19 have led to nationwide and global lockdowns, resulting in the digitalization of many services. Mental health practitioners likewise adopted web-based strategies to continue supporting clients. Yet, in a time with many in isolation and a worldwide escalation in mental health concerns [7], the urgency for the mainstream adoption of scalable digital mental health services has intensified. In addition to serving as an ad hoc solution during this pandemic, digital mental health interventions through web-based or mobile platforms may provide unique benefits beyond traditional services such as in-person counseling or therapy [8]. For instance, they improve the accessibility and cost of mental health services since traditional mental health interventions are typically subject to standard clinic hours and may not be affordably priced. With digital interventions, individuals can receive on-demand support in moments of crisis, beyond the boundaries of the clinic or therapy room. Digital services can additionally circumvent the stigma-related barriers to mental health treatment since they can be accessed discreetly and at one’s own convenience. In these ways, digital interventions can serve as complements to traditional mental health services, provide interim services for those waiting to receive in-person treatment, and provide new services for those facing barriers to accessing existing clinical interventions [9,10]. Mental health care systems could thus benefit from incorporating digital interventions as part of standard care practices. As an acknowledgment of these advantages, a national tertiary health care division in Singapore piloted the use of Wysa, which is an artificial intelligence (AI)–enabled mental health app among their own health care staff during the COVID-19 pandemic.

Wysa is a mental health mobile app, accessible on the Apple App Store and Google Play Store, designed to address various mental health concerns through self-guided interventions. The app features rule-based AI through a conversational chatbot. This facilitates user interactions, offering personalized support and interventions by comprehending user responses and directing them to relevant resources. Wysa’s AI models operate within a rule-based framework, ensuring clinical safety and adherence to predefined scripts. There is also an option to connect with a human coach, which supplements the app’s fully automated functions.

Due to the primarily self-guided nature of the app, the interventions it provides can be administered to large populations at minimal cost. Moreover, Linardon et al [11] found in their meta-analysis that self-guided interventions significantly surpassed waitlist and active controls in efficacy for reducing symptoms of depression and anxiety. Previous studies examining the utility of mental health mobile apps found Wysa to have high usability and engagement ratings [12,13]. In previous studies, general populations and patients [14] with symptoms of anxiety and depression reported improvements with the use of Wysa. This study serves to evaluate the experience of using Wysa among a population of health care workers during the pandemic.

Wysa is DCB.0129 compliant (the National Health Services [NHS] Clinical Risk Management Standards under the Health and Social Care Act 2012), and its information security management system (ISMS) and privacy information management system (PIMS) have been audited and certified by the British Standards Institution (BSI) for ISO 27001 and ISO 27701. The parent Wysa Android and iOS apps are certified as CE (European Conformity) Medicines and Healthcare Products Regulatory Agency (MHRA) Class 1 Medical Devices. Additionally, Wysa ensures that all user data are encrypted at rest in the storage servers and are hosted by industry leaders who have stringent General Data Protection Regulation (GDPR) and HIPAA (Health Insurance Portability and Accountability Act)–compliant security programs and controls.

### Methods

#### Overview

In this pilot exercise, health care staff from a national tertiary health care cluster were provided access to Wysa through paid licenses. Data regarding this community’s use of Wysa and feedback were obtained over a period of 2 years, from July 2020 to June 2022. Advertisements were internally disseminated to the institutions’ staff to encourage them to activate their licenses and engage with the app to support their mental well-being.

#### Wysa App

While users interacted with the Wysa chatbot, Wysa would prompt them to indicate which challenges they would like support with. The life challenges disclosed by the present sample of health care workers are given in Table 1. These challenges were predetermined categories that users were able to select, and they present an overview of the key concerns that this population sought help with through the chatbot.

Table 2 displays various categories of emotions reported by users during organic interactions with the chatbot. For interpretative purposes, the specific emotions mentioned in users’ chat messages were identified and combined into categories based on similarity. This enables insight into the range of emotions experienced by these users, which yields granularity to better understand the state of mind they were in as they engaged with the app.
Table 1. Health care workers self-reported life challenges while using the Wysa Digital App during the COVID-19 pandemic over 2 years.

<table>
<thead>
<tr>
<th>User-reported life challenge</th>
<th>Unique users, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy and sleep</td>
<td>531 (97.2)</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>404 (73.9)</td>
</tr>
<tr>
<td>Relationships</td>
<td>317 (58)</td>
</tr>
<tr>
<td>Stress and anxiety</td>
<td>307 (56.1)</td>
</tr>
<tr>
<td>Life events</td>
<td>118 (21.6)</td>
</tr>
<tr>
<td>Low mood and depression</td>
<td>125 (22.9)</td>
</tr>
<tr>
<td>Trauma and loss</td>
<td>71 (13)</td>
</tr>
</tbody>
</table>

Table 2. Emotions self-disclosed by health care workers during Wysa chatbot interaction over 2 years during the COVID-19 pandemic.

<table>
<thead>
<tr>
<th>Emotion valence and emotion category</th>
<th>Count, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negative valence emotions</strong></td>
<td></td>
</tr>
<tr>
<td>Low mood and loneliness</td>
<td>368</td>
</tr>
<tr>
<td>Stress and worry</td>
<td>321</td>
</tr>
<tr>
<td>Anger</td>
<td>215</td>
</tr>
<tr>
<td>Fatigue and pain</td>
<td>150</td>
</tr>
<tr>
<td><strong>Positive valence emotions</strong></td>
<td></td>
</tr>
<tr>
<td>Happiness</td>
<td>466</td>
</tr>
<tr>
<td>Confidence</td>
<td>222</td>
</tr>
<tr>
<td>Improved mood</td>
<td>177</td>
</tr>
<tr>
<td>Excitement</td>
<td>88</td>
</tr>
<tr>
<td>Gratitude</td>
<td>39</td>
</tr>
</tbody>
</table>

Ethical Considerations

The conduct of this service evaluation was granted institutional review board exemption by the National University Health System, and the institutional review board’s requirement for informed consent was waived for this posthoc descriptive analysis. All user data analyzed as part of this service evaluation was anonymous, and user data remained fully confidential. The users who engaged with the app were not provided with additional compensation, as the app use was provided to them for free to engage with as they pleased. Users were not systematically recruited to engage with the app, which enabled an evaluation of genuine receptiveness to the app and engagement rate.

Results

Digital App Engagement

Out of the 500 licenses purchased by the health care institution, 527 users were onboarded to the app. The discrepancy in numbers is due to the Wysa system disregarding users who had logged in only briefly without further engagement. Of the 527 health care staff who were onboarded, 495 users completed at least 1 full session, and 422 completed 2 or more sessions. One session is defined as a full interaction with the Wysa chatbot or an intervention tool. This indicates a high engagement rate of 93.9% engaging at least once and 80.1% engaging at least twice beyond onboarding. On average, the number of in-app sessions completed by each of the 495 active app users over this period of 2 years was 10.9 (SD 23.3), and the number of weeks of use was 3.80 (SD 5.58). The average number of messages sent by a user was 66.3 (SD 147) and the average number of messages exchanged in each session was 6.25 (SD 8.10). The weeks of use were not necessarily consecutive, although 40.5% of users used Wysa consecutively for at least 3 weeks.

From October 28-30, 2020, a “gratitude challenge” was run in which daily emails prompted users to show gratitude to important people in their lives. Wysa also offered shareable gratitude cards for users to share with their loved ones. This “challenge” was designed to facilitate a boost in engagement with mental health support. During this period, 53 users engaged with the app for an average of 2.30 (SD 1.60) sessions. In the 3 days before the challenge period (October 25-27), there were 31 users who used the app for 2.03 (SD 1.69) sessions on average. Out of these users, 16 of them returned during the period of the challenge. In the 3 days after the challenge period (October 31-November 2), 44 users used the app (mean 2.26, SD 2.38). Of these 44 users, 25 of them were returning users who had engaged during the challenge period. This data may suggest that the challenge was successful in bolstering user engagement, since there was a 71% increase in engagement during the challenge compared to 3 days before, and the number of users who engaged with Wysa after the challenge period had increased by 41.9% compared to before. Initiatives similar to
this “challenge” implemented recurrently may have the potential to promote sustained engagement.

While engaging in sessions, users were prompted to provide feedback on their present interaction and experience on a scale of 1-5. Feedback was given by 99 users on 234 sessions at a response rate of 20% (mean 4.07, SD 0.95). Of these, 218 sessions received scores of 3 and above, while 213 sessions received scores of 4 and above, suggesting a strong positive feedback rate of 93.2% and 91% receiving highly positive feedback, respectively. Of the 495 active app users, 158 individuals responded to the Generalized Anxiety Disorder Assessment-7 (GAD-7) and the Patient Health Questionnaire-2 (PHQ-2) screening questionnaires within the app. User-experience feedback was given by 59 out of the 158 users. This response rate of 37.3% is higher than the overall sample’s rate of 20%. The user-experience feedback given by these users as a function of their GAD-7 and PHQ-2 scores is shown in Table S1 in Multimedia Appendix 1. Approximately 30%-40% of individuals falling into each severity category gave feedback, with no obvious differences in the rate of response between categories.

**Clinical Utility**

Wysa contains an assortment of self-guided interventions, which each have a collection of interventions categorized by relevance to a circumstance or emotion. The interventions consist of AI-enabled conversations that target specific interventions. Table 3 gives a breakdown of how the Wysa interventions were used among this community of health care workers. Descriptions of the various interventions can be found in Multimedia Appendix 2. The data are based on users who actively sought help from these interventions by directly clicking on them. The number of unique users provides an indication of use pattern, that is, how many distinct individuals contributed to the total use compared to repeat users. Figure 1 illustrates this ratio (use factor) of unique users to total use. The higher the use factor, the higher the rate of repeat use. For instance, the number of unique users for sleep meditation interventions was 84, suggesting that this handful of users repeatedly engaged with this intervention a total of 564 times, with a high use factor of 6.7. In comparison, although guided visualization interventions were used a total of 502 times (an overall use comparable to that of sleep meditations), these interventions were used repeatedly by a larger number of 172 users with a use factor of 2.9. As such, it appears that a smaller subset of individuals repeatedly used interventions for sleep meditation, whereas guided visualizations were used by a wider range of individuals with a lower use factor.

Unsurprisingly, since sleep difficulties tend to co-occur alongside distressing events, interventions for sleep were overwhelmingly sought after, with high total use levels (Table 3). The next most popular interventions were for anxiety, especially health anxiety, stress, and self-esteem. These use patterns generally coincide with the top life challenges, as reported in Table 1. Furthermore, the popularity of the health anxiety intervention aligns with the backdrop of the COVID-19 pandemic.

GAD-7 and PHQ-2 screening tools were embedded within the app to screen for anxiety and depressive symptoms, respectively, which users were prompted to respond to after their very first chatbot conversation. This initial screening helped to capture baseline indications of the present population. Of the 495 active app users, 158 individuals responded to these questionnaires at this time point (158/495, 31.9% response rate). Table S1 in Multimedia Appendix 1 presents a breakdown of the prevalence of anxiety and depressive symptoms in this sample according to severity (number of users). Overall, 46.2% (73/158) reported symptoms of anxiety, and 15.2% (24/158) were likely to have symptoms of depression.

<table>
<thead>
<tr>
<th>Interventions</th>
<th>Total use, n</th>
<th>Unique users, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mindfulness</td>
<td>928</td>
<td>159</td>
</tr>
<tr>
<td>Sleep meditation</td>
<td>564</td>
<td>84</td>
</tr>
<tr>
<td>Guided visualization</td>
<td>502</td>
<td>172</td>
</tr>
<tr>
<td>Thought recording</td>
<td>227</td>
<td>120</td>
</tr>
<tr>
<td>Behavioral activation</td>
<td>128</td>
<td>74</td>
</tr>
<tr>
<td>Psychoeducation</td>
<td>163</td>
<td>104</td>
</tr>
<tr>
<td>Breathing exercises</td>
<td>46</td>
<td>28</td>
</tr>
<tr>
<td>Cognitive restructuring</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>Acceptance</td>
<td>54</td>
<td>30</td>
</tr>
<tr>
<td>Grounding</td>
<td>34</td>
<td>20</td>
</tr>
<tr>
<td>Social support</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Problem solving</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Habit building</td>
<td>58</td>
<td>45</td>
</tr>
</tbody>
</table>
The average use rates (number of weeks and sessions) according to users’ GAD-7 and PHQ-2 scores are also displayed in Table S1 in Multimedia Appendix 1. The average number of weeks that users engaged with Wysa provides an indication of how long their symptoms or interest in using Wysa lasted, while the average number of sessions gives an estimate of use frequency. To illustrate, this data demonstrates that health care workers in this sample with moderate anxiety engaged with Wysa for an average of 6.45 weeks. Over these 6.45 weeks, users completed an average of 20.2 sessions.

Individuals with no, mild, or moderate anxiety used Wysa for a high mean number of sessions (mean_total 19.5) and weeks of use (mean_total 6.04), almost 2 times higher than the overall active app user average of 10.9 and 3.80, respectively, while there was a limited sample of individuals with symptoms of severe anxiety. Users who were likely to have symptoms of depression engaged with Wysa more and for a marginally longer duration on average than those who were unlikely, as seen in Table S1 in Multimedia Appendix 1. This finding is promising, as it suggests that Wysa may be more beneficial and motivating to use for individuals with symptoms of depression.

Regardless of GAD-7 or PHQ-2 severity, the most popular interventions used were for sleep, especially for deep sleep, and anxiety, especially for health anxiety (Table S1 in Multimedia Appendix 1). Users falling into the moderate and severe GAD-7 categories were more targeted in using Wysa, specifically to help with sleep and anxiety, and did not explore the other interventions as much as other users. On the other hand, users with no or mild anxiety explored more of Wysa’s other interventions. Likewise, users who were unlikely to have symptoms of depression engaged with a wider variety of interventions than those who were screened as likely. In general, this data emphasizes the popularity of sleep and anxiety interventions among this community and implies that greater severity of mental health symptoms is associated with more targeted engagement with the app.

**Discussion**

**Overview**

Due to heightened restrictions on in-person contact, the COVID-19 pandemic has brought to the forefront the current limitations of mental health care delivery. As such, the advent of the pandemic has proved a catalyst for new digital initiatives to be facilitated [15]. The present community case study is a service evaluation of a pilot exercise within a national tertiary health care division in Singapore using a self-guided, AI-enabled mental health app (Wysa) during the COVID-19 pandemic. The availability of Wysa for this population of health care staff to use was well-received, particularly against this backdrop of pandemic-related anxiety. Even beyond the height of the pandemic (2020-21), the continued use of Wysa by this community suggests that this digital mental health solution has merits surpassing those of temporary pandemic-specific support. The high initial engagement rate of 93.9% of onboarded users completing at least 1 in-app session may have been due to this intervention being introduced through a trusted source—a health care institution or employer [14]. Nevertheless, 80.1% of users returned to engage with the app voluntarily for a full second time, and the mean number of in-app sessions completed was 10.9 for an average of 3.80 weeks. In-app use reflects recurring engagement with Wysa, but users could also engage with the chatbot through the website through a widget. As a whole, the data demonstrates that users may have experienced some real or perceived benefit, which motivated their continued engagement with the app.

Among this community, 46.2% reported symptoms of anxiety, and 15.2% were likely to have symptoms of depression. These prevalence rates are comparable to those of other health care
communities in other parts of the world during the COVID-19 pandemic. For instance, these rates are similar to the 44.6% that had symptoms of anxiety (GAD-7) and 14.8% that had moderate to severe depression (PHQ-9) in a sample of health care workers from China [16]. However, compared to a population of health care workers in the United Kingdom following the pandemic’s first peak, the present sample’s anxiety prevalence was higher than the UK sample’s 34.3%, but depression prevalence was lower than the UK sample’s 31.2% (PHQ-4) [17]. Following a similar trend, the present sample’s rate of anxiety was slightly higher than the pooled prevalence rate of 40%, but depression prevalence was lower than the pooled prevalence of 37% from a meta-analytic study of health care workers across 19 countries during this pandemic [18]. Compared to the general population during the COVID-19 pandemic, this community of health care workers’ anxiety and depression symptom prevalence rates were similarly higher and lower than the global prevalence rates of around 27.3%-31.9% and 33.7%, respectively [19,20].

In general, the intervention use patterns (Table 3) concur with the user-reported life challenges (Table 1). The data regarding the patterns of intervention use indicates that “for deep sleep” was the most popular intervention used by the greatest number of users repeatedly for the most number of times. “Sleep sounds” was used the second-largest number of times, but only by a handful of unique individuals. This suggests that an intervention for sleep sounds may not be attractive to a wide range of users, yet for those who decide to try it, it appears motivating to use, thus resulting in its high repeat-use rate. The “health anxiety” intervention was the next most popular intervention, used by almost as many unique users who engaged with the intervention “for deep sleep.” The popularity of this intervention may be idiosyncratic to the COVID-19 pandemic period. Overall, this data implies that this community of health care workers primarily sought help from Wysa for sleep concerns and anxiety or found these interventions to be the most beneficial. The analysis of most used interventions according to GAD-7 and PHQ-2 severity further indicates that this is true regardless of their mental health status.

Users in this community with moderate to severe symptoms of anxiety were more targeted in using Wysa specifically to help with sleep and anxiety and did not explore the other interventions as much as other users. This was reasonable to expect, as individuals with anxiety often have co-occurring sleep difficulties and would likely allot more attention to addressing these key concerns. Users with no or mild anxiety explored more of Wysa’s other interventions. Their engagement in a greater range of interventions could explain why these individuals continued to use Wysa despite their lack of or few anxiety symptoms. Likewise, those who were unlikely to have symptoms of depression engaged with a wider variety of interventions than those who were likely. Although this may partially be due to a large number of individuals falling under the unlikely category, thus contributing to the heterogeneity of use patterns, engagement in a large range of interventions by users with fewer mental health concerns may reflect their use of this app in a manner relating to principles of positive psychology [21] to further enhance one’s quality of life. This contrasts with the ostensibly treatment-targeted approach taken by users with greater psychopathological challenges. The dichotomous manner of use noted here may attest to the breadth of what the Wysa app could support users with.

Individuals with mild or moderate anxiety symptoms may sufficiently benefit from using Wysa in a self-directed manner, thus leading to high use and positive feedback rates. This hypothesis may be supported by Karyotaki and colleagues’ review [22], which found that individuals with more severe depression may benefit more from guided interventions involving human support, whereas individuals with mild or subthreshold symptoms may be adequately supported by self-guided interventions. Although, in the present sample, this does not hold true with regard to depressive symptoms, since users likely to have depression engaged with Wysa at comparable rates to those who were unlikely (Table S1 in Multimedia Appendix 1). This is an encouraging finding, as it suggests that individuals with symptoms of depression may have derived benefit from, or at least enjoyed, continuing to use this app to support their well-being. Yet, the depression severity in this study was determined using the PHQ-2. More accurate and potentially different results might have been obtained if the PHQ-9 had been used. In addition, individuals with no symptoms of anxiety as reported on the GAD-7 engaged with Wysa at high rates, on par with users with mild to moderate symptoms. This may indicate that Wysa is motivating to use for individuals that may not have any overt mental health struggles, which could imply that Wysa is able to support users in maintaining their positive mental health or to flourish further in eudaimonia, as proposed in a previous paragraph. Nonetheless, grounded conclusions could only be determined through direct qualitative inquiry or more rigorous quantitative pre-post analyses, which this evaluation was not able to perform due to methodological constraints. To summarize, Wysa appears to be beneficial or motivating to use for individuals with no, mild, or moderate anxiety, whereas individuals with severe anxiety symptoms may need to seek other forms of guided clinical care.

As there was a 71.0% increase in use of the app during the “gratitude challenge” period, the challenge appeared to be successful in boosting app engagement. Extant literature provides support for the advantages of practicing gratitude, from experimental research that found gratitude to enhance the psychological benefits of social support [23] to a randomized controlled trial of a gratitude intervention in which gratitude effectively improved well-being [24]. A review by Bono and Sender [25] suggests that gratitude’s ability to amplify social benefits and motivate self-improvement makes it foundational to human development. This powerful potential may have propelled the increase in app engagement. Future endeavors to implement similar digital mental health interventions may benefit from a greater volume of similar initiatives at regular intervals. However, to assess the efficacy of such initiatives more accurately, more robust research is needed to isolate the effect of the challenge from other confounding variables. In addition, the present sample of users who responded to the GAD-7 and PHQ-2 screening questionnaires was too small to draw any conclusions regarding the mental health status of...
individuals that these types of “challenges” may work best on. Future studies should evaluate this further.

Altogether, the present service evaluation shows promising evidence to endorse the implementation of digital mental health interventions as standard practice, as the high repeat-use rates for Wysa’s interventions corroborate with the strong positive feedback and user engagement rates. Health care systems could consider implementing similar digital interventions as part of routine care for patients or to support their own staff internally. Clinicians may leverage these interventions to improve the capacity of mental health care service delivery, since these digital interventions have the potential to greatly increase access to evidence-based psychological support. For such ventures to be implemented, a systematic approach could be taken by first determining the appropriate digital solution for the community in question [9]. Using an integrated tool such as Wysa with its diverse assortment of interventions would be able to cover a wide variety of needs, possibly ranging from treatment-targeting solutions to the eudaimonic enhancement of well-being. Yet, as seen in this study, the majority of this community’s users were chiefly interested in the interventions for sleep and anxiety, regardless of the severity of their symptoms. If the community has more specific needs, perhaps more targeted apps or other digital interventions could be considered.

Limitations
The evaluation of the pilot exercise outlined in this study has several limitations. First, intervention use data from this population was based on users directly clicking on the interventions. As such, this was not able to capture the interventions used that were suggested by Wysa through chatbot conversations. Next, the PHQ-2 was used to collect data on users’ depressive symptoms. However, the PHQ-9 may have been able to yield more precise information. In addition, comparison between baseline and subsequent scores would be required to generate more reliable quantitative results that elucidate Wysa’s ability to moderate symptoms of poor mental health. In the present exercise, insufficient users responded at follow-up screenings in order to conduct meaningful pre-post analyses. Moreover, there was only a moderate response rate of 31.9% to the screening questionnaires at baseline. Conclusions drawn from the present data may thus not be representative of the entire community that participated, and the small sample size of respondents is a limitation in itself. The users who responded to the screening interventions also appeared to contribute to a large proportion of the overall population’s engagement with Wysa, since this sample’s average use rates were approximately twice as high as the overall population’s average rates. This may further indicate a biased sample. Furthermore, although the present data demonstrates the propitiousness of the “gratitude challenge,” its true efficacy in boosting engagement could not be ascertained since a controlled experimental design would be necessary.

Another limitation is that assumptions regarding reasons for engagement with Wysa are speculative; for example, in assuming that users who repeatedly used the app benefitted from its use or that the number of weeks of use reflects how long their symptoms lasted. However, it is possible that these users might not have seen much benefit, yet they were patient in attempting to engage with Wysa. This would thus be more telling of the user’s character or personality than of the app’s effectiveness. Similarly, it is not possible to determine whether the ubiquitous popularity of the sleep and anxiety interventions was due to users being mainly concerned with these challenges or whether this phenomenon was a result of users’ expectations regarding what an app could plausibly help with. To yield more reliable conclusions, qualitative data would be required, collected through feedback through the app or a follow-up focus group discussion. Conducting a focus group may further enable the generation of insight into the lived experiences of these users. This may be especially useful to illuminate the experience of those with severe anxiety symptoms as they attempted using Wysa, to deduce the actual reasons for their low rates of engagement, and for those with no anxiety, to discern whether Wysa simply sustained their mental health or was able to promote even greater degrees of well-being. Nevertheless, due to the anonymous nature of the data, reidentification of the users to follow up with would not be possible. For more rigorous research, a dedicated study would have to be designed with ethical approval and consent obtained from health care workers to participate. Future studies could thus be developed based on the promising findings from this community service evaluation.

Acknowledgments
The authors would like to thank Ms Priscilla Teo (the group chief human resource officer) of the National University Health System, Singapore, for permission to conduct this evaluation and for the purchase of licenses to use Wysa, and Ramakant Vempati and Tanya Malik from Wysa Inc for their support in the conduct of this evaluation. No additional funding was obtained for the running of this formative service evaluation.

Authors’ Contributions
All authors contributed to this study and approved the submitted version. CLC was responsible for the overall preparation of this manuscript. CS and MR assisted with analyzing the data. JCMW implemented the Wysa program with Nicholas Sii, Mae Kng, and Suhana Alwi and directed the data analysis and manuscript preparation as senior author.

Conflicts of Interest
CS and MR are employed by Wysa Inc. All other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Multimedia Appendix 1
Table S1. Health care workers’ anxiety and depressive symptom prevalence and associated Wysa app use rates, feedback, and interventions used over 2 years during the COVID-19 pandemic.

[DOCX File, 91 KB - formative_v8i1e51858_app2.docx]

References


Abbreviations

AI: artificial intelligence
BSI: British Standards Institution
CE: European Conformity
GAD-7: Generalized Anxiety Disorder Assessment-7
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
ISMS: information security management system
MHRA: Medicines and Healthcare Products Regulatory Agency
NHS: National Health Services
PHQ-2: Patient Health Questionnaire-2
PIMS: privacy information management system

Edited by A Mavragani; submitted 22.08.23; peer-reviewed by J Shumake, K Muroi; comments to author 02.12.23; revised version received 20.12.23; accepted 06.03.24; published 19.04.24.

Please cite as:
Chang CL, Sinha C, Roy M, Wong JCM.
JMIR Form Res 2024;8:e51858
URL: https://formative.jmir.org/2024/1/e51858
doi:10.2196/51858
PMID:38640476

©Christel Lynne Chang, Chaitali Sinha, Madhavi Roy, John Chee Meng Wong. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Remote Symptom Monitoring Using Patient-Reported Outcomes in Patients With Chronic Kidney Disease: Process Evaluation of a Randomized Controlled Trial

Birgith Engelst Grove1,2, RN, MHSc, PhD; Annette de Thurah2,3, MPH, PhD; Per Ivarsen2,4, MD, PhD; Ann Katrine Kvisgaard1, MSc; Niels Henrik Hjollund1,2,5, MD, PhD; Regine Grytnes1, PhD; Liv Marit Valen Schougaard1, PhD

1AmbuFlex, Centre for Patient-Reported Outcomes, Gødstrup Hospital, Herning, Denmark
2Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
3Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark
4Department of Renal Medicine, Aarhus University Hospital, Aarhus, Denmark
5Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark

Corresponding Author:
Birgith Engelst Grove, RN, MHSc, PhD
AmbuFlex, Centre for Patient-Reported Outcomes
Gødstrup Hospital
Møllegade 16
Herning, 7400
Denmark
Phone: 45 28904835
Email: bigcri@rm.dk

Abstract

Background: In Denmark, outpatient follow-up for patients with chronic kidney disease (CKD) is changing from in-hospital visits toward more remote health care delivery. The nonuse of remote patient-reported outcomes (PROs) is a well-known challenge, and it can be difficult to explain which mechanisms of interventions influence the outcome. Process evaluation may, therefore, be used to answer important questions on how and why interventions work, aiming to enhance the implications for clinical practice.

Objective: This study aimed to provide insight into the intervention process by evaluating (1) the representativity of the study population, (2) patient and physician use patterns, (3) patient adherence to the intervention, and (4) clinical engagement.

Methods: A process evaluation determining the reach, dose, fidelity, and clinical engagement was carried out, alongside a multicenter randomized controlled trial (RCT). We developed and implemented an intervention using PRO measures to monitor outpatients remotely. Data were collected for the PRO intervention arms in the RCT from 4 sources: (1) PRO data from the participants to determine personal factors, (2) the web-based PRO system to identify key usage intervention patterns, (3) medical records to identify clinical factors relating to the use of the intervention, and (4) semistructured interviews conducted with involved physicians.

Results: Of the 320 patients invited, 152 (47.5%) accepted to participate. The study population reflected the target population. The mean adherence rate to the PRO intervention arms was 82% (95% CI 76-87). The questionnaire response rate was 539/544 (99.1%). A minority of 13 (12.9%) of 101 patients needed assistance to complete study procedures. Physicians assessed 477/539 (88.5%) of the questionnaires. Contact was established in 417/539 (77.4%) of the cases, and 122/539 (22.6%) of the patients did not have contact. Physicians initiated 288/417 (69.1%) and patients requested 129/417 (30.9%) of all the contacts. The primary causes of contact were clinical data (242/417, 58%), PRO data (92/417, 22.1%), and medication concerns and precautionary reasons (83/417, 19.9%). Physicians found the use of PRO measures in remote follow-up beneficial for assessing the patient’s health. The inclusion of self-reported clinical data in the questionnaire motivated physicians to assess patient responses. However, some barriers were emphasized, such as loss of a personal relationship with the patient and the risk of missing important symptoms in the absence of a face-to-face assessment.

Conclusions: This study demonstrates the importance and practical use of remote monitoring among patients with CKD. Overall, the intervention was implemented as intended. We observed high patient adherence rates, and the physicians managed most
questionnaires. Some physicians worried that distance from the patients made it unfeasible to use their “clinical glance,” posing a potential risk of overlooking crucial patients’ symptoms. These findings underscore key considerations for the implementation of remote follow-up. Introducing a hybrid approach combining remote and face-to-face consultations may address these concerns.

**Trial Registration:** ClinicalTrials.gov NCT03847766; https://clinicaltrials.gov/study/NCT03847766

**KEYWORDS**
chronic kidney disease; pragmatic randomized controlled trial; process evaluation; patient-reported outcome measures; remote monitoring; monitoring; patient-reported outcome; chronic kidney; intervention

**Introduction**

Lifestyle and a growing elderly population in Denmark have increased the number of patients with chronic diseases to approximately 1 million in a population of 6 million, with growing health care expenditure as a consequence [1]. To improve efficacy, new ways of delivering health care to people with chronic kidney disease (CKD) have been posited, such as remote monitoring, which is defined as using technology to monitor patients at a distance [2,3]. One way of remotely monitoring patients is to collect information about their symptoms by obtaining repetitive patient-reported outcomes (PROs) while the patients are at home. This allows for frequent capture of important disease-specific outcomes and also allows clinicians direct access to the patients’ health status [4,5]. PROs are measures of a patient’s health conveyed directly by the patient, without interpretation by a clinician or anyone else [6], often gathered from electronic questionnaires. Remote monitoring using PROs is increasingly applied among patients with chronic conditions in Denmark [4,7]. However, several personal and external factors may affect a person’s ability to engage with and use digital health interventions [8,9]. Prior studies have shown sociodemographic and economic inequality in patients who attend [10,11] and adhere to PRO-based remote follow-up [12].

A pragmatic randomized controlled trial (RCT) named PROKID (PRO measures in Kidney care) is currently investigating whether PRO-based remote follow-up is a safe and effective alternative to health care delivery to patients with CKD [7]. The primary outcome in the PROKID trial is renal function measured using the estimated glomerular filtration rate (eGFR) as this is the single-most accurate measurement for CKD progression [13] and thereby indicates the level of safety in remote monitoring. The effectiveness of remote monitoring is measured using clinical data, resource use, and PROs, such as the quality of life and illness perception [14]; also see Grove et al (unpublished data, 2024). PRO-based interventions in other patient populations have shown the ability to (1) reduce the number of outpatient visits [15,16], (2) improve communication between health care professionals and patients [17-20], and (3) offer patients a more comprehensive understanding and greater knowledge of their condition [21-23]. However, several knowledge gaps regarding the use and acceptability of remote PRO interventions are still present. Trials are vulnerable to various biases that may undermine external validity [24]. Evidence suggests that often, this risk of bias is related to patients not using the technology as intended [25]. Thus, examining the quality (fidelity) and quantity (dose) of what was implemented in practice and the extent to which the intervention reached its intended users is essential when evaluating trials [26]. The degree to which patients engage with and use the intervention as intended is a crucial component in evaluations [26]. The nonuse of remote PRO is a well-known challenge [16,27] and may blur the overall interpretation of using PRO as the basis for remote follow-up. However, little is known about the components that might enhance patient and clinical engagement and thus might have the most significant impact in terms of use and adherence to the intervention. The PROKID trial [14] cannot explain which mechanisms of the intervention have an influence on the outcome. Process evaluation may therefore be used to answer important questions on how and why interventions work, aiming to enhance the implications for clinical practice [26].

Thus, we aimed to conduct a process evaluation to obtain insights into the PROKID intervention’s working mechanisms by evaluating (1) the representativity of the study population, (2) patient and physician use patterns, (3) patient adherence to the intervention, and (4) clinical engagement.

**Methods**

**Study Setting**

In 2019, the PROKID trial, a Danish multicenter RCT, was initiated to evaluate the noninferiority of PRO-based remote follow-up compared to usual outpatient clinic visits in managing the decline in renal function and maintaining patients’ quality of life [7]. The intervention was implemented in a real-life setting, and clinicians and patients were involved in the development process. We provided short oral seminars for the staff in the outpatient clinics, where the intervention was discussed and refined. We invited patients to provide input into the design and the patient information sheets of the intervention. Participants were eligible for the trial if they attended follow-up from January 2019 to August 2021 at a renal outpatient clinic in Central Denmark Region and had a renal function of eGFR≤40 mL/minute. In addition, participants had to be ≥18 years old and able to complete a questionnaire. Newly referred outpatients were randomized into the following 3 groups: (1) PRO-based follow-up, (2) PRO-based telephone follow-up, or (3) usual outpatient follow-up (Figure 1).
In the intervention arms, each patient participated in 6 PRO consultations (when PROs were used in a consultation) during 18 months of follow-up. Prior to each PRO consultation, patients needed to undergo blood tests, measure their blood pressure and weight, and complete a disease-specific questionnaire. The questionnaires were sent to patients 7 days ahead of each consultation, and in the case of nonresponse, they were sent reminders on the fourth day and the day just prior to the consultation. The questionnaire included information about blood pressure, weight, self-rated health, renal-specific symptoms, and a free-text box [28]. The physicians accessed the patients’ PRO responses through a graphical overview embedded in the electronic health record system [5]. In the PRO-based follow-up intervention group, physicians had to approve each patient’s responses, manage them according to a color-coded algorithm, and determine whether the patient needed contact. A clinical expert group has assigned a color to each item response according to the severity of the symptom, as previously described: red, yellow, or green [14,28]. Each patient’s need for contact was evaluated based on their PRO responses and other clinical data, such as blood samples and blood pressure. Therefore, it was up to the physicians to decide whether contact with the patient was needed. This applied not only to instances where the color code was red or yellow but also to instances in which the questionnaire displayed a green color code. In the PRO-based telephone follow-up group, the physicians had to approve each patient’s response, call the patient, and document the conversation as conducted. Figure 2 outlines the groups, the content, and the purpose of using PRO.
Figure 2. Overview of the intervention groups in the PROKID trial including patients with CKD at Aarhus University Hospital, Gødstrup Hospital, and Viborg Regional Hospital from January 2019 to August 2021. CKD: chronic kidney disease; PROKID: PRO measures in Kidney care.

Study Design
The process evaluation was an integral component of the PROKID trial and focused on adding information not provided in the trial and clarifying how the intervention was received in practice to increase accuracy in the trial results. The design and methods of the PROKID trial are described elsewhere [7]. The design of this evaluation was influenced by Steckler and Linnan’s [26] process evaluation framework, building on 4 themes: reach, dose, fidelity, and clinical engagement. This study used qualitative and quantitative methods to answer the research questions. The study was reported in accordance with Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-EHEALTH) guidelines [29].

Participants
We obtained data from patients attending outpatient follow-up at the renal outpatient clinics at Aarhus University Hospital, Gødstrup Hospital, and Viborg Regional Hospital in Central Denmark Region. In the quantitative phase of the process evaluation, aimed to identify the reach of the intervention, the sample consisted of 320 patients, all eligible for randomization. The PROKID trial recruited 152 (47.5%) patients, and the 105 (69.1%) patients allocated to the PRO intervention groups informed the dose and fidelity of the trial. The qualitative phase of the process evaluation consisted of semistructured interviews with physicians (N=10) involved in the PROKID trial (Figure 1). Physicians were purposefully sampled from all involved outpatient clinics. Of the 10 physicians, 5 (50%) were female, 6 (60%) were senior consultants, and 4 (40%) were consultants, and their ages ranged from 37 to 67 years. Most of them had under 2 years of experience in using PRO in clinical practice.

Process Evaluation Components
The research team identified core process questions for the evaluation stages. These related to describing the quantity and quality of what was delivered, covered by the following 4 components: reach, fidelity, dose, and clinical engagement [26]. An overview of the evaluation components and methods is outlined in Table 1.
## Table 1. Research questions and key components in the process evaluation following the PROKID\(^a\) trial, Central Denmark Region.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Research questions</th>
<th>Applied method</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reach</td>
<td>- Who receives the intervention, and is the sample representative?</td>
<td>- Quantitative personal and clinical profiling of participants and nonparticipants</td>
<td>- Hospital Business Intelligence Register</td>
</tr>
<tr>
<td></td>
<td>- Why do patients disagree to participate?</td>
<td>- Patient-reported reasons for nonparticipation</td>
<td>- AmbuFlex database</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Questionnaire data/PROs(^b)</td>
<td>- Questionnaire data/PROs(^b)</td>
</tr>
<tr>
<td>Dose</td>
<td>- Who adheres to the intervention?</td>
<td>- Quantitative personal and clinical profiling in the level of adherence</td>
<td>- Hospital Business Intelligence Register</td>
</tr>
<tr>
<td></td>
<td>- To what extent have the patients received and engaged in the intervention?</td>
<td>- Number of distributed questionnaires and patient responses</td>
<td>- AmbuFlex database</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Number of item responses</td>
<td>- REDCap(^c) database</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Time spent completing the questionnaire</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Color-coded algorithm</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paper/web distribution</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Number of reminders</td>
<td></td>
</tr>
<tr>
<td>Fidelity</td>
<td>- Is the intervention delivered as intended?</td>
<td>- Number of clinics and physicians using the system</td>
<td>- AmbuFlex database</td>
</tr>
<tr>
<td></td>
<td>- Do the physicians incorporate the patients’ PRO responses in the consultation?</td>
<td>- Physician assessments of patient responses (contact/no contact)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Number of physicians involved in each patient pathway</td>
<td></td>
</tr>
<tr>
<td>Clinical engagement</td>
<td>- How do the physicians perceive the intervention in clinical practice?</td>
<td>- Individual semistructured interviews</td>
<td>- Involved physicians from all centers</td>
</tr>
</tbody>
</table>

\(^a\)PROKID: PRO measures in Kidney care.

\(^b\)PRO: patient-reported outcome.

\(^c\)REDCap: Research Electronic Data Capture.

### Reach

Reach was measured by the degree to which the intended target population participated in the intervention. Participation was defined as patients being eligible and willing to participate in the PROKID trial. Participants and nonparticipants were compared according to demographic information and PROs, measured using questionnaires and clinical data collected before entrance to the trial. All patients who declined to participate were encouraged to provide a reason for refusal. A complete overview of the data sources and the PROs and clinical outcomes is presented in Table S1 in Multimedia Appendix 1.

### Dose

The assessment encompassed measuring the quantity of the intervention components given to the patients, as well as gauging the extent to which patients engaged in and adhered to the intervention. Intervention adherence to PRO-based remote follow-up was exclusively calculated among patients in the intervention groups (N=105). Adherence was defined as the patient being able to complete the questionnaires, measure their blood pressure, and have blood samples taken prior to each of the 6 PRO consultations without assistance from the clinicians or the researcher. The percentage of successfully completed PRO consultations was used as a proxy measure for adherence. An adherence rate of 100% indicated that the patients successfully attended each consultation without assistance from anyone. Engagement in the intervention was measured by calculating as number of questionnaires distributed to patients/number of questionnaire responses received from patients, including the need for sending reminders. The extent of nonresponders was calculated. Furthermore, the item response rate was measured as total number of items distributed to patients/number of item responses. The time spent completing the questionnaire and paper/electronic version distribution was automatically logged in the AmbuFlex database [30].

### Fidelity

Fidelity measured the extent to which the intervention was delivered as planned. Fidelity was measured by counting the number of physicians using the system, including the extent to which the physicians assessed the patients’ responses. Furthermore, the mean number of physicians involved in each patient pathway was reported. The patients’ preferences for mode of contact with their physicians, as well as the extent and nature of the clinical assessment of their responses, was obtained. This information was automatically registered in the AmbuFlex database [30]. A study nurse or the project researcher obtained a quantification of patients who needed course support during the trial and registered in the Research Electronic Data Capture (REDCap) system [31]. Course support was defined as the patient asking for help from the researcher or the project nurse regarding completion of the study procedures. Evaluation components of dose and fidelity were described separately for each PRO intervention group.
Clinical Engagement

There is no universal definition of clinical engagement; it may be an attitude, a behavior, or an outcome [32]. We defined clinical engagement as the physicians’ perception, attitude, and satisfaction of using PRO actively in the decision-making processes in remote renal care. The clinical engagement toward using PRO-based remote follow-up in clinical practice was explored among the renal physicians who delivered the intervention. We used a qualitative approach inspired by Braun and Clarke’s [33] 6-phase thematic analysis. Individual semistructured interviews with purposively sampled physicians (N=10) most experienced with PRO-based follow-up were performed. The interviews were conducted by the project researcher (author BEG) with experience in qualitative research and occurred immediately after the physicians completed the patients’ final PRO consultation in one of the involved outpatient clinics. An interview guide was used to elicit clinical engagement, asking about the physicians’ views toward the PRO intervention and how they had experienced following the patients remotely (Table S2 in Multimedia Appendix 2). During each interview, notes were written down by BEG, and statements were summed up at the end of each interview and confirmed by the interviewee. These notes constituted the data material. For analysis of all the notes, an analytic coding scheme was developed based on the interview questions and an initial reading of the notes from the interviews (Table S3 in Multimedia Appendix 3) [34]. Citations used in this paper have been translated from Danish.

Data Analyses

Quantitative data were analyzed using descriptive statistics, such as frequencies, means (SDs), or median (IQRs), as appropriate. Descriptive quantitative information on reach, dose, and fidelity was provided. The presentation of variations between patients in the 2 intervention arms in terms of dose and fidelity were presented. The participation rate was calculated as number of patients consenting to participate/number of eligible patients. Reasons for nonparticipation were synthesized into categories by 2 researchers. We classified patients with low adherence by using the lower quartile. Thus, adherence<83% was the threshold defining low adherence to PRO-based remote follow-up. The adherence threshold was estimated from patients in both PRO intervention arms. Deceased and patients who withdrew due to ending follow-up at the hospital were excluded from the analyses. Participation and adherence data were linked to demographic data collected prebaseline of the trial. Personal and clinical characteristics of high or low adherence were described. Differences were determined using the χ² test.

Notes from the semistructured interviews constituted the qualitative data. We performed a thematic analysis to extract important aspects that influenced the physicians’ engagement toward PRO-based remote follow-up. First, all notes were read and re-read several times for familiarization with the data. Second, for an overview of the complete data, statements for each patient were entered into a Microsoft Excel spreadsheet by interview question. Third, preliminary themes that were relevant to the aim of evaluation were identified. Fourth, a coding framework was developed based on the initial themes. Finally, themes were identified and reviewed against the original statements and context. The quantitative and qualitative analyses and reporting were conducted prior to knowing the trial outcomes to avoid biased interpretation of the results [26].

Ethical Considerations

Patients provided written consent to participate in the trial, including the use of information from their medical records, the AmbuFlex database, and registers. Additionally, verbal consent was obtained from the physicians participating in the interviews. The study was approved by the Danish Authorities for Health Research (number 1-45-70-8-22).

Results

Reach

In total, 1060 patients with CKD were screened for participation in the PROKID trial (Figure 1). Of these, 320 (30.2%) patients were found eligible, and 152 (47.5%) agreed to participate in the PRO-based remote follow-up intervention. No statistically significant differences in patient and clinical factors were found between participants and nonparticipants (Table 2). A tendency toward lower participation by older age and lower health literacy was seen. The reasons for not participating in the PRO-based intervention are outlined in Table 3.
Table 2. Characteristics of patients with CKD⁹ (N=320) among participants and nonparticipants from Aarhus University Hospital, Gødstrup Hospital, and Viborg Regional Hospital in the PROKID⁹ trial from January 2019 to August 2021.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total patients (N=320)</th>
<th>Participants (n=152)</th>
<th>Nonparticipants (n=168)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), median (IQR)</strong></td>
<td>74 (12)</td>
<td>74 (11)</td>
<td>75 (11)</td>
</tr>
<tr>
<td>≤69, n (%)</td>
<td>91 (28.4)</td>
<td>48 (31.6)</td>
<td>44 (26.2)</td>
</tr>
<tr>
<td>70-79, n (%)</td>
<td>147 (45.9)</td>
<td>70 (46.1)</td>
<td>76 (45.2)</td>
</tr>
<tr>
<td>≥80, n (%)</td>
<td>82 (25.6)</td>
<td>34 (22.4)</td>
<td>48 (28.6)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>200 (62.5)</td>
<td>98 (64.5)</td>
<td>102 (60.7)</td>
</tr>
<tr>
<td><strong>Renal function</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eGFR⁹, mean (SD)</td>
<td>29.2 (6.7)</td>
<td>28.6 (6.2)</td>
<td>29.8 (7.1)</td>
</tr>
<tr>
<td>CKD 3b, n (%)</td>
<td>151 (47.2)</td>
<td>70 (46.1)</td>
<td>81 (48.2)</td>
</tr>
<tr>
<td>CKD 4/5, n (%)</td>
<td>155 (48.4)</td>
<td>82 (53.9)</td>
<td>73 (43.5)</td>
</tr>
<tr>
<td>Missing values, n (%)</td>
<td>14 (4.4)</td>
<td>N/A⁹</td>
<td>14 (8.3)</td>
</tr>
<tr>
<td><strong>Educational level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (&lt;10 years): none/short &lt;1 year</td>
<td>93 (29.1)</td>
<td>37 (24.3)</td>
<td>56 (33.3)</td>
</tr>
<tr>
<td>Medium (10-12 years): skilled worker/short</td>
<td>148 (46.3)</td>
<td>71 (46.7)</td>
<td>77 (45.8)</td>
</tr>
<tr>
<td>Long (&gt;12 years): middle/long higher</td>
<td>58 (18.1)</td>
<td>29 (19.1)</td>
<td>29 (17.3)</td>
</tr>
<tr>
<td>Missing values</td>
<td>21 (6.6)</td>
<td>15 (9.9)</td>
<td>6 (3.6)</td>
</tr>
<tr>
<td><strong>Labor market affiliation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>43 (13.4)</td>
<td>20 (13.2)</td>
<td>26 (15.5)</td>
</tr>
<tr>
<td>Unemployed (retirement, early retirement)</td>
<td>259 (80.9)</td>
<td>119 (78.3)</td>
<td>140 (83.3)</td>
</tr>
<tr>
<td>Missing values</td>
<td>18 (5.6)</td>
<td>13 (8.6)</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td><strong>Comorbidity (Charlson index), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (&gt;2)</td>
<td>110 (34.4)</td>
<td>52 (34.2)</td>
<td>58 (34.5)</td>
</tr>
<tr>
<td>Medium (1-2)</td>
<td>195 (60.9)</td>
<td>94 (61.8)</td>
<td>101 (60.1)</td>
</tr>
<tr>
<td>Low (0)</td>
<td>15 (4.7)</td>
<td>6 (3.9)</td>
<td>9 (5.4)</td>
</tr>
<tr>
<td><strong>Health Literacy Questionnaire (HLQ) 4: social support for health, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.3 (0.55)</td>
<td>3.2 (0.50)</td>
<td>3.2 (0.59)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>3.2 (0.6)</td>
<td>3.2 (0.6)</td>
<td>3.2 (1)</td>
</tr>
<tr>
<td>Missing values, n (%)</td>
<td>21 (6.6)</td>
<td>15 (9.9)</td>
<td>6 (3.6)</td>
</tr>
<tr>
<td><strong>HLQ 6: ability to actively engage with health care providers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.9 (0.71)</td>
<td>3.9 (0.70)</td>
<td>3.8 (0.71)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4 (0.8)</td>
<td>4 (0.8)</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Missing values, n (%)</td>
<td>21 (6.6)</td>
<td>14 (9.2)</td>
<td>7 (4.2)</td>
</tr>
<tr>
<td><strong>HLQ 9: understanding health information well enough to know what to do</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.9 (0.67)</td>
<td>3.9 (0.60)</td>
<td>3.9 (0.72)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4 (0.8)</td>
<td>3.8 (0.8)</td>
<td>4 (0.8)</td>
</tr>
<tr>
<td>Missing values, n (%)</td>
<td>21 (6.6)</td>
<td>14 (9.2)</td>
<td>7 (4.2)</td>
</tr>
<tr>
<td><strong>Self-efficacy (General Self-Efficacy [GSE] scale)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>29.8 (5.4)</td>
<td>30 (5)</td>
<td>29.6 (5.7)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>30 (8)</td>
<td>30 (7)</td>
<td>30 (8)</td>
</tr>
<tr>
<td>Missing values, n (%)</td>
<td>23 (7.2)</td>
<td>14 (9.2)</td>
<td>9 (5.4)</td>
</tr>
</tbody>
</table>
Table 3. Reasons for nonparticipating in the PRO-based remote follow-up intervention among patients with CKD (n=168) at Aarhus University Hospital, Godstrup Hospital, and Viborg Regional Hospital.

<table>
<thead>
<tr>
<th>Reasons</th>
<th>Nonparticipants (n=168), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient-reported reasons</strong></td>
<td></td>
</tr>
<tr>
<td>Unknown (information unavailable)</td>
<td>17 (10.1)</td>
</tr>
<tr>
<td>Did not want to participate (no further reasons reported)</td>
<td>15 (9.1)</td>
</tr>
<tr>
<td>Could not cope with participating</td>
<td>14 (8.3)</td>
</tr>
<tr>
<td>Preferred telephone consultations/no attendance</td>
<td>13 (7.7)</td>
</tr>
<tr>
<td>Had enough of coping with comorbidity</td>
<td>13 (7.7)</td>
</tr>
<tr>
<td>Visual or hearing disability</td>
<td>11 (6.5)</td>
</tr>
<tr>
<td>Memory issues</td>
<td>7 (4.2)</td>
</tr>
<tr>
<td>Preferred the standard follow-up program</td>
<td>6 (3.6)</td>
</tr>
<tr>
<td>Attends another outpatient follow-up (requiring attendance)</td>
<td>5 (3.0)</td>
</tr>
<tr>
<td>Did not wish to complete the questionnaire</td>
<td>4 (2.4)</td>
</tr>
<tr>
<td><strong>Clinical reported reasons</strong></td>
<td></td>
</tr>
<tr>
<td>Health care professional reasons&lt;sup&gt;c&lt;/sup&gt;</td>
<td>52 (31.0)</td>
</tr>
<tr>
<td>Other&lt;sup&gt;d&lt;/sup&gt;</td>
<td>10 (6.0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>PRO: patient-reported outcome.
<sup>b</sup>CKD: chronic kidney disease.
<sup>c</sup>Ended follow-up, rapid illness progression, compliance issues, comorbidity.
<sup>d</sup>Died before enrollment, departed, other study participation.

Dose

Of the 152 patients, 105 (69.1%) were allocated to the 2 PRO-based intervention arms. Of those, 8 (7.6%) patients died during follow-up and 4 (3.8%) left the study due to termination of outpatient follow-up, leaving 93 (88.6%) patients adhering to PRO-based remote follow-up (Figure 1). The mean adherence rate to the intervention was 82% (95% CI 76-87). In total, 70 (75.3%) of 93 patients had an adherence rate ≥83%, as indicated by the dotted vertical line in Figure 3. The distribution of patients according to the level of adherence rates is shown in Figure 3.
Overall, no statistical differences were found in patient and clinical factors between patients with high or low intervention adherence, albeit a tendency toward lower adherence by poor self-reported health status and lower patient activation was seen (Table S4 in Multimedia Appendix 4). Patients in the PRO-based follow-up group had a significantly lower adherence rate than patients in the PRO-based telephone follow-up group (Table 4).

The overall response rate was 99.1% (539/544). Accordingly, the item response rates were relatively high, with a mean response rate of 96.7% (14,314/14,788; Table 4). In total, 453 (84%) of the 539 patients responded electronically, and they spent a median of 9.22, IQR (6.87) minutes completing the questionnaire.
Table 4. Results of the process evaluation of the PROKID\textsuperscript{a} trial among patients with CKD\textsuperscript{b} (N=101) from Aarhus University Hospital, Godstrup Hospital, and Viborg Regional Hospital.

<table>
<thead>
<tr>
<th>Elements of process evaluation</th>
<th>Total patients (N=101)</th>
<th>PRO\textsuperscript{c}-based follow-up (n=48)</th>
<th>PRO-based telephone follow-up (n=53)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dose</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence rate, mean % (95% CI)</td>
<td>82 (76-87)</td>
<td>73 (63-83)</td>
<td>90 (84-95)</td>
</tr>
<tr>
<td>Questionnaire response rate response/to-</td>
<td>539/544 (99.1)</td>
<td>249/250 (99.6)</td>
<td>290/294 (98.6)</td>
</tr>
<tr>
<td>total, n/N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item response rate total, n/N (%)</td>
<td>14,314/14,788 (96.7)</td>
<td>6329/6696 (95.5)</td>
<td>7985/8092 (98.7)</td>
</tr>
<tr>
<td>Time completing questionnaire (minutes), median (IQR)</td>
<td>9.22 (6.87)</td>
<td>9.37 (6.97)</td>
<td>9.20 (6.78)</td>
</tr>
<tr>
<td>Reminders, n/N (%)</td>
<td>367/367 (100)</td>
<td>174/367 (47.4)</td>
<td>193/367 (52.6)</td>
</tr>
<tr>
<td>Web responses, n/N (%)</td>
<td>453/539 (84.0)</td>
<td>199/249 (89.9)</td>
<td>254/290 (87.6)</td>
</tr>
<tr>
<td>Assessed questionnaires, n/N (%)</td>
<td>477/539 (88.5)</td>
<td>215/249 (86.3)</td>
<td>262/290 (90.3)</td>
</tr>
<tr>
<td>Course support\textsuperscript{d}, n/N (%)</td>
<td>13/101 (12.9)</td>
<td>9/48 (18.8)</td>
<td>4/53 (7.5)</td>
</tr>
<tr>
<td>Involved physician per patient, median (range)</td>
<td>3 (1-7)</td>
<td>3 (1-6)</td>
<td>3 (1-7)</td>
</tr>
<tr>
<td><strong>Color coded algorithm, n/N (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red\textsuperscript{e}</td>
<td>N/A\textsuperscript{f}</td>
<td>151/249 (60.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Yellow\textsuperscript{g}</td>
<td>N/A</td>
<td>83/249 (33.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Green\textsuperscript{h}</td>
<td>N/A</td>
<td>15/249 (6.0)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Clinical assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact, n/N (%)</td>
<td>417/539 (77.4)</td>
<td>137/249 (55.0)</td>
<td>280/290 (96.6)</td>
</tr>
<tr>
<td>No contact, n/N (%)</td>
<td>122/539 (22.6)</td>
<td>112/249 (45.0)</td>
<td>10/290 (3.4)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PROKID: PRO measures in Kidney care.
\textsuperscript{b}PRO: patient-reported outcome.
\textsuperscript{c}CKD: chronic kidney disease.
\textsuperscript{d}The patient asked for help from the researcher or the project nurse regarding completion of study procedures.
\textsuperscript{e}High symptom burden or the patient wishes for contact.
\textsuperscript{f}N/A: not applicable.
\textsuperscript{g}Some symptom burden or the patient may need contact.
\textsuperscript{h}No symptom burden and the patient needs no contact.

Fidelity

The patients completed 539 questionnaires, and in 477 (88.5%) of the 539 cases, a physician assessed their responses (Table 4), leaving 11.5% (62/539) of the questionnaires unnoticed by a physician. The algorithm indicated that 234 (94%) of the 249 patients in the PRO-based follow-up group experienced some disease burden or needed clinical contact flagged by a red or yellow color code. When this was assessed, the physicians found that 137 (55%) of the 249 patients needed to be contacted and 112 (45%) did not need to be contacted. In 288 (69.1%) of all the 417 cases, the physician initiated the contact, and in 129 (30.9%) of the cases, the patient requested contact. The primary causes of contact (242/417, 58%) were attributed to clinical data, including blood samples and blood pressure. Contact based on PRO answers alone accounted for 22.1% (92/417) of the contacts, while concerns related to medication and a precautionary approach prompted the remaining contacts (83/417, 19.9%). In total, 10 (3.4%) of the 290 patients in the PRO-based telephone follow-up group were unavailable when the physician contacted them by phone. On average, 3 (range 1-7) physicians were involved in each patient pathway. A minority of 13 (12.9%) of 101 patients needed assistance to complete study procedures. Common reasons for needing assistance were measuring blood pressure and obtaining information about how to undergo blood tests or how to access the questionnaire. A study nurse or the researcher contacted the patient to address the reasons behind noncompliance and collaboratively devise a plan to ensure future adherence. The majority of those needing assistance were in the PRO-based follow-up group.

Clinical Engagement

The physicians emphasized that their perception and commitment to the use of PRO-based remote follow-up was influenced by the fact that the PROKID intervention enriched their consultations, as a physician explained:

https://formative.jmir.org/2024/1/e48173
I believe patients receive a more thorough and attentive consultation through this approach, where all symptoms are carefully examined. [Physician 9]

However, PRO-based follow-up also seemed to limit their ability to assess the patients’ health status, as explained in the following by a physician:

I find it quite straightforward to assess individuals over the phone. However, what I miss is observing their movement from the waiting room to the consultation room; it reveals a great deal, such as being able to better evaluate characteristics for instance like skin color and psychical impairment. [Physician 4]

On the one hand, all physicians felt that the PROKID intervention is beneficial for their ability to manage and identify patients’ symptoms and estimate their need for contact. The clinicians’ motivation to review the patients’ responses was heightened because the questionnaire included clinical data, such as information about blood pressure and weight.

The clear advantage lies in having everything compiled in this questionnaire, which significantly aids in retaining and managing all the information. [Physician 3]

On the other hand, the physicians were also concerned that they might overlook something in the patients’ health status. This 2-sided notion was described in an interview in the following way:

You might lose some connection with the patients, the familiarity that comes with knowing them, and eventually find yourself relying more on questionnaires and blood tests. This shift can make the interaction feel less personal…Conversely, we receive something self-reported (PRO) that delves into deeply personal matters. The risk lies in potentially overlooking something, especially when the patient prefers minimal contact. [Physician 10]

To remedy this risk of overlooking symptoms, the physicians wished to combine remote and face-to-face consultations in the future implementation of the PROKID intervention’s remote follow-up. Specifically, the assessment of patients in the PRO-based follow-up group troubled the physicians, as these patients were not automatically contacted by phone as opposed to patients in the PRO-based telephone follow-up group. This was described by a physician in the following way:

I sometimes feel concerned about patients in PRO-based follow-up, who does not visit the hospital. For instance, does this woman truly comprehend the extent of her illness? Is she receiving the necessary information? She may be unaware of what she does not know. [Physician 5]

The thematic analysis revealed important aspects that influenced the physicians’ engagement in using PRO. These aspects represented facilitators of and barriers to the implementation of PRO-based remote follow-up in clinical practice. As shown in Textbox 1, the physicians’ engagement in the implementation of PRO-based remote follow-up was supported by several facilitating factors—for instance, PRO enabled an excellent overview of the patients’ health status and increased the possibility for visualization and clarity, as described in this quote:

I wouldn’t have inquired about this if the patient had not communicated it to me. It has completely altered the course of our conversation. [Physician 5]
Facilitators of PRO use in remote care:
• Better overview of the patient’s overall health condition
• Ensured relevant data are available and lined up
• A great basis for visualizing the patient provided by the combination of PRO and blood tests
• Qualified phone consultations
• Enhanced the patients’ ability to reflect on their symptoms
• Enabled patient-centered focus
• More convenient for the physician and the patient compared to usual follow-up
• An excellent supportive tool for patients during a period of stable illness

Barriers to using PRO in remote care:
• Decrease in physicians’ ability to sense the patients’ condition through clinical observation
• Decrease in physicians’ knowledge of patients
• Decrease in crucial knowledge on patient medical history
• Induced risk of forgetting the responsibility of the patient
• Provoked a lower personal relationship with the patient
• Training needed to manage patients’ health status remotely
• Specific patient flow needed to maintain clinical skills of using PRO
• Incentive required by physicians to assess patients’ PRO responses

PRO was also perceived as enabling patient-focused care, to be convenient for both the patient and the clinic, as described by this physician:

_It must be convenient for someone like him, who does not have to make the one-hour trip every third month just to let me pressure his angles a bit. We can just as easily handle that over the phone._ [Physician 2]

However, some barriers to the physicians’ engagement in implementing PRO-based follow-up worked against some of the facilitating factors. Specifically, the physicians faced challenges in accurately evaluating the patients’ health condition and determining their requirement for medical attention and intervention. This was mainly due to the lack of face-to-face interaction and the inability to use their clinical observation skills.

### Discussion

**Principal Findings**

In this process evaluation, the intervention processes of PRO-based remote follow-up (the PROKID trial) were evaluated [7]. In total, 47.5% of the eligible patients participated, and they reflected the target population. Overall, we found high percentages of the utility of the system. The dose ranged from 73% to 99%, with a mean adherence rate of 82%. Fidelity showed high usage by physicians, with 88.5% of the total questionnaires being assessed. In total, 12.9% of patients needed assistance to comply with study procedures; this need was predominantly observed in the PRO-based follow-up group. Ten physicians were interviewed about their engagement in the PROKID intervention; most of them rated PRO as being of additional value. However, some barriers, such as decreased knowledge of patients and a lack of using their clinical observational skills, were highlighted.

Overall, the percentages of utility were high in the PRO-based intervention. We believe that several reasons may explain this success. First, an internal pilot study was conducted in an outpatient clinic before trial onset. The pilot study showed that patients were willing to participate and that it was possible to deliver the intervention. We spent time observing the routines and activities in the outpatient clinic to inform our intervention. We had continuous dialogues and meetings with staff members and patients to target the intervention to existing clinical practice. Several studies have highlighted the importance of patients and health care providers knowing why PROs are used and the relevance of the questions asked [35,36]. The fact that the intervention was implemented in a real-life setting and that the users were involved in its development may have impacted patient and clinical engagement with PRO-based remote follow-up [4,37].

Another possible explanation for the high usage percentages may be the procedures used to approach and recruit patients. We included newly referred patients, and it became clear that most preferred no follow-up or remote follow-up, probably due to a feeling of being more ill when entering a hospital [38] or a general lack of experience attending outpatient follow-up. Patient-reported reasons for nonparticipation confirmed this finding, as 7% of the eligible patients declined participation due to preference for telephone consultation or no attendance. Thus, trial participation had the advantage that patients could attend
remote follow-up, which was not part of the routine practice at the time. Finally, the degree of involvement of the project researcher and the clinical study nurse in the recruitment phase and during the trial may have motivated patients to participate and made them adhere to and use the intervention. In total, 12.9% of the patients needed course support, highlighting the benefit of having someone to call when problems occur. Prior findings from a parallel qualitative study, in which we interviewed 15 patients in the PROKID trial, support this finding [38].

A well-known concern in RCTs is the challenge of inducing selection bias caused by participation of a selected group failing to represent the target population [24]. We did not find any personal or clinical differences between participants and nonparticipants in our study, which adds value to the generalization of the results of the PROKID trial. Wiegel et al [12] found, in a recently published systematic review, an overall adherence to repetitive electronic PROs in populations with chronic diseases ranging between 61% and 96%, which corresponds well with the mean adherence of 82% found in our study. However, the definition of adherence is known to vary between studies [12]. We believe that having a predefined cutoff point for low adherence may help researchers and clinicians identify patients needing further assistance. Presumably, due to low power, we found no differences in personal or clinical factors between patients with low or high adherence. This might challenge clinicians seeking to identify patients most suitable for PRO-based remote follow-up. The definition and threshold for adherence correspond to the goal of using remote PRO in the outpatient clinic, as in the real-life setting, patients should be capable of completing the PRO questionnaire by themselves.

We found a high response rate in our study, contrary to other previous findings [35]. We believe the 2 main reasons for this were the high use of reminders and the fact that a study nurse called the patient if questionnaire responses had not been submitted at the date of the PRO consultations. The use of reminders is known to increase the response rate [39]. Knowledge of the extent of this phenomenon is important to convey to the outpatient clinics when the intervention is widely implemented.

Another interesting finding is the number of physicians involved in each patient pathway. Preferably, patients should have 1 contact physician during follow-up, but we saw that each patient had contact with 3 (range 1-7) physicians during the 18 months’ trial. Evidence from a meta-analysis suggests that for the safety and continuity of care, it is essential to have as few different health care professionals as possible [40]. A qualitative study integral to the PROKID intervention concluded that barriers to patient engagement in PRO-based remote follow-up are unfamiliarity with the physician and remote follow-up challenging the patient-physician relationship [38]. Furthermore, the involvement of multiple physicians in reviewing each patient over the 18 months posed a challenge, leading to diminishing familiarity with the medical history of the patients. Remote monitoring is increasingly substituting traditional hospital visits [41]. A recent study argued that telephone follow-up is as safe as visits to the outpatient clinic, especially if the patient is acquainted with the physician [42]. Thus, it seems crucial to incorporate individual patient-physician consultations to the extent possible.

A prior study of field observations during the PRO consultations showed that the extent to which the physicians incorporated the patient responses into the consultations reached almost 100% [38]. However, data from this process evaluation study found that the physicians merely assessed 88.5% of the questionnaire responses received. This might indicate observer bias in the qualitative study [38], showing a falsely high result, although it might also indicate a false lower result in our study because of recall bias, as it seemed as if the physicians actually opened and assessed the patients’ PRO responses but forgot to approve and document the responses in the AmbuFlex database. However, it is a well-known challenge to get clinicians to assess and approve patients’ PRO responses [19,43]. The qualitative analyses showed that the physicians found PRO-based remote follow-up to enrich the consultations, but it also induced a risk of failing to notice important symptoms. Prior research has shown that engagement decreases when activities or tasks are perceived as wasteful of resources or creating more harm than benefit [44]. Therefore, enhancing the physicians’ satisfaction and perception of the usability of implementing PROs in remote care seems crucial. The physicians emphasized that they needed an incentive to open the patients’ PRO responses. During the interviews, it became clear that the patients’ blood pressure and weight were the physicians’ primary objects of interest. This result was supported in a qualitative study on patient perspectives [38] and has also been found to be an issue in other studies [22,45]. Hence, an important factor driving the assessment of patients’ PRO responses was the incorporation of the self-reported clinical outcomes (blood pressure and weight) in the questionnaire.

The color-coded algorithm showed that most of the patients needed contact with a physician or were burdened by disease-specific symptoms. Thus, the proportion of patients contacted by physicians was relatively high, which may indicate a high claim for clinical attention in this patient group. However, it may also indicate a sensitive algorithm [4]. The clinical experts aimed to develop an algorithm with high sensitivity since a low level of false-negative cases was more important than a high level of false-positive cases. Thus, the physicians’ ability to assess whether to contact patients was pivotal for the follow-up process. Further studies that consider the results of the green-yellow-red color-coded algorithm will need to be undertaken.

Strengths and Limitations

Process evaluation was considered integral to the PROKID trial, and the results of this process evaluation were analyzed prior to knowledge of the trial outcomes. Thereby, the risk of bias was reduced [46]. This study enabled us to perform a more detailed examination of the process of recruitment and intervention adherence to inform the interpretation of trial results and generalizability [47]. A main strength of this study is that it included patients from a well-defined population covering all centers following patients with CKD in Central Denmark Region. Quantitative and qualitative methods complemented one another, and a combination of self-reported, qualitative data
and register-based data was used to link different data sources. Another key strength of this study was the use of Steckler and Linnan’s [26] process evaluation framework, which allowed us to structure the process evaluation appropriately. Observations and interviews had the potential to influence the involved physicians’ and patients’ behavior and commitment. However, we believe that the timing of the evaluation in the finishing part of the trial limited this effect. The evaluator responsible was also the principal investigator in the PROKID trial, which provided insight into and understanding of the context, as an outsider could not have added to the findings.

This study also has some limitations. Initial development and use of a logic model may have better illustrated the underpinning theory [26]. The absence of predefined core questions was also a notable limitation. In the qualitative data analyses, the researcher’s familiarity with the participating departments may have induced a certain risk of bias. This may have led to a more positive attitude toward PRO-based remote follow-up among the physicians. However, it may also have made them confident that critical comments would be received constructively.

Conclusion

The PRO-based PROKID intervention was generally well received by patients and physicians, as we found high percentages of utility across all process evaluation components. Our findings highlight that remote PRO-based follow-up is feasible in a clinical setting and may be a relevant and valuable tool when remotely monitoring patients with CKD. We found high adherence and response rates among the patients, and the physicians assessed the majority of patient responses. However, it seems important to have a key figure who helps patients in need of assistance to comply with PRO-based remote follow-up. The physicians emphasized both enabling factors, such as improved overview and patient-focused care, and barriers, particularly the complexities of evaluating health conditions without direct interaction. Suggestions for future implementations included a combination of remote and face-to-face consultations to address this concern.

Acknowledgments

BEG, AdT, and LMVS conceived the study. BEG collected data. BEG and AKK performed the analyses. All authors contributed to the data interpretation and critical revision of the manuscript. All authors have read and approved the final manuscript and stand by the integrity of the entire work. This paper was written without the use of artificial intelligence technology. The Karen Elise Jensen Foundation and the AmbuFlex – Centre for Patient-reported Outcomes kindly funded this study.

Data Availability

The data sets generated and analyzed during this study are available in anonymous form from the corresponding author on request or by contacting the AmbuFlex – Centre for Patient-reported Outcomes.

Conflicts of Interest

BEG, RG, NHH, AKK, and LMVS are employees of AmbuFlex – Centre for Patient-reported Outcomes. However, this is a public-funded department with no financial benefit.

Multimedia Appendix 1
Investigated patient and clinical factors.
[DOCX File, 15 KB - formative_v8i1e48173_app1.docx ]

Multimedia Appendix 2
Interview guide.
[DOCX File, 24 KB - formative_v8i1e48173_app2.docx ]

Multimedia Appendix 3
Preliminary coding scheme aiming to explore the physicians' perceptions and experiences using PRO-based remote follow-up. PRO: patient-reported outcome.
[PDF File (Adobe PDF File), 565 KB - formative_v8i1e48173_app3.pdf ]

Multimedia Appendix 4
Characteristics of patients adhering to PRO-based follow-up according to demographic, disease-related, and individual determinants divided into high (≥83%) and low (<83%) adherence (n=93). PRO: patient-reported outcome.
[PDF File (Adobe PDF File), 441 KB - formative_v8i1e48173_app4.pdf ]

Multimedia Appendix 5
CONSORT-EHEALTH checklist.
References


Abbreviations

CKD: chronic kidney disease
eGFR: estimated glomerular filtration rate
PRO: patient-reported outcome
PROKID: PRO measures in Kidney care
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

©Birgith Engelst Grove, Annette de Thurah, Per Ivarsen, Ann Katrine Kvigsgaard, Niels Henrik Hjollund, Regine Grytnes, Liv Marit Valen Schougaard. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e48173
Exploring the Use of Customized Links to Improve Electronic Engagement With Sexual and Reproductive Health Care Among Young African American Male Individuals: Web-Based Survey Study

Sandy Arena¹, BSN; Mackenzie Adams¹, MPH; Jade Burns¹, PhD
School of Nursing, University of Michigan, Ann Arbor, MI, United States

Abstract

Background: Research has shown that heterosexual African American male individuals aged 18-24 years have a higher prevalence of sexually transmitted infections (STIs) and are more likely to engage in risky sexual behavior. There is a critical need to promote sexual reproductive health (SRH) services among this population, especially in urban settings. Young African American male individuals use social media platforms to access health information, showcasing the potential of social media and web-based links as tools to leverage electronic engagement with this population to promote SRH care.

Objective: This study aims to explore electronic engagement with young African American male individuals in discussions about SRH care. This paper focuses on the recruitment and social media marketing methods used to recruit young, heterosexual African American male individuals aged 18-24 years for the Stay Safe Project, a larger study that aims to promote SRH services among this population in Detroit, Michigan. We investigate the use of TinyURL, a URL shortener and customized tool, and culturally informed social media marketing strategies to promote electronic engagement within this population.

Methods: Participants were recruited between December 2021 and February 2022 through various modes, including email listserves, Mailchimp, the UMHealthResearch website, X (formerly Twitter), Facebook, and Instagram. Images and vector graphics of African American male individuals were used to create social media advertisements that directed participants to click on a TinyURL that led to a recruitment survey for the study.

Results: TinyURL metrics were used to monitor demographic and user data, analyzing the top countries, browsers, operating systems, and devices of individuals who engaged with the customized TinyURL links and the total human and unique clicks from various social media platforms. Mailchimp was the most successful platform for electronic engagement with human and unique clicks on the custom TinyURL link, followed by Instagram and Facebook. In contrast, X, traditional email, and research recruiting websites had the least engagement among our population. Success was determined based on the type of user and follower for each platform, whether gained in the community through sign-ups or promoted at peak user time and embedded and spotlighted on nontraditional media (eg, social media sites, blogs, and podcasts) for the user. Low engagement (eg, traditional email) from the target population, limited visibility, and fewer followers contributed to decreased engagement.

Conclusions: This study provides insight into leveraging customized, shortened URLs, TinyURL metrics, and social media platforms to improve electronic engagement with young African American male individuals seeking information and resources about SRH care. The results of this study have been used to develop a pilot intervention for this population that will contribute to strategies for encouraging sexual well-being, clinic use, and appropriate linkage to SRH care services among young, heterosexual African American male individuals.
Introduction

Nearly half of the 26 million people who are diagnosed with sexually transmitted infections (STIs) in the United States are younger than 25 years old [1]. Within this population, heterosexual African American male individuals aged between 18 and 24 engage in higher rates of risky sexual behavior (eg, early sexual initiation, inconsistent condom use, sex while under the influence, and multiple partners) and face a disproportionate burden of STIs compared to other adolescents and young adults [2-4].

In urban areas, this problem is particularly acute. African American male individuals who engage in low-risk sexual behavior, such as using condoms consistently or having sex with only 1 partner, are still at an increased risk of STIs due to risky social and sexual networks within urban communities that have high rates of violence, incarceration, and poverty [5-7]. A study by Marcell et al [8] found that among an urban sample of 70 African American and Hispanic young male individuals aged between 15 and 24 years, 67% reported having visited a health care provider in the past year, yet only 50% reported receiving HIV testing and only 39% reported receiving STI testing. These daunting statistics regarding sexual health indicate that young African American male individuals have a substantially greater need for sexual reproductive health (SRH) services (eg, preventative services, vaccination, diagnostic testing, and treatment) than other groups [9].

Studies show the significant factors in a young African American male individual’s decision to seek SRH services include concerns about confidentiality, self-confidence, and choice of provider, in addition to stigma, cost of care and contraceptives, and long wait times [8-10]. To explore the central phenomenon of SRH use and clinic access among young African American male individuals, community engagement with this population is crucial. Community engagement is particularly important for reducing HIV and STI disparities among young African American male individuals, and a lack of engagement with SRH care among this population can increase the risk of STIs and other health disparities [11]. Burns et al [12] found that social media can be used to effectively engage with young African American male individuals and provide them with information and resources regarding SRH care. Therefore, engaging with young African American male individuals electronically through social media may be a helpful strategy to ensure that they have access to the information and resources needed to make informed decisions about their SRH.

Social media platforms, particularly among African American adolescents and young adults, are already being used to seek general and sensitive health information due to their accessibility and anonymity [13]. Teadt et al [13] highlight the potential of social media platforms to serve as valuable tools for promoting sexual health awareness, prevention, and intervention development aimed at this population. For instance, in 2018, researchers from the Tulane School of Public Health and Tropical Medicine launched “Check It Nola” [14], a website for their research study geared toward using novel methods in promoting SRH care among heterosexual African American male individuals aged between 15 and 24 years in New Orleans, Louisiana. The “Check It” program used culturally informed marketing strategies (eg, using images of African American male individuals wearing crowns and the tagline “a king knows his status”) to resonate with young African American male individuals on popular social media platforms such as Instagram (Meta Platforms), X (formerly Twitter; X Corp), and Facebook (Meta Platforms), in addition to a website, as a means to engage them in SRH care [15]. This study seeks to create novel strategies similar to the “Check It” program by further exploring electronic engagement with SRH and preventative care services by young African American male individuals.

Following these existing approaches, we developed this study to explore an innovative approach to engaging with young African American male individuals about SRH care through culturally informed advertising on social media platforms (Instagram, Facebook, and X) in addition to traditional methods such as email, Mailchimp, and a participant recruiting website (UMHealthResearch). We used the TinyURL web service to create customized, shortened URLs that allowed us to track which social media and web-based platforms young African American male individuals were engaging with most. This study is part of the Stay Safe Project, a larger study which used a mixed method approach of surveying and focus groups to identify the positive and negative experiences of young African American male individuals when accessing SRH services at community health centers in a large city, the results of which will be reported elsewhere. It is important to recognize that community engagement is essential in order to reduce sexual health disparities among young African American male individuals. The use of social media and web-based links to promote SRH services and increase awareness may be used as an effective way to reach this population and be used to mediate health care–seeking behavior. This study describes the process for electronic engagement with young African American male individuals with regard to SRH services and may be used to promote long-term risk-reduction methods among young male individuals in the community.

Methods

Overview

This study examines electronic recruitment methods for young, heterosexual African American male individuals as part of a larger project, the Stay Safe Project, that seeks to promote SRH services for this population.
Study Population

The priority population for this study was young, heterosexual African American male individuals. Inclusion criteria for the Stay Safe Project were: (1) aged between 18 and 24 years; (2) self-identify as male; (3) live in Detroit, Michigan; (4) self-identify as African American or Black; (5) self-identify as heterosexual and report having sex only with women in the last year; and (6) speak, read, and write in English.

Recruitment

Participants were recruited between December 2021 and February 2022 through traditional email listserves, Mailchimp (an email marketing service or newsletter; Intuit), the UMHealthResearch website for web-based recruitment [16], X, and through Facebook and Instagram using paid advertising.

For each recruitment platform, we created a unique, shortened TinyURL link that led to a web-based recruitment survey for the study so we could track electronic engagement and user data. URL, which stands for Uniform Resource Locator, was first defined in 1994 as “the syntax and semantics for a compact string representation of a resource available through the internet” [17]. In modern terms, a URL is a unique address that identifies a specific web page on the internet. TinyURL is a URL shortening service that allows customers to create branded, customized URLs and track analytics for those URLs [18]. TinyURL allows customers to create shortened, customized links. Today, the terms “URL” and “link” are often used interchangeably. However, a link is defined as a clickable element on a web page that will lead to a specific URL [19]. Each TinyURL link was customized to the platform. All customized links directed participants to the same URL for our recruitment survey.

We offered a US $20 gift card to those who completed the survey, met the study criteria, and participated in the focus groups to incentivize participants to click on the links to our recruitment survey.

Social Media Marketing and Analysis

Photo and banner advertisements were designed to capture the attention of young African American male individuals with the aim of recruiting eligible individuals for this study. Advertisements included images or vector graphics of African American male individuals with headlines calling for the participation of young African American male individuals and visible information on requirements and details about the focus group study, in addition to the US $20 gift card compensation. Advertisements were designed to grab the attention of social media users and were created with bright, appealing colors. The customized TinyURL links, which participants were directed to through our social media advertisements, helped the research team track total clicks, total unique clicks, and total human clicks on our survey link for each recruitment platform (Instagram, Facebook, X, traditional email, Mailchimp, and UMHealthResearch). TinyURL also allowed the research team to track world traffic (where in the world people were clicking on our link), the top browsers and devices people were using, and the popular days and times people were accessing our recruitment survey through Qualtrics (Silver Lake) from each custom link. The survey was developed to evaluate demographic data to determine eligibility for the study in addition to assessing satisfaction with SRH care and behavioral health risk characteristics (eg, sexual activity and substance use).

Additionally, we used Facebook and Instagram Insights (within Meta Business Suite) to further monitor and evaluate the performance of the recruitment campaigns, as well as the audience’s engagement with them. Insights is a tool offered by Meta that provides data and metrics about the performance of profile pages and individual posts on Facebook and Instagram. Using Insights, we were able to track content interactions such as the number of views, likes, comments, and shares, in addition to impressions, reach, link clicks, and profile visits on Facebook and Instagram. We were able to tailor our paid advertisements by sex, age range, and location (male individuals aged between 18 and 24 in Detroit). We tracked the times and days people were the most active on Facebook and Instagram and tailored our posting schedule accordingly with the hopes of increasing engagement with young African American male individuals in our community. Table 1 shows the definitions of terms we used for tracking electronic engagement.
Table 1. Terms and definitions.

<table>
<thead>
<tr>
<th>Terms</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total clicks</td>
<td>The total number of times a link was clicked on [20].</td>
</tr>
<tr>
<td>Unique clicks</td>
<td>The first click of a link by a unique user [20].</td>
</tr>
<tr>
<td>Human clicks</td>
<td>The total number of clicks on a link identified as authentic (from humans, as opposed to bots) [20].</td>
</tr>
<tr>
<td>World traffic</td>
<td>The countries and subregions where people were clicking on a link [20].</td>
</tr>
<tr>
<td>Top browser</td>
<td>The top browser used when a user clicks on a link to access and view websites (eg, Google Chrome [Google LLC], Safari [Apple Inc], Firefox [Mozilla], and Microsoft Edge [Microsoft Corporation]) [20].</td>
</tr>
<tr>
<td>Top operating system</td>
<td>The top operating system on a device used to click a link (eg, Microsoft Windows, macOS, iOS, and Android) [20].</td>
</tr>
<tr>
<td>Top device</td>
<td>The top device used to click on a link (eg, smartphones, computers, and tablets) [20].</td>
</tr>
<tr>
<td>Insights</td>
<td>Social media performance metric tool available for business accounts on Instagram and Facebook that shows the number of views, likes, comments, shares, impressions, reach, profile visits, and link clicks [21,22].</td>
</tr>
<tr>
<td>Impressions</td>
<td>The total number of times a post or ad has been seen. This includes both paid and unpaid impressions and may include multiple views by the same person [21,22].</td>
</tr>
<tr>
<td>Reach</td>
<td>The number of unique people that have seen your post or ad on a screen. This means that if an ad is seen by a person multiple times, it will only count as one impression [21,22].</td>
</tr>
<tr>
<td>Profile visits</td>
<td>The number of times a user has clicked on a post or ad to visit our Instagram or Facebook page [21,22].</td>
</tr>
</tbody>
</table>

**Ethical Considerations**

The University of Michigan Institutional Review Board approved this study (HUM#00156338). All participants were provided with detailed information about the study before participating and were asked to provide their electronic informed consent before beginning the survey. Participants were assured their information would remain confidential and be stored on a secure, password-protected cloud service at the University of Michigan (U-M Dropbox).

**Results**

**Overview**

Overall, using TinyURL metrics allowed us to track important demographic and user information on individuals who clicked on our customized TinyURL links. This information, which is displayed in Table 2, includes the top countries, browsers, operating systems, and devices of the individuals who engaged with our customized TinyURL links. This information was collected in order to gain insight into electronic engagement with our customized TinyURL links and measure the success of our web-based marketing campaigns in reaching our target population. We have chosen to report both total human clicks and total unique clicks on our customized TinyURL links. When measuring successful electronic engagement, we have chosen to focus on total human clicks, as it provides a comprehensive overview of the activity generated by our links and represents whether the same individual clicked on a link more than once. Table 2 reports demographic and electronic user information based on total human clicks, as captured by the TinyURL platform. However, in our analysis we also discuss unique clicks since this information is measured only once, the first time that an individual clicks on the link.

Mailchimp was found to be our most successful platform for electronic engagement, with 245 human clicks and 106 unique clicks on the custom TinyURL link from December 2021 to March 2022. The most popular day users clicked on the TinyURL link for Mailchimp was Thursday, and the most popular times for users were morning and afternoon, occurring between 8 AM and 4 PM. The second-most successful platform was Instagram, which received 190 human clicks and 158 unique clicks from December 2021 to March 2022. The most popular day users clicked on the TinyURL link for Instagram was Sunday, with the clicks being evenly distributed throughout the day. The custom TinyURL link for Facebook found moderate success, receiving 56 human clicks and 46 unique clicks from December 2021 to March 2022. The most popular day users clicked on the custom TinyURL link for Facebook was Monday. Clicks for this link were evenly distributed throughout the day. The custom TinyURL link for X received 39 human clicks and 28 unique clicks from December 2021 to March 2022. Days and times users clicked on the TinyURL link for X were evenly distributed throughout the week and the day. The TinyURL link for traditional email received 31 human clicks and 24 unique clicks from December 2021 to March 2022. These were evenly distributed throughout the week and throughout the day. The custom TinyURL link for UMHealthResearch received 15 human clicks and 10 unique clicks from December 2021 to March 2022. These clicks were evenly distributed throughout the week, falling mostly during weekdays, with the most popular time being in the afternoon and night, occurring between 12 PM and 12 AM.
Table 2. TinyURL metrics for Instagram, Facebook, X, traditional email, Mailchimp, and the UMHealthResearch website.

<table>
<thead>
<tr>
<th>Metrics and ranks</th>
<th>Instagram</th>
<th>Facebook</th>
<th>X</th>
<th>Traditional email</th>
<th>Mailchimp</th>
<th>UMHealthResearch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Clicks, n (%)</td>
<td>Value</td>
<td>Clicks, n (%)</td>
<td>Value</td>
<td>Clicks, n (%)</td>
<td>Value</td>
</tr>
<tr>
<td>Top countries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>United States</td>
<td>96/190 (50.5)</td>
<td>United States</td>
<td>29/56 (50)</td>
<td>United States</td>
<td>21/34 (61.8)</td>
</tr>
<tr>
<td>2</td>
<td>Ireland</td>
<td>38/190 (20)</td>
<td>Ireland</td>
<td>16/56 (28.6)</td>
<td>Germany</td>
<td>3/34 (8.8)</td>
</tr>
<tr>
<td>3</td>
<td>Kenya</td>
<td>28/190 (14.7)</td>
<td>Kenya</td>
<td>4/56 (7.1)</td>
<td>France</td>
<td>1/34 (2.9)</td>
</tr>
<tr>
<td>Top browsers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Chrome</td>
<td>109/188 (58)</td>
<td>Chrome</td>
<td>22/56 (43)</td>
<td>Chrome</td>
<td>12/34 (35.3)</td>
</tr>
<tr>
<td>2</td>
<td>Instagram</td>
<td>33/188 (17.6)</td>
<td>Facebook</td>
<td>11/56 (17.9)</td>
<td>cURL</td>
<td>10/34 (29.4)</td>
</tr>
<tr>
<td>3</td>
<td>Facebook</td>
<td>20/188 (10.6)</td>
<td>Safari</td>
<td>7/56 (12.5)</td>
<td>Internet Explorer</td>
<td>7/34 (20.6)</td>
</tr>
<tr>
<td>Top operating systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Android</td>
<td>100/184 (54.4)</td>
<td>Android</td>
<td>27/56 (55.8)</td>
<td>Windows</td>
<td>11/24 (45.8)</td>
</tr>
<tr>
<td>2</td>
<td>iOS</td>
<td>38/184 (20.7)</td>
<td>iOS</td>
<td>11/56 (22)</td>
<td>Mac</td>
<td>5/24 (20.8)</td>
</tr>
<tr>
<td>3</td>
<td>Windows</td>
<td>33/184 (18)</td>
<td>Windows</td>
<td>11/56 (22)</td>
<td>Android and Linux (tied)</td>
<td>3/34 (12.5)</td>
</tr>
<tr>
<td>Top devices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Smartphone</td>
<td>88/184 (47.8)</td>
<td>Smartphone</td>
<td>27/50 (54)</td>
<td>Desktop</td>
<td>20/24 (83.3)</td>
</tr>
<tr>
<td>2</td>
<td>Tablet</td>
<td>50/184 (27.2)</td>
<td>Tablet</td>
<td>12/50 (24)</td>
<td>Tablet</td>
<td>3/24 (12.5)</td>
</tr>
<tr>
<td>3</td>
<td>Desktop</td>
<td>46/184 (25)</td>
<td>Tablet</td>
<td>11/50 (22)</td>
<td>Smartphone</td>
<td>1/24 (4.2)</td>
</tr>
</tbody>
</table>

aNot available.
bCURL: client for URL.

Instagram and Facebook Insights

In addition to TinyURL metrics, we also used Instagram Insights, Meta Business Suite, and Facebook Insights to measure the performance of our paid advertisements on the 2 platforms. The Instagram Insights feature for businesses allowed us to measure impressions, reach, profile visits, and link clicks for the paid advertisements that we posted on the platform. Using Instagram Insights, we measured impressions, reach, profile visits, and link clicks for 5 paid advertisements from December 2021 to February 2022 (Table 3). Additionally, the use of Meta Business Suite and Facebook Insights allowed us to measure impressions, reach, and link clicks for the posts that we paid to be boosted on the platform. Using Meta Business Suite and Facebook Insights, we measured impressions, reach, and link clicks for 2 paid advertisements in February 2022 (Table 4).
Overview
Our findings suggest that Mailchimp was the most popular platform for electronic engagement, likely due to the fact that we had a well-curated list of email addresses from people who have previously engaged with our research, including colleagues, students, social media site followers, and community members (i.e., individuals residing in Detroit, Michigan, or Southeastern Michigan). The use of Mailchimp allowed us to directly reach this audience and provide them with relevant and targeted content about our SRH care campaign and access to the customized link, resulting in high community engagement. In addition to Mailchimp, we also found that Instagram and Facebook were popular social media platforms for engaging with young people. Our results coincide with the fact that Instagram and Facebook were popular social media platforms for engaging with young people. Instagram is reported to be an especially popular social media platform among young adults and teens, and Facebook remains a widely used platform by a majority of Americans [23].

On the other hand, we found X, traditional email, and UMHealthResearch to be our least successful platforms for electronic engagement. This may have been attributed to several reasons. First, our following on X was smaller than on our other platforms. Second, for emails that were sent, messages may have been filtered as spam and never opened. Traditional emails can be tracked after being sent but only after being opened if connected to an email tracking software system such as Mailchimp [24]. Finally, our advertisements on UMHealthResearch may have faced challenges in attracting engagement due to participants being required to create an account and undergo verification, resulting in a limited sample. UMHealthResearch promoted this study to only individuals who were already using the third-party recruitment platform, so the study advertisement was not as visible compared to other platforms.

Survey Eligibility Results
The survey yielded 525 responses, of which 489 finished the survey and 32 were eligible for the study for focus groups. The survey respondents were removed from the eligible sample if they did not identify as African American or Black (16 removed), were outside of the age range of between 18 and 24 years old (27 removed), did not identify as heterosexual (23 removed), or did not reside in the Southeast Michigan area (391 removed). These exclusions were made to identify an eligible sample.

Discussion
Overview
Our findings suggest that Mailchimp was the most popular platform for electronic engagement, likely due to the fact that we had a well-curated list of email addresses from people who have previously engaged with our research, including colleagues, students, social media site followers, and community members (i.e., individuals residing in Detroit, Michigan, or Southeastern Michigan). The use of Mailchimp allowed us to directly reach this audience and provide them with relevant and targeted content about our SRH care campaign and access to the customized link, resulting in high community engagement. In addition to Mailchimp, we also found that Instagram and Facebook were popular social media platforms for engaging with young people. Our results coincide with the fact that Instagram and Facebook were popular social media platforms for engaging with young people. Instagram is reported to be an especially popular social media platform among young adults and teens, and Facebook remains a widely used platform by a majority of Americans [23].

On the other hand, we found X, traditional email, and UMHealthResearch to be our least successful platforms for electronic engagement. This may have been attributed to several reasons. First, our following on X was smaller than on our other platforms. Second, for emails that were sent, messages may have been filtered as spam and never opened. Traditional emails can be tracked after being sent but only after being opened if connected to an email tracking software system such as Mailchimp [24]. Finally, our advertisements on UMHealthResearch may have faced challenges in attracting engagement due to participants being required to create an account and undergo verification, resulting in a limited sample. UMHealthResearch promoted this study to only individuals who were already using the third-party recruitment platform, so the study advertisement was not as visible compared to other platforms.

The use of customized TinyURL links helped us track which social media platforms were the most effective in reaching young African American male individuals regarding SRH care. TinyURL allowed us to track clicks on each custom URL we created and see information such as country, browser, device, etc from each person. This information can be used to tailor future social media marketing campaigns about SRH to these platforms, which can help to ensure that young African American male individuals have access to the information and resources they need to make informed decisions about their sexual health. Additionally, customized TinyURL links are easy to share and remember, which can lead to stronger engagement, especially with younger populations. Overall, custom TinyURL links allowed us to track clicks on each specific post and advertisement that were the most engaging and which platforms were most effective in reaching people. The information found through tracking TinyURL metrics can then be used to create more effective and engaging content about SRH for young African American male individuals. In addition to informing strategic marketing and content creation, the information from TinyURL metrics can be used to engage young African American male individuals in events and research about SRH care advertised on social media.

By using culturally informed social media advertisements and customized TinyURL links, we were able to effectively engage with young African American male individuals, provide them with information and resources regarding SRH care, and encourage them to seek health care. The results of this study may be useful for improving the effectiveness of future social media marketing campaigns using culturally informed advertising to engage young African American male individuals in SRH care.

Implications
Our findings suggest that customized TinyURL links can be an additional way to promote long-term risk-reduction methods

Table 3. Instagram Insights metrics.

<table>
<thead>
<tr>
<th>Dates</th>
<th>US $/day</th>
<th>Impressions</th>
<th>Reach</th>
<th>Profile visits</th>
<th>Link clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 18, 2021, to December 22, 2021</td>
<td>2</td>
<td>1437</td>
<td>738</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>January 27, 2022, to January 30, 2022</td>
<td>3</td>
<td>1850</td>
<td>1157</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>January 30, 2022, to February 2, 2022</td>
<td>3</td>
<td>1868</td>
<td>1214</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>February 10, 2022, to February 13, 2022</td>
<td>3</td>
<td>1355</td>
<td>1018</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>February 16, 2022, to February 18, 2022</td>
<td>4</td>
<td>1290</td>
<td>936</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 4. Meta Business Suite and Facebook Insights metrics from paid advertisements.

<table>
<thead>
<tr>
<th>Dates</th>
<th>US $/day</th>
<th>Impressions</th>
<th>Reach</th>
<th>Link clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td>February 10, 2022, to February 13, 2022</td>
<td>5</td>
<td>2001</td>
<td>1485</td>
<td>11</td>
</tr>
<tr>
<td>February 16, 2022, to February 18, 2022</td>
<td>5</td>
<td>2097</td>
<td>1427</td>
<td>9</td>
</tr>
</tbody>
</table>
among young male individuals in the community. Creating a TinyURL link for each platform we used to post content about SRH care allowed us to provide young African American male individuals with an easy way to access information and resources about SRH. By using customized TinyURL links, researchers can engage with young African American male individuals in a memorable, convenient, and trackable way on social media to help this population make informed decisions about their SRH. Custom TinyURL links are convenient and memorable enough to be easily shared between young male individuals, in person or using technology. Researchers using TinyURL can gain insight on how young African American male individuals engage with content and resources provided through those links. Additionally, by tracking which social media platforms are most effective in reaching young male individuals, researchers can target future social media marketing campaigns to these platforms. For example, a clinic seeking to engage with young African American male individuals in their community could use TinyURL to create custom links for their different social media platforms and track which links generate the most engagement. Using customized TinyURL links could inform the clinic’s social media marketing strategy on how to best engage with young African American male individuals about ongoing SRH care, from anonymous surveys to making an annual appointment to chatting with a health care provider virtually.

Limitations
One limitation of this study was the unexpected engagement from users outside the United States, particularly in countries such as Ireland and Kenya, clicking on our links. In the case of Instagram and Facebook, despite us setting target audience parameters for our paid advertisements, such as location, age, and sex, we found that the platforms could not always accurately control who sees the advertisements and limit it to the demographic characteristics of our target population. Factors such as IP address blockers or users intentionally altering their location could also result in engagement from unexpected locations. Additionally, we promoted a US $20 gift card for eligible participants who completed the survey, which could have attracted users from different countries to click on our links, potentially resulting in engagement from a wider audience than initially intended. Further research would be needed to draw definitive conclusions about the reasons for engagement with specific countries. As a result, the generalizability of our findings may be limited.

Another limitation of this study was the high number of survey responses that were ineligible for the study. Out of 525 total responses, only 489 participants completed the survey, and only 32 were ultimately eligible for the study to participate in focus groups. This may be due to web-based social media targeting not being geographically limited enough or the focus on the city of Detroit not being prominent enough in the advertisements, leading to responses from participants who did not meet the study criteria. Additionally, social media platforms such as Instagram and Facebook, where we ran paid advertisements, did not offer the option to specify ethnicity or race when targeting an audience for advertisements, which may have resulted in a large number of responses from participants who did not align with our target population. Meta, the parent company of Facebook and Instagram, removed certain advertisement targeting options that included race, religion, sexual orientation, and political affiliation in January 2022 [25]. Meta’s detailed targeting advertising options are often changing and being refined, and options to target certain specific groups may not always be available [26]. Additionally, there is no guarantee that your advertisement will only be shown within your specific geographic target area; this poses an inherent challenge when using a recruitment strategy for your target population that is centered on social media [27]. This limitation makes it difficult to tailor our recruitment methods, making it difficult to ensure we reach the correct pool of participants. This makes social media advertising somewhat unpredictable, and the effectiveness of this method may be limited by this factor. This inherent limitation calls for a more tailored approach in the design and use of participant recruiting social media advertisements, in addition to rigorous screening to ensure the internal validity of the results.

In reviewing TinyURL metrics for our links, there were missing data as we noticed that TinyURL could not capture all information on country, browser, operating system, and device for all recorded total human clicks for the links we created. For example, with Instagram, while TinyURL had data regarding the top countries for all 190 total human clicks when it came to reviewing the top browsers, TinyURL only had information for 188 of those clicks. For top operating systems, TinyURL only had information for 184 of those clicks. TinyURL identifies these data using user agent strings [20], which may be blocked by specific devices or browsers due to privacy settings. This aspect is important as it limits the collection of these data from users who click on customized links. However, this limitation may have had minimal impact on the insights and conclusions, as information was still viewable to the majority of the users who were able to access and click our links.

In addition, our use of social media as a recruitment strategy may have limited the pool of participants for this study. By using this method, we may have inadvertently excluded those without internet access and those who do not use or do not have social media accounts. It is important to note that social media approaches typically come with this selection bias, and it is important to acknowledge that this may have recruited from a skewed pool of applicants.

Furthermore, we encountered issues with spam responses and participants who may have been solely interested in winning the gift card, which could have affected the validity of our results. A solution to this would be setting up multiple captchas at different phases of the survey (eg, eligibility or payment) or advertisement to assess for human use.

Conclusion
Using customized, shortened URLs and TinyURL metrics to study social media engagement specifically to promote SRH care among young African American male individuals is innovative and provides future researchers with the potential to discover valuable insights into how to reach this population both locally and globally with sexual health information and resources. Social media is a powerful tool that can be leveraged
to reach young people with information about SRH care. TinyURL links offer a modern, quick means to engage with people about SRH care and track their electronic engagement. Customized TinyURL links, in conjunction with culturally informed advertisements and content, can be shared on popular social media platforms to engage with young African American male individuals seeking information and resources about SRH care. The findings from this study have the potential to inform future literature on the topic of improving electronic engagement and content creation with SRH care in unique spaces among young African American male individuals.

Acknowledgments
The authors want to thank the University of Michigan School of Nursing Center for Sexuality and Health Disparities, the Ginsberg Center, and Detroit Community Health Connection for their support. A special thank you to research staff Nidhi Kandi, Julian Lanum, Yi-Chia Hwang, and Freda Frimpong for their contributions to the Stay Safe Project.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to ethical concerns about maintaining the confidentiality of survey participants’ personal information but are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References

https://formative.jmir.org/2024/1/e48371 JMIR Form Res 2024 | vol. 8 | e48371 | p.1908 (page number not for citation purposes)


Abbreviations
SRH: sexual and reproductive health
STI: sexually transmitted infection

©Sandy Arena, Mackenzie Adams, Jade Burns. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Cancer Care Supportive Text Messaging Program (Text4Hope) for People Living With Cancer and Their Caregivers During the COVID-19 Pandemic: Longitudinal Observational Study

Reham Shalaby\(^1\), MD, PhD; Wesley Vuong\(^2\), MPH; Belinda Agyapong\(^1\), MSc; April Gusnowski\(^2\), BA; Shireen Surood\(^2\), PhD; Vincent Agyapong\(^3\), MD, PhD

\(^1\)Department of Psychiatry, University of Alberta, Edmonton, AB, Canada
\(^2\)Alberta Health Services, Addiction & Mental Health, Edmonton, Canada, Edmonton, AB, Canada
\(^3\)Department of Psychiatry, Dalhousie University, Halifax, NS, Canada

**Corresponding Author:**
Vincent Agyapong, MD, PhD
Department of Psychiatry
Dalhousie University
5909 Veterans Memorial Lane, 8th Floor Abbie J Lane Memorial Building
QEI Health Sciences Centre
Halifax, NS, B3H 2E2
Canada
Phone: 1 7802157771
Fax: 1 9024734887
Email: vn602367@dal.ca

**Abstract**

**Background:** Cancer is the leading cause of death in Canada, and living with cancer generates psychological demands, including depression and anxiety among cancer survivors and caregivers. Text4Hope-Cancer Care SMS text messaging–based service was provided to people with cancer and caregivers during the COVID-19 pandemic to support their mental health.

**Objective:** The aim of this study is to examine the clinical effectiveness of and satisfaction with Text4Hope-Cancer Care in addressing mental health conditions among people living with cancer and caregivers.

**Methods:** The study was conducted in Alberta, Canada. People who were diagnosed or receiving cancer treatment and caregivers self-subscribed to receive 3-months daily supportive cognitive behavioral therapy–based SMS text messages and a web-based survey was sent at designated time points to collect clinical and nonclinical data. The Hospital Anxiety and Depression scale (HADS) was used to examine changes in anxiety and depression symptoms after receiving the service. Satisfaction with the service was assessed using a survey with a Likert scale. Descriptive and inferential statistics were used, and test significance was considered with \(P \leq 0.05\).

**Results:** Overall, 107 individuals subscribed to the service, and 93 completed the program (completion rate 93/107, 86.9%). A significant improvement in the anxiety symptoms (HADS-Anxiety [HADS-A] subscale) was reported after 3 months of Text4Hope-Cancer Care \((t_{11}=2.62; \ P=0.02)\), with medium effect size (Hedges \(g=0.7\)), but not depression symptoms (HADS-Depression [HADS-D] subscale). Subscribers expressed high satisfaction and agreed that the service has helped them to cope with mental health symptoms and improve their quality of life. Most subscribers read the SMS text messages more than once (30/30, 100%); took time to reflect or took a beneficial action after reading the messages (27/30, 90%); and highly agreed (27/30, >80%) with the value of the received supportive SMS text messages as being relevant, succinct, affirmative, and positive. All subscribers recommended SMS text messaging for stress, anxiety, and depression and for cancer care support (30/30, 100%).

**Conclusions:** Text4Hope-Cancer Care was well-perceived and effectively addressed anxiety symptoms among people living with cancer and caregivers during the peak of the COVID-19 pandemic. This study provides evidence-based support and insight for policy and stakeholders to implement similar convenient, economic, and accessible mental health services that support vulnerable populations during crises.

**International Registered Report Identifier (IRRID):** RR2-10.2196/20240
Introduction

Cancer is the leading cause of death and is responsible for 28.2% of all deaths in Canada [1]. The disease mostly affects Canadians aged 50 years and older, but it can occur at any age [1]. In 2018, 1.5 million Canadians were living with or survived cancer [1,2]. However, newly diagnosed cases are projected to increase by 40% between 2015 and 2030 [1].

Living with cancer generates ongoing psychological demands that impose a heavy toll on patients and caregivers (a person who takes care of a patient with cancer) [3]. Depression and anxiety are common mental health conditions among cancer survivors, with nearly 30% reporting clinical symptoms of depression and 40% for anxiety [4,5]. These numbers vary according to different types of cancers [4,5]. Similar results were found for caregivers, where about 1 in 2 experienced symptoms of depression and anxiety. These rates may be as high as 65% and 91% for mothers of children living with cancer [3,6].

Cancer survivors often report depression and anxiety symptoms, including the fear of the recurrence of cancer, low emotional support, social isolation, and feeling lonely or abandoned after the intensive support during the phase of the treatment [4,7-9]. Depression and anxiety symptoms negatively impact the health of cancer survivors and caregivers by eroding quality of life, decreasing treatment compliance, and increasing the risk of morbidity and mortality [4,8,10].

Several interventions have been proposed to address psychological symptoms among cancer survivors and caregivers. These interventions can include educational, cognitive, or mindfulness-based therapies which are effective in addressing mental health symptoms [11,12]. However, these interventions (eg, in-person or facilitated by trained personnel) can be costly and time-consuming when compared to mobile-based interventions (eg, internet-based cognitive behavioral therapy (iCBT) or SMS text messages) that do not require a specialist facilitator [10,13,14].

Furthermore, a major body of literature suggests that remotely delivered interventions are impactful and effective in improving mental health compared to face-to-face services [15,16]. Web-based resources have been embraced as potential solutions that can cross the likely gap between evidence-based therapies and community practice, the gap which has thrived by stigma, underrecognition of psychological symptoms and lack of trained professionals [15,17,18].

SMS text messaging has demonstrated effectiveness in the treatment and prevention of mental disorders among diverse populations [19-22]. SMS text messaging services such as Text4Support effectively reduced the risk of harm to self and other harm symptoms after 6 months of intervention in a randomized controlled trial [23] and improved distress, anxiety, and depression in clinical samples [24]. Similarly, the Text4PTSI program has achieved fidelity and significantly reduced psychological distress among public safety personnel post intervention [25,26]. SMS text message–based population-level messaging programs have reported user satisfaction rates of well over 80%, and most subscribers have reported feeling connected to support systems and improving their ability to manage anxiety, depression, and general life issues, suggesting an improvement in mental health literacy [15,27-29].

In this project, Text4Hope-Cancer Care SMS text message–based service was provided to people living with cancer and caregivers during the COVID-19 pandemic. The service aimed to support their mental health by providing a daily supportive SMS text message based on cognitive behavioral therapy (CBT) principles. The service is designed to reduce stress, anxiety, and depression symptoms associated with the pandemic among this vulnerable population. In this project, we report the results of the Text4Hope-Cancer Care program regarding its clinical effectiveness in addressing mental health conditions and the satisfaction among people living with cancer and caregivers 3 months after receiving the service.

Methods

Study Design and Setting

The study was conducted in Alberta, Canada, where patients diagnosed or receiving cancer treatment and caregivers were eligible to self-subscribe to receive a daily supportive SMS text message for 3 months. The messages were crafted with a CBT framework and were developed by a clinical psychologist, and then reviewed and revised by a multidisciplinary team of psychiatrists, mental health therapists, and counselors who work with patients with cancer and caregivers (Text4Hope-Cancer Care is part of the primary Text4Hope service, which was introduced to support the mental health of the general public during the COVID-19 pandemic among Albertans [30]).

Examples of the Text4Hope-Cancer Care messages are (1) do things you enjoy—these activities can remind you who you are and take your mind off cancer for a while; (2) advocate for your needs using assertiveness—assertiveness is being respectful to you and the other person. Be direct, nonaggressive, and specific; and (3) cancer affects the whole family—to help you and your family cope, try to maximize quality time together, communicate, and create a schedule together.

Ethical Considerations

Participation in the program was voluntary, and completion of the survey did not preclude receipt of subsequent supportive SMS text messages. Subscribers could opt out of the service at any time, and no incentives or inducements for participating in...
the program were offered. Further details of the methodology are described in the study protocol [31]. The study was conducted according to the guidelines of the Declaration of Helsinki. Ethics approval has been granted by the University of Alberta Health Research Ethics Board (Pro00086163).

**Data Collection**

Data collection was run at baseline, 6 weeks, and 3 months, and subscribers were invited to complete these surveys via a link embedded in the received SMS text message at the designated time points. Survey questions were programmed into Select Survey, a web-based survey tool [31].

Overall, 107 subscribers participated in Text4Hope-Cancer Care program during the period between March 19, 2020, and December 19, 2022, and 14 subscribers opted out of the service (14/107, 13.1%), leaving 93 who completed the service (program completion rate of 93/107, 86.9%). From Figure 1, 49 subscribers responded to the surveys across all study time points, yielding a response rate of (49/107, 45.8%). There were a total of 65 eligible surveys received (35 at baseline and 30 at follow-up time points 6 weeks and 3 months), and only 12 subscribers have completed both the baseline and follow-up surveys. Survey questions were designed based on study objectives and available evidence from peer-reviewed literature [28,29]. The survey consisted predominantly of categorical items with Likert-scale responses and few continuously rated items. The items included sociodemographic characteristics such as gender, age, ethnic background, educational level, relationship status, employment status, and housing condition; and COVID-19–related data, such as self-isolation or quarantine due to the pandemic and perspectives regarding the likelihood of contracting COVID-19. The survey additionally included items related to cancer conditions, such as the duration of cancer diagnosis, treatment received for cancer therapy, and treatment postponed due to the COVID-19 pandemic for the participants and the loved ones who were diagnosed with cancer, if any. Clinical and service satisfaction data were also collected.

**Figure 1.** Study flowchart.

**Outcome Measures**

**Hospital Anxiety and Depression Scale**

To assess the effectiveness of the Text4Hope-Cancer Care service, the Hospital Anxiety and Depression Scale (HADS) was used and presented with the 2 subscales HADS-Depression (HADS-D) and HADS-Anxiety (HADS-A) [32]. Each HADS subscale contains 7 items; each item has 4 responses with an ordinal score from 0 to 3 and a total score between 0 and 21 in each subscale, with higher scores denoting higher levels of anxious or depressive state [33]. Some items have reversed responses (ie, from 3 to 0 [3 items for HADS-D and 5 items for HADS-A]). The scale reports how participants felt and behaved during the past week [33]. Since the scale is devoid of items related to somatic symptoms or symptoms of severe psychiatric illness, it was designed to detect the common mental disorders of depression and anxiety in the adult population attending medical (nonpsychiatric) outpatient clinics [33-35]. A total score of 0-7 is considered normal, 8-10 is considered borderline abnormal (borderline case), and 11-21 is considered abnormal (case) [32].
Based on previous studies, the sensitivity and specificity of approximately (0.80) were achieved when a cutoff score of 8 or above was used to define caseness (defined as the possible and probable anxiety disorders and depression among patients in nonpsychiatric hospital clinics [32]), on both HADS-A and HADS-D subscales [34]. Cronbach $\alpha$ for HADS-A was 0.83 (ranging from 0.67 to 0.90), and for HADS-D was 0.82 (ranging from 0.67 to 0.90), and the correlation between the 2 subscales was 0.56 [34]. For this study, the mean scores of HADS subscales were used to detect the differences between the study groups and also to trace changes over time.

**Satisfaction Measures**

A satisfaction questionnaire was provided to collect data related to the dimensions of the study. The dimensions considered included acceptability of the service (subscriber satisfaction or experience), appropriateness (subscriber feedback related to how helpful the daily messages have been in relation to their mental health and cancer diagnosis during the COVID-19 pandemic), and subscribers’ opinion about the use of technology-based services as part of health care during crises times.

**Sample Size Calculation**

Using a web-based script [36], we estimated that the sample size needed to assess the effects of the daily supportive SMS text messages among patients with cancer and caregivers on the outcome variables would be 34 with a projection that the effect size for the reduction in mean HADS-D and HADS-A scores at 3 months from baseline would be 0.2, a population variance of 1.0 for each subscale mean score, a 2-sided significance level $\alpha=0.05$, and a power of 80% ($\beta=0.2$).

**Statistical Analysis**

Descriptive statistics were performed, and chi-square or Fisher exact tests were used to define the distribution of categorical sociodemographic and cancer-related factors against 2 groups of subscribers. The first group (completers) refers to those who completed baseline and follow-up surveys, while the second group (noncompleters) refers to those who responded only to 1 survey. A post hoc analysis was run for the categorical variables with more than 2 response options when it showed a significant association, using adjusted residuals and $z$ score test. A corrected $P$ value was reported on this occasion ($P$ value $\times$ number of comparisons). An independent 2-tailed $t$ test was used to compare the continuous data of baseline mean scores of HADS subscales between the completers and noncompleters.

To examine the changes before and after receiving Text4Hope-Cancer Care Subscription Information text messages for the completers group, a paired $t$ test was used to find the difference in the mean scores of anxiety and depression HADS subscales from baseline to follow-up time points. Imputation of the missing data was applied using the Last Observation Carried Forward method, that is, the last observation (6 weeks) was carried forward (imputed) when the latest response (3 months) was missing [37]. Due to the small sample size of the completers ($n=20$), Hedges $g$ effect size was reported when the result showed significance.

For the satisfaction data, descriptive analysis was run with frequency and percentage for reporting data, using bar graphs. SPSS (version 28; IBM Corp) [38] was used to perform the statistical analysis and statistical test significance was considered with $P\leq0.05$.

**Results**

**Text4Hope-Cancer Care Subscription Information**

At baseline, subscribers were asked (multiple response questions) about their source of knowledge about the Text4Hope-Cancer Care service. The majority heard about the service from a website, such as Alberta Health Services or Mental Health Foundation websites (13/35, 37.1%), followed by social media platforms, such as Facebook and Instagram (Meta Platforms: 7/35, 20.1%). The news (eg, newspaper, web-based articles, and television), advice from a health care professional, and word of mouth (eg, family, friend, and coworker) were equally reported by the subscribers (5/35, 14.3%) each. Other sources such as employers, posters, and public libraries were also reported.

**Baseline and Sociodemographic Characteristics**

Table 1 presents the results of sociodemographic, cancer-related, and clinical characteristics of the study sample. From the table, (16/48, 33.3%) of the participants were between 50 and 59 years inclusive. Participants were predominantly females (42/48, 87.5%), White (47/48, 97.9%), had attained postsecondary education (40/44, 90.9%), were in a relationship (35/44, 79.5%), had employment (25/44, 56.8%), and owned their homes (30/44, 68.2%).

Regarding COVID-19–related questions, nearly a third of the subscribers reported needing to self-isolate or quarantine during the pandemic (15/48, 31.3%). The majority were neutral regarding the likelihood that they will contract COVID-19 (16/35, 45.7%); however, compared to someone who has not been diagnosed with cancer and is not receiving treatment for cancer, half of the subscribers reported that they would more likely contract COVID-19 (9/18, 50%).

Regarding cancer-related data, there were 18 cancer cases and 17 caregivers responded to the survey at baseline. The duration of having a diagnosis of cancer was majorly split between less than 6 months, and 1 year or more duration (7/18, 38.9%), each. In terms of the received treatment for cancer, the highest proportions were reported for surgery (10/18, 55.6%) and chemotherapy (9/18, 50.0%). Having cancer treatment or surgery postponed due to the COVID-19 pandemic was reported by a few participants (3/18, 16.7%).

In terms of the questions related to the loved ones of the participants, most participants reported having had 3 or more loved ones diagnosed with cancer and received treatment for cancer (9/16, 56.3%); having a loved one passed away due to cancer (14/17, 82.4%); mostly were a partner or a parent (6/17, 35.3%, each); the loved one was diagnosed with cancer for 1 year or more (10/17, 58.8%); having received either surgery and chemotherapy (6/17, 35.3%, each); and the majority reported having their loved ones’ cancer treatment or surgery postponed due to the COVID-19 pandemic (15/17, 88.2%).
Table 1. Distribution of the characteristics of the participants who responded to baseline and follow surveys (completers) and the participants who responded to only 1 survey (noncompleters).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Completers (n=12)</th>
<th>Noncompleters (n=37)</th>
<th>Total (N=49)</th>
<th>Chi-square (df)/t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>0 (0)</td>
<td>10 (28)</td>
<td>10 (21)</td>
<td>a</td>
<td>.12</td>
</tr>
<tr>
<td>40-49</td>
<td>4 (33)</td>
<td>5 (14)</td>
<td>9 (19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-59</td>
<td>5 (42)</td>
<td>11 (31)</td>
<td>16 (33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥60</td>
<td>3 (25)</td>
<td>10 (28)</td>
<td>13 (27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
</tr>
<tr>
<td>Male</td>
<td>1 (8)</td>
<td>5 (14)</td>
<td>6 (13)</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11 (92)</td>
<td>31 (86)</td>
<td>42 (88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
</tr>
<tr>
<td>White</td>
<td>12 (100)</td>
<td>35 (97)</td>
<td>47 (98)</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>1 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.74</td>
</tr>
<tr>
<td>Less than high school</td>
<td>0 (0)</td>
<td>2 (6)</td>
<td>2 (5)</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>High school diploma</td>
<td>1 (8)</td>
<td>1 (3)</td>
<td>2 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postsecondary education</td>
<td>11 (92)</td>
<td>29 (91)</td>
<td>40 (91)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.70</td>
</tr>
<tr>
<td>In a relationship</td>
<td>10 (83)</td>
<td>25 (78)</td>
<td>35 (80)</td>
<td>0.2 (1)</td>
<td></td>
</tr>
<tr>
<td>Not in a relationship</td>
<td>2 (17)</td>
<td>7 (22)</td>
<td>9 (21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.21</td>
</tr>
<tr>
<td>Employed</td>
<td>5 (42)</td>
<td>20 (63)</td>
<td>25 (57)</td>
<td>1.5 (1)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>7 (58)</td>
<td>12 (38)</td>
<td>19 (43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.77</td>
</tr>
<tr>
<td>Own a home</td>
<td>8 (67)</td>
<td>22 (69)</td>
<td>30 (68)</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>Renting</td>
<td>3 (25)</td>
<td>5 (16)</td>
<td>8 (18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living with family</td>
<td>1 (8)</td>
<td>5 (16)</td>
<td>6 (14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19 related questions, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-isolate or self-quarantine due to COVID-19 symptoms, recent travel, or because being in contact with someone who may have COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.59</td>
</tr>
<tr>
<td>No</td>
<td>9 (75)</td>
<td>24 (67)</td>
<td>33 (69)</td>
<td>0.3 (1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3 (25)</td>
<td>12 (33)</td>
<td>15 (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscribers’ perspective regarding the likelihood that they will contract COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.53</td>
</tr>
<tr>
<td>Likely</td>
<td>1 (8)</td>
<td>6 (26)</td>
<td>7 (20)</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>Unlikely</td>
<td>5 (42)</td>
<td>7 (30)</td>
<td>12 (34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>6 (50)</td>
<td>10 (44)</td>
<td>16 (46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscribers’ perspective regarding the likelihood that they will contract COVID-19 compared to someone who has not been diagnosed with cancer and is not receiving treatment for cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.51</td>
</tr>
<tr>
<td>More likely</td>
<td>3 (43)</td>
<td>6 (55)</td>
<td>9 (50)</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>As likely</td>
<td>4 (57)</td>
<td>3 (27)</td>
<td>7 (39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less likely</td>
<td>0 (0)</td>
<td>2 (18)</td>
<td>2 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer-related questions reported at baseline (12 completers and 23 noncompleters)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of subscribers, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.56</td>
</tr>
<tr>
<td>Patients with cancer</td>
<td>7 (58)</td>
<td>11 (48)</td>
<td>18 (51)</td>
<td>0.4 (1)</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Completers (n=12)</td>
<td>Noncompleters (n=37)</td>
<td>Total (N=49)</td>
<td>Chi-square (df) / t test (df)</td>
<td>P value</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>-------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Caregivers</td>
<td>5 (42)</td>
<td>12 (52)</td>
<td>17 (49)</td>
<td></td>
<td>.85</td>
</tr>
<tr>
<td><strong>Duration of the diagnosis with cancer, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 6 months ago</td>
<td>2 (29)</td>
<td>5 (46)</td>
<td>7 (39)</td>
<td></td>
<td>.85</td>
</tr>
<tr>
<td>6 months to less than 1 year ago</td>
<td>2 (29)</td>
<td>2 (18)</td>
<td>4 (22)</td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td>1 year or more</td>
<td>3 (43)</td>
<td>4 (36)</td>
<td>7 (39)</td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td><strong>Cancer treatment received, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>6 (86)</td>
<td>3 (27)</td>
<td>9 (50)</td>
<td>a</td>
<td>.049</td>
</tr>
<tr>
<td>Hormone therapy</td>
<td>2 (29)</td>
<td>0 (0)</td>
<td>2 (11)</td>
<td>a</td>
<td>.14</td>
</tr>
<tr>
<td>Immunotherapy</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Radiation therapy</td>
<td>3 (43)</td>
<td>4 (36)</td>
<td>7 (39)</td>
<td>a</td>
<td>.99</td>
</tr>
<tr>
<td>Stem cell therapy</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Surgery</td>
<td>4 (57)</td>
<td>6 (55)</td>
<td>10 (56)</td>
<td>a</td>
<td>.99</td>
</tr>
<tr>
<td>Targeted therapy</td>
<td>2 (29)</td>
<td>0 (0)</td>
<td>2 (11)</td>
<td>a</td>
<td>.14</td>
</tr>
<tr>
<td>None</td>
<td>0 (0)</td>
<td>4 (36)</td>
<td>4 (22)</td>
<td>a</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Having cancer treatment or surgery postponed due to the COVID-19 pandemic, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
</tr>
<tr>
<td>Yes</td>
<td>1 (14)</td>
<td>2 (18)</td>
<td>3 (17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>5 (71)</td>
<td>8 (73)</td>
<td>13 (72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not applicable</td>
<td>1 (14)</td>
<td>1 (9)</td>
<td>2 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total number of loved ones who have been diagnosed with cancer and received treatment for cancer, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.55</td>
</tr>
<tr>
<td>1-2</td>
<td>2 (67)</td>
<td>5 (39)</td>
<td>7 (44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>1 (33)</td>
<td>8 (62)</td>
<td>9 (56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Having a loved one passed away due to cancer, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
</tr>
<tr>
<td>No</td>
<td>1 (25)</td>
<td>2 (15)</td>
<td>3 (18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3 (75)</td>
<td>11 (85)</td>
<td>14 (82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Loved one diagnosed with cancer, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.047</td>
</tr>
<tr>
<td>Partner</td>
<td>0 (0)</td>
<td>6 (46)</td>
<td>6 (35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent</td>
<td>1 (25)</td>
<td>5 (39)</td>
<td>6 (35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friend</td>
<td>1 (25)</td>
<td>2 (15)</td>
<td>3 (18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child</td>
<td>2 (50)</td>
<td>0 (0)</td>
<td>2 (12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>When was your loved one diagnosed with cancer? n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.41</td>
</tr>
<tr>
<td>Less than 6 months ago</td>
<td>1 (25)</td>
<td>5 (39)</td>
<td>6 (35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months to less than 1 year ago</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>1 (6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year or more</td>
<td>2 (50)</td>
<td>8 (62)</td>
<td>10 (59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cancer treatment received by the loved one, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>3 (60)</td>
<td>3 (25)</td>
<td>6 (35)</td>
<td>a</td>
<td>.28</td>
</tr>
<tr>
<td>Hormone therapy</td>
<td>0 (0)</td>
<td>3 (25)</td>
<td>3 (9)</td>
<td>a</td>
<td>.52</td>
</tr>
<tr>
<td>Immunotherapy</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>1 (6)</td>
<td>a</td>
<td>.99</td>
</tr>
<tr>
<td>Radiation therapy</td>
<td>2 (40)</td>
<td>0 (0)</td>
<td>2 (12)</td>
<td>a</td>
<td>.07</td>
</tr>
<tr>
<td>Surgery</td>
<td>3 (60)</td>
<td>2 (17)</td>
<td>5 (29)</td>
<td>a</td>
<td>.12</td>
</tr>
<tr>
<td>Targeted therapy</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>1 (6)</td>
<td>a</td>
<td>.99</td>
</tr>
<tr>
<td>Do not know</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>1 (6)</td>
<td>a</td>
<td>.99</td>
</tr>
<tr>
<td><strong>Loved one having cancer treatment or surgery postponed due to the COVID-19 pandemic, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
</tr>
</tbody>
</table>
Regarding group differences, only 2 questions from above showed significant differences between completers and noncompleters. A significantly higher proportion of completers reported receiving chemotherapy (6/12, 50.0%) compared to noncompleters (3/23, 30.4%; \( P = .049 \)). Similarly, there was a significant difference regarding the loved ones who were diagnosed with cancer \( (P = .047) \). Post hoc analysis revealed that the proportion of the participants who had a child diagnosed with cancer who completed 2 (50%) surveys was significantly more than expected when compared to those who completed only 1 (0%) survey among the same group \( (z \text{ score} = 2.7; \text{ adjusted } P = .03) \).

Regarding the clinical assessment, the independent t test revealed no significant difference between completers and noncompleters at their baseline scores of HADS-D \( (t_{31} = 0.09; P = .93) \) and HADS-A \( (t_{31} = 0.28; P = .78) \) subscales.

### Clinical Outcome Results

Table 2 shows the changes in the mean scores of the primary outcome subscales, HADS-D and HADS-A, after 3 months from the baseline for participants who completed both the baseline and other follow-up surveys (completers). As illustrated in Table 2, the participants recorded lower mean scores on both HADS-D and HADS-A subscales at follow-up compared to their mean scores at baseline. However, only the HADS-A subscale was statistically significant \( (t_{11} = 2.62; P = .02) \), with a medium effect size \( (g = 0.7) \).

### Satisfaction Data

#### Overview

The feedback regarding the receptivity and satisfaction with Text4Hope cancer care service was reported by the subscribers at the follow-up time points \( (n = 30) \), that is, after receiving the service.

#### Benefits of Text4Hope-Cancer Care SMS Text Messages

Figure 2 demonstrates participants’ satisfaction measure data in relation to their perceived mental and physical health. The figure showed that the majority of the subscribers agreed that the messages have helped them to cope with symptoms related to cancer diagnosis, including stress (23/30, 76.7%), anxiety (21/30, 70%), depression (18/30, 60%), and loneliness (17/30, 56.7%). Similarly, most participants agreed that the supportive SMS text messages helped them feel connected to a support system (25/30, 83.3%), feel hopeful that they can manage issues related to cancer diagnosis (21/30, 72.4%), and that the messages improved their overall mental well-being (23/30, 76.7%). However, only 16 (53.3%) participants agreed that the messages enhanced their quality of life.
Figure 2. Appropriateness of Text4Hope-Cancer Care SMS text messages with patients’ mental well-being and cancer diagnosis.

Receptivity of the Supportive SMS Text Messages

Figure 3A demonstrates that most subscribers agreed that the SMS text messages were always or often relevant (20/29, 69%), succinct (24/28, 85.7%), affirmative (24/29, 82.8%), and positive (27/30, 90%).

In regard to reading the SMS text messages and actions taken after, Figures 3B and 3C show that all subscribers reported reading the messages (30/30, 100%). Similarly, the majority reported their return to read the messages (27/30, 90.0%), and they agreed that after reading the messages, they took time to reflect on the messages (23/30, 76.7%).

Figure 3. Receptivity of the supportive SMS text messages. (A) Agreement that the text messages were relevant, succinct, affirmative, and positive. (B) Reading the text messages. (C) After receiving the Text4Hope cancer care text messages.
**Overall Satisfaction With the Received Supportive SMS Text Messages and Preference of Their Schedule**

Figure 4 shows that most subscribers were satisfied with the service (28/30, 93.3%), preferred receiving the messages once per day (22/30, 73.3%), and the majority preferred the SMS text messages for a 12-month period (13/30, 43.3%).

**Opinions About the Use of Technology-Based Services as Part of Health Care**

Figure 5 demonstrates that most subscribers were in agreement that everyone struggling with cancer in times of crisis, such as the COVID-19 pandemic, would recommend technology-based services as part of their health care. The top-rated recommended items were SMS text messaging for stress, anxiety, and depression (13/13, 100%); SMS text messaging for cancer care support (13/13, 100%); and telephone counseling for stress, anxiety, and depression (13/13, 100%). These were followed by consultation via telephone conferencing for physical (12/13, 92.3%) and mental health care (12/13, 92.3%); consultation via videoconferencing for mental health care (12/13, 92.3%); telephone counseling for cancer care support (12/13, 92.3%); and web-based counseling for stress, anxiety, and depression (12/13, 92.3%). Following that, web-based counseling for cancer care support was recommended by 11 (84.6%) participants. Notably, email services were relatively less preferred by the study subscribers, including email messaging and support for stress, anxiety, and depression (8/12, 66.7%), and email messaging and support for cancer care support (8/12, 66.7%).
Discussion

Principal Findings

This study was designed to address mental health burdens among people living with cancer and caregivers in Alberta, Canada, during the COVID-19 pandemic, using the Text4Hope-Cancer Care service. The study also examined the satisfaction of the end users with the provided service. Overall, 93 subscribers completed the Text4Hope-Cancer Care program, and 49 responded to all surveys across study time points. The principal finding of this study was the significant improvement in anxiety symptoms among the people living with cancer and caregivers who received a daily message for 3 months of Text4Hope-Cancer Care service. Although there was an observed improvement in depressive symptoms after 3 months of receiving daily supportive SMS text messages, the difference was not statistically significant.

Text4Hope cancer care represents a supportive approach provided to a vulnerable population who have been already disadvantaged by their cancer diagnosis during the pandemic time, where operations were canceled, appointments and medical care was disrupted or delayed, and hospitals were closed due to national lockdowns and pandemic obligations [3,39,40]. Other research has tapped into the need for mental health support for people living with cancer and caregivers using diverse modalities, such as cognitive or mindfulness-based, acceptance and commitment, supportive-expressive, educational, and meaning-centered psychotherapy [11,41,42]. Our findings were not different from other SMS text messaging services provided in different contexts. Text4Support, for example, has demonstrated clinical effectiveness in reducing the risk of harm to self and other harm symptoms after 6 months of intervention in a randomized controlled trial [23,24]. The findings from this study, however, contradict findings from other studies in relation to addressing depressive symptoms. In 2 randomized controlled trials that used supportive SMS text messages, patients with depression showed symptom reduction on standardized self-report scales compared to a patient group not receiving messages (with large effect sizes: Cohen $d=0.85$ and 0.67) [21,22]. This could be explained by the heavy toll that those who live with cancer or caregivers are experiencing, particularly, during the COVID-19 pandemic. It is, however, possible, that the participants’ baseline depression would have deteriorated across the study time points without the daily supportive SMS text messages. Due to ethical considerations, this study did not include a control group who would not have received the supportive messages during the pandemic, to compare them to those who have received the service.

During the COVID-19 pandemic, people need to feel connected to a health support system. Therefore, several initiatives that included SMS text messaging services were provided and have reported comparative effectiveness to our study. Text4Hope service was provided to support the mental health of the general public in Alberta; the service reported quite similar findings to our study, where a significant reduction in stress, anxiety, depression symptoms, and suicidal ideation was achieved [14,30,43,44]. Another related service, the Text4Hope-Addiction program, was introduced to people living with substance use disorders; the service reported significant improvement in
standardized measures for craving, anxiety, and depression in subscribers [45]. Similarly, the Text4PTSI program significantly reduced psychological distress among public safety personnel postintervention [25,26].

The gap in the health care service experienced by patients living with cancer seems evident and could be attributed to the lack of social and mental health support. A recent systematic review of 33 studies has highlighted the lack of social support, among others, as an exacerbating factor within the social construct for developing depression and anxiety among older patients living with cancer [9]. Such mental health symptoms, particularly depression, are linked to increased risk of mortality among people living with cancer due to psychosocial reasons such as reduced adherence to therapy and appointments; thus, the accessibility to cancer control activities and mental health support can control cancer incidence, morbidity, and mortality rates [1,9].

Our study noticed that most of the subscribers were recently diagnosed and living with cancer for less than 1 year, compared to the caregiver group, whose loved ones were living with cancer for more than 1 year. This may reflect the critical and intense periods in which these vulnerable groups may need mental health support that can span over the duration of the disease, during its early stages, as seen with people living with cancer, or later as seen with caregivers. From the literature, it was reported that the third to fifth year after diagnosis is a period that typically requires close clinical follow-up, as well as supportive care, compared to 5 years after the diagnosis, where most survivors would likely have completed their treatment [1].

Another observation in this study is that caregivers whose children were diagnosed with cancer were more committed to completing the follow-up surveys. This was supported by the fact that childhood cancer diagnosis impacts the entire family, who in turn need mental health support [46]. Aligned with this observation, a population-based cohort study in Ontario reported that the mothers of children with cancer had an increased rate of mental health outpatient visits compared to mothers in the general population [46]. This observation has persisted over decades of follow-up, reflecting the critical needs along with the commitment of the mothers of children with cancer to the provided mental health support services.

Regarding the satisfaction with the service, study subscribers expressed a high degree of satisfaction with the service in terms of coping with mental health symptoms, such as anxiety and depression and improved overall mental well-being. The subscribers highly agreed with the value of the received supportive SMS text messages, as being relevant, succinct, affirmative, and positive. These results were not different from the satisfaction levels achieved with similar services. Among the users of Text4Mood and Text4Hope, more than 3 in 4 have also endorsed an improvement in their ability to manage anxiety, depression, and general life issues, suggesting an improvement in their resilience and mental health literacy [28,29]. Several SMS text message–based population-level SMS text messaging programs have reported user satisfaction rates of well over 80% [15,27,28]. Similarly, most of Text4Mood and Text4Hope respondents reported feeling connected to support systems, as compared to our study [28,29]. In the same context, well above 50% (16/30) of the subscribers reported that the service has improved their quality of life. This parameter is of particular importance in people living with cancer who often report marked unmet survivorship care needs that consequently compromise their quality of life [8].

Participants’ responses showed satisfaction with the SMS text messages’ content and frequency. Most subscribers read the SMS text messages more than once and took time to reflect or take beneficial action after reading the messages. A similar result was obtained with Text4Hope and Text4Mood services, where more than 1 in 5 reported returning to read the SMS text messages [28,29]. Most subscribers (13/30, 43%) preferred receiving the SMS text messages for a 12-month period, as compared to lower populations who opted for 3 (23%) or 6 (27%) months. This may reflect the significant psychological burden imposed by cancer diagnosis and the ongoing need for psychological support services that give them a sense of belonging and attachment to a health care system, particularly during the tough time of the COVID-19 pandemic. Therefore, the study participants probably opted for an extended length of the service to maintain the same sense of security and belonging.

Seeing the technology services, that could be part of the health care system, support people struggling with cancer during crisis times, SMS text messaging and phone counseling were among the top-rated care services either for mental health issues or cancer-related care. This was not surprising since, nowadays, phones represent an integral part of almost everyone’s life that could facilitate communication and provide channels of support when other conventional lines are obstructed or disrupted due to crises, such as the COVID-19 pandemic. Notably, compared to phone services, email services were relatively less preferred by the study subscribers. This could be attributed to the lack of accessibility to the internet required to access emails, particularly compared to the widely spread phones and available cellular plans.

Digital technology, therefore, introduces future opportunities to support the development of a scalable mental health workforce with the potential to leverage these technologies for integrating into mental health care, particularly for people with severe mental illness and economically disadvantaged communities [19,47,48].

Strengths and Limitations

The study is not without limitations. The low number of subscribers to the program was one of the limitations that curbed the evidence of Text4Hope cancer care to be evaluated among the rest of the community of people living with cancer and caregivers. Notwithstanding, the study reached the desired sample size as previously determined. The dropout rate of this study was aligned with other SMS text messaging services, such as Text4Hope Canada, where the dropout rate was similar (13%) and relatively higher than other web-based services (eg, iCBT) with a reported adherence rate of (52.8%) [10,43]. It is also of note that this study did not report on the time points at which the subscribers dropped out of the study (n=14); this information could have been helpful in better understanding the receptivity of the service among the targeted population and supporting...
the planning of future support services. Additionally, although the response rate was relatively low (49/107, 45.8%), as compared to the program completion rate (93/107, 87%), lower response rates are usually encountered with such web-based surveys, particularly when no monetary incentives were used, as the case in this study [43,49,50]. On the other hand, the relatively high response rate is one of the strengths of the study which reflects subscribers’ commitment to provide their feedback; thus, the results of the study could be generalized to all cohorts of the study.

Conclusions

Text4Hope-Cancer Care represents the health system and community response to close the gap during the pandemic time via providing mental health support to a vulnerable sector of the community during the pandemic time. The service was well-perceived and has successfully achieved significant effectiveness in addressing anxiety symptoms among people living with cancer and caregivers during the peak time of the COVID-19 pandemic. Compared to the costly and time-consuming conventional interventions, the SMS text messaging service represents a scalable, cost-effective intervention that can be used on any mobile technology, does not require technical skills, and does not require expensive data plans. This study provides the necessary evidence-based support and insight for policy and stakeholders to implement and guide future resource allocation to convenient, economic, and accessible mental health services that support vulnerable populations during crises.

Acknowledgments

This study was supported by grants from the Mental Health Foundation, the Calgary Health Trust, the University Hospital Foundation, the Alberta Children’s Hospital Foundation, the Royal Alexandra Hospital Foundation, and the Alberta Cancer Foundation. The funder had no role in the design and conduct of the study, collection, management, analysis, interpretation of the data; preparation, review, and approval of the study; or the decision to submit the results for publication. Support for the project was received from Alberta Health Services and the University of Alberta.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

VA performed the conceptualization. RS, WV, AG, SS, and VA performed data curation. RS and VA performed the formal analysis. VA contributed to the funding acquisition, investigation, and project administration. VA and RS contributed to the methodology. VA performed the supervision. RS contributed to writing—original draft. RS, BA, WV, AG, SS, and VA contributed to writing—review and editing.

Conflicts of Interest

None declared.

References


20. ResilienceNHope suite of supportive text messaging programs reduce stress, anxiety and depression. ResilienceNHope. URL: https://www.youtube.com/watch?v=KmaPvICNti0 [accessed 2023-03-06]


48. Shalaby et al. JMIR FORMATIVE RESEARCH
Abbreviations

CBT: cognitive behavioral therapy
HADS: Hospital Anxiety and Depression Scale
HADS-A: Hospital Anxiety and Depression Scale-Anxiety
HADS-D: Hospital Anxiety and Depression Scale-Depression
iCBT: internet-based cognitive behavioral therapy

©Reham Shalaby, Wesley Vuong, Belinda Agyapong, April Gusnowski, Shireen Surood, Vincent Agyapong. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Objective Assessment of Physical Activity at Home Using a Novel Floor-Vibration Monitoring System: Validation and Comparison With Wearable Activity Trackers and Indirect Calorimetry Measurements

Yuki Nakajima\(^1\)*, BSc; Asami Kitayama\(^1\)*, MSc; Yuji Ohta\(^1\), DEng; Nobuhisa Motooka\(^1\), PhD; Mayumi Kuno-Mizumura\(^2\), PhD; Motohiko Miyachi\(^3,4\), PhD; Shigeho Tanaka\(^4,5\), PhD; Kazuko Ishikawa-Takata\(^4,6\), PhD; Julien Tripette\(^1,4,7\), PhD

\(^1\)Department of Human-Environmental Sciences, Ochanomizu University, Bunkyo, Japan
\(^2\)Department of Performing Arts, Ochanomizu University, Bunkyo, Japan
\(^3\)Faculty of Sport Sciences, Waseda University, Tokorozawa, Japan
\(^4\)National Institute of Health and Nutrition, National Institutes of Biomedical Innovation, Health and Nutrition, Settsu, Japan
\(^5\)Faculty of Nutrition, Kagawa Nutrition University, Sakado, Japan
\(^6\)Faculty of Applied Biosciences, Tokyo University of Agriculture, Setagaya, Japan
\(^7\)Center for Interdisciplinary AI and Data Science, Ochanomizu University, Bunkyo, Japan

*these authors contributed equally

Corresponding Author:
Julien Tripette, PhD
Center for Interdisciplinary AI and Data Science
Ochanomizu University
2-1-1 Otsuka
Bunkyo, 112-8610
Japan
Phone: 81 03 5978 2032 ext 2032
Email: tripette.julien@ocha.ac.jp

Abstract

Background: The self-monitoring of physical activity is an effective strategy for promoting active lifestyles. However, accurately assessing physical activity remains challenging in certain situations. This study evaluates a novel floor-vibration monitoring system to quantify housework-related physical activity.

Objective: This study aims to assess the validity of step-count and physical behavior intensity predictions of a novel floor-vibration monitoring system in comparison with the actual number of steps and indirect calorimetry measurements. The accuracy of the predictions is also compared with that of research-grade devices (ActiGraph GT9X).

Methods: The Ocha-House, located in Tokyo, serves as an independent experimental facility equipped with high-sensitivity accelerometers installed on the floor to monitor vibrations. Dedicated data processing software was developed to analyze floor-vibration signals and calculate 3 quantitative indices: floor-vibration quantity, step count, and moving distance. In total, 10 participants performed 4 different housework-related activities, wearing ActiGraph GT9X monitors on both the waist and wrist for 6 minutes each. Concurrently, floor-vibration data were collected, and the energy expenditure was measured using the Douglas bag method to determine the actual intensity of activities.

Results: Significant correlations \((P<.001)\) were found between the quantity of floor vibrations, the estimated step count, the estimated moving distance, and the actual activity intensities. The step-count parameter extracted from the floor-vibration signal emerged as the most robust predictor \((r^2=0.82; P<.001)\). Multiple regression models incorporating several floor-vibration–extracted parameters showed a strong association with actual activity intensities \((r^2=0.88; P<.001)\). Both the step-count and intensity predictions made by the floor-vibration monitoring system exhibited greater accuracy than those of the ActiGraph monitor.
Conclusions: Floor-vibration monitoring systems seem able to produce valid quantitative assessments of physical activity for selected housework-related activities. In the future, connected smart home systems that integrate this type of technology could be used to perform continuous and accurate evaluations of physical behaviors throughout the day.

(JMIR Form Res 2024;8:e51874) doi: 10.2196/51874

KEYWORDS
smart home system; physical behavior; physical activity; activity tracker; floor vibration; housework-related activity; home-based activity; mobile phone

Introduction

There is evidence associating regular physical activity with lower risks for mortality and noncommunicable diseases [1], encouraging researchers, policy makers, and health care companies to develop strategies for the promotion of active lifestyles. The self-monitoring of physical behavior is one approach that has been described as effective for helping people increase their level of physical activity [2-5]. The recent expansion of the activity tracker device market has opened new perspectives for the promotion of active behavior [3,6].

Technology that enables the objective assessment of physical behaviors has evolved considerably over the past decades [7]. Modern activity trackers are generally worn at the waist or the wrist, and while the most recent devices usually feature multiple sensing abilities, the evaluation of physical behaviors mostly results from the treatment of acceleration data acquired by an integrated 3-axis microelectromechanical accelerometer sensor chip [8]. Software tools are able to compute a wide range of parameters related to physical behaviors, such as sedentary time, step count, and estimated energy expenditure [9]. Activity tracker devices can be paired with smartphone apps, transforming smartphone handsets into genuine hubs for the monitoring of physical activity and sedentary behaviors. Although waist- and wrist-worn activity tracker devices have been linked to inaccurate energy expenditure predictions [10-12], the emergence of the 5G and Internet of Things devices opens up room for more accurate and continuous monitoring, where multiple connected devices can collect a wealth of information about people’s physical behaviors throughout the day. In such a connected environment, and while housework-related activities account for a substantial proportion of daily physical activity in some populations [13], smart home systems could provide crucial information to (1) support the continuity of the monitoring of physical activity and sedentary behaviors when people stay at home by allowing them to remove their wearable activity tracker device and (2) improve the accuracy of energy expenditure predictions related to home-based activities. However, although various smart home projects have already included technological features allowing monitoring of the physical behaviors of the occupants, to date, the information has mainly been used as input to smart appliances to adapt to the living environment, assist occupants with disabilities, or optimize domestic energy consumption [14-17]. In these projects, various monitoring technology devices have been considered, including motion sensors, low-resolution video cameras, Kinect systems (Microsoft), and accelerometer-based wearable monitors [16,18-20]. Floors with sensing capabilities have also been developed. For instance, binary pressure detection floor systems have been used to detect the position of occupants [21,22]. Floor geophones and accelerometer sensors have been used to locate footsteps or evaluate room occupancy [23,24]. Unfortunately, none of these projects have prioritized the objective and quantitative assessment of physical behaviors with the ultimate goal of providing lifestyle-oriented feedback to the occupants. Nevertheless, smart home systems capable of monitoring physical activity and sedentary behaviors while providing feedback to the occupants have the potential to encourage individuals to adopt more active and healthier behaviors throughout the day [2-5].

In this study, the floor-vibration measurement system of the experimental Ocha-House was used to collect information about the floor vibrations generated by the occupant and estimate the energy expenditure and step count. A structured experiment consisting of the completion of 4 typical home-based activities was conducted to assess the validity of these estimations with respect to the actual measurements performed by indirect calorimetry (energy expenditure) and direct observation (step count). The estimations of the floor-vibration monitoring system were also compared with those of wrist- and wrist-worn research-grade activity tracker devices. It is hypothesized that the floor-vibration monitoring system is capable of correctly predicting energy expenditure.

Methods

The Ocha-House Project

The experimental Ocha-House is in the Bunkyo district in the central area of the Tokyo metropolitan region. The project was originally designed as a ubiquitous computing house that allows the mounting of various sensing devices [25]. According to Japanese standards, the Ocha-House corresponds to an extended 1LDK dwelling, that is, a 1-bedroom apartment with a kitchen separated from the living and dining areas. An overview of the building and experimental area is shown in Figure 1.
Figure 1. Overview of the Ocha-House and the experimental area. (A) External view of the building (image captured from southwest corner of the yard). (B) Plan of the house and the experimental area, with furniture indicated in light gray, and green squares (labeled 1-8) indicating sensor positions. The house comprises 2 distinct areas separated by a wall. The west side features a fully open space housing a bedroom corner and a living room without any additional partition wall. On the east side, the kitchen and dining room are interconnected through an open space. The toilet and bathroom corners are situated in enclosed spaces on the east side of the building.

The experimental Ocha-House is in the Bunkyo district in the central area of the Tokyo metropolitan region (Figure 1A). The total experimental surface area was 42 m$^2$ (Figure 1B). A total of 8 high-sensitivity uniaxial accelerometers (shear-type pickup PV-87; Rion Co. Ltd) were installed on the floor to measure the floor vibrations occurring on the experimental surface. The PV-87 sensor characteristics were specified as follows by the manufacturer: charge sensitivity $+40 \text{ pC to } -40 \text{ pC/m/s}^2$; range of detection 1 to 3000 Hz; dimensions 24 mm (hex) x 30.5 (H) mm; mass 115 g. The optimal number of sensor units and their locations were determined through a series of preliminary experiments. These experiments involved progressively increasing the sensitivity setting of the sensors and the number of units placed on both the west and east sides of the Ocha-House. The operation continued until the coverage was deemed sufficient to detect human motion across all parts of the experimental surface. The data related to these preliminary experiments are not presented here. The sensors were mounted on the floor using double-sided tape, as recommended by the manufacturer, and connected to 4 UV-16 2-channel charge amplifiers (Rion Co Ltd) configured in accordance with the manufacturer’s recommendations. The floor-vibration data acquisition was performed using USB-6008 data acquisition devices (National Instrument Corp), a laptop equipped with MATLAB 2015b, and the necessary data acquisition toolbox (MathWorks Inc). The signal was digitized at a 100-Hz sampling rate with a resolution of 12 bits. The abovementioned system is described hereafter as the floor-vibration monitoring system.
Study Protocol
In total, 10 female participants engaged in 4 activities in the Ocha-House. Participants were recruited from the Ochanomizu University campus, which is a women’s university. They were selected based on the inclusion criterion of being aged $\geq 18$ years, with the exclusion criterion being physical imbalance. Participant characteristics are summarized in Table 1, and the details of the 4 activities performed at the Ocha-House are presented in Figure 2. Before the commencement of the experiment, each participant completed a brief walking trial in the Ocha-House, lasting approximately 1 minute. In total, 2 researchers with expertise in gait analysis visually determined the gait type of each participant, specifically identifying whether they exhibited a heel strike or a lighter mid-strike or forefoot-strike landing. The walking trial revealed that all participants could be categorized into either heel-strike landing or mid-strike or forefoot-strike landing categories, with no other gait types observed. All experiments were conducted without any footwear, including slippers. In total, 9 participants wore socks. Moreover, 1 participant did not wear socks on the day of the experiment and completed the protocol barefoot.

Table 1. Participant characteristics (N=10).

<table>
<thead>
<tr>
<th></th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>24 (7)</td>
</tr>
<tr>
<td>Body weight (kg), mean (SD)</td>
<td>47 (6)</td>
</tr>
<tr>
<td>BMI (kg/m(^2)), mean (SD)</td>
<td>19 (1)</td>
</tr>
</tbody>
</table>

**Gait type**
- Heel strike: 3
- Mid- or forefoot strike\(^a\): 7

\(^a\)The column “gait type” refers to the number of participants presenting a strong heel strike during the walking gait cycle, as opposed to participants who presented a lighter mid-strike or forefoot-strike landing. The gait type of the participants was visually determined during the walking calibration trial.

Figure 2. Images of experiments. (A) Sitting and watching videos; (B) ironing, folding, and hanging clothes; (C) cooking, setting the table, and serving food; and (D) cleaning the room.
In total, 4 activities were selected from the “inactivity quiet/light” or “home activities” categories of the compendium of physical activities [26]. The metabolic equivalent of task (MET) values, indicating the intensity of each activity, were reported in the compendium as follows:

- Sitting and watching videos, hereafter referred to as *sitting* (approximately 1.3 MET, activity code 07020).
- Ironing, folding, and hanging clothes, hereafter referred to as *ironing* (1.8-2 METs, activity codes 05070 and 05090).
- Cooking, setting the table, and serving food, hereafter referred to as *cooking* (approximately 2.5 METs, activity codes 05051 and 05052).
- Cleaning the room, hereafter referred to as *cleaning* (approximately 3.3 METs, activity code 05030).

Each activity lasted 6 minutes. To balance the contribution of the 3 tasks within “cooking, setting the table, and serving food” and considering the volume of the Douglas bag used for energy expenditure measurement (see the section Indirect Calorimetry), participants were orally given time information. This ensured that they spent approximately 40 seconds on each task every 2 minutes. The floor-vibration monitoring system recorded floor vibrations, estimate the number of steps, and compute quantitative parameters as described in the Floor Vibration Signal Treatment and Data Feature Extraction section. The participants wore 2 ActiGraph GT9X monitors (ActiGraph LLC) at the waist and wrist. The 10-second epoch activity count and the step-count prediction were recorded for each activity. Finally, Douglas bags were used during the last 2 minutes of each activity to collect the air expired by the participants, perform indirect calorimetry measurements, and obtain the actual energy expenditure. Throughout the experiments, the researchers stood quietly on an insulated part of the floor outside the experimental area to avoid producing confounding vibrations. The experiments were video recorded (data not shown).

**Ethical Considerations**

The experimental protocol was approved by the Ochanomizu University Research Ethics Committee (2018-18). All the participants provided written informed consent and they did not receive any compensation. The individual shown in Figure 2 provided informed consent for the publication of their image.

**Floor-Vibration Signal Treatment and Data Feature Extraction**

The 8-sensor floor-vibration data of each 6-minute activity corresponded to 8 time series of 36,000 samples. Raw data were expressed in volt. For each series, the floor-vibration signal was rectified and smoothed using a Butterworth filter. The vector norms of the 8 sensors were then computed. A location-based calibration coefficient was applied to the vector norm for each data sample to uniformize the vibration magnitudes throughout the experimental surface (Multimedia Appendix 1). The 3 following data features were extracted:

- Floor count—for each participant and each activity, 36,000 sample values (6-min ×100 Hz sampling rate) of the uniformized vector norm time series were summed to obtain the *floor count* parameter.
- Step count—data from the uniformized vector norm time series were cut in windows of 1 second with an overlap of 50%. For each window, the number of steps was computed using a standard peak detection algorithm configured to detect vibration peaks with a minimum prominence of 1.05 SD and a minimum interval of 250 ms [27]. The *step count* parameter was computed for each activity of each participant by summing the number of unique peaks throughout the activity (6 min). The step count parameter was used as both a data feature, allowing the prediction of energy expenditure and a physical activity parameter to be compared with the direct observations and the outcomes of the waist- and wrist-worn ActiGraph devices.
- Moving distance—the uniformized vector norm time series was cut into windows of 1 second and averaged for each window. The window average was compared with a criterion value calculated for each individual from the data collected during the walking trial to determine whether the participant was moving. Subsequently, the filtered data of the 8-sensor time series corresponding to the windows where the participant was moving were used to compute their location in the house. The distances between all locations taken sequentially were summed for each activity to obtain the *moving distance* parameter (6 min). This parameter was tested and validated against the moving distance estimated from the observation of the video records (Figure S1 in Multimedia Appendix 2).

The signal treatment and computation of the 3 previously described parameters were performed using the toolboxes included in the SciPy library [28]. An overview of the entire data processing process is shown in Figure 3.
Figure 3. Floor-vibration signal flow processing chart. From 8 time series of raw signals to the extraction of 3 floor-vibration–based data features.

Actigraphy
The participants were equipped with 2 ActiGraph GT9X monitors worn on the waist and the wrist. The waist-worn device was positioned near the right hip of the participant, on the belt, or on the upper edge of the skirt or the bottom of the trousers. An elastic belt provided by the manufacturer was used when the clothes worn by the participant did not allow the tight mounting of the monitor. The wrist-worn GT9X device was mounted tightly on the nondominant hand always in the same direction. The 2 monitors were mounted by the same experimenter for all participants. The ActiGraph monitor data were collected in 1-second epochs. The 10-second epoch activity count data were extracted from the wrist-worn as well as waist-worn devices, and the “Crouter adult (2010)” equation was used to compute energy expenditure predictions (ActiLife 6; ActiGraph LLC) [29]. This algorithm uses a refined 2-regression model to distinguish between walking and lifestyle activities. In this study, the activity intensities were expressed in MET (ie, energy expenditure/participant weight/6 min). The number of steps estimated by the waist- and wrist-worn monitors was also recorded.

Indirect Calorimetry
The actual energy expenditure was measured using the Douglas bag method during the last 2 minutes of each activity. The air composition of the bags was analyzed using a mass spectrometry gas analyzer (ARCO-1000; Arco System) calibrated on each experimental day in accordance with the manufacturer’s instructions. The gas volume was determined using a gas meter (DC-5; Shinagawa). The energy expenditure was estimated from oxygen consumption (VO₂) and carbon dioxide production (VCO₂) using the Weir formula (ie, 3.9VO₂+1.1VCO₂). In addition, the resting metabolic rate of each participant was evaluated before the experiment. The participant lay for 15 minutes on the bed of the Ocha-House, and the expired air was collected during the last 2 minutes. The actual MET value for each activity was calculated as the activity energy expenditure divided by the resting metabolic rate.

Video Recording
The experimental sessions were video recorded using an Arrows M03 smartphone (Fujitsu Ltd) or an iPad Mini 3 (Apple Inc). In total, 2 independent investigators inspected the videos and counted the number of actual steps for the 4 activities of each participant. In this study, a “step” is defined as the shift of the body weight support from one leg to the other, which includes a single-leg support phase and occurs at least partially on the anterior-posterior axis.

Statistical Analysis
The 4 activities were compared for the floor count, step count, moving distance parameters using an ANOVA or the
Kruskal-Wallis test, and multiple comparison procedures (Tukey or Nemenyi tests) were performed to locate differences. The same analysis was conducted for parameters extracted from the waist- and wrist-worn GT9X monitors and for the indirect calorimetry measurements. The relationships between the actual intensities measured by indirect calorimetry and floor count, step count, and moving distance were investigated using single linear regression tests. Multiple regression models were used to explore the relationship between different combinations of descriptors, including floor count, step count, and moving distance, and the actual intensities measured using indirect calorimetry. This analysis was conducted hierarchically. First, the regressions were only performed on the floor-vibration–extracted parameters. Second, the participant characteristics (ie, body weight and gait type) were included in the model descriptors. Additional hierarchical models are presented in Multimedia Appendix 2. The model with the highest $R^2$ value best explained the variation in the data and was selected for subsequent testing. Finally, mixed model ANOVA and post hoc pairwise operations were used to compare the performance of the best models built on data obtained with the floor-vibration–based monitoring system, the waist- and wrist-worn GT9X monitors, against the actual intensity and the actual number of steps.

The underlying assumptions for each test were evaluated before conducting the analyses. Data are presented as mean (SD). The statistical analysis was performed using the following Python libraries: StatsModels (0.13.2), Pingouin (0.5.3), and Scikit-Posthocs (0.7.0) [30,31]. Data used for the statistical analysis are shared in the Multimedia Appendix 3.

**Results**

### Actual Intensities and Number of Steps

Activity intensities calculated from indirect calorimetry measurements were as follows: 1.2 (SD 0.2) MET for the sitting behavior, 1.9 (SD 0.4) MET for the ironing activity, 2.5 (SD 0.4) MET for the cooking activity, and 3.7 (SD 0.6) MET for the cleaning activity (Figure 4A). Pairwise comparisons indicated significant differences in intensity between sitting and cooking, between sitting and cleaning, and between ironing and cleaning ($P=.004$, $P<.001$, and $P=.01$, respectively). The actual number of steps was 0 (SD 0) for sitting, 48 (SD 36) for ironing, 133 (SD 35) for cooking, and 281 (SD 48) for cleaning (Figure 4B). Pairwise comparisons indicated significant differences in steps between sitting and cooking, between sitting and cleaning, and between ironing and cleaning ($P=.003$, $P<.001$, and $P=.003$, respectively).
Figure 4. Comparison between the 4 experimental home activities. (A) actual intensities (indirect calorimetry evaluation); (B) actual number of steps (video observation); (C) activity intensity predicted by the waist-worn ActiGraph GT9X device; (D) activity intensity predicted by the wrist-worn ActiGraph GT9X device; (E) floor-vibration–based computed floor-count; (F) floor-vibration–based computed step-count; (G) floor-vibration–based computed moving-distance. The intensity predictions of the GT9X monitors were computed using the “Crouter adult (2010)” equation [29]. Yellow line: median. Green point: average. Outliers are not depicted. MET: metabolic equivalent of task. *P<.05, **P<.001.

Floor Count, Step Count, and Moving Distance Parameters Computed From Floor Vibrations

The floor count parameter (arbitrary units) was as follows for each activity: 113 (SD 58) for sitting, 610 (SD 277) for ironing, 1046 (SD 748) for cooking, and 2323 (SD 1255) for cleaning the room (Figure 4E). Pairwise comparisons indicated significant differences in intensity between sitting and cooking, between sitting and cleaning, and between ironing and cleaning (P=.006, P<.001, and P=.04, respectively). The estimated number of steps was as follows: 12 (SD 7.3) for sitting, 78 (SD 27) for ironing, 133 (SD 53) for cooking, and 251 (SD 59) for cleaning.
The mixed model ANOVA showed a significant effect of the measurement method and a significant interaction with the activity for both the predictions of the number of steps and activity intensities ($P<.001$).

**Activity Intensity Predictions**

The wrist-worn activity tracker estimated the activity intensities as follows: 1.8 (SD 0.4) MET for *sitting*, 6.4 (SD 0.5) MET for *ironing*, 5.6 (SD 0.7) MET for *cooking*, and 6.8 (SD 0.4) MET for *cleaning* (Figure 4D). Pairwise comparisons indicated significant differences between *sitting* and *ironing* and between *sitting* and *cleaning* ($P=.002$ and $P<.001$, respectively). The estimated number of steps was as follows: 11 (SD 5.7) for *sitting*, 177 (SD 39) for *ironing*, 131 (SD 37) for *cooking*, and 197 (SD 35) for *cleaning* (Multimedia Appendix 2). Pairwise comparison analyses indicated significant differences between *sitting* and *ironing* and between *sitting* and *cleaning* ($P<.001$ for both).

The results for the activity count parameters of the wrist- and wrist-worn devices are shown in Multimedia Appendix 2.

### Relationship Between the Floor-Vibration–Based Outcomes and the Actual Activity Intensities

As presented in Table 2, *floor count*, *step count*, and *moving distance* were significantly associated with the intensity of physical behavior ($r^2=.56$, $r^2=.82$, $r^2=.66$, respectively; $P<.001$). Combining the 3 parameters resulted in an $r^2$ value of 0.84. Combining *floor count*, *step count*, and *moving distance* with participant personal characteristics, such as body weight and gait type, allowed predicting the intensity of physical behaviors with an accuracy of 88% (Table 1). The results of the additional hierarchical models are presented in Multimedia Appendix 2.

### Table 2. Relationship between floor-vibration–based parameters and actual activity intensities evaluated by indirect calorimetry.

<table>
<thead>
<tr>
<th>Models and predictor variables</th>
<th>Standardized partial regression coefficient</th>
<th>$P$ value</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single regressions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Floor count</td>
<td>0.745</td>
<td>&lt;.001</td>
<td>0.56</td>
</tr>
<tr>
<td>2 Step count</td>
<td>0.904</td>
<td>&lt;.001</td>
<td>0.82</td>
</tr>
<tr>
<td>3 Moving distance</td>
<td>0.815</td>
<td>&lt;.001</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Multiple regressions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Step count</td>
<td>0.971</td>
<td>&lt;.001</td>
<td>0.85</td>
</tr>
<tr>
<td>Floor count</td>
<td>−0.43</td>
<td>.01</td>
<td>0.85</td>
</tr>
<tr>
<td>Moving distance</td>
<td>0.361</td>
<td>.02</td>
<td>0.85</td>
</tr>
<tr>
<td>5 Step count</td>
<td>0.916</td>
<td>&lt;.001</td>
<td>0.88</td>
</tr>
<tr>
<td>Moving distance</td>
<td>0.259</td>
<td>.09</td>
<td>0.88</td>
</tr>
<tr>
<td>Floor count</td>
<td>−0.224</td>
<td>.20</td>
<td>0.88</td>
</tr>
<tr>
<td>Body weight</td>
<td>−0.08</td>
<td>.21</td>
<td>0.88</td>
</tr>
<tr>
<td>Gait type</td>
<td>−0.155</td>
<td>.03</td>
<td>0.88</td>
</tr>
</tbody>
</table>

*aRegression models combining 2 vibration parameters with and without participant characteristic parameters are shown in Multimedia Appendix 2. “Gait type” is binary data (mid- or forefoot strike vs heel-strike foot landing).*

**Step Count and Activity Intensity Predictions**

The mixed model ANOVA showed a significant effect of the measurement method and a significant interaction with the

\[ \text{Step Count} = -0.155 \times \text{Gait type} - 0.08 \times \text{Body weight} - 0.224 \times \text{Floor count} + 0.259 \times \text{Moving distance} - 0.43 \times \text{Floor count} + 0.971 \times \text{Step count} - 0.361 \times \text{Moving distance} - 0.745 \times \text{Floor count} + 0.904 \times \text{Step count} + 0.815 \times \text{Moving distance} \]

### Actigraphy

The waist-worn activity tracker estimated the activity intensities as follows: 1.1 (SD 0.2) MET for *sitting*, 1.2 (SD 0.3) MET for *ironing*, 2.1 (SD 0.4) MET for *cooking*, and 4.2 (SD 0.7) MET for *cleaning* (Figure 4C). Pairwise comparison analyses indicated significant differences between *sitting* and *cooking*, between *sitting* and *cleaning*, and between *ironing* and *cleaning* ($P=.03$, $P<.001$, and $P=.03$, respectively). Furthermore, a significant correlation was found between the *moving distance* outcomes computed from the floor-vibration signal and the *moving distances* estimated from the video records (Figure S1 in Multimedia Appendix 2).

### Relationship Between the Floor-Vibration–Based Parameters and Actual Activity Intensities Evaluated by Indirect Calorimetry

As presented in Table 2, *floor count*, *step count*, and *moving distance* were significantly associated with the intensity of physical behavior ($r^2=.56$, $r^2=.82$, $r^2=.66$, respectively; $P<.001$). Combining the 3 parameters resulted in an $r^2$ value of 0.84. Combining *floor count*, *step count*, and *moving distance* with participant personal characteristics, such as body weight and gait type, allowed predicting the intensity of physical behaviors with an accuracy of 88% (Table 1). The results of the additional hierarchical models are presented in Multimedia Appendix 2.

### Table 2. Relationship between floor-vibration–based parameters and actual activity intensities evaluated by indirect calorimetry.

<table>
<thead>
<tr>
<th>Models and predictor variables</th>
<th>Standardized partial regression coefficient</th>
<th>$P$ value</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single regressions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Floor count</td>
<td>0.745</td>
<td>&lt;.001</td>
<td>0.56</td>
</tr>
<tr>
<td>2 Step count</td>
<td>0.904</td>
<td>&lt;.001</td>
<td>0.82</td>
</tr>
<tr>
<td>3 Moving distance</td>
<td>0.815</td>
<td>&lt;.001</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Multiple regressions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Step count</td>
<td>0.971</td>
<td>&lt;.001</td>
<td>0.85</td>
</tr>
<tr>
<td>Floor count</td>
<td>−0.43</td>
<td>.01</td>
<td>0.85</td>
</tr>
<tr>
<td>Moving distance</td>
<td>0.361</td>
<td>.02</td>
<td>0.85</td>
</tr>
<tr>
<td>5 Step count</td>
<td>0.916</td>
<td>&lt;.001</td>
<td>0.88</td>
</tr>
<tr>
<td>Moving distance</td>
<td>0.259</td>
<td>.09</td>
<td>0.88</td>
</tr>
<tr>
<td>Floor count</td>
<td>−0.224</td>
<td>.20</td>
<td>0.88</td>
</tr>
<tr>
<td>Body weight</td>
<td>−0.08</td>
<td>.21</td>
<td>0.88</td>
</tr>
<tr>
<td>Gait type</td>
<td>−0.155</td>
<td>.03</td>
<td>0.88</td>
</tr>
</tbody>
</table>

*aRegression models combining 2 vibration parameters with and without participant characteristic parameters are shown in Multimedia Appendix 2. “Gait type” is binary data (mid- or forefoot strike vs heel-strike foot landing).*

**Step Count and Activity Intensity Predictions**

The mixed model ANOVA showed a significant effect of the measurement method and a significant interaction with the activity for both the predictions of the number of steps and activity intensities ($P<.001$).
Significant underestimations in the number of steps were noted for the wrist-worn ActiGraph device predictions when all activities were considered together (Figure 5A, "total").

Regarding the prediction of activity intensities (Figure 5B), the pairwise comparison analyses revealed statistically significant overestimations for the wrist-worn ActiGraph device across all activities. The wrist-worn device and the floor-vibration system did not exhibit any difference with the actual number of steps. When activities are considered separately, pairwise comparisons indicate that the wrist-worn ActiGraph device underestimated the number of steps completed during cleaning but overestimated it for sitting and ironing activities. The floor-vibration–based predictions overestimated the number of steps required for sitting.

Regarding the prediction of activity intensities, the pairwise comparison analyses revealed statistically significant overestimations for the wrist-worn ActiGraph device across all activities. The floor-vibration system did not show any differences with the actual intensities evaluated by indirect calorimetry. Finally, the wrist-worn ActiGraph device showed a slight but significant underestimation for the estimations of the ironing activity intensities. The results are shown in Figure 5.

Figure 5. Comparison of prediction methods for each activity. (A) number of steps; (B) activity intensity. Energy expenditure measurements were collected for 2 minutes over the 6 minutes of each activity. Therefore, comparisons of measurement methods for the total energy expenditure are not presented in this figure. In panel B, the green bar shows the prediction of the model with the highest coefficient of determination (model 5; Table 2). The marks indicate a significant difference against the actual number of steps. MET: metabolic equivalent of task. *P<.05, **P<.001.

Discussion

Principal Findings

This study presented a novel quantitative method that uses the monitoring of floor vibrations for the evaluation of physical behaviors at home. The floor count, step count, and moving distance parameters were computed from the floor-vibration signal. Statistical models combining these 3 parameters showed significant correlations with the actual energy expenditure measured in 10 participants in a structured experiment that included 4 common home-based activities. In addition, the step count parameter did not show any significant difference with the number of steps completed by the participants during the experiment. The predictions of the floor-vibration monitoring system for both the activity intensity and number of steps were equal to or more accurate than those obtained by the Actigraphy method using the refined 2-regression model.

Floor Vibration for the Quantification of Physical Behaviors

The actual activity intensities measured by the indirect calorimetry for the 4 activities increased in accordance with the intensities presented in the compendium of physical activities [26], that is, 1.2 versus 1.3 MET; 1.9 versus 1.8 to 2.0 MET; 2.5 versus 2.5 MET; and 3.7 versus 3.3 MET for sitting (and watching video), ironing (and folding clothes), cooking (and setting the table), and cleaning the room, respectively (Figure 4A). The floor count, step count, and moving distance parameters also increased gradually for the 4 activities, indicating the feasibility of evaluating the intensity of physical behaviors in the home environment using the information provided by floor vibrations (Figures 4E-4G). Among them, step count was the only parameter that showed significant differences between all 4 activities. The single regression analyses also indicated a strong association between the step count and the actual activity intensities ($r^2=0.82; P<.001$), whereas floor count and moving distance showed a weaker but still significant association with the actual activity intensities ($r^2=0.56$ and 0.66, respectively; $P<.001$ for both). These pieces of information taken together may suggest that the estimated number of steps extracted from the floor-vibration signal could be used to make a reliable quantitative estimation of physical behavior in home settings. All the 3 parameters were designed to describe the inhabitant’s motion. Although the step count and floor count parameters can capture the physical dimension of the movement, the moving distance parameter adds a spatial dimension to the evaluation. The better performance of step count alone compared with moving distance alone may be due to the location approximation inherent to the limited number of sensors used to cover the entire house surface (Multimedia Appendix 2). A step is detected and not affected by any approximation. On the other hand, floor count may also be susceptible to inaccuracies, potentially due to variations in floor-vibration wave attenuation. Factors influencing this attenuation include the proximity of furniture, its weight and
contact surface with the floor, proximity of support beams and walls, and the irregular geometry of the Ocha-House floor area.

Despite the possible weaknesses of the floor count and moving distance parameters, multiple linear regression models combining step count, floor count, and moving distance still exhibited stronger associations ($r^2 \geq 0.85$; Table 1; Table S1 in Multimedia Appendix 2). As presented in Table 2, the inclusion of body weight and gait type parameters as descriptors greatly lowered the contribution of floor count to the model. Indeed, although floor count does not correlate with either body weight or gait type (Table S2 in Multimedia Appendix 2), it is still the only parameter extracted from floor vibrations that can quantitatively capture the forces applied on the floor. Given that the actual body weight can be easily inputted into any smart home system, the question of the relevance of extracting and using the floor count parameter to make accurate predictions of energy expenditure remains open. Additional studies, including a population with more heterogeneous anthropometric characteristics, may be needed to address this question further.

Finally, the ANOVA revealed that the predictions made by the floor-vibration monitoring system also showed less deviation than those of the 2 research-grade ActiGraph GT9X monitors for both the number of steps and activity intensity end points (Figure 5). The underestimation of the number of steps noted for the wrist-worn ActiGraph device may be related to walking gait characteristics when movements are performed in closed and narrow spaces. Such environments may not allow sufficient acceleration to meet the necessary signal processing threshold criteria required to count a step, as suggested elsewhere [32]. In contrast, the overestimated number of steps observed for the wrist-worn device may be the result of confounding upper limb movements performed in a frequency range similar to that of walking gait, which may occur in the course of completing housework-related activities.

Regardless of the performance of the ActiGraph GT9X monitors and although no external validation experiment has been conducted, taken together, these observations emphasize the good performance of the floor monitoring system for the quantitative evaluation of physical behaviors performed in home settings.

**Perspectives**

Although the market for wearable activity trackers is still in its growing phase [6], waist- and wrist-worn physical activity monitors, including research- or consumer-grade devices, have been associated with inaccurate predictions of daily physical activity [10-12]. During the past decade, the computation of accurate predictions for housework-related activities using traditional accelerometer-based activity tracker devices has been the object of specific software development [33,34]. However, this study still showed statistical differences between the predictions made by the ActiGraph GT9X monitors and the actual values for both number of steps and activity intensities. These observations emphasize the necessity of developing new methods that can accurately evaluate physical behaviors at home to improve the computation of daily physical activity metrics. When considering long-term use, it is crucial to distinguish between consumer- and research-grade devices. This study used 2 ActiGraph GT9X monitors, recognized as research-grade devices, in the context of a short semistructured experiment. However, consumer-grade physical activity tracking devices used in everyday life are subject to additional extrinsic limitations that can impede their ability to provide continuous monitoring. For instance, the common practice of removing watches and other wearables at home can have significant impacts on the evaluation of physical behaviors in home settings.

Considering the current limitations of wearable activity tracker devices, smart home systems, such as floor-vibration monitoring technologies similar to those used in this study, present a suitable opportunity to enhance the self-monitoring of daily physical activity. Such systems offer a novel approach to improve the accuracy of estimating energy expenditure and the number of steps performed at home, especially when considering their integration with a 5G network composed of interconnected devices dedicated to the evaluation of daily physical activity. By ensuring accurate and continuous measurements when individuals are at home, smart home systems could help maintain people’s interest in self-monitoring, making them a pivotal factor in promoting and sustaining active and healthy lifestyles. However, the potential widespread adoption of such systems should not only be considered from a technological perspective but should also acknowledge the role of sociocultural factors in shaping user acceptance and usability.

**Limitations and Strengths**

The main limitation of this study is that the proposed method only assesses the physical behaviors of 1 inhabitant at a time. Quantifying the physical behaviors of multiple individuals would require additional signal processing tools to link vibration events with the individuals generating them. Although each individual may exhibit a unique gait signature reflected in the floor-vibration signal, extracting such information is beyond the scope of this study. This limitation could also be addressed by analyzing the sequence of interactions with smart and connected home furniture devices, similar to what has already been described elsewhere [35]. Furthermore, it is important to note that the experimental Ocha-House used in this study was originally designed for a single inhabitant, aligning with the living environment of millions of Japanese people. The structured nature of the experimental protocol may be cited as a second limitation of the study, which restricts the generalization of the observations to what may happen under free-living conditions. To further evaluate the feasibility of using the floor-vibration–based monitoring method, semistructured experiments using a portable breath-by-breath gas exchange analyzer could be conducted to assess energy expenditure during longer periods of activity. A third limitation is that the external validity of the floor-vibration parameter–based activity intensity prediction models (Table 2) was not tested, thus mitigating the interpretations of the comparison test performed against the actigraphy method. The cost of the system is also considered to be a limitation. In this study, expensive, high-sensitivity shear-type accelerometer sensors were used. Further studies are necessary to explore the feasibility of using cheaper accelerometer sensors similar to those commonly used in wearable devices. The results of the
multiple regression analyses (Table 2: Multimedia Appendix 2) indicated that the floor-vibration–based step count and moving distance parameters contributed more to the activity intensity prediction models. These 2 parameters may not require the computation of a high-resolution signal. Fourth, the quasi-absence of responses in the floor count, step count, and moving distance parameters during sitting and watching videos may suggest that the floor-vibration monitoring system may also be capable of evaluating sitting behaviors (Figures 4E–Figures 4G). However, owing to the structured nature of the protocol, which involves short observation windows, further interpretations regarding the accuracy of the system in predicting energy expenditure for home-based sitting activities cannot be made. Given the importance, complexity, and intricacies of sedentary behaviors that can occur at home, specific studies should be designed to understand how floor-vibration monitoring systems may contribute to the objective assessment of home-based sedentary behaviors. Finally, the participants in this study were all women and exhibited relatively homogeneous characteristics in terms of age and weight. The BMI scores indicated a limited variability in physical fitness. These observations constrain the generalizability of our results to a more diverse population. Given that age and physical fitness are recognized determinants of energy expenditure, future investigations should aim to recruit a more diverse sample of participants and consider a broader spectrum of personal characteristics in the development of energy expenditure prediction models.

This study has several strengths and originalities. First, the results of the present experiment are in line with those of previous studies, which described a good relationship between the force exerted by an individual on the floor of a small squared 6.25 m² metabolic chamber equipped with a force transducer and the actual energy expenditure [36-38]. They allow extending the previous observation to a different sensing technology, larger non–squared living surfaces, and a wider range of activities that are usually performed at home. Another strength of this study is that it is the first to compare the outcomes of a smart home system with those of research-grade activity trackers.

Conclusions

This study presents a novel floor-vibration monitoring system that can be used in smart home settings to quantify physical activity at home. The hardware included high-sensitivity accelerometers. In this case, 8 sensors were required to cover a surface area of 42 m². The software includes a simple data processing workflow for the computation of the floor count parameter, which is a quantitative index of the floor vibrations, and the step count and moving distance parameters. Regression models combining the information of these 3 parameters showed a strong association with the actual intensities measured using indirect calorimetry for the 4 tested home-based activities. A significant association was also observed between the step count parameter computed using the floor-vibration signal and the actual number of steps. Further studies, conducted under real-life conditions or using semistructured experimental protocols, are necessary to extend the results of this study and validate the monitoring of floor vibrations as a surrogate method for evaluating physical behaviors at home. Considering the current evolution of 5G technologies and IoT devices, smart home systems are expected to contribute to a more continuous evaluation of daily physical activity.

Acknowledgments

The authors would like to thank the individuals who participated in this study. NM, YO, and JT would like to acknowledge all the research students who contributed to the Ocha-House project in the past: Chihiro Nakajima, Saki Kaneko, Mana Akao, and Mio Sasaki. This study has received research grants from the Japan Society for the Promotion of Science (Grant-in-Aid for Young Scientists-B 15K20997 directed to JT and Grant-in-Aid for Scientific Research-C 21K11335 directed to JT and YO) and from the Japanese Foundation for the Promotion of Precision Measurement Technology (精密測定技術振興財団研究費).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary descriptions and results of the calibration and walking trials.
[DOCX File, 841 KB - formative_v8i1e51874_app1.docx]

Multimedia Appendix 2
Supplementary analyses and results.
[DOCX File, 293 KB - formative_v8i1e51874_app2.docx]

Multimedia Appendix 3
Data used for the statistical analysis.
[XLSX File (Microsoft Excel File), 20 KB - formative_v8i1e51874_app3.xlsx]

References


Abbreviations

MET: metabolic equivalent of task

Edited by A Mavragani; submitted 16.08.23; peer-reviewed by A Martinko, H Namba; comments to author 25.09.23; revised version received 24.12.23; accepted 03.01.24; published 25.04.24.

Please cite as:
URL: https://formative.jmir.org/2024/1/e51874
doi:10.2196/51874
PMID:38662415
Developing Social Enhancements for a Web-Based, Positive Emotion Intervention for Alzheimer Disease Caregivers: Qualitative Focus Group and Interview Study

Ian Kwok¹, PhD; Emily Gardiner Lattie¹, PhD; Dershung Yang², PhD; Amanda Summers¹, MA; Paul Cotten³, BA; Caroline Alina Leong¹, BA; Judith Tedlie Moskowitz¹, MPH, PhD

¹Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
²BrightOutcome, Buffalo Grove, IL, United States
³University of California San Francisco, San Francisco, CA, United States

Abstract

Background: Alzheimer disease is a degenerative neurological condition that requires long-term care. The cost of these responsibilities is often borne by informal caregivers, who experience an elevated risk of negative physical and psychological outcomes. Previously, we designed a positive emotion regulation intervention that was shown to improve well-being among dementia caregivers when delivered through one-on-one videoconferencing lessons with a trained facilitator. However, the format required significant resources in terms of logistics and facilitator time. To broaden the reach of the intervention, we aimed to develop the Social Augmentation of Self-Guided Electronic Delivery of the Life Enhancing Activities for Family Caregivers (SAGE LEAF) program, an iteration of the intervention in a self-guided, web-based format with enhanced opportunities for social connection.

Objective: The aim of this study was to gather feedback to inform the design of social features for the SAGE LEAF intervention. In the absence of a facilitator, our goal with the self-guided SAGE LEAF intervention was to integrate various social features (eg, discussion board, automated support, and profiles) to maximize engagement among participants.

Methods: Qualitative data were collected from 26 individuals through (1) interviews with participants who completed a previous version of the intervention via videoconferencing with a facilitator, (2) focus groups with dementia caregivers who had not previously experienced the intervention, and (3) focus groups with Alzheimer disease clinical care providers. We conducted a qualitative thematic analysis to identify which social features would be the most helpful and how they could be implemented in a way that would be best received by caregivers.

Results: Interview and focus group feedback indicated that participants generally liked the potential features suggested, including the discussion boards, multimedia content, and informational support. They had valuable suggestions for optimal implementation. For example, participants liked the idea of a buddy system where they would be matched up with another caregiver for the duration of the study. However, they expressed concern about differing expectations among caregivers and the possibility of matched caregivers not getting along. Participants also expressed interest in giving caregivers access to a podcast on the skills, which would allow them to review additional content when they wished.

Conclusions: Taken together, the discussions with caregivers and providers offered unique insights into the types of social features that may be integrated into the SAGE LEAF intervention, as well as implementation suggestions to improve the acceptability of the features among caregivers. These insights will allow us to design social features for the intervention that are optimally engaging and helpful for caregivers.

(JMIR Form Res 2024;8:e50234) doi:10.2196/50234

https://formative.jmir.org/2024/1/e50234
KEYWORDS
Alzheimer disease; dementia; caregiving; eHealth; web-based interventions; positive emotion; stress; coping

Introduction

Background
The impact of Alzheimer disease (AD) continues to broaden as the global average life expectancy grows [1,2]. Consequently, the number of individuals who will assume the role of a primary caregiver of a friend or family member with AD is expected to rise, with estimates indicating that informal care accounts for 40% of the total cost of care [3]. In the United States alone, this amounts to an estimated annual total of 18.6 billion hours of unpaid care [4].

At the individual level, the protracted nature of AD results in an extended caregiving role that intensifies as the care recipient’s health gradually declines [5]. For example, initial caregiving responsibilities may include assisting with activities of daily living such as providing transportation, preparing meals, and helping with chores [6]. However, in more advanced stages of the disease, caregivers often have to cope with agitation, aggression, and wandering [7] while shouldering an increasing logistical and financial burden of coordinating care [8,9].

The weight of these responsibilities comes at a cost to caregivers, who experience adverse mental health outcomes such as increased depression, anxiety, and suicidality [10-13] as well as negative consequences for physical health, demonstrated by increased sleep disturbance, fatigue, and undernutrition [14,15]. This, in turn, may lead to a decline in the quality of care and, subsequently, poorer outcomes for the care recipient [5].

In light of the growing recognition of the stress of AD caregiving, researchers are developing targeted interventions that offer a combination of psychoeducation, social support, and psychological support for caregivers [16,17]. Of note, researchers are increasingly using eHealth technologies that leverage electronic information and communication (eg, telehealth, mobile apps, and web-based applications) to broaden the dissemination of these resources [18,19]. After the onset of the COVID-19 pandemic, eHealth technologies found renewed significance when in-person AD support services were suspended, necessitating a rapid shift to telehealth offerings [20].

This study was the first step in the adaptation of an existing caregiver intervention into a socially enhanced web-based intervention—Social Augmentation of Self-Guided Electronic Delivery of the Life Enhancing Activities for Family Caregivers (SAGE LEAF). SAGE LEAF will comprise a positive emotion regulation curriculum that has been shown to be helpful for individuals experiencing significant life stress, including those with type 2 diabetes, metastatic breast cancer, HIV, and depression [21-24]. In a previous study, the intervention was also tailored specifically for dementia caregivers and delivered through videoconferencing by trained facilitators [25]. The intervention was effective at reducing symptoms of depression while improving self-reported physical health and positive emotion outcomes. However, the one-on-one facilitation that was provided for participants required a significant commitment of resources in terms of recruiting and training facilitators as well as an estimated 6 to 8 individual contact hours per participant over the course of 6 weeks. Hence, our goal was to tailor the self-guided, web-based version of the intervention for caregivers while incorporating unique social features that may help foster engagement among participants.

Objectives
Specifically, our aim was to enhance social presence, defined by computer-human interaction researchers as the perception of others in a virtual environment [26-28]. The construct has been shown to be associated with enhanced perceived learning and satisfaction in e-learning environments [29]. However, to identify potential social features that may be helpful for AD caregivers, it is also necessary to first understand how they currently use social technologies to support their caregiving activities and emotional well-being. For example, caregivers may use discussion boards hosted by the Alzheimer’s Association [30] or on social platforms such as Facebook or Reddit [31]. Furthermore, in the context of the COVID-19 pandemic, caregivers are now increasingly reliant on these social technologies with the rapid shift from in-person to virtual support resources, which include videoconferencing support groups for caregivers [32].

Hence, to examine the preferences and requirements of AD caregivers, this study aimed to solicit feedback on a set of potential social features for the SAGE LEAF intervention. These were identified from a review that we conducted on social features that were being implemented on research-focused and commercial eHealth applications (I Kwok, unpublished data, May 2021) in consultation with study team members and developers who were involved in the design of previous versions of the intervention (Textbox 1).

We collected feedback through (1) individual interviews with caregivers who completed the previous version of the intervention [25], (2) focus groups with dementia caregivers who had not yet been exposed to the intervention, and (3) focus groups with clinical providers of people with AD. The findings will inform the types of social features to be included in the SAGE LEAF intervention and how they can be implemented in a way that is most beneficial for caregivers.
Individual interviews: dementia caregivers who participated in a previous version of the intervention [36].

Profiles: a profile page that may be shared with others. Examples of profile content include being able to choose an avatar and answering some questions about themselves. Such features involve varying levels of self-disclosure, data management, and personalization that may enhance the sense of the presence of others in the intervention [36-38].

Private messaging: participants are able to send each other private messages. Some examples include messaging through SMS text messages [39,40] or commercially available applications such as Facebook and WhatsApp [41] or built into the eHealth intervention [42,43].

Buddy system or matching: pairing participants who are going through the intervention at the same time. Some examples include matching participants who are going through the intervention at the same time or with someone who has previously completed the program as a “peer mentor” for each other or enrolling a partner whom the participant has an existing relationship with [44-46].

Videoconferencing group: a facilitated group with participants through videoconference. Similar formats include support group and education or training videoconferencing sessions [47,48].

Discussion board: a web-based discussion board on the content of the lessons. Possible discussion board enhancements include notifications for when other users like or comment on their posts [49-51].

Automated support: this may include reminders or notifications for participants who are not logging in or feedback collected at the end of lessons. Such features have been shown to enhance adherence in eHealth interventions [52,53].

Multimedia, videos, or podcasts: these forms of multimedia are commonly used to disseminate educational material in a way that engages participants, thereby promoting literacy and enhancing health-related outcomes [54]. This might include testimonials and quotes from previous participants or messages from the study team.

Informational support: frequently asked questions or other information about how to connect with caregiver organizations, online support groups, and mental health care providers. Such resources may enhance a participant’s sense of perceived support and information competence [55].

Methods

Design

A combination of focus groups and interviews was conducted to solicit feedback on the short list of potential social features for the SAGE LEAF intervention. Subsequently, a qualitative analysis was conducted on the transcripts to identify and gather feedback on each feature.

Sample and Sampling

Three groups of participants were recruited (Tables 1 and 2):

1. Individual interviews: dementia caregivers who participated in a previous version of the intervention [36]. We emailed individuals who had previously provided consent for recontact and provided them with information about the interviews. If they wished to participate, they completed a screener survey to determine whether they met the following eligibility criteria: (1) currently identifying as the primary family caregiver of a person with dementia, (2) ability to speak and read English, (3) access to high-speed internet, and (4) access to a webcam for videoconferencing. The interviews were conducted by the lead author (IK), who adhered to a semistructured interview protocol to guide the discussions. They lasted approximately 45 to 60 minutes, and topics included (1) the types of social connection technologies that participants use in their everyday life, (2) reactions to potential social features that may be implemented for SAGE LEAF (eg, private messaging, discussion board, and virtual profiles), and (3) solicitation of suggestions for other social features not previously mentioned.

2. Caregiver focus groups: dementia caregivers from Northwestern Memorial Hospital’s Cognitive Neurology and Alzheimer’s Disease Center (CNADC) were recruited for 2 focus groups comprising 5 caregivers each (n=10). We emailed caregivers who had previously provided consent to be contacted for research purposes through the CNADC and provided them with information about the focus groups. If they wished to participate, they were asked to complete a web-based screener survey to determine whether they met the following eligibility criteria: (1) currently identifying as the primary family caregiver of a person with dementia, (2) ability to speak and read English, (3) access to high-speed internet, and (4) access to a webcam for videoconferencing. The focus groups lasted approximately 90 to 120 minutes and were similar in content to the interviews.

3. Clinician focus groups: we contacted clinicians who provided care for people with AD or their family or informal caregivers (eg, physicians, nurses, and social workers) from the CNADC and the University of California, San Francisco’s Memory and Aging Center via email. Both are comprehensive research and care centers that treat AD; hence, clinicians are involved in a broad range of AD programs that integrate patient care, training, and research. Interested clinicians completed a screener survey where they could indicate their professional experience to determine whether they met the following eligibility criteria: (1) current employment as a clinician for patients with AD and their caregivers (eg, physicians, nurses, and social workers), (2) access to high-speed internet, and (3) access to a webcam for videoconferencing. One focus group was conducted for AD care providers (n=6).
Table 1. Caregiver participant characteristics (n=20).

<table>
<thead>
<tr>
<th>Caregiver characteristics</th>
<th>Interview participants (n=10)</th>
<th>Focus group participants (n=10)</th>
<th>All caregivers (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8 (80)</td>
<td>7 (70)</td>
<td>15 (75)</td>
</tr>
<tr>
<td>Male</td>
<td>2 (20)</td>
<td>3 (30)</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>0 (0)</td>
<td>1 (10)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>White</td>
<td>10 (100)</td>
<td>9 (90)</td>
<td>19 (95)</td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>70.20 (9.48)</td>
<td>63.40 (6.52)</td>
<td>66.80 (8.65)</td>
</tr>
<tr>
<td>Years of caregiving, mean (SD)</td>
<td>7.50 (2.50)</td>
<td>4.80 (3.19)</td>
<td>6.15 (3.17)</td>
</tr>
</tbody>
</table>

Table 2. Provider participant characteristics (n=6).

<table>
<thead>
<tr>
<th>Provider characteristics</th>
<th>Focus group participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
</tr>
<tr>
<td>Nursing</td>
<td>3 (50)</td>
</tr>
<tr>
<td>Social work</td>
<td>3 (50)</td>
</tr>
<tr>
<td>Years in practice, mean (SD)</td>
<td>17.67 (11.33)</td>
</tr>
<tr>
<td>Patients with AD³ (%)</td>
<td>77.67 (31.89)</td>
</tr>
</tbody>
</table>

³AD: Alzheimer disease.

Data Extraction Procedure

We conducted a qualitative thematic analysis to identify (1) positive and negative feedback and implementation suggestions for each of the potential social features and (2) additional social features suggested by participants.

The interview and focus group recordings were transcribed and uploaded to the Dedoose qualitative and mixed methods research application (SocioCultural Research Consultants) in preparation for thematic analysis. Because the interviews and focus groups followed a similar structure, with questions asked about the same set of features, all transcripts were collated and analyzed together. In total, 2 independent coders performed the coding (IK and CL). First, they conducted open-ended coding of a test transcript to extract a preliminary list of codes. The coders met to discuss their observations, resolve differences in these observations, and refine the list of codes. An additional study team member (JM) was involved in reviewing the list to ensure that it captured the potential range of feedback. Next, to ensure a high level of agreement, the coders performed an interrater reliability test on 3 transcripts. While there is debate about the applicability of interrater reliability in interpretive qualitative research, its measurement allows for greater transparency and motivates adherence to the established coding guidelines [57]. During the test, a Cohen κ of 0.68 was achieved, which is regarded as a substantial level of agreement [58]. Subsequently, the coders met to resolve the discrepancies highlighted by the test and refined the codebook once again.

Next, the coders proceeded with a first round of independent coding that comprised 5 transcripts. They randomly selected one of the transcripts to be double coded and met to discuss any discrepancies in their coding. Finally, they proceeded with a second round of independent coding of the 5 remaining transcripts.

The codes were organized based on the type of social feature that they referenced. For each social feature, subcodes were created for positive feedback, negative feedback or barriers, and implementation suggestions. Additional codes were created for other social technologies mentioned (eg, WhatsApp and Facebook), web-based resources, and additional ideas suggested by participants. In the Results section, the results are categorized by social feature type. Quotes from participants are presented with the following identifiers: (1) “PX” to denote participants who completed a previous version of the intervention, (2) “CR” to denote caregivers who participated in the focus groups, and (3) “PR” to denote providers who participated in the focus groups. Due to the personal nature of the interviews and focus groups, the full transcripts are not publicly available.

Ethical Considerations

The protocol for this study was approved by Northwestern University’s institutional review board (STU00215548) before conducting the study. Eligible participants were sent an electronic consent form via email delineating the risks and benefits of participation, at which point they could confirm their preference to participate or decline participation in the study. Data collected from REDCap (Research Electronic Data...
Peer Groups or Cohort

Participants generally liked the idea of being in a peer group, and caregivers drew comparisons to their previous participation in caregiving support groups. However, they expressed concern as to how they would be matched into groups:

I think if you want to have peers, you do need to match them up closely. By the progression of disease maybe. [CR2]

I think that you have to be careful as to who’s in the peer group...I’ve been in other support groups where [caregivers were caring for] somebody had dementia, somebody with Alzheimer’s, and [another caregiver was caring for someone with] PPA (primary progressive aphasia). And they’ve got similarities, but they’re so drastically different. [CR3]

Furthermore, participants had varying opinions on how caregivers should be grouped together. Some suggestions included grouping by age, relationship to the care recipient, or type of dementia. One participant also suggested that it may be helpful to group individuals based on their recreational interests instead of their caregiving status:

I think it’s also good to have in a peer group...And I don’t think gender or anything like that matters whatsoever, it’s more about where they are and what they’re doing at the time, with the same type of diagnosis. [CR1]

Well, I think that general background is so important...You know, educational—the number of degrees you get isn’t as important as interests. Are you interested in art? Are you interested in gardening?...It’s one way that certainly people get together. [PX1]

Hence, type of dementia or progression of the disease were the most preferred methods of grouping individuals. Participants in the focus groups and interviews were able to readily articulate their grouping preferences by tapping into their own experience of caregiving and their specific needs based on their care recipient’s diagnosis. This underscores the utility of segmenting future participants by type of dementia diagnosis.

Profiles

There is a high degree of variability in how user profiles are used in eHealth platforms, ranging from toggling basic settings to more extensive social features such as managing invites, notifications, and social groups. Our proposed enhanced user profiles would allow participants to customize the way in which they present themselves in the virtual space and would help lay the groundwork for future social interaction on the SAGE LEAF platform. Such user profiles might include the ability to select an avatar or allow participants to display their name and other personal information if they choose to do so.

Similar to the feedback on the peer groups, focus group and interview participants were primarily interested in the caregiving-related characteristics of other users, such as the relationship between the caregiver and care recipient and type of dementia diagnosis:

Family relationship. If they’re the adult child or spouse/partner. I’ve also—if it’s the younger onset versus later onset. [PR1]

I think people would be interested in knowing the diagnosis of the person the other caregivers are caring for. So, if someone is an adult child caring for their parent with Alzheimer’s, I feel like they would be interested in meeting other people in similar situations. [PR3]

I think age is a big factor. I’ve had a lot of people who are younger, like maybe family caregivers, who are more interested in talking with somebody their own age. Or around—near their own age. [PR4]

Taken together, this feedback suggests that the information that is shared in the profiles may be similar to the variables by which participants might be grouped together. This emphasizes the synergy between these features and how they may foster a sense of shared identity among participants.

Private Messaging

A private messaging function would allow participants to contact each other individually on their own and would function similarly to social networking platforms such as Facebook and Instagram that allow for direct messaging. Overall, feedback was positive, but one clinician articulated their concerns about privacy and security:

...in my support group there are people who request, you know, to be connected to each other. And I think I always try to make sure that I ask both parties before I connect them...So maybe the option to stay private or to be public. [PR5]

Contrary to our expectations that caregivers may share similar concerns about security and privacy regarding being contacted, the feedback suggests that caregivers were open to the idea of being able to send and receive messages and expressed minimal hesitation toward the private messaging feature:

Oh, I think that would be fine. You know, I see nothing wrong with that. I think, you know, friendships could be formed out of that. And private support and like-minded thinking people...I see absolutely nothing wrong with that. [PX10]

I think that’s fine. I think that would be good because you’re going to have to find a way to build trust, and then it’s also somebody who is in a similar circumstance than yourself. So, yeah, so that’s how I would look at it. [PX12]
Hence, feedback was generally positive on the ability to connect with other caregivers through a private messaging feature. Participants emphasized that their potential willingness to use this feature was based on the assumption that other participants would be in a similar caregiving situation to theirs. For instance, participants indicated that the similarity of the diagnosis of their care recipient or whether they were also spousal or other family caregivers were important attributes that might influence their use of this feature.

**Buddy System or Matching**

To maximize the sense of social presence that future participants might experience, we initially proposed a buddy system where participants would be paired up with a peer or “buddy”—either with someone who was going through the program at the same time as them or with participants who had previously completed the intervention. This social feature would complement the peer groups in that participants would be able to feel like they were part of a group while being able to connect individually with other participants. Alternatively, this could be deployed as a stand-alone feature in the event that there were not enough participants to form a cohort. Overall, we received mixed feedback from caregivers and clinicians on the concept of a buddy system.

In their feedback, caregivers expressed interest in this idea because it would provide accountability for progressing through the program to their potential buddies and enhance their motivation to engage with the content:

> Oh, it totally would [be helpful]. Because I would be more concerned about disappointing the other person. “Oh, they need me! I have to check my email,” or “I have to check that text. I don’t want to disappoint them.” [CR3]

> I think it’s a great idea. You know, it would have been nice, if I had had one, but I was just flying by the skin of my teeth and sometimes I crash landed. [PX7]

In contrast, clinicians expressed significant concerns about the implementation of a buddy system. They described past experiences with similar efforts where the matching was unsuccessful or burdensome and led to a disappointing experience for the caregivers involved:

> I’ve tried connecting caregivers that I work with, and unless they really hit it off naturally in most cases it doesn’t work out. [PR5]

> I think there can be a problem in the two caregivers having really different expectations about what the relationship is going to be...I think it would add a level of burden to caregivers too... [PR2]

Hence, while caregivers expressed enthusiasm for this idea, clinicians spoke from their own past experiences and were strongly against the idea of matching because they found it challenging to establish shared expectations and to anticipate whether caregivers who were matched would get along well. Therefore, it is unclear whether the benefits of the buddy system may be outweighed by potential complications that arise from these unanticipated social dynamics.

**Videoconferencing Group**

At present, support groups form a crucial resource for AD caregivers, as demonstrated by the wide range of group programming in both virtual and in-person formats [17,59]. Hence, another possible feature was a videoconferencing group where participants would be able to log in at a given time during the week to connect with other caregivers with the specific focus on discussing the skills being taught in the program.

The onset of the COVID-19 pandemic hastened the transition of in-person support groups to web-based videoconferencing groups. This transition was demonstrated in the readiness that caregivers expressed in adopting these videoconferencing technologies. It should be noted that all the interviews and focus groups took place at the start of the pandemic:

> I have been surprised that the Zoom meetings—I’ve gone to many of them...I should point out I’m 83 years old, okay?...via online, that sort of thing, would be very good for a person like me. [PX1]

In terms of implementation suggestions, one participant highlighted that these videoconferencing groups would be a good addition to the program as long as participation was optional. This underscores the importance of building flexibility into the social features being offered as caregivers have competing demands or may simply prefer different features:

> I think if you could make it as an offering but not a requirement...But I think you have to be understanding of the fact that not everybody’s going to be able to do that at the same time...it’s hard for me to commit weekly to a certain time. [PX10]

One concern that was expressed by both caregivers and clinicians was the importance of making sure that the videoconferencing group discussions stayed on topic. Caregivers articulated various past experiences where their time was not spent efficiently because other participants deviated from the focus of the discussions:

> ...I would go and check out other groups, and that was always a real disappointment. And I would not go back to those when, you know, somebody would just insist on eating up the entire hour with their issues. And so that’s a problem... [PX11]

> I would gravitate toward anything where there was some real-time moderation or facilitation, just to help keep the learning on track. [CR5a]

The feedback suggests that videoconferencing groups can be helpful for caregivers. However, there was concern about the efficiency of these meetings, which could be addressed by having a facilitator who is able to moderate and guide the discussions. A facilitated group would allow participants to discuss the topics freely while ensuring that the time is directed toward the topics and skills taught in the program. Participants also liked the idea of having a portion of the videoconferencing group sessions be not necessarily related to caregiving or the positive emotion skills taught in the program, with several participants expressing interest in an informal happy hour where they could connect with each other casually.
Discussion Group

Caregivers often seek information about their care recipient’s diagnosis, behaviors, and symptoms through the internet. Hence, many already participate in AD-specific discussion groups that are associated with the Alzheimer’s Association or informal groups that proliferate on social media platforms such as Facebook and Reddit. In line with our expectations, participants were generally open to the idea of using a discussion board. One clinician suggested that there may be some overlap with these existing platforms, which could present a barrier to participants using the discussion board:

...some of the feedback we get from caregivers is that, “You’re asking me to do something I already have a mechanism for doing that. So, I already have a way to share photos with people that I’m close to, it’s called Facebook or whatever. But you’re asking me to sort of do it in this different venue.” So that’s been a negative when you’re asking somebody to do something, that they already have a way to do that. [PR2]

Furthermore, both participants and providers emphasized that the use of the discussion board would be contingent on how the benefits of engaging with it were conveyed to participants. Some of their suggestions included highlighting how the discussion board could amplify their practice of the skills or allow them to feel more connected with other participants in the study:

I think there are some advantages, and that if you really say the discussion board is to really talk about the skills or share examples of where you use the skills...And if you framed it so that—I could even see it as being a way to amplify the skills. [PR3]

...to encourage people and say, “Hey, look at, you know, it’s normal for you to feel isolated and trying to get questions answered. It’s worth it to try and work with these tools.” [PX12]

This feedback suggests that caregivers may be open to using the discussion boards, yet there were concerns about how these discussion boards might duplicate existing resources. Thus, it is essential to highlight the benefits of engaging with the discussion board associated with the positive emotion skills program to encourage its use. This may be in the form of prompts or reminders to participants about these benefits.

Automated Support

As described in Textbox 1, automated support would comprise notifications or reminders that are sent out based on certain triggers, for example, if a participant does not log on to the platform for a certain number of days or if they endorse poor mood for an extended period. With automated support, the intention is to provide caregivers with a sense that their participation is valued and that we would be responsive to their level of engagement. Similar to the feedback collected in previous versions of the intervention, participants found the concept of reminders helpful but expressed the need for these messages to be framed in a way that was supportive and encouraging:

I guess that’s where I would give them points, and like, more like entice them rather than nag them. [PR5]

Because when you first said it [automated support], it was totally irritating to me. I thought, “I’m doing this to take care of myself, and now you’re making me accountable?!” I don’t have time today!” And then after you talked a little more, then I felt better about it...I think it’s how you frame it. Or how I frame it for myself. [CR3]

You could try to be really empathetic and kind of understand why they didn’t get to it, versus the risk that if someone got an automated message that might just add to their sense of everything negative about why they haven’t done the skills. [PR2]

Across the board, providers and caregivers reiterated the importance of supportive and encouraging messaging when implementing the automated support features. This underscores the importance of emphasizing the rewarding aspects of participation—instead of reprimanding or penalizing caregivers for not using the various features. Furthermore, it may be helpful for this supportive language to be integrated not only into the automated support features but throughout the intervention as well—for example, using the registration emails, videos, and podcasts as opportunities for cheerleading and supporting participants.

Multimedia, Videos, or Podcasts

Multimedia content such as videos and podcasts may help enhance the perception that there are study staff members behind the program and other caregivers who are involved in the study. In previous versions of the intervention, caregivers worked one-on-one with a facilitator to learn the skills. To compensate for a lack of face time in this self-guided format, we proposed the addition of multimedia content to make the skill-building lessons more engaging by hearing directly from the team members involved in the development of the intervention. In their feedback, participants unanimously liked the idea of including this multimedia content:

I think that would be good. I mean, again, it takes the program out of being a program and puts it into a dialogue with someone. And I think it’s always good to see the face of the people who are running the program. [PX1]

It might be encouraging for them to hear and see that they’re not alone, that others have gone through it and have come out on the other side. [PX14]

Some participants suggested the idea of including a podcast as part of the program. This would allow caregivers to review the material at a time that is most convenient for them. This is consistent with feedback on other social features, in which participants suggested that flexibility may be helpful for caregivers who are busy:

And I like the idea of the podcast, so that you can do it on your time and when it’s convenient for you...ten, less than ten minutes here and there throughout the week... [CR1]
I participate in a 30-day class right now...It is a five-minute podcast that she sends, along with a list of daily activities and a curriculum has been provided in advance. So, you know that the five-minute podcast is five minutes out of your day, and you can do that, it's pretty easy to find five minutes. [CR4]

The overwhelmingly positive feedback on the proposed multimedia social features demonstrates that participants are interested in the sharing of insights from the study team as well as from previous participants. In the absence of live communication, their feedback suggests that such multimedia features may be central to developing a sense of social presence.

Informational Support

Caregivers often use web-based resources to seek information about providing care for their loved ones with dementia. While participants felt that informational support could be helpful, their feedback suggests that it was important for it to be targeted and provide specific information that was useful for caregivers:

But be real specific...the specific information is way more helpful. [PX6]

...my husband’s diagnosis is not specifically Alzheimer’s...a lot of the things that have to do with the Alzheimer's Association don’t apply to him... [PX10]

...the referral...You know, a piece of paper with 20 different organizations on it were not helpful. [CR5a]

Their feedback also suggests that many caregivers are discerning about such resources and sophisticated in their information search methods. Hence, the informational support provided by the intervention should be thorough and specific for it to be meaningful for participants. For example, participants indicated that it would be helpful if such resources were organized by geographic location or if they could be sorted in a way that would make it easy for participants to find the resources that are most helpful to them. Another approach would be to provide additional resources that relate specifically to the skills that are being taught.

Other Social Features

We also collected feedback on other social features that might be helpful for caregivers. One participant suggested that the study team solicit participants’ input throughout the program to foster a sense of involvement. This has some similarities to the brief survey that we will provide at the end of each lesson asking participants to rate how they felt about the lesson on a scale from 1 to 5 stars. While this feature was not previously considered a social feature, the act of soliciting feedback provides participants with an opportunity to express their thoughts about the program and reinforces the sense that there is a study team who is collecting the feedback and trying to improve the intervention for the benefit of caregivers:

I think asking for opinions...getting involved in just what you’re doing and asking what I think. “Okay, what do you think of the program?” It’s certainly one way to get involved, as long as it’s done in such a way that it’s meaningful. [PX1]

Participants also mentioned how the use of other platforms such as Instagram or Facebook may complement the intervention. The feedback suggests that creating a parallel dialogue on these already used platforms could foster an enhanced sense of social connection. One benefit to using these platforms is that participants would be able to connect with each other regarding the positive emotion skills across multiple platforms, which may enhance their learning. Participants also mentioned how the COVID-19 pandemic heightened their sense of social isolation; hence, the integration of these popular social networking platforms may help caregivers feel more connected as they complete the study:

I think if you had something that allowed people to respond to one another, whether it was a chat room...they create a Facebook group that is specific to that course...And that those people during that course can talk to each other, and every now and then the instructor chimes in if she feels that there’s something that she can add to it or some guidance. But I think something where people could connect would be nice. [PX10]

I was thinking along the same lines of connecting on a specific theme, you mentioned gardening or cooking, I think those are the kinds of things that people do on Instagram or Facebook. But the WhatsApp group can be more private, so it can be formed just with the people who meet each other, and then they can share... [PR5]

Discussion

Principal Findings

In this study, we collected feedback on social features that may be implemented for a web-based positive emotion skills intervention for AD and other dementia caregivers. Through (1) individual interviews with participants who completed a previous version of the intervention, (2) focus groups with dementia caregivers, and (3) focus groups with AD clinicians, we collected information about the specific needs and preferences of caregivers in the implementation of these social features. Participants provided a number of insights into how to implement these features in a way that may be best received by caregivers.

Overall, participants provided extensive feedback on the proposed features and how they could be best implemented. However, they had fewer suggestions for additional features that might enhance a sense of social connection. This may be because we asked participants open-ended questions about additional features toward the end of the interviews and focus groups, at which point they may have exhausted their ideas or there may have been overlap with our proposed features. Nonetheless, participants were engaged throughout the discussions and provided unique insights into how we could refine our feature set.

Participants generally liked the proposed social features and provided valuable suggestions for how they could be improved. One such example is the multimedia content proposed for the
intervention, for which participants suggested a podcast format to allow caregivers to review this additional material at their convenience. This is similar to other exploratory eHealth interventions that have used this delivery format to enhance health literacy [60], weight loss [61], and self-compassion [62].

Other feedback related to the automated support features, in which participants emphasized the importance of providing encouragement to caregivers instead of shaming them for nonadherence. In a study of a physical activity intervention for older adults, it was found that, when the messaging was positively framed (ie, described in terms of the rewards and benefits of exercise as opposed to the costs of inactivity), participants’ pedometer readings indicated that they had walked more compared to those who received negative or neutral messaging [63]. Therefore, we could incorporate this positive framing, for example, if participants have not logged into the website for several days, and send personalized email messages letting them know that their participation is missed, while recognizing that caregivers have busy schedules, and reminding them that they might receive a boost in positive emotion by spending just a couple of minutes completing the home practice activities.

There were certain instances in which caregivers and clinicians differed in their feedback. For example, caregiver participants were generally open to the idea of being paired up with a buddy in the program. However, clinicians who had implemented similar programs were able to speak from their own experiences with attempts to match participants that were not successful based on differences in life experiences and expectations for engaging with a buddy program. Thus, although caregivers thought that they would enjoy a buddy feature, clinicians noted significant barriers to the implementation of this feature. Another example of disagreement between caregivers and clinicians was the private messaging feature, where one clinician highlighted concerns about privacy and security that caregivers did not report. Across eHealth interventions, researchers have far more information about how these platforms work and the accompanying benefits and risks compared to their participants [64]. Hence, researchers have an ethical responsibility to convey this information to participants. Recognizing these differing perspectives underscores the importance of integrating feedback from both caregivers and clinicians in refining these social features.

Participants were asked for additional suggestions for features that would enhance social connection or a sense of social presence. Their suggestions included soliciting feedback from caregivers as they progress through the program and using existing social media platforms to foster a sense of social connection beyond the SAGE LEAF intervention.

Further Research and Implications

The feedback collected from the focus groups and interviews will be used to inform the development of the social features for the SAGE LEAF intervention. This will include developing a list of “trigger events” for the automated support features and wording the notifications or reminders in a way that would be supportive to participants. We will also include enhanced user profiles where participants can toggle how they would like to receive notifications and share more detailed information about their caregiving status to other users if they wish to. We will also include videos and podcasts where study team members will introduce each positive emotion skill and suggest methods for mastering it.

The feedback from the focus groups and interviews also helped clarify which social features may be potentially challenging to implement, such as the buddy system. In future versions of the intervention, a study team member could facilitate a matching process among participants. However, this may require additional resources to implement.

The feedback made clear that informational resources were extremely helpful for caregivers. However, it was apparent that caregivers already seek these resources through web-based groups or informational websites hosted by caregiver organizations. Furthermore, it appeared that this information is most helpful when it is specific and tailored for the caregiver and care recipient. Recognizing that the primary aim of the intervention was to deliver the positive emotion skills and not more general caregiving skills per se and acknowledging that it would take significant resources to successfully implement these informational support features, this feature is less likely to be prioritized for inclusion in future iterations of SAGE LEAF.

While this study focused on all the potential social feature enhancements (ie, discussion boards, podcasts, and automated notifications) intended for SAGE LEAF, future versions of the intervention could explore which enhancements are most effective by using a factorial design where participants are randomly assigned to different combinations of the features to determine which can be most helpful or may best enhance a sense of social connection. In a randomized controlled trial of a previous version of the intervention designed for individuals with depressive symptoms [50], we randomly assigned participants to different combinations of enhancements and found that facilitator contact in combination with virtual badges yielded the highest participant engagement. Given that SAGE LEAF will be entirely self-guided, future research should explore which features, individually and in combination, lead to the biggest impact on caregiver engagement and well-being. For example, user profiles may help caregivers disclose more information about themselves and their caregiving circumstances, which may then enhance the quality of the interactions that take place on the discussion boards. Additional research may also involve measuring the extent to which these combinations of features lead to measurable increases in social presence—which is hypothesized to mediate the relationship between the application of these social features and desired intervention outcomes.

Strengths and Limitations

The feedback collected from caregivers and providers offered valuable perspectives not only on features that may be helpful and engaging but also on ways in which they may be implemented to best benefit caregivers. In addition, the combination of interview and focus group formats allowed for both individual feedback and group discussions to aid in the generation of ideas.
However, the semistructured format of the interviews and focus groups potentially limited the range of feedback collected. With our questions focusing primarily on the proposed features, this may have constrained the participants’ ability to provide novel ideas for new social features.

Another limitation to our study is the lack of ethnic diversity in our caregiver sample, which consisted of primarily White participants. To achieve a more diverse sample and perspectives in future studies, future research should oversample for underrepresented ethnic groups if needed.

Conclusions
This study involved a qualitative analysis of focus groups and interviews with caregivers and clinicians to determine which social features might be most helpful in tailoring a self-guided positive emotion intervention for AD caregivers. The feedback collected suggests that the participants were mostly open and receptive to the innovative social features we proposed. However, their lived and professional experiences provided unique insights into how best to implement these features in a way that would be helpful and engaging for caregivers participating in future versions of SAGE LEAF.

Acknowledgments
The authors would like to express their gratitude to the Mesulam Center for Cognitive Neurology and Alzheimer’s Disease for their collaboration in this research. This study was supported in part by a Small Business Innovation Research grant (R43AG065080) from the National Institute on Aging to BrightOutcome. In addition, this study was supported in part by an Alzheimer’s Disease Core Center grant (P30 AG013854) from the National Institute on Aging to Northwestern University, Chicago, Illinois. The authors gratefully acknowledge the assistance of the Clinical Core and its participants. Generative artificial intelligence was not used in the writing of this manuscript.

Data Availability
The data collected in the coding process may be obtained by contacting the first author (IK).

Conflicts of Interest
None declared.

References


37. Standard E. Human Factors (HF); personalization of eHealth systems by using eHealth user profiles (eHealth). European Telecommunications Standards Institute. 2010. URL: https://www.etsi.org/deliver/etsi_es/202600_202699/202642/01.01.01_50/es.202642v010101m.pdf [accessed 2023-04-04]


54. Aida A, Svensson T, Svensson AK, Chung UI, Yamauchi T. eHealth delivery of educational content using selected visual methods to improve health literacy on lifestyle-related diseases: literature review. JMIR Mhealth Uhealth 2020 Dec 09;8(12):e18316 [FREE Full text] [doi: 10.2196/18316] [Medline: 33295296]


Abbreviations

AD: Alzheimer disease
CNADC: Cognitive Neurology and Alzheimer’s Disease Center
REDCap: Research Electronic Data Capture
SAGE LEAF: Social Augmentation of Self-Guided Electronic Delivery of the Life Enhancing Activities for Family Caregivers

Edited by A Mavragani; submitted 23.06.23; peer-reviewed by A AL Sawafi; comments to author 06.02.24; revised version received 26.02.24; accepted 06.03.24; published 25.04.24.

Please cite as:

©Ian Kwok, Emily Gardiner Lattie, Dershung Yang, Amanda Summers, Paul Cotten, Caroline Alina Leong, Judith Tedlie Moskowitz. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Characterizing Technology Use and Preferences for Health Communication in South Asian Immigrants With Prediabetes or Diabetes: Cross-Sectional Descriptive Study

Lu Hu1, PhD; Laura C Wyatt2, MPH; Farhan Mohsin2, MPH; Sahnah Lim2, PhD; Jennifer Zanowiak2, MA; Shinu Mammen3, MPH; Sarah Hussain2, MBBS; Shahmir H Ali3, PhD; Deborah Onakomaiya4, MPH; Hayley M Belli2, PhD; Angela Aifah2, PhD; Nadia S Islam2, PhD

1Department of Population Health, Center for Healthful Behavior Change, Institute for Excellence in Health Equity, New York University Grossman School of Medicine, New York, NY, United States
2Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States
3Department of Social and Behavioral Sciences, School of Global Public Health, New York University, New York, NY, United States
4Vilcek Institute of Graduate Biomedical Sciences, New York University Grossman School of Medicine, New York, NY, United States

Corresponding Author:
Lu Hu, PhD
Department of Population Health
Center for Healthful Behavior Change, Institute for Excellence in Health Equity
New York University Grossman School of Medicine
180 Madison Ave
New York, NY, 10016
United States
Phone: 1 646 501 3438
Email: lu.hu@nyulangone.org

Abstract

Background: Type 2 diabetes disproportionately affects South Asian subgroups. Lifestyle prevention programs help prevent and manage diabetes; however, there is a need to tailor these programs for mobile health (mHealth).

Objective: This study examined technology access, current use, and preferences for health communication among South Asian immigrants diagnosed with or at risk for diabetes, overall and by sex. We examined factors associated with interest in receiving diabetes information by (1) text message, (2) online (videos, voice notes, online forums), and (3) none or skipped, adjusting for sociodemographic characteristics and technology access.

Methods: We used baseline data collected in 2019-2021 from two clinical trials among South Asian immigrants in New York City (NYC), with one trial focused on diabetes prevention and the other focused on diabetes management. Descriptive statistics were used to examine overall and sex-stratified impacts of sociodemographics on technology use. Overall logistic regression was used to examine the preference for diabetes information by text message, online (videos, voice notes, or forums), and no interest/skipped response.

Results: The overall sample (N=816) had a mean age of 51.8 years (SD 11.0), and was mostly female (462/816, 56.6%), married (756/816, 92.6%), with below high school education (476/816, 58.3%) and limited English proficiency (731/816, 89.6%). Most participants had a smartphone (611/816, 74.9%) and reported interest in receiving diabetes information via text message (609/816, 74.6%). Compared to male participants, female participants were significantly less likely to own smartphones (317/462, 68.6% vs 294/354, 83.1%) or use social media apps (Viber: 102/462, 22.1% vs 111/354, 31.4%; WhatsApp: 279/462, 60.4% vs 255/354, 72.0%; Facebook: Messenger 72/462, 15.6% vs 150/354, 42.4%). A preference for receiving diabetes information via text messaging was associated with male sex (adjusted odds ratio [AOR] 1.63, 95% CI 1.01-2.55; P=.04), current unemployment (AOR 1.62, 95% CI 1.03-2.53; P=.04), above high school education (AOR 2.17, 95% CI 1.41-3.32; P<.001), and owning a smart device (AOR 3.35, 95% CI 2.17-5.18; P<.001). A preference for videos, voice notes, or online forums was associated with male sex (AOR 2.38, 95% CI 1.59-3.57; P<.001) and ownership of a smart device (AOR 5.19, 95% CI 2.83-9.51; P<.001). No interest/skipping the question was associated with female sex (AOR 2.66, 95% CI 1.55-4.56; P<.001), high school education or below (AOR 2.02, 95% CI 1.22-3.36; P=.01), not being married (AOR 2.26, 95% CI 1.13-4.52; P=.02), current employment (AOR 1.96, 95% CI 1.18-3.29; P=.01), and not owning a smart device (AOR 2.06, 95% CI 2.06-5.44; P<.001).
Conclusions: Technology access and social media usage were moderately high in primarily low-income South Asian immigrants in NYC with prediabetes or diabetes. Sex, education, marital status, and employment were associated with interest in mHealth interventions. Additional support to South Asian women may be required when designing and developing mHealth interventions.

Trial Registration: ClinicalTrials.gov NCT03333044; https://classic.clinicaltrials.gov/ct2/show/NCT03333044, ClinicalTrials.gov NCT03188094; https://classic.clinicaltrials.gov/ct2/show/NCT03188094

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-019-3711-y

KEYWORDS
South Asian immigrants; type 2 diabetes; technology access; technology use; prediabetes; health disparities; mHealth; health equity; immigrant health; mobile health; smartphone; diabetes; diabetic; DM; diabetes mellitus; immigrants; prevention; regression; regression model; logistic regression; mobile health interventions

Introduction

The South Asian population represents one of the fastest-growing minoritized groups in the United States [1]. Between 2010 and 2017, the population size grew from 3.5 million to 5.4 million, representing an increase of 40% [2]. South Asians include a diverse group of individuals with ancestry from India, Bangladesh, Pakistan, Nepal, Bhutan, Sri Lanka, and the Maldives, and the majority of South Asian individuals in the United States are foreign-born immigrants [1,2]. Approximately 10% of South Asian immigrants in the United States live in poverty, with immigrants from Bhutan (33.3%), Bangladesh (24.2%), Nepal (23.9%), and Pakistan (15.8%) having the highest poverty rates [3]. In New York City (NYC), compared to the 13% poverty rate among the non-Hispanic White population, poverty rates were reported to be much higher for the Bangladeshi (32%) and Pakistani (29%) populations [4]. A significant proportion of South Asian individuals have limited English proficiency (LEP) and are engaged in low-wage jobs (eg, taxi drivers, cashiers, and restaurant workers) [3].

Type 2 diabetes (T2D) disproportionately affects many racial/ethnic minorities in the United States, including the South Asian population [5]. The MASALA study, conducted between 2010 and 2013, found that the age-adjusted prevalence of T2D was 23% among individuals of South Asian ethnicity, compared to a prevalence of 9.4%, 13.3%, and 10.3% in the White, African American, and Hispanic American populations, respectively [5,6]. Similar patterns were observed in another population-based study, which reported that T2D prevalence was 17% in US Asian Indian, 15% in Native American/Alaskan native, 13% in non-Hispanic Black, 10% in Hispanic, and 8% in non-Hispanic White populations [7]. Given the high economic and societal costs of T2D, there is an urgent need to develop interventions to prevent and manage the T2D burden in the South Asian population.

Lifestyle counseling programs such as the National Diabetes Prevention Program or Diabetes Self-Management Education and Support program are evidence-based interventions that provide important counseling and support to individuals to manage or prevent T2D [8,9]. However, access to these evidence-based programs has been significantly limited among underserved immigrant communities [10]. First, these programs are often offered during in-person visits, which can be a significant barrier for many low-income South Asian immigrants who do not have a driver’s license or do not know how to navigate the public transportation system [11,12]. Second, most of these existing programs lack cultural and linguistic tailoring of the content [13]. A significant percentage of South Asian immigrants reported LEP and had difficulty understanding their health care providers during office visits [11,13]. In addition, the lack of cultural tailoring can make it challenging for this group of immigrants to initiate dietary and behavioral changes. Lastly, a lack of insurance or other financial barriers can further prevent South Asian immigrants from benefitting from these programs [14].

Given such barriers, mobile health (mHealth) has been rapidly growing over the past few decades and could be a promising approach to increase access to these evidence-based programs among South Asian immigrants [15,16]. Several studies suggest that text message–based interventions among minoritized populations can improve glycemic control for patients with T2D [17-19]. Using text messaging and social media platforms may be promising for immigrant populations, owing to their social needs to stay connected with friends and families in their home countries [20-22]. One mHealth study conducted in India found that participants receiving motivational messages to reduce diabetes risk behaviors had a greater improvement in health behaviors when compared to that of control participants [23].

The inclusion of culturally tailored mHealth interventions among South Asian individuals living in the United States has been largely limited [24,25]. While several research teams have reported using telephones to follow up or conduct motivational interviews with participants [26-28], the use of smartphone or remote technologies remains understudied among South Asian immigrants and there is little data characterizing the use and access of mHealth tools in this population. In a post–COVID-19 world, such information is urgently needed, given the increasing interest in telehealth. To address this knowledge gap, this study leveraged baseline data from the intervention groups of two parent clinical trials whose primary aims were to examine the effectiveness of a community-clinical linkage intervention on diabetes prevention and control in a primarily low-income South Asian immigrant population.

For this study, we examined technology access, current use, and factors associated with preferences for health
communication among South Asian immigrants in NYC. Given that South Asian women bear a higher T2D burden [29] than their male counterparts and face significant cultural barriers to health care–seeking behaviors [30], we also examined whether there were sex differences in technology access, current use, and preferences for health communication. Finally, we examined factors associated with interest in receiving diabetes information by (1) text message, (2) online (videos, voice notes, online forums), and (3) none or skipped, adjusting for sociodemographics and technology access. This information will serve as a critical first step to the development of tailored mHealth interventions for the South Asian immigrant population in the United States.

**Methods**

**Study Design**

We used data collected in two ongoing clinical trials focused on diabetes prevention and management among South Asian immigrants living in NYC (ClinicalTrials.gov NCT03333044 and NCT03188094) [31,32]. Both trials aimed to examine the effectiveness of an integrated community-clinical linkage intervention on health outcomes among low-income South Asian immigrants. For the diabetes prevention trial, the primary outcome was the proportion of participants achieving 5% weight loss at the 6-month follow-up. For the diabetes management trial, the primary outcome was the proportion of participants achieving a hemoglobin A1c (HbA1c) value of 7% or less at the 6-month follow-up. For both trials, primary data outcomes included measurements taken from electronic health records and survey data were only collected from intervention group participants; control group participants were not contacted. We conducted a cross-sectional analysis of deidentified data collected during the baseline period of the intervention for the participants from these two parent studies.

**Sample Recruitment**

Seven bilingual community health workers (CHWs) led the recruitment effort of the parent clinical trials. Participants were recruited from 18 primary care practice sites in NYC serving primarily South Asian patients (>70%). Registry reports were generated at each site to identify potentially eligible patients. Following the randomization of potential participants on the registry list, an introduction letter was mailed to all patients randomized to the treatment group. After 1-2 weeks, a CHW followed up with a phone call to provide further information.

To be eligible for the parent study, participants had to: (1) self-identify as South Asian; (2) be between 21-75 years old; (3) have an appointment with a physician for routine nonemergency primary care in the last 12 months; (4) have a diagnosis of diabetes for at least 12 months (for diabetes study participants); (5) have an HbA1c of at least 7% in the last 12 months (for diabetes study participants); and (6) have a BMI of ≥23 kg/m², as the threshold for overweight for Asian individuals [33] (for prediabetes study participants).

The study team first finalized all materials (flyers, consent forms, surveys) in English. One of the bilingual CHWs performed the translation and then conducted back-translation to account for discrepancies. Translated documents were then reviewed by separate bilingual CHWs to ensure accuracy and use of lay language. If there was any discrepancy, the group of CHWs met with the administrative team to reach an agreement. All study materials were available in English, Bengali, Urdu, and Punjabi. Surveys were administered by bilingual CHWs in the participant’s preferred language via face-to-face interviews or over the phone between 2019 and 2021.

**Ethical Considerations**

The parent studies (S17-00693 and S17-01479) were approved by the New York University Grossman School of Medicine Institutional Review Board (IRB). All participants provided informed consent. The parent IRB approvals cover the secondary analysis of deidentified data without additional consent from participants. In the parent studies, participants received a US $20 gift card for completion of the 6-month follow-up survey. The data used in the analyses were deidentified.

**Outcome Variables**

The outcome of interest was the participants’ preference in receiving diabetes information through the following modalities: (1) text message, (2) online (videos, voice notes, online forum), or (3) none/skipped question (suggesting no interest).

**Predictor Variables**

Sociodemographic information included age (continuous), sex, education level (high school or below vs above high school), marital status (married vs not married), employment status (employed, not employed, retired), LEP (speaking English less than very well vs well, not well, or not at all), and number of years lived in the United States (continuous).

Questions were adapted from the National Cancer Institute Health Information National Trends Survey [34] to assess technology access, current use of technology for health management, current social media use, and interest in mHealth interventions. The questions were reviewed by bilingual CHWs and our community advisory board for cultural relevance. Participants were asked if they owned a basic mobile phone, a smartphone, a tablet, or none of these (all options that apply could be selected). Participants owning a smartphone or tablet were asked whether any health-related apps were installed on the device (“yes” vs “no” or “don’t know”). All participants were asked if Wi-Fi was installed at home (“yes” vs “no”) and if they had used an electronic device (eg, Fitbit, blood glucose meter, or blood pressure monitor) to track their health in the past 12 months. Participants were asked if, over the past 30 days, they had used: (1) Viber, (2) WhatsApp, (3) basic text messaging through a phone carrier, (4) iMessage, (5) Facebook Messenger, and (6) IMO (a messaging app that is popular in South Asian countries).

**Statistical Analyses**

Descriptive statistics were used to examine the distribution of sociodemographic variables, technology access, current use, and interest in receiving diabetes information in the future in the overall sample and stratified by sex. Means and SDs are reported for continuous variables, whereas frequencies and

https://formative.jmir.org/2024/1/e52687
percentages are reported for categorical variables. Differences by sex were assessed using *t* tests and $\chi^2$ tests.

Bivariate analyses were run to examine each outcome and all potential predictor variables. The $\chi^2$ test was used for comparisons of categorical variables and ANOVA was used for comparisons of continuous variables. All variables found to be significant at $P<.01$ (with any of the three outcomes) were included in the initial regression models, along with variables that were not statistically significant in bivariate analyses but with theoretical significance. A total of three multivariable logistic regression models were used to examine factors associated with interest in receiving diabetes information via different mHealth modalities (dependent variable), including (1) by text message, (2) via combined online format (videos, voice notes, or online forum), and (3) not interested in receiving any diabetes information or skipped the question. In these multivariable logistic regression models, independent variables included age, sex, education, marital status, employment status, years in the United States, English proficiency, access to Wi-Fi at home, and access to a smart device. Adjusted odds ratios and 95% CIs were calculated. All analyses were performed with SPSS version 28.

## Results

### Participant Characteristics

The sample included 816 participants, with 413 (50.6%) in the prediabetes study and 403 (49.4%) in the diabetes study. The mean age of the participants was 51.8 (SD 11.0, range 22-75) years; 92.6% (756/816) of the participants were married and 89.6% (731/816) reported LEP. More than half the sample identified as female (462/816, 56.6%), reported a high school education or less (476/816, 58.3%), and were unemployed or retired (431/816, 52.8%). The majority of individuals were Bangladeshi, followed by Indian, Pakistani, Indo-Caribbean, and Nepali. All were born outside of the United States and the mean time lived in the United States was 12.9 (SD 8.7) years (Table 1).

There were sex differences in most sociodemographic characteristics (Table 1). Compared to male participants, female participants were significantly more likely to be younger, living in the United States for less time, have a high school education or less, be unemployed or retired (334/462, 72.3%; 97/354, 27.4%), and report LEP (Table 1).
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total sample (N=816)</th>
<th>Female (n=462)</th>
<th>Male (n=354)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>51.8 (11.0)</td>
<td>51.0 (10.2)</td>
<td>52.7 (12.0)</td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>23 (22-75)</td>
<td>52 (25-74)</td>
<td>54 (22-75)</td>
<td></td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Married</td>
<td>756 (92.6)</td>
<td>419 (90.7)</td>
<td>337 (95.2)</td>
<td></td>
</tr>
<tr>
<td>Not married</td>
<td>54 (6.6)</td>
<td>39 (8.4)</td>
<td>15 (4.2)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>6 (0.7)</td>
<td>4 (0.9)</td>
<td>2 (0.6)</td>
<td></td>
</tr>
<tr>
<td>Highest education level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≤High school</td>
<td>476 (58.3)</td>
<td>329 (71.2)</td>
<td>147 (41.5)</td>
<td></td>
</tr>
<tr>
<td>&gt;High school</td>
<td>334 (40.9)</td>
<td>129 (27.9)</td>
<td>205 (57.9)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>6 (0.7)</td>
<td>4 (0.9)</td>
<td>2 (0.6)</td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Currently employed</td>
<td>381 (46.7)</td>
<td>125 (27.1)</td>
<td>256 (72.3)</td>
<td></td>
</tr>
<tr>
<td>Not employed, not working</td>
<td>381 (46.7)</td>
<td>324 (70.1)</td>
<td>57 (16.1)</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>50 (6.1)</td>
<td>10 (2.2)</td>
<td>40 (11.3)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>4 (0.5)</td>
<td>3 (0.6)</td>
<td>1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Bangladeshi</td>
<td>522 (64.0)</td>
<td>283 (61.3)</td>
<td>239 (67.5)</td>
<td></td>
</tr>
<tr>
<td>Indian</td>
<td>119 (14.6)</td>
<td>67 (14.5)</td>
<td>52 (14.7)</td>
<td></td>
</tr>
<tr>
<td>Pakistani</td>
<td>87 (10.7)</td>
<td>67 (14.5)</td>
<td>20 (5.6)</td>
<td></td>
</tr>
<tr>
<td>Nepali</td>
<td>27 (3.3)</td>
<td>8 (1.7)</td>
<td>19 (5.4)</td>
<td></td>
</tr>
<tr>
<td>Indo-Caribbean</td>
<td>61 (7.5)</td>
<td>37 (8.0)</td>
<td>24 (6.8)</td>
<td></td>
</tr>
<tr>
<td>Number of years lived in the United States</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Valid responses, n</td>
<td>799</td>
<td>449</td>
<td>350</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>12.9 (8.7)</td>
<td>11.6 (7.8)</td>
<td>13.9 (9.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Median (range)</td>
<td>11 (0.3-45)</td>
<td>10 (0.3-37)</td>
<td>12 (0.8-45)</td>
<td></td>
</tr>
<tr>
<td>English proficiency, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>Speaks English less than very well (LEPb)</td>
<td>731 (89.6)</td>
<td>421 (91.1)</td>
<td>310 (87.6)</td>
<td></td>
</tr>
<tr>
<td>Speaks English very well</td>
<td>75 (9.2)</td>
<td>33 (7.1)</td>
<td>42 (11.9)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>10 (1.2)</td>
<td>8 (1.7)</td>
<td>2 (0.6)</td>
<td></td>
</tr>
</tbody>
</table>

aP values do not include missing responses.
bLEP: limited English proficiency.

Technology Ownership and Current Use

As shown in Table 2, nearly three-quarters of the participants reported having a smartphone, whereas less than 2% reported having a tablet. Of those with a smart device (smartphone or tablet), 22.1% (136/615) reported having apps related to health or wellness. Approximately one-quarter of the sample used an electronic device or monitor (eg, blood glucose meter, Fitbit band) to track their health in the past 12 months. The majority reported having Wi-Fi at home. The most common social media forms used in the past 30 days included basic text message via a cellular carrier and WhatsApp, followed by Viber and Facebook Messenger. When asked about interest in receiving diabetes information, three-quarters of the participants indicated interest in receiving diabetes-related information via text message, followed by videos, while 17.6% (144/816) of the participants were not interested in receiving diabetes information or skipped the question.
Table 2. Technology access, current social media use, and interest in mobile health interventions, stratified by sex.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total sample (N=816), n (%)</th>
<th>Female (n=462), n (%)</th>
<th>Male (n=354), n (%)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology access</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mobile device ownership</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic mobile phone</td>
<td>144 (17.6)</td>
<td>106 (22.9)</td>
<td>38 (10.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Smartphone</td>
<td>611 (74.9)</td>
<td>317 (68.6)</td>
<td>294 (83.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tablet</td>
<td>15 (1.8)</td>
<td>7 (1.5)</td>
<td>8 (2.3)</td>
<td>.45</td>
</tr>
<tr>
<td>Missing</td>
<td>5 (0.6)</td>
<td>4 (0.9)</td>
<td>1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Has a mobile device (basic mobile phone or smartphone)</td>
<td></td>
<td></td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td>Yes</td>
<td>751 (92.0)</td>
<td>417 (90.3)</td>
<td>334 (94.4)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>60 (7.4)</td>
<td>41 (8.9)</td>
<td>19 (5.4)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>5 (0.6)</td>
<td>4 (0.9)</td>
<td>5 (0.6)</td>
<td></td>
</tr>
<tr>
<td>Has a smart device (smartphone or tablet)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>615 (75.4)</td>
<td>319 (69.0)</td>
<td>296 (83.6)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>196 (24.0)</td>
<td>139 (30.1)</td>
<td>57 (16.1)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>5 (0.6)</td>
<td>4 (0.9)</td>
<td>1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Has apps related to health and wellness on smartphones or tablets (among those with smart devices)</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Yes</td>
<td>136 (22.1)</td>
<td>55 (17.3)</td>
<td>81 (27.3)</td>
<td></td>
</tr>
<tr>
<td>No or don’t know</td>
<td>443 (72.0)</td>
<td>240 (75.2)</td>
<td>203 (68.6)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>36 (5.9)</td>
<td>24 (7.5)</td>
<td>12 (4.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Other than a tablet or smartphone, have you used an electronic device or monitor to track your health within the last 12 months?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Yes</td>
<td>211 (25.9)</td>
<td>127 (27.5)</td>
<td>84 (23.7)</td>
<td></td>
</tr>
<tr>
<td>No or don’t know</td>
<td>531 (65.1)</td>
<td>283 (61.3)</td>
<td>248 (70.1)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>74 (9.1)</td>
<td>52 (11.3)</td>
<td>22 (6.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Current social media use</strong></td>
<td></td>
<td></td>
<td></td>
<td>.33</td>
</tr>
<tr>
<td>Has Wi-Fi installed at home</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>647 (79.3)</td>
<td>362 (78.4)</td>
<td>285 (80.5)</td>
<td></td>
</tr>
<tr>
<td>No or don’t know</td>
<td>98 (12.0)</td>
<td>62 (13.4)</td>
<td>36 (10.2)</td>
<td></td>
</tr>
<tr>
<td>Skipped</td>
<td>71 (8.7)</td>
<td>38 (8.2)</td>
<td>33 (9.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Types of social media used in the past 30 days</strong></td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Viber</td>
<td>213 (26.1)</td>
<td>102 (22.1)</td>
<td>111 (31.4)</td>
<td></td>
</tr>
<tr>
<td>Basic text messages via cellular carrier</td>
<td>561 (68.8)</td>
<td>299 (64.7)</td>
<td>262 (74.0)</td>
<td>.01</td>
</tr>
<tr>
<td>iMessage</td>
<td>114 (14.0)</td>
<td>64 (13.9)</td>
<td>50 (14.1)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>WhatsApp</td>
<td>534 (65.4)</td>
<td>279 (60.4)</td>
<td>255 (72.0)</td>
<td>.001</td>
</tr>
<tr>
<td>Facebook Messenger</td>
<td>222 (27.2)</td>
<td>72 (15.6)</td>
<td>150 (42.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>IMO</td>
<td>71 (8.7)</td>
<td>31 (6.7)</td>
<td>40 (11.3)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Interest in mobile health interventionsb</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text message (eg, through phone carrier or WhatsApp)</td>
<td>609 (74.6)</td>
<td>323 (69.9)</td>
<td>286 (80.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Videos (eg, through WhatsApp or Viber)</td>
<td>225 (27.6)</td>
<td>88 (19.0)</td>
<td>137 (38.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Voice notes (eg, through WhatsApp)</td>
<td>45 (5.5)</td>
<td>22 (4.8)</td>
<td>23 (6.5)</td>
<td>.28</td>
</tr>
<tr>
<td>Online forum or online support group (eg, on Facebook)</td>
<td>24 (2.9)</td>
<td>10 (2.2)</td>
<td>14 (4.0)</td>
<td>.15</td>
</tr>
</tbody>
</table>
Similar to the sample characteristics, we also noted differences by sex (Table 2). Compared to male participants, female participants were significantly less likely to own a smartphone; have health-related apps on their smart devices; use Viber, basic text messages, WhatsApp, IMO, and Facebook Messenger among those with a smartphone or tablet; or report interest in receiving diabetes intervention via text message or videos. Conversely, female participants were significantly more likely to have a basic mobile phone and to report no interest or skipped the question about receiving diabetes intervention information compared to male participants.

**Factors Associated With Interest in Receiving Diabetes Intervention**

Table 3 provides a summary of the regression analysis to identify factors associated with an interest in receiving diabetes information. The significant factors associated with receiving information via text messaging included male sex, current unemployment, above high school education, and owning a smartphone or tablet.

When combining responses on videos, voice notes, and online forums/online support groups, participants who identified as male and who had a smartphone or tablet were more likely to report an interest in receiving diabetes information through these online formats.

Factors associated with no interest in receiving diabetes information or skipping the question included female sex, not married, currently employed, high school education or below, and not owning a smart device.
Table 3. Factors associated with interest in receiving diabetes information by modality based on multivariable regression.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Text message AOR(^a) (95% CI)</th>
<th>P value</th>
<th>Online (videos, voice notes, online forum) AOR (95% CI)</th>
<th>P value</th>
<th>None/skipped AOR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Reference</td>
<td>N/A(^b)</td>
<td>Reference</td>
<td>N/A</td>
<td>2.66 (1.55-4.56)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>1.63 (1.04-2.55)(^c)</td>
<td>.04</td>
<td>2.38 (1.59-3.57)</td>
<td>&lt;.001</td>
<td>Reference</td>
<td>N/A</td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.00 (0.98-1.02)</td>
<td>.66</td>
<td>1.00 (0.98-1.02)</td>
<td>.86</td>
<td>1.00 (0.97-1.02)</td>
<td>.65</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1.92 (0.99-3.72)</td>
<td>.05</td>
<td>1.99 (0.91-4.37)</td>
<td>.09</td>
<td>Reference</td>
<td>N/A</td>
</tr>
<tr>
<td>Not married</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>2.26 (1.13-4.52)</td>
<td>.02</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤High school</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>2.02 (1.22-3.36)</td>
<td>.01</td>
</tr>
<tr>
<td>&gt;High school</td>
<td>2.17 (1.41-3.32)</td>
<td>&lt;.001</td>
<td>1.15 (0.80-1.65)</td>
<td>.47</td>
<td>Reference</td>
<td>N/A</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>1.96 (1.18-3.29)</td>
<td>.01</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1.62 (1.03-2.53)</td>
<td>.04</td>
<td>1.02 (0.68-1.52)</td>
<td>.94</td>
<td>Reference</td>
<td>N/A</td>
</tr>
<tr>
<td>Years in the United States</td>
<td>0.99 (0.97-1.01)</td>
<td>.20</td>
<td>1.00 (0.98-1.02)</td>
<td>.99</td>
<td>1.00 (0.98-1.03)</td>
<td>.90</td>
</tr>
<tr>
<td>English fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;Very good</td>
<td>0.70 (0.36-1.34)</td>
<td>.28</td>
<td>1.20 (0.66-2.18)</td>
<td>.55</td>
<td>1.50 (0.71-3.19)</td>
<td>.29</td>
</tr>
<tr>
<td>Very good</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
</tr>
<tr>
<td>Has a smartphone or tablet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.35 (2.17-5.18)</td>
<td>&lt;.001</td>
<td>5.19 (2.83-9.51)</td>
<td>&lt;.001</td>
<td>Reference</td>
<td>N/A</td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>3.35 (2.06-5.44)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wi-Fi at home</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.83 (0.48-1.43)</td>
<td>.50</td>
<td>1.85 (0.97-3.55)</td>
<td>.06</td>
<td>1.59 (0.83-3.05)</td>
<td>.16</td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\)AOR: adjusted odds ratio.

\(^b\)N/A: not applicable.

\(^c\)Significant relationships are italicized.

Discussion

Principal Findings

This study examined technology access, current use, and factors associated with preferences for health communication in a large sample of predominantly low-income South Asian immigrants with prediabetes or T2D. We found a moderately high level of technology ownership and social media usage, and that participants who were of the female sex, currently employed, not married, had high school education or below, and no smart device were less likely to report interest in mHealth programs.

Our data revealed that approximately three-quarters of South Asian immigrants in NYC had a smartphone and over 60% used social media or text messaging apps. These results revealed some gaps in access to smart devices between South Asian immigrants and the general US population. According to national mobile device ownership data compiled by the Pew Research Center, 85% of the overall US general population had a smartphone, including 85% of non-Hispanic White adults, 85% of non-Hispanic Black adults, and 83% of Hispanic adults [35]. While over 50% of the US general population owns a tablet, less than 2% of the South Asian immigrants in our sample reported owning a tablet. These data suggest that underserved low-income South Asian immigrants may need additional resources and infrastructure support to benefit from telemedicine or telehealth programs in the post–COVID-19 era.

Compared to male participants, female participants were less likely to own a smartphone, have health-related apps on the smartphone, use social media apps, or report interest in text messages– or video-based diabetes information. These differences may be explained by the different sociodemographic characteristics of female and male participants in our study. Compared to male participants, female participants were more likely to have a lower level of education and to be unemployed, which have been reported to be significant predictors of technology engagement and interest in mHealth-related interventions [36-39]. In addition, the gender norms in South
Asian culture may also contribute to these disparities. Past studies have noted patriarchal cultural norms in South Asian families, contributing to differences in health care–seeking and behaviors between men and women [30]. Future research studies may need to explore culturally tailored strategies that may better engage South Asian women in health and mHealth interventions [40].

Our findings also suggested that participants with less than a high school education, currently employed, and not having a smart device were more likely to report no interest in mHealth interventions or skip the question. These factors were consistent with several prior reports, which found that education, employment status, sex, and access to technologies significantly affect people’s current and future interest in mHealth interventions [36,38,40], as well as a study among South Asians in Canada [41]. Those with limited education and access to technology may have less exposure and experience with technologies and have limited digital literacy, and thus show no or little interest in technology-based interventions [38,39,42]. Future studies need to consider how to reach this group and design an appropriate intervention that meets their needs, which can help to ensure equity in the implementation of diabetes interventions and ultimately equity in health disparities driven by social factors.

Strengths and Limitations

There are several limitations of this study to note. We used available data from two clinical trials that aimed to test the effectiveness of an integrated intervention on diabetes prevention and management in South Asian immigrants in the United States [31,32]. To minimize participant burden, we did not include an extensive list of questions about digital literacy, prior experience with various technologies, and current use of various apps or health websites. In addition, some of the data were collected during the COVID-19 pandemic. Due to remote work and learning, some participants may have purchased a smartphone or Wi-Fi or used telemedicine services due to COVID-19 [43,44]. This may lead to higher rates of smart device ownership or interest in mHealth interventions. Finally, the study sample consisted of predominately low-income South Asian immigrants, largely those from Bangladesh, who participated as volunteers in our clinical trials; therefore, our findings related to technology use may not be generalizable to the broader South Asian community in NYC.

We also want to highlight several strengths of the study. This is a relatively large sample size with a sizeable number of male participants for a health intervention. All participants were recruited from clinical settings, which greatly enhances the representation of our sample. In addition, bilingual and bicultural CHWs took the lead in recruitment, which makes the inclusion of non-English–speaking participants possible. These populations were often overlooked or not included in clinical trials that focus on English-speaking populations [45,46].

Practice Implications

Compared to those of the general US population, the rates of technology access and adoption are relatively lower among underserved low-income South Asian immigrants, particularly among low-income South Asian women. When working with this community, it may be helpful to consider basic and simple technologies such as telephone calls and basic text messages. Although it is possible to engage this community in telemedicine or telehealth programs, this requires hands-on technology training and community engagement strategies. Due to COVID-19, our group transformed several in-person diabetes intervention programs into remote formats (eg, Zoom-based group sessions). Our preliminary data demonstrated great engagement and retention [15]. Our data also revealed that South Asian women were less likely to have access to technology or interest in mHealth interventions. Future work needs to explore barriers and develop tailored strategies to engage this unique group.

Conclusions

We found that technology ownership, access, and social media usage were moderately high in primarily low-income South Asian immigrants living in NYC with a diagnosis of prediabetes or diabetes. Significantly higher usage and technology access was seen among male participants than among female participants. Most participants reported interest in receiving diabetes-related information via text messaging or videos. However, disparities existed within our study sample. Female participants who had limited education and no smart device were more likely to report no interest in mHealth interventions. Future studies may consider strategies to engage South Asian women and leverage text messaging and social media apps as a potential strategy to disseminate culturally and linguistically adapted diabetes information in this underserved immigrant population.

Acknowledgments

We are grateful to all study participants for their time and participation in the parent studies. Parts of this study were presented at the 43rd Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine, Baltimore, MD, April 6-9, 2022. NSI’s time is supported in part by the National Institutes of Health (NIH) Grant R01DK11048; the NIH National Institute on Minority Health and Health Disparities (NIMHD) Grants U54MD000538 and R01MD018528; the Centers for Disease Control and Prevention Grant U48DP001904; the National Institute of Diabetes and Digestive Kidney Diseases Grants R18DK110740 and 1U2CDK137135; the National Heart Lung and Blood Institute Grant 1UG3HL15310; and the NIH National Center for the Advancement of Translational Science Grant UL1TR0001445. LH’s time is partially supported by the NIMHD Grants R01MD017579-01A1, R18HS029813, and R00MD012811. The views of this study are those of the authors and do not necessarily represent the views of the NIMHD, NIH, and US Department of Health and Human Services.

https://formative.jmir.org/2024/1/e52687
Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

Conceptualization was completed by LH, LCW, SL, and NSI. Data management, analysis, and visualization were completed by LCW. Writing of the original draft was completed by LH and FM. Writing and editing were completed by LH, LCW, FM, SL, JZ, SM, SH, SHA, DO, HMB, AA, and NSI. Funding acquisition was completed by NSI.

Conflicts of Interest

None declared.

References


Abbreviations

- AOR: adjusted odds ratio
- CHW: community health worker
- HbA1c: hemoglobin A1c
- IRB: institutional review board
- LEP: limited English proficiency
- mHealth: mobile health
- NYC: New York City
- T2D: type 2 diabetes

Edited by A Mavragani; submitted 12.09.23; peer-reviewed by C Nguyen-Truong, M Makowsky; comments to author 27.02.24; revised version received 12.03.24; accepted 14.03.24; published 26.04.24.

Please cite as:
URL: https://formative.jmir.org/2024/1/e52687
doi:10.2196/52687
PMID:38669062

©Lu Hu, Laura C Wyatt, Farhan Mohsin, Sahnah Lim, Jennifer Zanowiak, Shinu Mammen, Sarah Hussain, Shahmir H Ali, Deborah Onakomaiya, Hayley M Belli, Angela Aifah, Nadia S Islam. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Acceptance, Satisfaction, and Preference With Telemedicine During the COVID-19 Pandemic in 2021-2022: Survey Among Patients With Chronic Pain

Michael Alexander Harnik¹,², MD; Alina Scheidegger³, PhD; Larissa Blättler¹, MSc; Zdenek Nemecek¹, MD; Thomas C Sauter⁴,⁵, MME, MD; Andreas Limacher⁶, PhD; Florian Reisig¹, MD; Martin grosse Holtforth³,⁷, PhD; Konrad Streitberger¹, MD

¹Department of Anaesthesiology and Pain Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
²University Hospital Würzburg, Department of Anaesthesiology, Intensive Care, Emergency and Pain Medicine, Centre for Interdisciplinary Pain Medicine, Würzburg, Germany
³Psychosomatic Medicine, Department of Neurology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
⁴Department of Emergency Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland
⁵Emergency Telemedicine, University of Bern, Bern, Switzerland
⁶Department of Clinical Research, University of Bern, Bern, Switzerland
⁷Clinical Psychology and Psychotherapy, Department of Psychology, University of Bern, Bern, Switzerland

Corresponding Author:
Michael Alexander Harnik, MD
Department of Anaesthesiology and Pain Medicine
Inselspital, Bern University Hospital
University of Bern
Freiburgstrasse 18
Bern, 3010
Switzerland
Phone: 41 31 632 39 65
Email: michael.harnik@insel.ch

Abstract

Background: The COVID-19 pandemic has forced many health care providers to make changes in their treatment, with telemedicine being expanded on a large scale. An earlier study investigated the acceptance of telephone calls but did not record satisfaction with treatment or patients’ preferences. This warranted a follow-up study to investigate acceptance, satisfaction, and preferences regarding telemedicine, comprising of phone consultations, among health care recipients.

Objective: The primary aim was to assess the acceptance and satisfaction of telemedicine during the subsequent months of 2021-2022, after the initial wave of the COVID-19 pandemic in Switzerland. Furthermore, we aimed to assess patients’ preferences and whether these differed in patients who had already experienced telemedicine in the past, as well as correlations between acceptance and satisfaction, pain intensity, general condition, perception of telemedicine, and catastrophizing. Finally, we aimed to investigate whether more governmental restrictions were correlated with higher acceptance.

Methods: An anonymous cross-sectional web-based survey was conducted between January 27, 2021, and February 4, 2022, enrolling patients undergoing outpatient pain therapy in a tertiary university clinic. We conducted a descriptive analysis of acceptance and satisfaction with telemedicine and investigated patients’ preferences. Further, we conducted a descriptive and correlational analysis of the COVID-19 stringency index. Spearman correlation analysis and a chi-square test for categorical data were used with Cramer V statistic to assess effect sizes.

Results: Our survey was completed by 60 patients. Telemedicine acceptance and satisfaction were high, with an average score of 7.6 (SD 3.3; on an 11-point Numeric Rating Scale from 0=not at all to 10=completely), and 8.8 (SD 1.8), respectively. Respondents generally preferred on-site consultations to telemedicine (n=35, 58% vs n=24, 40%). A subgroup analysis revealed that respondents who already had received phone consultation, showed a higher preference for telemedicine (n/N=21/42, 50% vs n/N=3/18, 17%; χ² [N=60]=7.5, P=.02, Cramer V=0.354), as well as those who had been treated for more than 3 months (n/N=17/31, 55% vs n/N=7/29, 24%; χ² [N=60]=6.5, P=.04, Cramer V=0.329). Acceptance of telemedicine showed a moderate...
positive correlation with satisfaction ($r_s=0.41, P<.05$), but there were no correlations between the COVID-19 stringency index and the other variables.

**Conclusions:** Despite high acceptance of and satisfaction with telemedicine, patients preferred on-site consultations. Preference for telemedicine was markedly higher in patients who had already received phone consultations or had been treated for longer than 3 months. This highlights the need to convey knowledge of eHealth services to patients and the value of building meaningful relationships with patients at the beginning of treatment. During the COVID-19 pandemic, the modality of patient care should be discussed individually.

(JMIR Form Res 2024;8:e53154) doi:10.2196/53154

**KEYWORDS**

acceptance; satisfaction; patient preferences; COVID-19 pandemic; health care providers; phone consultations; pain therapy; eHealth services; patient care; health care delivery; telemedicine; chronic pain; preference

**Introduction**

Throughout the COVID-19 pandemic, there have been various measures to curb the spread of the virus, including restrictions on nonurgent care. In Switzerland, this resulted in the temporary suspension of most outpatient visits and the switch to telemedicine, that is, mostly phone consultations. This trend has been encouraged by various interest groups and associations advocating the rapid transition to remote services [1], especially because it can be argued that the COVID-19 pandemic not only exacerbates already persistent pain conditions but also increases the incidence of new-onset pain [2]. While some studies have examined the efficacy of eHealth for the treatment of several conditions (eg, osteoarthritis [3], musculoskeletal conditions [4,5], and chronic pain [6]), evidence from the patients’ perspective remains scarce, so that patients’ preferences are largely unknown, especially during this new era of the COVID-19 pandemic.

A previous survey conducted with 60 patients at our clinic showed that patients’ acceptance of telemedicine during the first wave of the pandemic was high [7]. However, at the time, satisfaction was not recorded. As this is an important aspect of patient-centered care [8], satisfaction has become part of the core outcome domains for chronic pain clinical trials [9] and serves as an anchor for clinically important differences in pain intensity and other outcomes [10]. Further, high satisfaction is associated with the appropriateness of care [11] and contributes to therapy adherence [12]. Likewise, preferences were not assessed in our first investigation. However, as fulfillment of patients’ preferences is linked to better adherence to therapy and better allocation of resources [13], it is important to know what patients actually favor.

Furthermore, psychological factors have a big impact on pain-related outcomes: generalized anxiety disorders can be linked to poor global well-being and life satisfaction [14] and hypochondriacal attitudes and beliefs have been shown to predict patients’ satisfaction [15]; therefore, these factors are important to take into account.

Finally, various studies provide evidence of the detrimental effect of physical distancing during the pandemic on mental health [16-19]. It is conceivable that governments’ restrictions influence patients’ acceptance and satisfaction with telemedicine.

This warranted a follow-up study to investigate (1) the levels of acceptance and satisfaction in the subsequent months, whether levels of acceptance changed after the first wave of the COVID-19 pandemic and what levels of satisfaction were achieved with telemedicine; (2) patients’ preferences for care during the pandemic and whether preferences differ in patients who had already experienced telemedicine; (3) correlations between acceptance and satisfaction, current pain intensity, general condition, as well as several measures of pain-related self-reports (anxiety, hypochondriacal worries, and catastrophizing) to assess whether these are associated with acceptance itself and patients’ preferences; and finally (4) correlation of acceptance with the severity of restrictive measures ordered by the government, to investigate whether more restrictions relate to higher acceptance.

**Methods**

**Sample**

We conducted an anonymous, voluntary survey between January 27, 2021, and February 4, 2022, at the Pain Center of the Bern University Hospital, Switzerland. Patients who were seen in person in our clinic received an invitation to complete the digital survey. Inclusion criteria were aged at least 18 years, with chronic pain, all conditions included. Patients with an acute illness requiring emergency procedures [20] were excluded.

**Procedure and Study Design**

A cross-sectional web-based survey was created to assess patients’ satisfaction with telemedicine consisting of phone calls (for the full questionnaire see Table S1 in Multimedia Appendix 1). In contrast to our previous study [7], the current investigation was conducted after the shutdown and with patients who were seen in person during routine outpatient treatment at our Pain Center. Whereas our previous study only accounted for first contacts with patients over the phone, the follow-up survey also involved patients who had already received a phone consultation. These respondents additionally rated their satisfaction as well as their perceptions of various aspects of telemedicine. A questionnaire was placed at the reception of our outpatient pain service, and newly referred patients were encouraged to complete it by scanning a QR code. We conducted this study solely by relying on departmental resources, without external funding.
Ethical Considerations

All the included patients provided informed consent for the reuse of their data. A jurisdictional inquiry was submitted to the Ethics Committee of the Canton of Bern, Switzerland (BASEC Req-2020-01406). The Ethics Committee decided that this study had no jurisdiction and was therefore exempted from ethics approval. Due to the anonymous assessment via QR code, it was impossible to identify individual patients. The participants received no financial compensation.

Materials

First, participants provided demographic data (sex and age) and reported the intensity and duration of their pain, their general condition, and previous interventional pain treatments.

Subsequently, the perceived acceptance of telemedicine and other aspects of patients’ expectations concerning the COVID-19 pandemic were assessed on a Numeric Rating Scale (0=not at all and 10=completely), that is, regarding their confidence in dealing with pain and with the pandemic, whether they feared a severe COVID-19 infection, and whether they thought the health care system and politics had taken the correct steps to cope with the consequences of the pandemic.

The validated German version of the generalized anxiety disorder 2-item [21,22] was used to assess the level of general anxiety with 2 items (eg, “nervousness, anxiety, or tension”) on a Likert scale from 0 to 3, ranging from “not at all” (0) to “nearly every day” (3), leading to a total score of 0 to 6, with a cutoff value of 3 for a generalized anxiety disorder.

To measure pain catastrophizing, the German short form of the Pain Catastrophizing Scale (PCS-6) was used [23]. The PCS-6 consists of 6 items (eg, “I keep thinking about how much it hurts”) that are rated on 5-point Likert scales, ranging from “not at all” (0) to “all the time” (4), leading to a total score of 0-24.

The Whiteley Index [24] was used to assess hypochondriacal worries and beliefs using 6 items (eg, “Do you often worry about the possibility that you have got a serious illness?”), rated on a 5-point Likert scale, ranging from “not at all” (1) to “very much” (5), leading to a total score of 6-30 [24,25].

Respondents were then asked to choose their preferred type of consultation during the COVID-19 pandemic: in person, by telephone, and/or by telemedicine. To check that our sample was representative, we compared the demographic data of our participants (sex, age, and pain intensity) with those of all patients who had been referred to the clinic during the study period. To analyze correlations with satisfaction, pain levels, general conditions, and other self-reported measures, as well as the severity of restrictive measures, a Spearman correlation was computed, with correlations of 0.2-0.4 considered weak, 0.4-0.6 moderate, and 0.6-0.8 strong [27].

Regarding preferences, we first analyzed them descriptively and then assessed whether they differed in participants who had previously received phone consultations, as well as whether there were differences between patients who had been treated for longer than three months at our clinic and newly treated participants. Using a chi-square test for categorical data, these groups were further evaluated with Cramer V statistic to assess effect sizes. A magnitude of 0.1 was considered small, 0.3 medium, and 0.5 large [28]. For continuous data, a 2-sided Student t test with equal variance or Welch t test in the case of unequal variance as indicated by the Levene test at P<.05 was calculated.

Descriptive analyses were performed with jamovi (version 2.2; jamovi project) [29]. Visualizations were performed using GraphPad Prism (version 8.0.0 for Windows, GraphPad Software), and R Studio Team (Posit). Correlations and corresponding graphs were computed in R Studio: Integrated Development for R (RStudio, PBC). Statistical significance was set at a P value of <.05.

Results

Demographic Data

The survey was accessible between January 27, 2021, and February 4, 2022. During this time, a total of 816 new patients were seen at our center. In total, 77 (9.4%) patients logged in, 60 (77.9%) patients of whom completed the questionnaire. Most participants were female (n=39, 65%) with an average age of 50.8 (SD 18.3) years (Table 1). About half of them (n=29, 48%) had been newly treated at the Pain Center, and the majority had already received phone consultations as part of their routine treatment (n=42, 70%). More than half of the respondents had previously received interventional pain therapy (n=31, 52%); however most procedures were either unsuccessful or had only provided short-term pain relief (n=24, 77%).

The mean pain intensity score was 6.3 (SD 2.3) and the general condition was reported to be 6.5 (SD 2.2; Table 2). A comparison with all patients who had been referred to the clinic (with acute and chronic conditions) during this study period revealed no statistically significant differences in terms of sex...
or age, indicating that our sample may be considered representative of all patients seen during this time (Table S2 in Multimedia Appendix 1).

Table 1. Baseline characteristics of all enrolled patients in 2021-2022.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean (SD)</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>50.8 (18.3)</td>
<td>N/A²</td>
</tr>
<tr>
<td>Sex, female</td>
<td>N/A</td>
<td>39 (65)</td>
</tr>
<tr>
<td>Pain duration (years)</td>
<td>2.7 (0.6)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Pain duration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below 3 months</td>
<td>N/A</td>
<td>3 (5)</td>
</tr>
<tr>
<td>3 m to 1 years</td>
<td>N/A</td>
<td>13 (22)</td>
</tr>
<tr>
<td>Over 1 years</td>
<td>N/A</td>
<td>44 (73)</td>
</tr>
<tr>
<td><strong>Interventional pain therapy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>25 (42)</td>
</tr>
<tr>
<td>Yes, without success</td>
<td>N/A</td>
<td>13 (22)</td>
</tr>
<tr>
<td>Yes, successful over the short term</td>
<td>N/A</td>
<td>11 (18)</td>
</tr>
<tr>
<td>Yes, successful over the long term</td>
<td>N/A</td>
<td>7 (12)</td>
</tr>
<tr>
<td>Not specified</td>
<td>N/A</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Phone consultation received (yes)</td>
<td>N/A</td>
<td>42 (70)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newly treated</td>
<td>N/A</td>
<td>29 (48)</td>
</tr>
<tr>
<td>Longer than 3 months</td>
<td>N/A</td>
<td>31 (52)</td>
</tr>
</tbody>
</table>

²N/A: not applicable.
Table 2. Descriptive analysis of patients’ responses.

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain&lt;sup&gt;a&lt;/sup&gt;, mean (SD)</td>
<td>6.3 (2.3)</td>
</tr>
<tr>
<td>General condition, mean (SD)</td>
<td>6.5 (2.2)</td>
</tr>
<tr>
<td>Accept&lt;sup&gt;b&lt;/sup&gt;, mean (SD)</td>
<td>7.6 (3.3)</td>
</tr>
<tr>
<td>PhonCov&lt;sup&gt;c&lt;/sup&gt;, mean (SD)</td>
<td>7.7 (3.2)</td>
</tr>
<tr>
<td>Satis&lt;sup&gt;d&lt;/sup&gt;,&lt;sup&gt;e&lt;/sup&gt;, mean (SD)</td>
<td>8.8 (1.8)</td>
</tr>
<tr>
<td>PercSinc&lt;sup&gt;d&lt;/sup&gt;,&lt;sup&gt;f&lt;/sup&gt;, mean (SD)</td>
<td>9.1 (1.6)</td>
</tr>
<tr>
<td>Quest&lt;sup&gt;g&lt;/sup&gt;, mean (SD)</td>
<td>9.2 (1.3)</td>
</tr>
<tr>
<td>HelpCons&lt;sup&gt;d&lt;/sup&gt;,&lt;sup&gt;h&lt;/sup&gt;, mean (SD)</td>
<td>7.9 (2.7)</td>
</tr>
<tr>
<td>Long-term improvement of pain, mean (SD)</td>
<td>6.8 (2.5)</td>
</tr>
<tr>
<td>Confidence in dealing with pain, mean (SD)</td>
<td>7.5 (2.2)</td>
</tr>
<tr>
<td>Confidence regarding COVID-19 pandemic, mean (SD)</td>
<td>9.0 (1.8)</td>
</tr>
<tr>
<td>Correct medical steps, mean (SD)</td>
<td>8.7 (1.9)</td>
</tr>
<tr>
<td>Correct political steps, mean (SD)</td>
<td>7.6 (2.4)</td>
</tr>
<tr>
<td>Adequate treatment of pain, mean (SD)</td>
<td>3.3 (3.4)</td>
</tr>
<tr>
<td>Fear of severe coronavirus infection, mean (SD)</td>
<td>4.0 (3.7)</td>
</tr>
</tbody>
</table>

**Preference, n (%)**
- Preferably a consultation on-site: 35 (58)
- Preferably a phone consultation: 24 (40)
- No consultation: 1 (2)

- Anxiety (GAD-2<sup>i</sup>, 0-6 points; n=17, 28%), mean (SD): 2.0 (1.7)
- Pain catastrophizing (PCS-6<sup>j</sup>, 0-24 points; PCS6), mean (SD): 10.3 (6.1)
- Frequent worries (WI-6<sup>k</sup>, 6-30 points; WI6), mean (SD): 7.0 (5.6)
- Strictness of government response (scale 0-100), median (IQR): 51 (48-56)

<sup>a</sup>Pain: Average pain intensity.
<sup>b</sup>Accept: Acceptance of telemedicine.
<sup>c</sup>PhonCov: Phone consultation without COVID-19.
<sup>d</sup>Only calculated if patients had already received phone consultations.
<sup>e</sup>Satis: Satisfaction with telemedicine.
<sup>f</sup>PercSinc: Perception of sincerity.
<sup>g</sup>Quest: Questions were addressed.
<sup>h</sup>HelpCons: Could be helped by phone consultation.
<sup>i</sup>GAD-2: generalized anxiety disorder 2-item.
<sup>j</sup>PCS-6: Pain Catastrophizing Scale 6-item.
<sup>k</sup>WI-6: Whiteley Index 6-item.

**Acceptance and Satisfaction With Telemedicine**

The acceptance of telemedicine varied among patients, with a mean score of 7.6 (SD 3.3). Most reported that they would also have made use of telemedicine without the necessity imposed by the COVID-19 pandemic restrictions (item “phone consultation without COVID-19,” mean 7.7, SD 3.2) and those patients who had received phone consultation reported high levels of satisfaction (mean 8.8, SD 1.8). These respondents felt that their concerns had been taken seriously (mean 9.1, SD 1.6), that their questions had been addressed (mean 9.2, SD 1.3), and that they could be helped (mean 7.9, SD 2.7). The distributions of acceptance, satisfaction, pain, and anxiety are displayed in Figure 1.
Figure 1. Baseline characteristics of satisfaction, acceptance, pain, and anxiety. Violin plots showing the distribution of respondents’ answers (N=60) on an 11-point scale for acceptance, satisfaction, and pain (left y-axis). On the right and separated, anxiety scores are indicated on the 6-point GAD-scale (right y-axis). Continuous line indicates the median and dotted lines the 1st and 3rd quartile. GAD: generalized anxiety disorder.

Affective State and Processing of Pain During the COVID-19 Pandemic

Participants tended to think their pain would improve over the long term (mean 6.8, SD 2.5). They were even more optimistic concerning their ability to cope with pain (mean 7.5, SD 2.2) and with the COVID-19 pandemic (mean 9.0, SD 1.8). There was high variability in responses to the “fear of pain not being addressed adequately in the future” and “fear of a severe COVID-19 infection.”

Screening for anxiety using the generalized anxiety disorder 2-item showed low scores (mean 2.0, SD 1.7; see Table 2 and Figure 1). However, using a cutoff of ≥3 points [22], we found about 1 quarter (n=17, 28%) of our patients had a high likelihood of having generalized anxiety or panic disorder. Catastrophizing scores using the PCS-6 were generally high (mean 10.3, SD 6.1), while the assessment of hypochondriacal worries and beliefs (operationalized with Whiteley Index 6-item) showed rather low values (mean 7.0, SD 5.6).

Patients’ Preferences for Care

In total, 24 participants (n=24, 40%) preferred telemedicine to on-site consultations (see Table 2). Nonetheless, a more detailed analysis revealed substantial differences between subgroups. The predilection for on-site consultation was distinctly higher in patients who had never had a phone consultation (14 of 18 patients, 78%), whereas half of the respondents who had already experienced telemedicine preferred on-site consultation over phone consultations (21 of 42 patients, 50%). A chi-square test showed a significant association with a medium effect size (χ² [N=60]=7.5, P=.02, Cramer V=0.354).

Likewise, participants who had been treated for less than three months tended to prefer on-site consultations (21 of 29 patients, 72%), whereas those treated for longer durations slightly preferred telemedicine over on-site consultations (17 of 31 patients, 55%). Here, the chi-square test was also significant with a medium effect size (χ² [N=60]=6.5, P=.04, Cramer V=0.329).

Correlation Analysis

Acceptance of telemedicine showed a moderate positive correlation with satisfaction (rₛ [58]=0.41, P<.05) and a strong correlation with patients’ intention to use phone consultation outside of the COVID-19 pandemic as well (rₛ [58]=0.65, P<.05). In addition, in patients who had already experienced telemedicine, strong correlations were observed between satisfaction and quality-related items, such as the perception of sincerity and patients’ opportunities to clarify their questions (rₛ [58]=0.85-0.91, P<.05). Pain and psychometric scores showed moderate to high correlations. Figure 2 provides an overview of statistically significant results (P<.05) of the correlation analyses with at least a moderate strength (rₛ ≥0.4). There were no statistically significant correlations below moderate strength.
Correlation of Acceptance With the Severity of Restrictive Measures

The COVID-19 stringency index reached a median of 50.93 (IQR 48.15-56), and measures remained largely the same over the course of this study period. The severity of restrictive measures did not correlate significantly with acceptance of telemedicine or any other investigated items (satisfaction, pain, general condition, catastrophizing, anxiety, hypochondriacal worries and beliefs, confidence regarding the COVID-19 pandemic, fear of severe coronavirus infection, or impressions regarding correct medical or political steps).

Discussion

Principal Findings

This study sheds light on the clinical reality of the COVID-19 pandemic, including how remote services are perceived and what patients actually prefer. We found high acceptance and satisfaction levels of telemedicine. Nevertheless, respondents generally favored on-site consultations, while our subgroup analysis revealed that this was associated with whether they had already received a phone consultation or had been treated for more than three months at our Pain Center. Acceptance was moderately associated with satisfaction and strongly associated with willingness to use telemedicine outside of the COVID-19 pandemic. None of the assessed items were associated with the strictness of measures against the COVID-19 pandemic.

Compared to our previous investigation [7], the general condition was rated slightly higher in our sample, while the average pain intensity was comparable.

Acceptance, Satisfaction, and Preferences

Our findings underscore a central point: acceptance and satisfaction are not the same as preference. This seems to be overlooked often, as most studies have not assessed what patients favor when offered several options for consultation.
Some might accept and be satisfied with 1 service but still prefer the other [30]. For example, a recent investigation found high acceptability of a telemedicine pilot initiative for patients with chronic noncancer pain, but patients’ preferences were not examined specifically [31]. It is important to note that in our study, patients who had never experienced telemedicine or had only been treated for several weeks, preferred on-site consultations. One potential explanation could be that telemedicine is still largely unknown or underused by many patients. This is in line with a previous study that showed that one of the main reasons for not using telehealth was a lack of knowledge about it [32]. The good health coverage in Switzerland may contribute to this factor, as it is easier to find a timely appointment in person than in countries with more remote areas. Another reason could be barriers to accessing telemedicine in the senior population [33], who may not be as familiar with modern or unconventional modes of care. An additional factor could be trust in the practitioner, which has been shown to improve satisfaction as well as other patient-reported outcomes, such as quality of life [34]. In our case, it is conceivable that patients who had already had a telephone consultation and had a good experience gained more trust. The same applies to people who have been in treatment for longer than 3 months; they had more time to gain trust in the treating doctor.

In summary, telemedicine acceptance in our study was high and comparable to other investigations that have examined acceptance in other collectives: for example, 1 study in patients who were urogynecologic found 87% acceptance [35], while 2 scoping reviews found high acceptance of telemedicine in palliative care [36] and older adult patients with tumors [37]. In another cross-sectional study of patients with tumors, high satisfaction was found, with a mean of 5.5 out of 7 [38]. Nevertheless, while acceptance of and satisfaction with telemedicine were high in these investigations, these findings cannot be applied to every patient as a “one-size-fits-all” rule, as patients’ individual preferences (and needs) might differ. This was exemplified in a recently published systematic review with meta-analysis on outcomes of web-based pain management for chronic widespread musculoskeletal conditions where the investigators found clinically significant improvements but failed to show a clinically relevant change [39]. Based on our findings, we would recommend asking patients specifically what service and mode of treatment they prefer and to build upon personal therapeutic relationships, thus improving patient-centered care.

Furthermore, the COVID-19 pandemic has led to barriers to seeking emergency care [40], and this should be considered as promoting the implementation of digital services, such as telemedicine and digital triage [41]. Nevertheless, it is currently an open question how telemedicine will continue to develop in the Swiss population after the first waves of the COVID-19 pandemic, and to what extent the initiated changes will last. Research on the future use of telehealth, which considers the experience of COVID-19 and the impact of the COVID-19 pandemic on health care providers and patients, is essential.

**Correlation Analyses**

The values regarding perception of care were highly correlated with each other and satisfaction. While this is an unsurprising result, it highlights the importance of maintaining high-quality services, directly translating into how satisfied patients feel. Concerning the COVID-19 stringency index, we were unable to show any meaningful influence on satisfaction, acceptance, preferences, or the perception of treatment, which does not support the idea that patients are substantially driven by their perception of the “strictness” of governmental measures. Previously, we hypothesized that due to government restrictions, acceptance of and preference for telemedicine would be higher (as it is a measure of reducing personal contacts). Another explanation for the lack of correlation could be that during this study period, government measures were largely stable and varied only mildly, whereas, in theory, they could range between scores of 0 and 100. Therefore, the stringency index may show associations as soon as it exhibits greater change.

**Limitations**

Our study is limited by the small sample size and the low response rate, which is a problem for reliability and generalizability and can lead to a nonresponse bias. We attempted to address this by comparing the general patient population seen at our clinic with our study respondents to ensure representativeness. One reason why so few patients participated could be the number of standard questionnaires that they received before an appointment at our clinic. It is conceivable that this could lead to reluctance to fill out additional forms if they are not necessary for treatment. Additionally, as this was a web-based survey only, this could exclude patients who are not familiar with QR code scanning and handling of technical devices [42]. It is also important to note that measuring acceptance and satisfaction on an 11-point scale only allows a very general impression with no further information on why telemedicine was accepted or rejected. By contrast, validated questionnaires [43,44] could paint a more accurate picture. This limitation was somewhat mitigated by the fact that we also asked whether the patients felt taken seriously, whether their questions were answered, whether they could be helped, and whether they would use telemedicine even without the COVID-19 pandemic. These items were all strongly correlated with satisfaction, which can be interpreted as an indication of their importance to satisfaction itself. Additionally, this study lacked information on socioeconomic status, income, level of education, digital competence, and literacy. This is important to consider in the overall context, as lower literacy, for example, can be seen as a significant factor in the failure of a medical intervention. Future studies should consider these points to evaluate a more comprehensive assessment of access to health care, including telemedicine. A possible template for such a study is the Levesque access framework, which, in addition to the determinants of health care (approachability, acceptability, availability or accommodation, affordability, and appropriateness), also takes into account the individual possibilities of patients and can thus capture possible barriers to health care in a differentiated way [45,46]. The qualitative or quantitative assessment of these aspects is likely to play an increasingly important role in the future, especially in the...
implementation of more modern technologies, such as chatbots [47] or other automated systems controlled by artificial intelligence [48]. In our view, the evaluation of patient acceptance and satisfaction in these areas is crucial for ensuring patient-centered care. In patients with chronic pain, it would be interesting to investigate whether there are patient groups that particularly benefit from telemedicine (for example, those with limited mobility, more comorbidities, those with frequent medication adjustments, or the need for a higher frequency of care, such as cancer pain). Furthermore, the term “telemedicine” is widely used nowadays, and in our case, it was applied only to phone consultations. It is possible that participants in other outpatient settings using other services such as video chats might experience telemedicine differently and prefer these methods. Owing to administrative and legal hurdles in Switzerland, the general introduction of video consultations remains a challenge in the near future. Finally, options for billing patients are restricted to a maximum time limit of 20 minutes for an appointment by phone or video call according to the Swiss tariff structure [49].

**Conclusions**

Although telemedicine was widely accepted in our patient population and patients were generally very satisfied with it, they nevertheless preferred on-site consultations. In contrast, patients who had already experienced telephone consultation or those who had already undergone treatment for more than three months showed significantly higher preferences for telemedicine. This highlights 2 points: first, knowledge of eHealth services needs to be conveyed to patients to make use of its many advantages. Second, our investigation stresses the importance of building meaningful relationships with patients at the beginning of treatment and how this can improve patients’ perception of care. During the COVID-19 pandemic, the modality of care should be discussed individually for each patient.

**Acknowledgments**

The authors thank Jeannie Wurz (Department of Anaesthesiology and Pain Medicine, Inselspital, Bern University Hospital) for reviewing and providing helpful feedback during the development of this paper. Only departmental funding was provided for this study. No generative artificial intelligence (AI) was used in the preparation of this paper.

**Data Availability**

The data used to support the findings of this study are restricted by the Cantonal Ethics Committee of Bern, Switzerland, to protect patient privacy. Data are available upon reasonable request from MAH, Doctor of Medicine, for researchers who meet the criteria for access to confidential data.

**Authors' Contributions**

MAH takes responsibility for the integrity of the work. He designed this study together with KS, acquired data, conducted the main analyses, wrote the interpretation, drafted this paper, implemented revisions by the other authors, and approved this version. AS acquired data, conducted the main analyses, revised this paper critically, and approved this version. LB helped with the design of this study, revised this paper critically, and approved this version. ZN acquired data, helped with data interpretation, revised this paper critically, and approved this version. TCS helped with data interpretation, revised this paper critically, and approved this version. AL helped with the design of this study, helped with data analysis and interpretation, revised this paper critically, and approved this version. MgH helped with the design of this study, revised this paper critically, and approved this version. KS takes responsibility for the integrity of the work. He designed this study together with MAH, acquired data, helped with data interpretation, revised this paper critically, and approved this version.

**Conflicts of Interest**

The authors have no conflicts of interest to declare. TCS holds the endowed professorship for emergency telemedicine at the University of Bern, Switzerland, founded by the Touring Club Switzerland. The founders do not influence the general direction of telemedicine research. In particular, there is no influence on the content of this publication or the decision to conduct or publish this study. All coauthors have seen and agreed with the contents of this paper and there are no financial interests to report.

Multimedia Appendix 1

Table S1 and Table S2.

[DOCX File, 36 KB - formative_v8i1e53154_app1.docx ]

**References**


40. Haußer WE, Sauter TC, Exadaktylos AK, Krummrey G, Schauer S, Müller M. Barriers to seeking emergency care during the COVID-19 pandemic may lead to higher morbidity and mortality—a retrospective study from a Swiss university hospital. Swiss Med Wkly 2020;150:w20331 [FREE Full text] [doi: 10.4414/smw.2020.20331] [Medline: 32799308]


**Abbreviations**

PCS-6: Pain Catastrophizing Scale 6-item
Attributes, Quality, and Downloads of Dementia-Related Mobile Apps for Patients With Dementia and Their Caregivers: App Review and Evaluation Study

Tzu Han Chen1; Shin-Da Lee2, PhD; Wei-Fen Ma3,4, PhD

1PhD Program for Health Science and Industry, China Medical University, Taichung, Taiwan
2PhD Program in Healthcare Science, Department of Physical Therapy, China Medical University, Taichung, Taiwan
3PhD Program in Healthcare Science, School of Nursing, China Medical University, Taichung, Taiwan
4Department of Nursing, China Medical University Hospital, Taichung, Taiwan

Corresponding Author:
Wei-Fen Ma, PhD
PhD Program in Healthcare Science
School of Nursing
China Medical University
No 100, Sec 1, Jingmao Road
Beitun District
Taichung, 406040
Taiwan
Phone: 886 4 22053366 ext 7107
Fax: 886 4 22053748
Email: lhdaisy@mail.cmu.edu.tw

Abstract

Background: The adoption of mobile health (mHealth) apps among older adults (>65 years) is rapidly increasing. However, use of such apps has not been fully effective in supporting people with dementia and their caregivers in their daily lives. This is mainly attributed to the heterogeneous quality of mHealth apps, highlighting the need for improved app quality in the development of dementia-related mHealth apps.

Objective: The aims of this study were (1) to assess the quality and content of mobile apps for dementia management and (2) to investigate the relationship between app quality and download numbers.

Methods: We reviewed dementia-related mHealth apps available in the Google Play Store and Apple App Store in Taiwan. The identified mobile apps were stratified according to a random sampling approach and evaluated by five independent reviewers with sufficient training and proficiency in the field of mHealth and the related health care sector. App quality was scored according to the user version of the Mobile Application Rating Scale. A correlation analysis was then performed between the app quality score and number of app downloads.

Results: Among the 17 apps that were evaluated, only one was specifically designed to provide dementia-related education. The mean score for the overall app quality was 3.35 (SD 0.56), with the engagement (mean 3.04, SD 0.82) and information (mean 3.14, SD 0.88) sections of the scale receiving the lowest ratings. Our analyses showed clear differences between the top three- and bottom three-rated apps, particularly in the entertainment and interest subsections of the engagement category where the ratings ranged from 1.4 to 5. The top three apps had a common feature in their interface, which included memory, attention, focus, calculation, and speed-training games, whereas the apps that received lower ratings were found to be deficient in providing adequate information. Although there was a correlation between the number of downloads (5000 or more) and app quality ($t_{15}=4.087, P<.001$), this may not be a significant determinant of the app’s perceived impact.

Conclusions: The quality of dementia-related mHealth apps is highly variable. In particular, our results show that the top three quality apps performed well in terms of engagement and information, and they all received more than 5000 downloads. The findings of this study are limited due to the small sample size and possibility of disregarding exceptional occurrences. Publicly available expert ratings of mobile apps could help people with dementia and their caregivers choose a quality mHealth app.

(JMIR Form Res 2024;8:e51076) doi:10.2196/51076)
KEYWORDS
app quality; caregiver; dementia; geriatrics; aging; technology; digital health; mHealth; mobile health; seniors; mobile app; patient; adoption; development; management

Introduction

Background

The global aging population is experiencing an astonishing surge, which will inevitably result in a significant rise in the prevalence of dementia [1]. Consequently, it has become crucial to identify efficacious strategies to support people affected by dementia and enhance the well-being of their caregivers [2]. In addition, numerous studies have shown that mobile health (mHealth) apps can effectively reduce medical costs and improve quality of life for middle-aged and older adults, especially after COVID-19 [3,4].

The use of technology among older adults (aged >65 years) has triggered noteworthy transformations in health care provision [5]. An area where technology has proven especially valuable is in the realm of dementia management, with mHealth apps dominating the forefront of this field [6]. In addition, the UK government has shown support for the advancement of intelligent assistive technology for individuals with dementia [7]. This includes endorsing the development of mHealth apps specifically tailored to patients with early-stage dementia and their caregivers [8]. These apps are believed to have significant potential in aiding cognitive function and facilitating self-care among those living with dementia [9].

However, the constant emergence of mHealth apps has made it challenging for both patients with dementia and their caregivers across different categories. The second goal was to assess the quality of individual apps using the user version of the Mobile Application Rating Scale (uMARS). The third objective was to perform a comparative analysis of the highest- and lowest-quality dementia-related mHealth apps, with the broader goal of establishing guidelines to facilitate future app development. Finally, the study aimed to explore the correlation between app quality and downloads. This was done to help identify the gaps in the currently available dementia-related mHealth apps and to provide recommendations for patients with dementia and their caregivers on how to select high-quality apps.

Objective

This study had several goals. The first goal was to analyze the content of mobile apps for people with dementia and their caregivers across different categories. The second goal was to assess the quality of individual apps using the user version of the Mobile Application Rating Scale (uMARS). The third objective was to perform a comparative analysis of the highest- and lowest-quality dementia-related mHealth apps, with the broader goal of establishing guidelines to facilitate future app development. Finally, the study aimed to explore the correlation between app quality and downloads. This was done to help identify the gaps in the currently available dementia-related mHealth apps and to provide recommendations for patients with dementia and their caregivers on how to select high-quality apps.

Methods

Search Strategy and Inclusion Criteria

Apps were identified from the Taiwan Apple App Store and Google Play Store. Between July 2022 and November 2022, the following search terms (in Mandarin and English) were used in the app stores: dementia, cognitive dysfunction, dementia caregiver, Alzheimer disease, dementia care, cognitive games, and memory games. The screening criteria and process are illustrated in Figure 1.

Apps were included if they met all of the following inclusion criteria: (1) exists in the Google Play Store for Android mobile devices and the App Store for Apple mobile devices; (2) addresses daily-life topics related to neurocognitive disorders [17], and (3) was purposefully developed with the primary goal of supporting patients or caregivers (including health care workers) with the topic of mild cognitive impairment; (4) can be downloaded and used for free; (5) mainly uses Mandarin or the English version can be translated into Mandarin and is easy to understand; and (6) has been updated within the last 5 years.
Stratified Random Sampling of Apps by Average Download Numbers

In November 2022, searches were conducted on the two platforms to find apps that met the above criteria. Of the 407 apps found, 332 were deemed ineligible after screening (Figure 1). The remaining 75 apps were thoroughly screened, resulting in 52 apps included for preliminary evaluation. Since the length of time an app has been available on a platform can affect its number of downloads, we calculated the ratio of download numbers with respect to time on the platform. Additionally, to consider uneven allocation and lack of continuity in stratification, the apps were sorted according to the ratio of downloads relative to the number of days since the release date on the platform. Thus, the average number of downloads was calculated as the total number of downloads/number of days on platform since the release date. The apps were then ranked according to the average number of downloads in ascending order, and we randomly selected 1 out of every 3 apps for a total of 17 apps that were subject to detailed quality assessment and review.

General Characteristics and Classification

Each app was used by two authors (THC and WFM) independently. According to their content subcategory, the selected apps were categorized into four different types using the guidelines provided by the National Institute for Health and Care Excellence and the National Health Service in the United Kingdom [18,19]. Any conflicts in app classification were adjudicated by discussion between the two reviewers regarding each domain within the extraction form to reach consensus. Details on the main characteristics and comments of the included apps are provided in Multimedia Appendix 1.

mHealth App Quality Evaluation

The uMARS is a tool that can be used to evaluate the quality of mHealth apps, including four objective subdomains: engagement, functionality, esthetics, and information. There is also a domain for subjective quality and another for perceived impact. Stoyanov et al [20] developed the uMARS in 2016, which showed excellent internal consistency (Cronbach $\alpha=0.90$). The uMARS scores are rated on a 5-point Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”).

The objective quality score is calculated as the average of the scores of the four dimensions. Engagement is defined as fun, interesting, customizable, interactive, and has prompts (eg, sends alerts, messages, reminders, feedback, allows sharing). Functionality refers to overall app functioning, easy to learn, navigation, flow logic, and gestural design of the app. Esthetics refers to the graphic design, overall visual appeal, color scheme, and stylistic consistency. Finally, the information domain assesses whether the app contains high-quality information (eg, text, feedback, measures, and references from a credible source). The subjective quality score reflects the rater’s personal interest in the app. The final uMARS subscale includes 6 items designed to assess the perceived impact of the app on the user’s awareness, knowledge, attitude, intention to change, help-seeking, and likelihood to change the target health behavior.
Reviewer Recruitment and Selection

Reviewers recruited for this study were required to have a professional background in clinical treatment, the health care industry, or information engineering. Additionally, they were required to have at least 3 years of work experience in elderly health care or health technology–related fields, as well as experience using digital mobile devices. Exclusion criteria included no relevant work experience in elderly health care or health technology–related fields in the past 5 years.

Five reviewers were recruited as an interdisciplinary group of experts. The initial reviewer possessed knowledge and had experience in creating a content management system for a dementia management app. The second reviewer was a health informatics researcher with sufficient training and expertise in the relevant health care technology fields focused on dementia. The third reviewer also had extensive experience in dementia and in the mHealth industry. The fourth reviewer was a psychiatric nurse with experience in caring for older adults along with clinical experience in dementia. The final reviewer was a nurse practitioner who has been providing care for older adults and patients with dementia for over a decade.

Evaluation Process

Each of the apps was assessed by the five reviewers and the evaluation process was conducted between December 17, 2022, and January 3, 2023. All 17 apps can be found on the Android platform; hence, the apps were reviewed when running on the same Android tablet. The experts were blinded to the download numbers, year, and country of development of the apps, and they were not allowed to discuss their assessments with each other to ensure independence in their ratings. We ensured an equal distribution of app assessments in each round by applying a ratio that took into account the download-to-time axis. Furthermore, each reviewer allocated a minimum of 30 minutes and a maximum of 1 hour to thoroughly evaluate the included apps.

Ethical Considerations

The study received ethical approval from the ethics committee of China Medical Hospital, Taiwan, on November 8, 2022 (approval number: CMUH111-REC2-151) and was conducted according to the guidelines of the Declaration of Helsinki.

The experts in this study were not compelled to take part and had the freedom to determine their involvement. Additionally, they possessed the ability to discontinue their participation at any juncture, without being required to supply a justification for their decision.

This study utilized legally obtained publicly available information, and it was ensured that the use of information aligns with its intended public knowledge purpose. Furthermore, data collected from research and expert evaluations are stored on a hard drive and encrypted. The evaluation process was fully anonymous, with no face-to-face interactions among experts, and the evaluation of the app was a non-nominal, noninteractive, and noninvasive study. Relevant original data regarding this research will be preserved for at least 3 years after the execution period, securely locked in the principal investigator’s office cabinet.

The clinical trial protocol developed by the research institute stipulates that in the event of adverse reactions resulting in damages, China Medical University Hospital is responsible for providing compensation. Nonetheless, adverse reactions explicitly disclosed in the informed consent form signed by the experts are not eligible for compensation. This study was not covered by liability insurance and the per-expert evaluation cost was US $170.

Statistical Analysis

The number and proportion of information displayed in the apps, including the country of app development, download number, and app type, were summarized using descriptive statistics. The uMARS scores, along with the scores for each domain and subscale, are presented as the mean and SD. The t test was used to examine the association between downloads and each domain of the uMARS. Statistical analyses were conducted using IBM SPSS Statistics v28 (IBM Corp). We considered \( P < .05 \) to indicate statistical significance in all analyses.

Results

App Attributes

The apps were primarily developed in the United States, and 11 out of the 17 dementia-related mobile apps were downloaded less than 5000 times. Among the 17 apps, 8 were classified as those designed to improve clinical outcomes from established treatment pathways through behavior change, and for enhancement of patient adherence and compliance with treatment; 5 were designed as standalone digital game therapeutics; 3 were classified for supporting clinical diagnosis and/or decision-making; and 1 app was primarily designed to provide disease-related education (Table 1).
Table 1. App characteristics (N=17).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Apps, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country of development</strong></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2 (12)</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1 (6)</td>
</tr>
<tr>
<td>United States</td>
<td>5 (29)</td>
</tr>
<tr>
<td>Taiwan</td>
<td>2 (12)</td>
</tr>
<tr>
<td>Germany</td>
<td>2 (12)</td>
</tr>
<tr>
<td>Australia</td>
<td>2 (12)</td>
</tr>
<tr>
<td>India</td>
<td>2 (12)</td>
</tr>
<tr>
<td>Canada</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>Number of downloads</strong></td>
<td></td>
</tr>
<tr>
<td>&gt;10</td>
<td>1 (6)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>1 (6)</td>
</tr>
<tr>
<td>&gt;100</td>
<td>2 (12)</td>
</tr>
<tr>
<td>&gt;500</td>
<td>3 (18)</td>
</tr>
<tr>
<td>&gt;1000</td>
<td>4 (24)</td>
</tr>
<tr>
<td>&gt;5000</td>
<td>2 (12)</td>
</tr>
<tr>
<td>&gt;10,000</td>
<td>1 (6)</td>
</tr>
<tr>
<td>&gt;100,000</td>
<td>1 (6)</td>
</tr>
<tr>
<td>&gt;1,000,000</td>
<td>1 (6)</td>
</tr>
<tr>
<td>&gt;10,000,000</td>
<td>1 (6)</td>
</tr>
<tr>
<td><strong>App type/function</strong></td>
<td></td>
</tr>
<tr>
<td>Improve clinical outcomes from established treatment pathways through behavior change, and enhancement of patient adherence and compliance with treatment</td>
<td>8 (47)</td>
</tr>
<tr>
<td>Standalone digital game therapeutic</td>
<td>5 (29)</td>
</tr>
<tr>
<td>Support clinical diagnosis or decision-making</td>
<td>3 (18)</td>
</tr>
<tr>
<td>Primarily to deliver disease-related education</td>
<td>1 (6)</td>
</tr>
</tbody>
</table>

App Quality Assessment by Interdisciplinary Experts

There was a notable level of agreement or correlation among the reviewers in their app evaluations, as indicated by the Kendall W statistic of 0.143, which was significant at P=.05.

Overall, the mean app quality score was 3.35 (SD 0.56), which ranged from 2.25 (worst-rated app) to 4.07 (best-rated app). For engagement, the mean score was 3.04 (SD 0.81). Furthermore, functionality had the highest mean score of 3.76 (SD 0.38) and showed the smallest variation in minimum and maximum scores among the apps evaluated. In other words, these apps were considered to have relatively high levels of functionality and usability by the interdisciplinary expert reviewers. The esthetic quality of the interface received a mean score of 3.45 (SD 0.65), indicating that visual design elements such as button size, icon clarity, and content arrangement were perceived as being well organized. Additionally, the information domain received a mean score of 3.14 (SD 0.88), suggesting that the presentation and accessibility of information on the screen could be improved. Multimedia Appendix 2 provides the complete details of app quality scores.

Top Three and Bottom Three Performers in App Quality Score

The apps ranked in the top three positions according to app quality scores included Memorad0 Brain Games, NeuroNation-Brain Training & Brain Games, and Brain Track. The common characteristic among these apps is that their interface consists of training games focused on memory, attention, concentration, calculation, and speed. Conversely, Alz Test, American Caregiver Association, and Dementia and Me ranked in the bottom three; these three apps performed poorly on both engagement and information.

The overall scores for each item for the top three and bottom three apps are provided in Table 2. The functionality domain received the highest average ratings, particularly for gestural design, navigation, and performance. The largest discrepancies in app quality ratings between the top three and bottom three apps were found in the areas of entertainment and interest, where the scores ranged from 1.4 (worst-rated app) to 5 (best-rated app). Similarly, in the subscale of perceived impact, there was
a significant difference in attitude, with ratings ranging from 1.2 (worst-rated app) to 4.2 (best-rated app).

Table 2. Comparison of quality between the top three- and bottom three-rated dementia-related apps.

<table>
<thead>
<tr>
<th>Metric</th>
<th>All apps (N=17)</th>
<th>Memorado</th>
<th>NeuroNation</th>
<th>Brain Track</th>
<th>Alz Test</th>
<th>American Caregiver Association</th>
<th>Dementia and Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking (out of 17)</td>
<td>N/A</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>15</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Number of downloads</td>
<td>N/A</td>
<td>&gt;1,000,000</td>
<td>&gt;10,000,000</td>
<td>&gt;5000</td>
<td>&gt;1000</td>
<td>&gt;100</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Engagement, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entertain</td>
<td>3 (1.08)</td>
<td>5 (0)</td>
<td>4.4 (0.55)</td>
<td>4.2 (0.84)</td>
<td>2.4 (0.89)</td>
<td>1.8 (0.45)</td>
<td>1.4 (0.55)</td>
</tr>
<tr>
<td>Interest</td>
<td>3 (1.09)</td>
<td>5 (0)</td>
<td>4.2 (0.45)</td>
<td>4.2 (0.45)</td>
<td>2.4 (0.89)</td>
<td>1.6 (0.55)</td>
<td>1.4 (0.55)</td>
</tr>
<tr>
<td>Customize</td>
<td>2.67 (0.49)</td>
<td>3.6 (0.89)</td>
<td>4 (1)</td>
<td>3.4 (1.14)</td>
<td>1.8 (0.45)</td>
<td>2.2 (0.45)</td>
<td>2.6 (0.89)</td>
</tr>
<tr>
<td>Interactivity</td>
<td>2.95 (0.72)</td>
<td>4 (1)</td>
<td>4.4 (0.89)</td>
<td>4 (0.71)</td>
<td>2.4 (0.89)</td>
<td>2 (1.41)</td>
<td>2.2 (0.45)</td>
</tr>
<tr>
<td>Target group</td>
<td>3.55 (0.65)</td>
<td>4.4 (0.89)</td>
<td>4.2 (0.84)</td>
<td>4.2 (0.45)</td>
<td>2.6 (0.55)</td>
<td>2.6 (0.89)</td>
<td>2.4 (0.55)</td>
</tr>
<tr>
<td>Functionality, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>3.72 (0.52)</td>
<td>4.6 (0.55)</td>
<td>3.8 (0.84)</td>
<td>4 (0.71)</td>
<td>3 (1.22)</td>
<td>3 (1.41)</td>
<td>2.6 (1.34)</td>
</tr>
<tr>
<td>Ease of use</td>
<td>3.62 (0.39)</td>
<td>4 (1)</td>
<td>3.4 (0.55)</td>
<td>3.6 (0.55)</td>
<td>3.4 (0.89)</td>
<td>3 (1.41)</td>
<td>3 (1.22)</td>
</tr>
<tr>
<td>Navigation</td>
<td>3.72 (0.29)</td>
<td>4 (0.71)</td>
<td>3.6 (0.55)</td>
<td>3.8 (0.45)</td>
<td>3.6 (0.55)</td>
<td>3.4 (0.89)</td>
<td>3.2 (0.84)</td>
</tr>
<tr>
<td>Gestural design</td>
<td>3.98 (0.37)</td>
<td>4.6 (0.55)</td>
<td>4.2 (0.84)</td>
<td>4.2 (0.45)</td>
<td>3.6 (0.55)</td>
<td>3.2 (1.1)</td>
<td>3.4 (0.89)</td>
</tr>
<tr>
<td>Esthetics, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layout</td>
<td>3.54 (0.58)</td>
<td>4.4 (0.89)</td>
<td>4 (1)</td>
<td>4.4 (0.89)</td>
<td>2.8 (1.1)</td>
<td>3 (1)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Graphics</td>
<td>3.51 (0.53)</td>
<td>4 (1)</td>
<td>4.2 (0.84)</td>
<td>4.2 (0.45)</td>
<td>3 (1)</td>
<td>2.6 (0.55)</td>
<td>3.2 (0.84)</td>
</tr>
<tr>
<td>Visual appeal</td>
<td>3.32 (0.65)</td>
<td>4.2 (0.84)</td>
<td>4.2 (0.84)</td>
<td>4 (0)</td>
<td>2.2 (0.84)</td>
<td>2.4 (0.55)</td>
<td>2.6 (0.55)</td>
</tr>
<tr>
<td>Information, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>3.55 (0.61)</td>
<td>4.5 (0.58)</td>
<td>4.2 (0.45)</td>
<td>4.2 (0.45)</td>
<td>2.75 (0.5)</td>
<td>3.5 (1.29)</td>
<td>2 (1.41)</td>
</tr>
<tr>
<td>Quantity</td>
<td>3.41 (0.74)</td>
<td>4.5 (0.58)</td>
<td>4 (1)</td>
<td>4 (0.7)</td>
<td>2.5 (1.29)</td>
<td>3.5 (1)</td>
<td>2 (0)</td>
</tr>
<tr>
<td>Visual information</td>
<td>3.4 (0.66)</td>
<td>4.25 (0.96)</td>
<td>4.2 (0.45)</td>
<td>4.2 (0.45)</td>
<td>2.6 (0.58)</td>
<td>3 (0.82)</td>
<td>1.67 (0.58)</td>
</tr>
<tr>
<td>Credibility</td>
<td>3.3 (0.58)</td>
<td>3.75 (0.5)</td>
<td>3.8 (0.84)</td>
<td>3.4 (0.55)</td>
<td>3.75 (1.26)</td>
<td>3.25 (0.96)</td>
<td>1.67 (0.58)</td>
</tr>
<tr>
<td>App subjective quality, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommend</td>
<td>3.2 (0.87)</td>
<td>4.4 (0.89)</td>
<td>4 (1)</td>
<td>4 (0.71)</td>
<td>2 (1)</td>
<td>2.2 (1.1)</td>
<td>1.4 (0.55)</td>
</tr>
<tr>
<td>Relevant</td>
<td>2.79 (0.88)</td>
<td>4.2 (0.84)</td>
<td>3.6 (1.14)</td>
<td>3.6 (0.55)</td>
<td>1.6 (0.55)</td>
<td>1.6 (0.55)</td>
<td>1.4 (0.55)</td>
</tr>
<tr>
<td>Pay</td>
<td>2.05 (0.7)</td>
<td>3.4 (0.89)</td>
<td>2.6 (1.52)</td>
<td>3.2 (0.84)</td>
<td>1.2 (0.45)</td>
<td>1 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Star rating</td>
<td>3.05 (0.88)</td>
<td>4.4 (0.89)</td>
<td>4 (1)</td>
<td>4 (1)</td>
<td>2.2 (0.84)</td>
<td>2 (1)</td>
<td>1.4 (0.55)</td>
</tr>
<tr>
<td>Perceived impact, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness</td>
<td>3.08 (0.73)</td>
<td>4 (1)</td>
<td>3.8 (1.3)</td>
<td>4 (0.71)</td>
<td>2.2 (1.33)</td>
<td>2.6 (1.14)</td>
<td>1.2 (0.45)</td>
</tr>
<tr>
<td>Knowledge</td>
<td>3.24 (0.81)</td>
<td>3.2 (1.64)</td>
<td>3.8 (1.3)</td>
<td>3.8 (0.84)</td>
<td>2.4 (1.67)</td>
<td>3.2 (1.64)</td>
<td>1.2 (0.45)</td>
</tr>
<tr>
<td>Attitudes</td>
<td>3.04 (0.8)</td>
<td>4.2 (0.84)</td>
<td>3.6 (1.34)</td>
<td>4 (0.71)</td>
<td>2 (1.22)</td>
<td>2.6 (1.14)</td>
<td>1.2 (0.45)</td>
</tr>
<tr>
<td>Intention to change</td>
<td>3.02 (0.77)</td>
<td>4.2 (0.84)</td>
<td>3.8 (1.3)</td>
<td>3.6 (0.89)</td>
<td>2 (1.22)</td>
<td>2.4 (1.14)</td>
<td>1.4 (0.89)</td>
</tr>
<tr>
<td>Help seeking</td>
<td>3.06 (0.69)</td>
<td>3 (1.58)</td>
<td>3.6 (0.89)</td>
<td>3.6 (0.89)</td>
<td>2.4 (1.44)</td>
<td>2.6 (1.82)</td>
<td>1.2 (0.45)</td>
</tr>
<tr>
<td>Behavior change</td>
<td>3.08 (0.77)</td>
<td>4 (0.71)</td>
<td>3.8 (0.84)</td>
<td>3.8 (0.84)</td>
<td>2 (1.22)</td>
<td>2.4 (1.34)</td>
<td>1.4 (0.89)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Association Between Downloads and Quality of Mobile Apps

The Connectivity in Digital Health survey of global mHealth apps reported that 55% of the apps available on the Google Play store, Apple App Store, Windows Phone Store, Amazon Appstore, and Blackberry World had fewer than 5000 total downloads [21]. Therefore, the 17 apps included in our study were divided into two subgroups based on the total number of downloads. The first subset consisted of 6 apps with more than 5000 total downloads, representing 35.3% of all apps. The mean app quality score for this subgroup was significantly higher than that of the group of apps with less than 5000 downloads (Table 3). In addition, apps with more than 5000 downloads generally had higher scores for each domain. However, neither information nor perceived impact scores were significantly correlated with the number of downloads (Table 3).

Table 3. Independent-samples t test of the correlation of the scores of app quality with download numbers.

<table>
<thead>
<tr>
<th>uMARS² domain</th>
<th>Score, mean (SD)</th>
<th>Apps downloaded &gt;5000 times (n=6)</th>
<th>Apps downloaded &lt;5000 times (n=11)</th>
<th>t statistic (df=15)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>App quality</td>
<td>3.88 (0.26)</td>
<td>3.06 (0.45)</td>
<td>4.087</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Engagement</td>
<td>3.95 (0.54)</td>
<td>2.55 (0.4)</td>
<td>6.054</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Functionality</td>
<td>4.05 (0.31)</td>
<td>3.6 (0.32)</td>
<td>2.785</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Esthetics</td>
<td>4.1 (0.35)</td>
<td>3.1 (0.48)</td>
<td>4.469</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>3.43 (0.53)</td>
<td>2.98 (1.00)</td>
<td>1.210</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Subjective quality</td>
<td>3.70 (0.44)</td>
<td>2.26 (0.40)</td>
<td>6.924</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Perceived impact</td>
<td>3.31 (0.50)</td>
<td>2.96 (0.79)</td>
<td>0.97</td>
<td>.35</td>
<td></td>
</tr>
</tbody>
</table>

²uMARS: user version of the Mobile Application Rating Scale.

Discussion

Principal Findings

According to our results, there was only one included app that primarily focused on delivering dementia-related education. Furthermore, the top three quality apps were all classified as the main app type, as they all served as standalone digital game therapeutics. In general, the dementia-related mHealth apps were of moderate quality with a common characteristic of high functionality. Nonetheless, these apps exhibited poor performance in engagement and the credibility of information domain. Although we found a correlation between the number of downloads and app quality, this may not be a significant determinant of the information provided and the app’s perceived impact.

Comparison With Prior Work

mHealth apps offer a new way to support people with dementia and their caregivers [22]. However, previous studies have pointed out that the scientific literature on the design and evaluation of web- and mobile-based health apps remains scarce [23,24]. To address this issue, our study directly assessed the app type in a practical setting and found the lack of a dementia management app that delivers disease-related education. A randomized controlled trial indicated that mHealth apps can be of educational value to patients by providing structured disease and treatment-related education; therefore, future app developers can focus on increasing the availability of this app type with educational value [25].

A previous study suggested that research collaboration between health care and software engineering experts could help advance our knowledge of app functionality and effectiveness [16]. Therefore, we established a panel of experts to obtain accurate results on the quality of currently available dementia-related mHealth apps and further identified their subjective quality and perceived impact. The pattern of high functionality and low information quality is in accordance with the findings of other studies on mobile apps designed for older adults [26]. Additionally, the inadequacy of credibility was associated with several risks, particularly in the areas of self-diagnosis, prevention, and health promotion [27].

High-quality mHealth apps offer self-management features, relaxation, recreation, and trustworthy information [28,29]. The uMARS consists of elements of usability and a broader range of areas that are used in the assessment of mHealth apps with superior quality. Notably, a consensus was reached among the reviewers in both the engagement and esthetics domains. However, there was no correlation or similarity among reviewers with respect to assessments on functionality and information of the apps. This discrepancy may be due to the different backgrounds of the reviewers [30]; health care providers may perceive the app’s information as inadequate, whereas experienced developers of dementia apps may find its functionality to be lacking.

Currently, little is known about why some health apps become popular and others do not, and researchers have demonstrated that the number of downloads on app marketplaces does not correlate with clinical utility or validity for mental health apps [31]. A study from the Netherlands and Portugal identified the predictors that might influence the number of downloads for urology apps [32]. However, there is little research on the predictors of app downloads for dementia-related mHealth apps in the PubMed database. Hence, to gain a more comprehensive understanding, the apps were stratified using a random sampling
approach. Due to the different themes of mHealth apps, our study found a positive relationship between app quality and number of downloads. Finally, the download number does only seem to be a limited orientation aid for the selection of an mHealth app, and future studies should consider this aspect.

Limitations

This study has several limitations. Initially, the search for mobile apps was conducted within a limited time frame and focused on apps that had been updated within the last 5 years. As such, the study fell short with respect to establishing causal relationships. In addition, rapidly expanding and ever-changing mobile app marketplaces are facing significant challenges in keeping pace with the dynamic landscape; hence, some of the apps evaluated in this study may have since changed or new alternatives may have been developed. Furthermore, the search for mobile apps was confined to app stores in Taiwan, which may not accurately represent app offerings in other countries due to regional disparities in developers’ decisions regarding app availability.

Previous research indicated that the cost associated with using mHealth apps acts as a major obstacle for older individuals when it comes to embracing mobile technologies [9,33]. Furthermore, a recent study discovered that 96% of mHealth apps that are accessible on the Chinese market can be downloaded without cost [34]. Consequently, one-quarter of the apps would have been overlooked if they required payment. Nonetheless, it is possible that within this group of paid apps, there may have been some high-quality apps that were unintentionally excluded from consideration.

Additionally, the stratification method represents both less popular and highly downloaded apps, mirroring real-world data [21]. However, this method resulted in a smaller sample size, which could potentially lead to some superior apps being overlooked by chance. With only 17 apps remaining for evaluation, it is possible that there may not have been sufficient statistical power to establish a significant relationship between app quality and download frequency.

Finally, to ensure a rigorous evaluation of the app content, experts from different fields were recruited to review the apps. However, the limited number of reviewers could potentially influence the results of the study, and the degree of agreement may not be strong given that the reviewers are from different disciplines and the time they allocated to evaluate each app could potentially impact the reliability of agreement.

Despite these limitations, this study helps to fill the gap in the evaluation of dementia-related mobile apps. The results can still be used to guide the selection of such apps in Taiwan and possibly other regions with similar app marketplaces, while also highlighting the need for ongoing evaluation of mobile apps for dementia care.

Conclusions

This study set out to gain a better understanding of the characteristics, quality, and downloads of dementia-related mHealth apps. In particular, the top three quality apps were all offered as standalone digital game therapeutics, which scored well on both engagement and information quality, and received more than 5000 total downloads. Nevertheless, the findings of our investigation do not offer a comprehensive solution due to the restricted scale of the sample and the potential for overlooking extraordinary instances. Consequently, annual reviews and publicly available expert ratings of mobile apps could help people with dementia and their caregivers choose a high-quality mobile app.

Acknowledgments

The authors acknowledge all staff and participants for their contributions to the study. This study was supported by the Ministry of Science and Technology (MOST 110-2314-B-039-041-MY2; NSTC112-2314-B-039-015) and China Medical University (CMU111-MF-108), Taiwan. The funders reviewed the study as part of the grant application but had no further role in study design; data collection, analysis, and interpretation; manuscript preparation; and paper publication.

Data Availability

The study data are identified participant data. The data that support the findings of this study will be available beginning 12 months and ending 36 months following the article publication from the corresponding author (WFM) upon reasonable request.

Authors’ Contributions

THC and WFM designed the study and were responsible for data collection and analysis. THC, SDL, and WFM all contributed to manuscript preparation and critical revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description, classification, and overall comments of reviewers after using the selected dementia-related mobile health apps.

[DOCX File, 30 KB - formative_v8i1e51076_app1.docx]
References


Abbreviations

mHealth: mobile health

uMARS: user version of the Mobile Application Rating Scale
Precision Assessment of Real-World Associations Between Stress and Sleep Duration Using Actigraphy Data Collected Continuously for an Academic Year: Individual-Level Modeling Study

Constanza M Vidal Bustamante1,2*, PhD; Garth Coombs III1,2*, PhD; Habiballah Rahimi-Eichi1,2,3,4, PhD; Patrick Mair1, PhD; Jukka-Pekka Onnela5, PhD; Justin T Baker3,4, PhD, MD; Randy L Buckner1,2,4,6, PhD

1Department of Psychology, Harvard University, Cambridge, MA, United States
2Center for Brain Science, Harvard University, Cambridge, MA, United States
3Institute for Technology in Psychiatry, McLean Hospital, Belmont, MA, United States
4Department of Psychiatry, Harvard Medical School, Boston, MA, United States
5Department of Biostatistics, Harvard University, Boston, MA, United States
6Athinoula A Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, United States

*these authors contributed equally

Corresponding Author:
Constanza M Vidal Bustamante, PhD
Department of Psychology
Harvard University
52 Oxford Street
Northwest Building, East Wing, Room 295.06
Cambridge, MA, 02138
United States
Phone: 1 617 384 8230
Email: constanzavidalbustamante@gmail.com

Abstract

Background: Heightened stress and insufficient sleep are common in the transition to college, often co-occur, and have both been linked to negative health outcomes. A challenge concerns disentangling whether perceived stress precedes or succeeds changes in sleep. These day-to-day associations may vary across individuals, but short study periods and group-level analyses in prior research may have obscured person-specific phenotypes.

Objective: This study aims to obtain stable estimates of lead-lag associations between perceived stress and objective sleep duration in the individual, unbiased by the group, by developing an individual-level linear model that can leverage intensive longitudinal data while remaining parsimonious.

Methods: In total, 55 college students (n=6, 11% second-year students and n=49, 89% first-year students) volunteered to provide daily self-reports of perceived stress via a smartphone app and wore an actigraphy wristband for the estimation of daily sleep duration continuously throughout the academic year (median usable daily observations per participant: 178, IQR 65.5). The individual-level linear model, developed in a Bayesian framework, included the predictor and outcome of interest and a covariate for the day of the week to account for weekly patterns. We validated the model on the cohort of second-year students (n=6, used as a pilot sample) by applying it to variables expected to correlate positively within individuals: objective sleep duration and self-reported sleep quality. The model was then applied to the fully independent target sample of first-year students (n=49) for the examination of bidirectional associations between daily stress levels and sleep duration.

Results: Proof-of-concept analyses captured expected associations between objective sleep duration and subjective sleep quality in every pilot participant. Target analyses revealed negative associations between sleep duration and perceived stress in most of the participants (45/49, 92%), but their temporal association varied. Of the 49 participants, 19 (39%) showed a significant association (probability of direction>0.975): 8 (16%) showed elevated stress in the day associated with shorter sleep later that night, 5 (10%) showed shorter sleep associated with elevated stress the next day, and 6 (12%) showed both directions of association. Of note, when analyzed using a group-based multilevel model, individual estimates were systematically attenuated, and some even reversed sign.
Conclusions: The dynamic interplay of stress and sleep in daily life is likely person specific. Paired with intensive longitudinal data, our individual-level linear model provides a precision framework for the estimation of stable real-world behavioral and psychological dynamics and may support the personalized prioritization of intervention targets for health and well-being.

(JMIR Form Res 2024;8:e53441) doi: 10.2196/53441

KEYWORDS
deep phenotyping; individualized models; intensive longitudinal data; sleep; stress; actigraphy; accelerometer; wearable; mobile phone; digital health

Introduction

Background

The transition to college is often accompanied by elevated stress and insufficient sleep [1-3], both of which have been found to impact daily functioning and, upon repeated exposure, to be associated with negative health outcomes ranging from internalizing disorders, including anxiety and depression, to cardiac and metabolic disease [4-9]. A better understanding of the dynamic day-to-day interplay between perceived stress and sleep duration could inform the mechanisms of their downstream impacts and the development of interventions to prevent or mitigate them [10-14]. Nevertheless, only a few studies have collected daily observations to evaluate within-person associations between stress responses and sleep duration, and their results are mixed: some studies report that heightened stress is followed by shorter sleep that night but not vice versa [15,16], others report that shorter sleep is followed by heightened stress the next day but not vice versa [17], and yet others report bidirectional relationships [18,19].

Inconsistent findings in stress-sleep associations might be at least partially explained by 2 interrelated methodological limitations, both of which are addressed by the individual-level modeling approach presented in this study. First, existing longitudinal studies have aggregated individual observations collected over short study periods of ≤14 days, thus limiting the estimation of stable associations that are robust to changing environmental demands (eg, first week of the semester vs final examinations period). Thanks to the adoption of digital phenotyping tools such as wearables and smartphones, research designs that sample individuals over much longer periods of time are increasingly feasible [20-22]. Sleep duration can now be passively tracked through continuous actigraphy sensing over months and even years, while perceived stress levels can be probed via brief daily smartphone-based surveys, with high compliance rates in student samples and small participant burden [23-25].

Moreover, the prevailing focus on group modeling and sample-level effects obscures the possibility that day-to-day associations between perceived stress and sleep duration may vary across individuals; for instance, it is possible that in certain individuals, getting less sleep than usual has no significant impact on stress levels the following day, but heightened stress during the day leads to shorter sleep duration later that night, while the opposite pattern might be true for others. Even when group-level models allow for individual-level estimates, such as in multilevel models with random effects, the degree to which the individual estimates are pulled toward the group (a shrinkage effect) is different for each individual based on the amount of data they provide, thus reducing the individual tailoring in a nonuniform manner [26].

Individual-level linear models (iLMs), which are fitted over a single individual’s intensive longitudinal observations, may offer a critical alternative to the estimation of stress-sleep associations. Compared to group-level approaches, iLMs provide estimates of phenotypes that are tailored to each person’s data and unbiased by the group [27-30]. They may also be more readily applicable in real-world contexts, where a clinician might use a precision health approach to evaluate and support each individual patient based on their personal data rather than on a hypothetical average patient [31-34]. Of note, individual-level approaches can also contribute to conclusions at the group or population level, such as by estimating the prevalence of each individually derived phenotype, rather than blurring across individuals, to arrive at a central tendency that might not accurately represent many of the included individuals.

Objectives

In this study, we introduce a novel iLM approach that leverages daily observations collected over a full academic year for the assessment of day-to-day associations between actigraphy-derived sleep duration and self-reported stress levels in first-year college students. Our aims were 2-fold. First, we used a pilot data set to develop and validate a parsimonious iLM that estimates concurrent or lagged associations between 2 daily variables of interest while accounting for the weekly structure in student behavior. We then leveraged this iLM for the target examination of bidirectional day-to-day associations between perceived stress and objective sleep duration in an independent data set of 49 first-year college students. We expected a negative association for most participants—such that higher stress levels are associated with shorter sleep—but we also anticipated that some participants might show positive or null associations. We further hypothesized that the lead-lag relationship between perceived stress and sleep duration (ie, which of the 2 temporally precedes the other) would vary across individuals such that, for some, elevated stress would associate with shorter sleep that night; for others, shorter sleep would associate with elevated stress the next day; and for still others, both directions of association would be identified.
Methods

Participants

Pilot Group

A total of 6 undergraduate students returning for their sophomore year volunteered for a year-long study (all aged 18-19 years; n=3, 50% women and n=3, 50% men). All had participated in a previous pilot study in our laboratory during their first year of college and were known to be compliant. These 6 pilot participants provided data to develop the statistical models that were then applied to the new group of target participants described in the next subsection. The pilot participants enrolled for this study during the first 2 weeks of their fall semester. As in the case of the target participants, they were required to be taking full-time classes and own a smartphone compatible with the study smartphone app, Beiwe, which is part of the open-source Beiwe platform for digital phenotyping [35]. Available and missing data information at the participant level is provided in Figure S1 in Multimedia Appendix 1. Given our focus on individual-level models where each person serves as their own baseline, and there is no aggregation across participants, students were not excluded for current or past psychiatric disorders or medication use, and nor were they excluded if they began treatment or medication for mental health issues during the study.

Target Group

In total, 49 undergraduate students beginning their first year of college volunteered for a year-long study (aged 18-19 years; mean age 18.1, SD 0.24 y; n=25, 51% women and n=24, 49% men). We have previously reported data from these participants [24]. Participants living on campus were recruited via flyers posted on campus boards and distributed via email lists and were enrolled during the first 2 weeks concurrently with the pilot participants. Enrollment criteria were the same as for the pilot participants other than the fact that the target first-year participants were all new to the university. Initially, 68 participants enrolled, of whom 19 (28%) were excluded based on issues with data acquisition, including early withdrawal from the study (n=7, 37%), technical failure of the actigraphy data (n=1, 5%), poor-quality actigraphy data (n=2, 11%), and completion of <50% of the daily surveys (n=9, 47%). Of the 49 participants in the final sample, 2 (4%) identified as American Indian, 5 (10%) as Asian, 7 (14%) as Black, 31 (63%) as White, and 2 (4%) as mixed race. Furthermore, 12% (6/49) reported prior diagnosis of a psychiatric disorder (including anxiety, depression, and attention-deficit/hyperactivity disorder), and 8% (4/49) maintained active diagnoses. The 49 first-year students had not yet declared their area of study, but they reported their desired future occupation to be medicine (n=15, 31%), business or finance (n=7, 14%), academia or other research (n=6, 12%), engineering (n=5, 10%), policy or government (n=5, 10%), law (n=4, 8%), and other or undecided (n=7, 14%). Of the 49 participants, 46 (94%) were iPhone users, whereas 3 (6%) used Android mobile phones.

Study Design

As previously reported in our study [24], this intensive longitudinal observational study collected passive and active data as participants engaged in their lives over a full academic year, extending a few days into the summer break. Both pilot and target participants completed smartphone-based daily surveys and a voice-recorded diary; wore an actigraphy wristband (GENEActiv Original; Activinsights Ltd) for continuous activity and sleep monitoring for the duration of the study; completed a battery of web-based questionnaires at the beginning, middle, and end of the study; and attended brief in-person check-ins every 3 to 4 weeks.

Ethical Considerations

Informed consent and all study procedures and methods were approved by the institutional review board of Harvard University (IRB16-1230). All participants completed an in-person informed consent session where study procedures were explained, and any questions were clarified. Participants were informed that they could withdraw their study participation at any time. Participants were compensated US $1 per each daily survey they submitted, US $1 per day for continuously wearing the actigraphy wristband, and US $20 per hour for web-based surveys and attending in-person visits. Milestone bonus payments for completing half of the study (US $100) and the full study (US $300) were also provided to compensate participants for their continued compliance.

Study data collected across devices were stored and automatically backed up in a secure data warehouse configured to automatically import data from various collection streams. All data were kept as securely as possible and were only accessible to study staff. Participants’ data were labeled with a randomly generated participant ID. Personally identifying information, such as names and contact information, were kept separate from all other collected data in a locked file cabinet (in a locked office, behind an ID card–secured suite during off-hours) and in a password-protected database.

Measures and Quality Control

Objective Sleep Duration

Daily sleep duration reflects the number of minutes between the estimated start and end of the day’s longest detected sleep episode. As redescribed from our original study [24], sleep duration was derived from the accelerometer data collected through the GENEActiv Original actigraphy wristband and analyzed via the deep phenotyping of sleep processing pipeline [36]. Participants wore the wristband on their nondominant wrist continuously, including during sleep and when bathing. Triaxial acceleration was collected with a sampling frequency of 30 Hz during the academic semesters and 10 Hz during the winter break (to extend battery life and memory while participants were away from campus). Participants were instructed to press the wristband’s button when they began trying to fall sleep at night and immediately after they awoke in the morning. Individuals exchanged their wristband for a fully charged one with reset memory at the in-person check-ins.
The deep phenotyping of sleep (DPSleep) processing pipeline was applied to the raw actigraphy data to detect the major sleep episode for each day [36]. The pipeline first converted the accelerometer data to minute-based activity estimates, removed the minutes when the individual was not wearing the device, and then estimated the major sleep episode based on a sliding window. Days where one of the boundaries of the sleep episode (ie, rises in relative activity both before and after a period of lower activity) could not be detected due to missing data were labeled as unusable. Two independent trained raters examined the automatically detected start and end times and the usability label of each sleep episode against the minute-based activity levels and the participant button presses when available. When necessary, they adjusted the automatic times and labels. Any disagreements between the 2 raters’ assessments were reviewed by the research team and resolved through discussion. A full description of the DPSleep processing pipelines applied to the actigraphy data, including quality control steps, can be found in the study by Rahimi-Eich et al [36].

All data that passed quality control were included in analysis, including days with no detected sleep episode (ie, with no extended periods of lower relative activity).

**Daily Telephone-Based Surveys**

Smartphone surveys were administered via the Beiwe app [35]. Each night before bed, participants completed a 46-item self-report survey related to their daily lives. As described originally in our study [24], the questions assessed a range of behaviors and internal states over the past 24 hours, including sleep quality, stress levels and sources, positive and negative affect, general physical health, daily consumption habits, studying behaviors, and sociability and support [24]. This paper reports analyses using 2 survey questions selected a priori that probed daily subjective sleep quality and perceived stress. The sleep quality question asked “How did you sleep last night?” and was answered on a 5-point scale (1 = terrible; little or no sleep, 2 = not so well: got some sleep but not enough, 3 = sufficient: got enough sleep to function, 4 = good: got a solid night’s sleep and felt well rested, and 5 = exceptional: one of my best nights of sleep). The perceived stress question asked “How much did you feel stressed over the past 24 hours?” and was also answered on a 5-point scale (1 = very slightly or not at all, 2 = a little, 3 = moderately, 4 = quite a bit, and 5 = extremely).

Surveys submitted between 5 PM (local time) on the day the survey opened and 6 AM the following day were considered to be on time. Surveys submitted after 6 AM the day after the survey was prompted were marked as missing. A participant was included in analysis if they were compliant with at least 100 daily surveys across the data collection period, and only on-time surveys from these participants were included.

**Analytical Approach**

**Development of the iLM**

An iLM was developed on the intensive day-level longitudinal data from the 6 pilot participants to test person-specific day-to-day behavioral associations. The iLM framework allows for individually tailored estimates by treating each day as the unit of observation and the individual as the population, as opposed to each individual as the unit of observation used to estimate the general population. An individual’s observed time series data can be understood as just one realization of a stochastic process whose data-generating process we are trying to model and understand [37], in this case through linear models. The intercept and slope estimates are unique to the individual, and we interpret them as a proxy for the individual’s phenotype. Of note, in this framework, each individual model can be interpreted as an independent test of the hypothesized association (eg, is shorter sleep than usual associated with heightened perceived stress the following day?).

The model was structured to be parsimonious while accounting for the nonindependence of the daily measures and the temporal structure imposed by the academic schedule. The day of the week was included as a categorical covariate to account for weekend versus weekday effects and other weekly structures imposed by the college schedule (eg, classes that meet on Monday-Wednesday-Friday vs those that meet on Tuesday-Thursday). In addition, because behavioral patterns during the semester vary substantially from those during the 5-week winter break (when students do not have classes and typically are away from campus), we decided a priori to only include in the model compliant data collected during the school semesters.

The final iLM took the following general form:

\[ y_t = \beta_0 + \beta_1 x_{\text{t-1}} + \beta_2 \text{DayOfWeek}_t + \epsilon_t \]

A daily observation on day \( t \) starts with the nighttime sleep episode and ends with the submission of the daily survey submitted in the evening before the next sleep episode. In the aforementioned formula, \( y_t \) is a single participant’s outcome variable (eg, sleep duration in min) at daily time point \( t \), \( \beta_0 \) is this participant’s individual intercept, \( x_t \) is the predictor variable (eg, sleep quality or stress measured on a 1-5 scale) at daily time point \( t \) or \( t-1 \) (depending on the lag of the tested association), \( \beta_1 \) is the participant’s individual slope for \( x_t \), and \( \epsilon_t \) is a normally distributed random error term.

All modeling was carried out in a Bayesian inference framework, which treats unknown parameters (eg, a slope) as random variables with a probability distribution rather than a discrete value; this distribution is updated based on the observed data (resulting in a posterior distribution) and serves as a measure of uncertainty around the parameter [38-40]. The Bayesian framework was favored for these analyses due to its flexibility in computing models with varied specifications (including complex random effect specifications in the multilevel models we fitted as part of our model validation process), robustness to sample data characteristics (eg, dispersion), and intuitive interpretation of the posterior distribution and 95% uncertainty interval (UI; ie, conditional on the data and the model, the probability that the parameter is contained in the interval is 0.95) [39,40]. For comparison, parallel iLMS fitted in a frequentist inference framework in the pilot validation stage yielded nearly identical point estimates.
to their Bayesian counterparts (refer to Figure S2 in Multimedia Appendix 1), suggesting that our model specification is robust across both statistical inference frameworks. All models were estimated in R (version 4.3.1: R Foundation for Statistical Computing) [41]. Bayesian models were estimated using the Stan modeling language [42] and the packages rstanarm (version 2.21.4 [43]), tidybayes (version 3.0.1 [44]), and bayesestR (version 0.13.1 [45]). Frequentist iLMs were estimated using the stats package included in base R [41].

Bayesian models were fit with default weakly informative priors specifying a gaussian distribution (mean 0, SD 2.5) to represent our diffuse prior knowledge. We estimated parameters using a Markov chain Monte Carlo (MCMC) approach. For each parameter, we sampled from 4 stationary Markov chains, each comprising 5000 sampling iterations, including a burn-in period of 2500 iterations that were discarded (for a total of 10,000 post–warm-up draws). Convergence of the 4 chains to a single stationary distribution was assessed quantitatively via the R-hat convergence diagnostic [46] (adequate convergence defined as R-hat <1.1) and qualitatively by visual inspection of trace plots showing the estimated parameter as a function of each chain’s iteration number (adequate convergence defined as the chains overlapping with each other throughout and a lack of structured patterns in each chain). Each model’s effective sample size (ESS) metric is reported. Each estimate in the MCMC process is serially correlated with the previous estimates: the higher the correlation, the more samples are needed to get to a stationary distribution. In the presence of nearly no autocorrelation, the ESS will be equal to the number of posterior draws requested (in this case, 10,000). Generally, the ESS should be at least 1000 to obtain stable estimates [38,47].

Adequacy of the model specification was assessed via 2 methods. First, posterior predictive checks entailed a visual comparison of the distribution of the observed outcome variable to the distribution of 100 simulated outcome data sets generated by applying 100 draws from the model parameters’ posterior distribution to our observed data set. Similarity in the distributions of the observed and model-generated outcomes suggests that the model specification captured the data well. Second, we inspected the model residuals against the model’s stationary distribution. In the presence of nearly no autocorrelation, the ESS will be equal to the number of posterior draws requested (in this case, 10,000). Generally, the ESS should be at least 1000 to obtain stable estimates [38,47].

We conducted model checks to evaluate the performance of the iLM framework. Posterior predictive checks (described earlier in this subsection) evaluated that a gaussian model specification captured the distribution of the data well. In addition, given the longitudinal design, we examined whether the model residuals lacked meaningful autocorrelation, which would suggest that our day-of-the-week covariate sufficiently captured the weekly structure of sleep duration. Finally, we compared the estimates obtained through the iLMs to the estimates obtained when the same data for the 6 pilot participants were fit within a single, group-based multilevel linear model (MLM), specifying fixed effects for intercepts and slopes and additional participant-level random intercepts and random slopes for the main predictor of interest (in this case, sleep quality). This comparison allowed us to assess our expectation that MLMs would provide individual estimates roughly comparable to those of the iLM, but with the critical difference that MLMs would systematically
attenuate these individual estimates, especially for individuals who deviate from the predominant association phenotype or when there is large heterogeneity in these phenotypes across individuals.

**Application of the iLM to the Novel Target Participants**

After developing the iLM and validating it extensively over the pilot data set, the modeling framework was then carried forward and applied to the independent target data set of first-year college students. All models were fit with a gaussian distribution. We first applied the same sanity check models as in the pilot data set. Subsequently, we used the iLM method to test a priori target hypotheses regarding the association between perceived daily stress and objective sleep duration.

Two target models were fit for each individual to examine bidirectional associations between sleep duration and perceived stress. A daily observation on day _t_ consists of last night’s sleep duration (ObjectiveSleepDuration, recorded passively via the actigraphy wristband) and the subjective rating of the present day’s overall stress levels (SubjectiveStressRating, reported by the participant in the evening). First, to test whether stress level during the day is related to sleep duration later that night (ie, a stress-then-sleep association), the model used the stress rating the day before sleep (day _t−1_ ) as the predictor of subsequent sleep duration (day _t_ ). Thus, the formula for this model was specified as follows:

\[
\text{ObjectiveSleepDuration}_t = \beta_0 + \beta_1 \text{SubjectiveStressRating}_{t-1} + \beta_2 \text{DayOfWeek}_t + \epsilon_t
\]

Next, to test whether sleep duration is related to stress levels the day after (ie, a sleep-then-stress association), the model used stress rating the day after sleep (day _t_ ) as predictor of the previous sleep duration (day _t−1_ , sleep duration last night):

\[
\text{ObjectiveSleepDuration}_t = \beta_0 + \beta_1 \text{SubjectiveStressRating}_t + \beta_2 \text{DayOfWeek}_t + \epsilon_t
\]

In this model, stress rating (predictor) is back-predicting sleep duration (outcome), which effectively tests the question “Is stress today associated with sleep last night?” As the objective sleep duration measure occurs temporally before the daily stress rating is submitted, results are interpreted as sleep duration being associated with increased or decreased stress the next day. Implementing the model in this way (rather than using sleep duration as the predictor and stress rating as the outcome) presented the advantage that the slope estimates across the 2 models are on equivalent units, namely, the change in the number of minutes in sleep duration per unit of change in stress rating. This allows us to directly compare the quantitative outputs for the models testing the questions “Is stress associated with subsequent sleep?” and “Is sleep associated with subsequent stress?”

As with the pilot data set, we conducted model checks to further evaluate the specification and performance of the iLM framework in our target data set, including posterior predictive checks; inspection of model residuals; and a comparison of the estimates obtained through the iLMs and the estimates obtained when the same data for the target sample were fit in a single, group-level MLM with random intercepts and random slopes per participant on the stress predictor. Additional inspection of model results against each participant’s raw time series data was conducted to complement our interpretation.

**Results**

**Participant-Level Descriptive Statistics**

Table 1 presents participant-level available data and summary statistics for sleep and stress variables used in analysis. The pilot participants provided a median of 178 (range 119-212 out of a total possible of 223) usable observations for modeling, that is, day-level observations collected during the school semesters with usable actigraphy and survey data. The target participants provided a median of 178 (range 84-214) usable observations. Participants’ total number of usable observations was not correlated with their mean sleep duration (r=0.09; _P_ =0.53) or mean stress levels (r=-0.11; _P_ =0.42). In addition, participants’ mean sleep duration did not differ on days with or without available actigraphy and survey data (paired 2-tailed _t_ test, _P_ =0.59), and participants’ mean stress levels did not differ on days with or without available actigraphy data (paired 2-tailed _t_ test, _P_ =0.70). These observations suggest that participants’ overall sleep and stress metrics did not introduce systematic missingness in the data (more details on participant-level missing data are presented in Figure S1 in Multimedia Appendix 1).
Table 1. Participant-level descriptive statistics of data used in analyses.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Daily observations</th>
<th>Sleep duration (min), mean (SD)</th>
<th>Sleep quality (1-5 Likert scale), mean (SD)</th>
<th>Stress (1-5 Likert scale), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>119</td>
<td>493.22 (101.89)</td>
<td>3.59 (0.64)</td>
<td>1.36 (0.74)</td>
</tr>
<tr>
<td>P2</td>
<td>212</td>
<td>412.08 (84.57)</td>
<td>2.50 (0.68)</td>
<td>2.95 (0.82)</td>
</tr>
<tr>
<td>P3</td>
<td>156</td>
<td>404.03 (68.81)</td>
<td>2.48 (0.55)</td>
<td>2.89 (0.93)</td>
</tr>
<tr>
<td>P4</td>
<td>200</td>
<td>425.20 (82.42)</td>
<td>2.64 (0.69)</td>
<td>3.15 (0.42)</td>
</tr>
<tr>
<td>P5</td>
<td>200</td>
<td>387.76 (58.32)</td>
<td>3.14 (0.74)</td>
<td>2.23 (1.14)</td>
</tr>
<tr>
<td>P6</td>
<td>147</td>
<td>379.29 (139.37)</td>
<td>2.65 (1.12)</td>
<td>3.85 (1.14)</td>
</tr>
<tr>
<td>T1</td>
<td>147</td>
<td>464.92 (96.66)</td>
<td>3.81 (0.94)</td>
<td>2.76 (1.06)</td>
</tr>
<tr>
<td>T2</td>
<td>212</td>
<td>460.47 (82.68)</td>
<td>3.08 (0.62)</td>
<td>3.16 (0.83)</td>
</tr>
<tr>
<td>T3</td>
<td>139</td>
<td>415.09 (114.00)</td>
<td>2.61 (0.71)</td>
<td>3.59 (0.82)</td>
</tr>
<tr>
<td>T4</td>
<td>103</td>
<td>349.96 (89.02)</td>
<td>2.81 (0.83)</td>
<td>1.69 (1.09)</td>
</tr>
<tr>
<td>T5</td>
<td>189</td>
<td>446.40 (60.47)</td>
<td>3.33 (0.62)</td>
<td>2.51 (0.83)</td>
</tr>
<tr>
<td>T6</td>
<td>199</td>
<td>465.13 (76.76)</td>
<td>3.07 (0.63)</td>
<td>2.68 (0.77)</td>
</tr>
<tr>
<td>T7</td>
<td>131</td>
<td>389.17 (56.96)</td>
<td>3.06 (0.91)</td>
<td>1.95 (1.19)</td>
</tr>
<tr>
<td>T8</td>
<td>173</td>
<td>402.33 (108.91)</td>
<td>2.87 (0.49)</td>
<td>2.69 (0.63)</td>
</tr>
<tr>
<td>T9</td>
<td>102</td>
<td>448.53 (144.97)</td>
<td>2.95 (0.79)</td>
<td>1.98 (1.01)</td>
</tr>
<tr>
<td>T10</td>
<td>208</td>
<td>432.88 (114.48)</td>
<td>2.71 (0.73)</td>
<td>3.07 (1.11)</td>
</tr>
<tr>
<td>T11</td>
<td>173</td>
<td>373.92 (94.30)</td>
<td>2.62 (0.73)</td>
<td>2.77 (0.78)</td>
</tr>
<tr>
<td>T12</td>
<td>87</td>
<td>475.03 (75.76)</td>
<td>2.83 (1.00)</td>
<td>3.57 (1.14)</td>
</tr>
<tr>
<td>T13</td>
<td>214</td>
<td>441.09 (73.22)</td>
<td>2.49 (0.56)</td>
<td>3.25 (0.70)</td>
</tr>
<tr>
<td>T14</td>
<td>84</td>
<td>454.23 (100.68)</td>
<td>3.40 (0.62)</td>
<td>2.52 (0.81)</td>
</tr>
<tr>
<td>T15</td>
<td>98</td>
<td>412.93 (64.32)</td>
<td>2.57 (0.66)</td>
<td>3.41 (0.87)</td>
</tr>
<tr>
<td>T16</td>
<td>152</td>
<td>414.31 (127.89)</td>
<td>2.70 (0.66)</td>
<td>3.08 (0.83)</td>
</tr>
<tr>
<td>T17</td>
<td>191</td>
<td>440.73 (74.20)</td>
<td>3.06 (0.50)</td>
<td>1.37 (0.68)</td>
</tr>
<tr>
<td>T18</td>
<td>92</td>
<td>385.57 (70.90)</td>
<td>3.09 (0.51)</td>
<td>2.87 (0.63)</td>
</tr>
<tr>
<td>T19</td>
<td>199</td>
<td>464.09 (74.95)</td>
<td>2.87 (0.44)</td>
<td>2.94 (0.59)</td>
</tr>
<tr>
<td>T20</td>
<td>192</td>
<td>497.65 (77.78)</td>
<td>3.69 (0.58)</td>
<td>2.30 (0.95)</td>
</tr>
<tr>
<td>T21</td>
<td>199</td>
<td>390.89 (82.58)</td>
<td>3.36 (0.76)</td>
<td>2.85 (1.05)</td>
</tr>
<tr>
<td>T22</td>
<td>126</td>
<td>464.25 (84.70)</td>
<td>3.33 (0.90)</td>
<td>2.53 (1.04)</td>
</tr>
<tr>
<td>T23</td>
<td>84</td>
<td>418.65 (73.60)</td>
<td>2.87 (0.62)</td>
<td>3.39 (0.79)</td>
</tr>
<tr>
<td>T24</td>
<td>148</td>
<td>482.32 (81.06)</td>
<td>3.19 (0.87)</td>
<td>2.01 (0.77)</td>
</tr>
<tr>
<td>T25</td>
<td>191</td>
<td>406.52 (95.58)</td>
<td>2.70 (0.78)</td>
<td>3.47 (1.00)</td>
</tr>
<tr>
<td>T26</td>
<td>205</td>
<td>446.68 (64.03)</td>
<td>3.01 (0.10)</td>
<td>2.94 (1.14)</td>
</tr>
<tr>
<td>T27</td>
<td>177</td>
<td>443.10 (114.21)</td>
<td>3.16 (0.82)</td>
<td>3.28 (0.76)</td>
</tr>
<tr>
<td>T28</td>
<td>209</td>
<td>409.12 (65.85)</td>
<td>2.92 (0.59)</td>
<td>2.67 (0.91)</td>
</tr>
<tr>
<td>T29</td>
<td>197</td>
<td>429.07 (74.85)</td>
<td>3.00 (0.47)</td>
<td>1.68 (0.88)</td>
</tr>
<tr>
<td>T30</td>
<td>195</td>
<td>396.16 (62.76)</td>
<td>2.67 (0.59)</td>
<td>2.85 (1.05)</td>
</tr>
<tr>
<td>T31</td>
<td>205</td>
<td>388.53 (94.91)</td>
<td>2.82 (0.83)</td>
<td>2.52 (0.71)</td>
</tr>
<tr>
<td>T32</td>
<td>117</td>
<td>397.49 (118.98)</td>
<td>2.61 (0.86)</td>
<td>3.73 (1.10)</td>
</tr>
<tr>
<td>T33</td>
<td>114</td>
<td>383.05 (135.02)</td>
<td>2.65 (0.89)</td>
<td>2.09 (1.27)</td>
</tr>
<tr>
<td>T34</td>
<td>206</td>
<td>436.76 (80.84)</td>
<td>2.88 (0.73)</td>
<td>2.43 (0.89)</td>
</tr>
<tr>
<td>T35</td>
<td>207</td>
<td>461.99 (71.50)</td>
<td>3.08 (0.78)</td>
<td>3.16 (1.23)</td>
</tr>
</tbody>
</table>
Sleep and Stress Fluctuate in Relation to the Academic Calendar in the Pilot Participants

There was a pattern of enhanced sleep and lower stress when students were released from the structured academic demands of the in-person school semesters. Participants were enrolled for a full academic year, including a fall semester and a spring semester (each lasting roughly 16 weeks), as well as a 5-week class-free winter break bridging the 2 semesters. During the winter break and weekends, pilot participants had longer objective (actigraphy-derived) sleep duration, better subjective sleep quality, and felt less stress compared to school semesters and weekdays (Figures 1A and 1B). The temporal structure of sleep and stress variables was further evidenced by their autocorrelation estimates. Autocorrelations were generally small (|r|<0.2; Figure 1C) but were strongest (highlighted with asterisks) at a 7-day lag (and again at a 14-day lag) for sleep duration, consistent with a weekly sleep schedule. By contrast, autocorrelations for stress were strongest at 1- and 2-day lags, suggesting that experiences of stress might come in chains of >1 day.

The school break and week-related changes observed in the data informed the design of the iLM seeking to capture stable person-level associations. Given that the winter break presents different environmental demands from the school semester, we decided a priori to exclude from the model the observations collected during this period. In addition, to account for weekly patterns in sleep behavior (outcome variable) within the school semesters, we added the day of the week as a covariate in the model.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Daily observations</th>
<th>Sleep duration (min), mean (SD)</th>
<th>Sleep quality (1-5 Likert scale), mean (SD)</th>
<th>Stress (1-5 Likert scale), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T36</td>
<td>188</td>
<td>470.90 (86.02)</td>
<td>3.36 (0.80)</td>
<td>2.70 (1.10)</td>
</tr>
<tr>
<td>T37</td>
<td>180</td>
<td>423.49 (78.52)</td>
<td>3.41 (0.63)</td>
<td>2.21 (0.85)</td>
</tr>
<tr>
<td>T38</td>
<td>139</td>
<td>409.22 (132.79)</td>
<td>2.56 (0.85)</td>
<td>3.39 (0.82)</td>
</tr>
<tr>
<td>T39</td>
<td>114</td>
<td>414.25 (99.79)</td>
<td>2.74 (0.67)</td>
<td>2.25 (1.07)</td>
</tr>
<tr>
<td>T40</td>
<td>102</td>
<td>483.04 (53.83)</td>
<td>3.37 (0.53)</td>
<td>2.84 (1.24)</td>
</tr>
<tr>
<td>T41</td>
<td>136</td>
<td>429.99 (91.77)</td>
<td>2.34 (0.67)</td>
<td>3.54 (0.74)</td>
</tr>
<tr>
<td>T42</td>
<td>193</td>
<td>416.79 (95.16)</td>
<td>3.26 (0.80)</td>
<td>1.21 (0.46)</td>
</tr>
<tr>
<td>T43</td>
<td>166</td>
<td>464.57 (116.07)</td>
<td>3.04 (0.35)</td>
<td>2.23 (0.82)</td>
</tr>
<tr>
<td>T44</td>
<td>179</td>
<td>456.01 (67.63)</td>
<td>3.34 (0.74)</td>
<td>1.96 (1.03)</td>
</tr>
<tr>
<td>T45</td>
<td>197</td>
<td>474.08 (64.90)</td>
<td>2.99 (0.60)</td>
<td>1.18 (0.54)</td>
</tr>
<tr>
<td>T46</td>
<td>203</td>
<td>478.48 (70.13)</td>
<td>1.99 (0.58)</td>
<td>2.35 (1.01)</td>
</tr>
<tr>
<td>T47</td>
<td>178</td>
<td>382.32 (78.80)</td>
<td>3.54 (0.72)</td>
<td>1.79 (0.95)</td>
</tr>
<tr>
<td>T48</td>
<td>189</td>
<td>406.65 (103.77)</td>
<td>2.81 (1.00)</td>
<td>2.17 (1.36)</td>
</tr>
<tr>
<td>T49</td>
<td>171</td>
<td>444.64 (87.14)</td>
<td>2.91 (0.86)</td>
<td>1.90 (1.11)</td>
</tr>
</tbody>
</table>

aIDs starting with “P” indicate pilot sample participants.
bIDs starting with “T” indicate target sample participants.
iLMs Capture Day-to-Day Associations in the Pilot Participants

To test the viability of the iLM approach, a proof-of-concept model examined the association between objective sleep duration and subjective sleep quality in the pilot participants (Table 2 and Figure 2). This model allowed for a test of construct validity, given that the tested association was intuitive and expected. The model tested the association between sleep duration and sleep quality for the same sleep episode (ie, a concurrent association). All pilot participants showed the expected positive association: when participants slept for a shorter period than usual, they also subjectively rated these same sleep events as worse quality (shown in orange in Figure 2A). The effect size was large for all individuals, and the 95% UI lay outside of the ROPE for all models. These results provided preliminary evidence that the iLM is a valid framework to reliably detect relations between psychological variables (in this case, subjective sleep quality rating) and objective behaviors (sleep duration) at the individual participant level.

A second, exploratory iLM tested the association between sleep duration (on day \( t \)) and sleep quality the day before (on day \( t-1 \)). We expected this model to show significant but weaker effects compared to the first model, given that the variables were referring to lagged sleep events. Moreover, we expected negative associations such that worse sleep quality on one night would associate with longer sleep duration the next night, that is, a compensatory sleep rebound effect. Half of the participants (3/6, 50%; P1, P4, and P6) showed this expected pattern of association (pd>0.975; shown in blue in Figure 2). Of the 6 participants, 1 (17%; P2) showed a positive association such that worse subjective sleep quality experienced on the preceding night was associated with shorter sleep durations on the subsequent night. This participant might experience chains of poor sleep over multiple days (eg, reduced sleep days ahead of a deadline to accommodate increased workload), rather than a sleep rebound effect immediately the following night. In addition, of the 6 participants, 2 (33%; P3 and P5) showed a positive slope estimate but without reaching statistical significance (pd<0.975). We did not observe structured patterns of association between individual slope estimates and individuals’ mean sleep quality, mean sleep duration, or total number of daily observations (Figure 2B). This suggests that the estimated slopes were not systematically influenced by person-level characteristics of the psychological and behavioral phenomena of interest.
### Table 2. Pilot sample results of individual-level linear models assessing sleep duration associated with concurrent sleep quality and assessing sleep duration associated with sleep quality the day before (all models included the day of the week as a covariate, but the table only shows model parameters for the main predictor, sleep quality).

<table>
<thead>
<tr>
<th>Model and pilot participant</th>
<th>Slope (median of the posterior distribution)</th>
<th>95% UI(^a)</th>
<th>pd(^b)</th>
<th>ROPE(^c) (+/−)</th>
<th>% UI in ROPE</th>
<th>R-hat</th>
<th>ESS(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sleep duration regressed on concurrent sleep quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>65.31(^e)</td>
<td>37.73 to 92.42</td>
<td>1.00</td>
<td>15.84</td>
<td>0.00</td>
<td>1.00</td>
<td>7291</td>
</tr>
<tr>
<td>P2</td>
<td>47.48(^e)</td>
<td>31.42 to 63.87</td>
<td>1.00</td>
<td>12.35</td>
<td>0.00</td>
<td>1.00</td>
<td>8177</td>
</tr>
<tr>
<td>P3</td>
<td>71.10(^e)</td>
<td>54.30 to 88.46</td>
<td>1.00</td>
<td>12.50</td>
<td>0.00</td>
<td>1.00</td>
<td>7484</td>
</tr>
<tr>
<td>P4</td>
<td>56.32(^e)</td>
<td>42.31 to 70.24</td>
<td>1.00</td>
<td>12.00</td>
<td>0.00</td>
<td>1.00</td>
<td>7649</td>
</tr>
<tr>
<td>P5</td>
<td>40.22(^e)</td>
<td>30.65 to 49.95</td>
<td>1.00</td>
<td>7.89</td>
<td>0.00</td>
<td>1.00</td>
<td>7030</td>
</tr>
<tr>
<td>P6</td>
<td>84.28(^e)</td>
<td>69.58 to 99.15</td>
<td>1.00</td>
<td>12.50</td>
<td>0.00</td>
<td>1.00</td>
<td>6578</td>
</tr>
<tr>
<td><strong>Sleep duration regressed on sleep quality the day before</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>−43.28(^e)</td>
<td>−67.61 to −18.99</td>
<td>1.00</td>
<td>15.84</td>
<td>0.00</td>
<td>1.00</td>
<td>7726</td>
</tr>
<tr>
<td>P2</td>
<td>25.05(^e)</td>
<td>8.84 to 41.57</td>
<td>1.00</td>
<td>12.35</td>
<td>3.93</td>
<td>1.00</td>
<td>8244</td>
</tr>
<tr>
<td>P3</td>
<td>6.80</td>
<td>−14.01 to 27.05</td>
<td>0.74</td>
<td>12.50</td>
<td>71.04</td>
<td>1.00</td>
<td>7478</td>
</tr>
<tr>
<td>P4</td>
<td>−18.92(^e)</td>
<td>−33.78 to −3.38</td>
<td>0.99</td>
<td>12.00</td>
<td>17.20</td>
<td>1.00</td>
<td>7823</td>
</tr>
<tr>
<td>P5</td>
<td>10.07</td>
<td>−1.05 to 21.18</td>
<td>0.96</td>
<td>7.89</td>
<td>34.29</td>
<td>1.00</td>
<td>8030</td>
</tr>
<tr>
<td>P6</td>
<td>−31.39(^e)</td>
<td>−51.90 to −10.44</td>
<td>1.00</td>
<td>12.50</td>
<td>1.36</td>
<td>1.00</td>
<td>6911</td>
</tr>
</tbody>
</table>

\(^a\)UI: uncertainty interval.  
\(^b\)pd: probability of direction.  
\(^c\)ROPE: region of practical equivalence.  
\(^d\)ESS: effective sample size.  
\(^e\)Statistically significant result.
Figure 2. Within-individual linear models capture real-world associations between objective sleep duration and subjective sleep quality. (A) Each row (from participant 1 [P1] to participant 6 [P6]) shows model results for an individual pilot participant. Individual-level models testing sleep duration associated with sleep quality the day before are displayed in blue; models testing sleep duration associated with concurrent sleep quality are displayed in orange. Column 1 shows the models’ estimated slopes (computed as the median of the posterior distribution) and uncertainty metrics. Symbol shading signifies statistically significant slopes. Error bars show 95% uncertainty intervals (UIs). Shaded density plots show the full posterior distributions of the slopes; gray shaded areas show the regions of practical equivalence (ROPEs). Column 2 shows the models’ predicted sleep duration as a function of sleep quality; shading around the lines indicates 95% UIs. Column 3 shows the participant-level distributions of sleep duration (in min) and sleep quality (1-5 Likert scale) across all daily observations used in analysis. (B) Slope estimates from column 1 in subpart A are plotted against participant-level estimates across the study period. Dur.: duration; n.s.: not statistically significant; pd: probability of direction; sig.: statistically significant.

Model Diagnostics Confirm Convergence and Adequate Specification

Both visual and quantitative MCMC diagnostic checks revealed that all iLMs converged successfully (Figure 3). Trace plots (Figure 3, column 1) revealed no structured pattern in the estimated slopes across sampling iterations. R-hat values were <1.1, and ESSs were >1000 for the estimated slopes of all models (Table 2). Posterior predictive checks comparing the observed distribution of the outcome variable (sleep duration) to 100 randomly sampled simulated data sets from the posterior predictive distribution confirmed that a gaussian model specification captured the observed data well (Figure 3, column 2). Inspection of the model residuals against the model’s predicted values confirmed homoscedasticity, with no structured pattern in the plots (Figure 3, columns 3 and 4). Finally, residual autocorrelation was generally low (|r|<0.2) for all models (Figure 3, column 5), suggesting that the model specification was able to account for the temporal structure in the data.
Figure 3. Within-individual model diagnostics suggest adequate convergence and specification. Each row (from participant 1 [P1] to participant 6 [P6]) shows model diagnostics for an individual pilot participant. Diagnostics for the models assessing the association between sleep duration and sleep quality the day before are displayed in blue; models assessing the association between sleep duration and concurrent sleep quality are displayed in orange. Column 1’s trace plots display time series of each Markov chain’s estimated slope (y-axis) as a function of postwarmup iterations (x-axis). Column 2 shows posterior predictive checks; black density lines show the observed distributions of the outcome (sleep duration), and the thin colored density lines show 100 replicated outcome distributions generated based on random samples from the models’ parameters’ posterior distributions. Columns 3 and 4 show the models’ predicted values (x-axis) against the models’ residuals (y-axis). Column 5 shows autocorrelation plots of model residuals.

iLMs Yield Similar, but Not Identical, Estimates to a Group-Based Multilevel Model

The slopes estimated from the iLM are similar to those obtained through a MLM testing the same associations between concurrent objective sleep duration and subjective sleep quality (Figure 4). This suggests that the individually tailored slopes obtained from our iLM are comparable to more traditional group-based approaches. Of note, although the estimates are similar, they are not identical. Comparison of the 2 models suggested attenuation of the individual-level estimates in the group models (refer to the part of Figure 4 highlighted with an asterisk). Even with random effects by participant, in the MLM, these estimates are, by design, biased by the group and may underestimate individual-level effects, especially for uncommon phenotypes [20]. As will be revealed later in the results from the larger sample of target participants, group-based estimation can even result in the reversal of the sign of the association for some individuals.
The iLM Is a Valid Framework to Identify Individual-Level Associations

Proof-of-concept analyses of the pilot data demonstrated that the iLM successfully captures expected associations between objective behavior and psychological variables, has an adequate model specification, and captures individual-level estimates unbiased by the group. Moreover, point estimates obtained with our Bayesian iLMs are nearly identical to those obtained through iLMs fitted with a frequentist inference framework (refer to Figure S2 in Multimedia Appendix 1), suggesting that our individual-level modeling approach is robust across both major statistical inference frameworks. In sum, the iLM is a valid framework to identify individual-level associations, which justified carrying over the model to test target hypotheses regarding the association between perceived stress and objective sleep duration in the independent target data set of first-year students (n=49).

Sleep and Stress Fluctuate With the Academic Calendar in the Target Participants

As with the pilot participants, during winter break and weekends, participants had longer objective sleep duration, better subjective sleep quality, and felt less stressed compared to the fall and spring semesters (columns 1 and 2 in Figure S3 in Multimedia Appendix 1). We replicated the autocorrelation pattern seen in the pilot data set: sleep duration showed the greatest autocorrelation at a 7-day lag (and then again at a 14-day lag), indicating that sleep patterns are tied to a weekly schedule, while stress showed the greatest autocorrelation at 1-day and 2-day lags, indicating that experiences of stress might come in chains of a few days (column 3 in Figure S3 in Multimedia Appendix 1). These school break– and week-related dynamics confirm that our target data set captured real-life dynamics associated with college life and reinforce our individual-level modeling decisions regarding the exclusion of winter break data and the addition of the day of the week as a covariate.

Shorter Objective Sleep Duration Is Associated With Worse Subjective Sleep Quality in the Target Participants

All participants showed a positive slope estimate for the association between concurrent sleep duration and sleep quality, and this effect was statistically significant in 46 (94%) of the 49 target participants (pd>0.975; Table 3 and Figure 5A, in orange). In other words, for only 10% of participants, sleep rated as worse quality was consistently followed by longer sleep the following night, suggesting that sleep rebound effects are perhaps less common or reliable than anticipated. Among these participants, a 1-point decrease in a 5-point sleep quality scale was associated with a subsequent sleep episode that was also longer by a median 59 (range 16-119) minutes, a substantial increase in sleep duration for most participants considering that the average sleep duration in the sample was 432 minutes (7.2 h; SD 34 min).

For the exploratory lagged model, the expected negative association between sleep duration and sleep quality the day before reached statistical significance in only 5 (10%) of the 49 participants (pd>0.975; Table 4 and Figure 5A, in blue). In other words, for only 10% of participants, sleep rated as worse quality was consistently followed by longer sleep the following night, suggesting that sleep rebound effects are perhaps less common or reliable than anticipated. Among these participants, a 1-point decrease in a 5-point sleep quality scale was associated with a subsequent sleep episode that was also longer by a median 18 (range 15-22) minutes across participants, a much smaller effect compared to the concurrent model. We did not observe structured patterns of association between individual slope estimates and individuals’ mean sleep quality, mean sleep duration, or total number of daily observations, suggesting that the model results were not systematically influenced by person-level aggregates of the variables that went into the model (Figure 5B).
Table 3. Target sample results of individual-level linear models assessing sleep duration associated with concurrent sleep quality (all individual-level models included the day of the week as a covariate, but the table only shows model parameters for the main predictor, sleep quality).

<table>
<thead>
<tr>
<th>Target participant</th>
<th>Slope (median of the posterior distribution)</th>
<th>95% UI^a</th>
<th>pd^b</th>
<th>ROPE^c (+/-)</th>
<th>% UI in ROPE</th>
<th>R-hat</th>
<th>ESS^d</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>58.57^e</td>
<td>43.90 to 72.99</td>
<td>1.00</td>
<td>10.30</td>
<td>0.00</td>
<td>1.00</td>
<td>7523</td>
</tr>
<tr>
<td>T2</td>
<td>78.35^e</td>
<td>62.80 to 93.58</td>
<td>1.00</td>
<td>13.36</td>
<td>0.00</td>
<td>1.00</td>
<td>8984</td>
</tr>
<tr>
<td>T3</td>
<td>83.88^e</td>
<td>60.42 to 107.40</td>
<td>1.00</td>
<td>16.12</td>
<td>0.00</td>
<td>1.00</td>
<td>7211</td>
</tr>
<tr>
<td>T4</td>
<td>68.08^e</td>
<td>51.48 to 84.48</td>
<td>1.00</td>
<td>10.74</td>
<td>0.00</td>
<td>1.00</td>
<td>7738</td>
</tr>
<tr>
<td>T5</td>
<td>59.57^e</td>
<td>48.38 to 70.46</td>
<td>1.00</td>
<td>9.77</td>
<td>0.00</td>
<td>1.00</td>
<td>8168</td>
</tr>
<tr>
<td>T6</td>
<td>43.51^e</td>
<td>27.26 to 59.18</td>
<td>1.00</td>
<td>12.22</td>
<td>0.00</td>
<td>1.00</td>
<td>8029</td>
</tr>
<tr>
<td>T7</td>
<td>16.10^e</td>
<td>5.49 to 26.60</td>
<td>1.00</td>
<td>6.26</td>
<td>0.82</td>
<td>1.00</td>
<td>7772</td>
</tr>
<tr>
<td>T8</td>
<td>84.78^e</td>
<td>53.43 to 115.84</td>
<td>1.00</td>
<td>22.05</td>
<td>0.00</td>
<td>1.00</td>
<td>8598</td>
</tr>
<tr>
<td>T9</td>
<td>115.91^e</td>
<td>86.49 to 144.67</td>
<td>1.00</td>
<td>18.39</td>
<td>0.00</td>
<td>1.00</td>
<td>6519</td>
</tr>
<tr>
<td>T10</td>
<td>89.47^e</td>
<td>71.20 to 108.46</td>
<td>1.00</td>
<td>15.77</td>
<td>0.00</td>
<td>1.00</td>
<td>7813</td>
</tr>
<tr>
<td>T11</td>
<td>73.07^e</td>
<td>57.17 to 89.42</td>
<td>1.00</td>
<td>12.98</td>
<td>0.00</td>
<td>1.00</td>
<td>7347</td>
</tr>
<tr>
<td>T12</td>
<td>27.69^e</td>
<td>12.94 to 43.04</td>
<td>1.00</td>
<td>7.56</td>
<td>0.00</td>
<td>1.00</td>
<td>7765</td>
</tr>
<tr>
<td>T13</td>
<td>45.99^e</td>
<td>30.31 to 61.92</td>
<td>1.00</td>
<td>13.01</td>
<td>0.00</td>
<td>1.00</td>
<td>7855</td>
</tr>
<tr>
<td>T14</td>
<td>119.30^e</td>
<td>93.16 to 145.87</td>
<td>1.00</td>
<td>16.15</td>
<td>0.00</td>
<td>1.00</td>
<td>6735</td>
</tr>
<tr>
<td>T15</td>
<td>56.74^e</td>
<td>40.07 to 72.96</td>
<td>1.00</td>
<td>9.77</td>
<td>0.00</td>
<td>1.00</td>
<td>8048</td>
</tr>
<tr>
<td>T16</td>
<td>107.86^e</td>
<td>81.17 to 133.61</td>
<td>1.00</td>
<td>19.39</td>
<td>0.00</td>
<td>1.00</td>
<td>8373</td>
</tr>
<tr>
<td>T17</td>
<td>57.72^e</td>
<td>37.47 to 77.20</td>
<td>1.00</td>
<td>14.88</td>
<td>0.00</td>
<td>1.00</td>
<td>8767</td>
</tr>
<tr>
<td>T18</td>
<td>66.72^e</td>
<td>40.75 to 91.96</td>
<td>1.00</td>
<td>14.01</td>
<td>0.00</td>
<td>1.00</td>
<td>7313</td>
</tr>
<tr>
<td>T19</td>
<td>94.56^e</td>
<td>74.90 to 113.90</td>
<td>1.00</td>
<td>16.98</td>
<td>0.00</td>
<td>1.00</td>
<td>7965</td>
</tr>
<tr>
<td>T20</td>
<td>62.54^e</td>
<td>45.72 to 79.42</td>
<td>1.00</td>
<td>13.31</td>
<td>0.00</td>
<td>1.00</td>
<td>9058</td>
</tr>
<tr>
<td>T21</td>
<td>44.44^e</td>
<td>30.24 to 58.59</td>
<td>1.00</td>
<td>10.88</td>
<td>0.00</td>
<td>1.00</td>
<td>8560</td>
</tr>
<tr>
<td>T22</td>
<td>57.24^e</td>
<td>43.62 to 70.66</td>
<td>1.00</td>
<td>9.38</td>
<td>0.00</td>
<td>1.00</td>
<td>7436</td>
</tr>
<tr>
<td>T23</td>
<td>45.24^e</td>
<td>20.23 to 71.17</td>
<td>1.00</td>
<td>11.94</td>
<td>0.00</td>
<td>1.00</td>
<td>7503</td>
</tr>
<tr>
<td>T24</td>
<td>48.20^e</td>
<td>34.92 to 61.39</td>
<td>1.00</td>
<td>9.34</td>
<td>0.00</td>
<td>1.00</td>
<td>8482</td>
</tr>
<tr>
<td>T25</td>
<td>47.93^e</td>
<td>32.46 to 63.25</td>
<td>1.00</td>
<td>12.21</td>
<td>0.00</td>
<td>1.00</td>
<td>8742</td>
</tr>
<tr>
<td>T26</td>
<td>22.51</td>
<td>–66.79 to 112.05</td>
<td>0.70</td>
<td>64.98</td>
<td>84.15</td>
<td>1.00</td>
<td>9231</td>
</tr>
<tr>
<td>T27</td>
<td>76.92^e</td>
<td>59.80 to 94.43</td>
<td>1.00</td>
<td>13.93</td>
<td>0.00</td>
<td>1.00</td>
<td>6829</td>
</tr>
<tr>
<td>T28</td>
<td>35.80^e</td>
<td>21.59 to 49.73</td>
<td>1.00</td>
<td>11.07</td>
<td>0.00</td>
<td>1.00</td>
<td>9366</td>
</tr>
<tr>
<td>T29</td>
<td>27.47^e</td>
<td>5.63 to 49.43</td>
<td>0.99</td>
<td>15.80</td>
<td>13.09</td>
<td>1.00</td>
<td>9333</td>
</tr>
<tr>
<td>T30</td>
<td>51.20^e</td>
<td>37.94 to 64.63</td>
<td>1.00</td>
<td>10.68</td>
<td>0.00</td>
<td>1.00</td>
<td>10463</td>
</tr>
<tr>
<td>T31</td>
<td>61.13^e</td>
<td>48.35 to 74.01</td>
<td>1.00</td>
<td>11.39</td>
<td>0.00</td>
<td>1.00</td>
<td>8859</td>
</tr>
<tr>
<td>T32</td>
<td>84.47^e</td>
<td>63.97 to 104.78</td>
<td>1.00</td>
<td>13.82</td>
<td>0.00</td>
<td>1.00</td>
<td>8314</td>
</tr>
<tr>
<td>T33</td>
<td>72.26^e</td>
<td>48.83 to 95.74</td>
<td>1.00</td>
<td>15.13</td>
<td>0.00</td>
<td>1.00</td>
<td>7530</td>
</tr>
<tr>
<td>T34</td>
<td>59.12^e</td>
<td>46.46 to 72.07</td>
<td>1.00</td>
<td>11.04</td>
<td>0.00</td>
<td>1.00</td>
<td>8186</td>
</tr>
<tr>
<td>T35</td>
<td>50.34^e</td>
<td>39.48 to 61.15</td>
<td>1.00</td>
<td>9.11</td>
<td>0.00</td>
<td>1.00</td>
<td>9152</td>
</tr>
</tbody>
</table>
TABLE 1. Estimated slopes (corresponding to the median of the posterior distribution) and their associated uncertainty intervals for each participant.

<table>
<thead>
<tr>
<th>Target participant</th>
<th>Slope (median of the posterior distribution)</th>
<th>95% UIa</th>
<th>pdb</th>
<th>ROPEc (±/−)</th>
<th>% UI in ROPE</th>
<th>R-hat</th>
<th>ESSd</th>
</tr>
</thead>
<tbody>
<tr>
<td>T36</td>
<td>45.89e</td>
<td>31.88 to 60.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T37</td>
<td>79.76e</td>
<td>65.18 to 94.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T38</td>
<td>96.75e</td>
<td>76.15 to 117.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T39</td>
<td>97.12e</td>
<td>74.12 to 119.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T40</td>
<td>21.66e</td>
<td>1.58 to 42.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T41</td>
<td>83.39e</td>
<td>64.82 to 102.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T42</td>
<td>67.41e</td>
<td>53.03 to 81.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T43</td>
<td>40.65</td>
<td>−11.52 to 90.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T44</td>
<td>48.21e</td>
<td>36.79 to 59.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T45</td>
<td>42.05e</td>
<td>27.71 to 56.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T46</td>
<td>11.97</td>
<td>−4.23 to 28.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T47</td>
<td>56.68e</td>
<td>42.98 to 70.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T48</td>
<td>58.27e</td>
<td>46.41 to 70.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T49</td>
<td>52.76e</td>
<td>40.08 to 65.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aUI: uncertainty interval.
bpd: probability of direction.
cROPE: region of practical equivalence.
dESS: effective sample size.
eStatistically significant result.

Figure 5. Longer objective sleep duration associates with higher subjective sleep quality in most individuals. (A) Estimated slopes (corresponding to the median of the posterior distribution) from the individual-level models that assess sleep duration associated with sleep quality the day before (triangles) and with concurrent sleep quality (circles) are plotted along the y-axis ordered by participant (x-axis). Statistically significant slope estimates are shaded blue (day before) or orange (concurrent). Error bars show 95% uncertainty intervals, and shaded density plots show the full posterior distributions of the slopes. (B) Slope estimates from subpart A are plotted against participant-level estimates across the study period.
Table 4. Target sample results of individual-level linear models assessing sleep duration associated with sleep quality the day before (all individual-level models included the day of the week as a covariate, but the table only shows model parameters for the main predictor, sleep quality).

<table>
<thead>
<tr>
<th>Target participant</th>
<th>Slope (median of the posterior distribution)</th>
<th>95% UI&lt;sup&gt;a&lt;/sup&gt;</th>
<th>pd&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ROPE&lt;sup&gt;c (+−)&lt;/sup&gt;</th>
<th>% UI in ROPE</th>
<th>R-hat</th>
<th>ESS&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>−6.18</td>
<td>−20.18 to 7.90</td>
<td>0.81</td>
<td>10.30</td>
<td>72.79</td>
<td>1.00</td>
<td>7331</td>
</tr>
<tr>
<td>T2</td>
<td>17.18</td>
<td>−1.23 to 35.01</td>
<td>0.97</td>
<td>13.36</td>
<td>33.68</td>
<td>1.00</td>
<td>7898</td>
</tr>
<tr>
<td>T3</td>
<td>4.86</td>
<td>−24.38 to 33.47</td>
<td>0.62</td>
<td>16.12</td>
<td>73.18</td>
<td>1.00</td>
<td>7339</td>
</tr>
<tr>
<td>T4</td>
<td>−8.14</td>
<td>−26.68 to 10.37</td>
<td>0.81</td>
<td>10.74</td>
<td>61.51</td>
<td>1.00</td>
<td>7692</td>
</tr>
<tr>
<td>T5</td>
<td>−11.59</td>
<td>−25.67 to 2.53</td>
<td>0.95</td>
<td>9.77</td>
<td>38.92</td>
<td>1.00</td>
<td>8057</td>
</tr>
<tr>
<td>T6</td>
<td>3.99</td>
<td>−12.52 to 20.73</td>
<td>0.68</td>
<td>12.22</td>
<td>84.16</td>
<td>1.00</td>
<td>8124</td>
</tr>
<tr>
<td>T7</td>
<td>1.79</td>
<td>−9.04 to 12.40</td>
<td>0.63</td>
<td>6.26</td>
<td>76.76</td>
<td>1.00</td>
<td>7715</td>
</tr>
<tr>
<td>T8</td>
<td>−6.92</td>
<td>−44.90 to 31.06</td>
<td>0.64</td>
<td>22.05</td>
<td>75.05</td>
<td>1.00</td>
<td>8294</td>
</tr>
<tr>
<td>T9</td>
<td>−6.64</td>
<td>−43.52 to 30.76</td>
<td>0.63</td>
<td>18.39</td>
<td>68.00</td>
<td>1.00</td>
<td>6542</td>
</tr>
<tr>
<td>T10</td>
<td>5.38</td>
<td>−16.89 to 29.41</td>
<td>0.68</td>
<td>15.77</td>
<td>82.23</td>
<td>1.00</td>
<td>8276</td>
</tr>
<tr>
<td>T11</td>
<td>11.88</td>
<td>−8.52 to 32.29</td>
<td>0.88</td>
<td>12.98</td>
<td>54.38</td>
<td>1.00</td>
<td>7592</td>
</tr>
<tr>
<td>T12</td>
<td>−15.31&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−29.24 to −1.46</td>
<td>0.98</td>
<td>7.56</td>
<td>11.92</td>
<td>1.00</td>
<td>6981</td>
</tr>
<tr>
<td>T13</td>
<td>1.30</td>
<td>−15.99 to 18.29</td>
<td>0.56</td>
<td>13.01</td>
<td>90.63</td>
<td>1.00</td>
<td>8803</td>
</tr>
<tr>
<td>T14</td>
<td>−0.44</td>
<td>−30.03 to 29.90</td>
<td>0.51</td>
<td>16.15</td>
<td>76.59</td>
<td>1.00</td>
<td>7033</td>
</tr>
<tr>
<td>T15</td>
<td>6.21</td>
<td>−13.53 to 25.51</td>
<td>0.73</td>
<td>9.77</td>
<td>60.83</td>
<td>1.00</td>
<td>7607</td>
</tr>
<tr>
<td>T16</td>
<td>−27.16</td>
<td>−56.68 to 2.91</td>
<td>0.96</td>
<td>19.39</td>
<td>29.00</td>
<td>1.00</td>
<td>7630</td>
</tr>
<tr>
<td>T17</td>
<td>−11.26</td>
<td>−30.90 to 8.36</td>
<td>0.87</td>
<td>14.88</td>
<td>64.72</td>
<td>1.00</td>
<td>8548</td>
</tr>
<tr>
<td>T18</td>
<td>−5.27</td>
<td>−29.15 to 18.45</td>
<td>0.67</td>
<td>14.01</td>
<td>74.54</td>
<td>1.00</td>
<td>7075</td>
</tr>
<tr>
<td>T19</td>
<td>2.84</td>
<td>−20.85 to 25.94</td>
<td>0.59</td>
<td>16.98</td>
<td>87.60</td>
<td>1.00</td>
<td>7754</td>
</tr>
<tr>
<td>T20</td>
<td>−15.15</td>
<td>−34.71 to 4.14</td>
<td>0.94</td>
<td>13.31</td>
<td>42.17</td>
<td>1.00</td>
<td>7930</td>
</tr>
<tr>
<td>T21</td>
<td>−13.72</td>
<td>−28.15 to 0.59</td>
<td>0.97</td>
<td>10.88</td>
<td>34.11</td>
<td>1.00</td>
<td>8292</td>
</tr>
<tr>
<td>T22</td>
<td>−14.97</td>
<td>−31.39 to 2.04</td>
<td>0.96</td>
<td>9.38</td>
<td>24.14</td>
<td>1.00</td>
<td>7859</td>
</tr>
<tr>
<td>T23</td>
<td>7.89</td>
<td>−17.17 to 32.62</td>
<td>0.73</td>
<td>11.94</td>
<td>59.32</td>
<td>1.00</td>
<td>7688</td>
</tr>
<tr>
<td>T24</td>
<td>−17.72&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−31.83 to −3.24</td>
<td>0.99</td>
<td>9.34</td>
<td>10.94</td>
<td>1.00</td>
<td>8332</td>
</tr>
<tr>
<td>T25</td>
<td>7.69</td>
<td>−9.26 to 24.56</td>
<td>0.82</td>
<td>12.21</td>
<td>71.23</td>
<td>1.00</td>
<td>7155</td>
</tr>
<tr>
<td>T26</td>
<td>−7.54</td>
<td>−80.37 to 64.79</td>
<td>0.58</td>
<td>64.98</td>
<td>96.56</td>
<td>1.00</td>
<td>8465</td>
</tr>
<tr>
<td>T27</td>
<td>−7.78</td>
<td>−27.82 to 12.37</td>
<td>0.78</td>
<td>13.93</td>
<td>74.72</td>
<td>1.00</td>
<td>8543</td>
</tr>
<tr>
<td>T28</td>
<td>−22.29&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−36.43 to −8.12</td>
<td>1.00</td>
<td>11.07</td>
<td>3.94</td>
<td>1.00</td>
<td>7913</td>
</tr>
<tr>
<td>T29</td>
<td>−15.76</td>
<td>−37.87 to 6.16</td>
<td>0.92</td>
<td>15.80</td>
<td>50.18</td>
<td>1.00</td>
<td>7036</td>
</tr>
<tr>
<td>T30</td>
<td>10.45</td>
<td>−4.82 to 25.99</td>
<td>0.91</td>
<td>10.68</td>
<td>51.31</td>
<td>1.00</td>
<td>6695</td>
</tr>
<tr>
<td>T31</td>
<td>2.76</td>
<td>−12.58 to 18.24</td>
<td>0.64</td>
<td>11.39</td>
<td>86.79</td>
<td>1.00</td>
<td>6494</td>
</tr>
<tr>
<td>T32</td>
<td>−2.39</td>
<td>−27.37 to 23.67</td>
<td>0.57</td>
<td>13.82</td>
<td>74.03</td>
<td>1.00</td>
<td>7438</td>
</tr>
<tr>
<td>T33</td>
<td>−1.18</td>
<td>−27.00 to 24.74</td>
<td>0.54</td>
<td>15.13</td>
<td>79.09</td>
<td>1.00</td>
<td>6429</td>
</tr>
<tr>
<td>T34</td>
<td>7.50</td>
<td>−7.80 to 23.05</td>
<td>0.83</td>
<td>11.04</td>
<td>68.41</td>
<td>1.00</td>
<td>8185</td>
</tr>
<tr>
<td>T35</td>
<td>2.35</td>
<td>−11.15 to 15.76</td>
<td>0.63</td>
<td>9.11</td>
<td>83.18</td>
<td>1.00</td>
<td>7304</td>
</tr>
<tr>
<td>T36</td>
<td>−7.26</td>
<td>−22.38 to 7.85</td>
<td>0.83</td>
<td>10.69</td>
<td>67.71</td>
<td>1.00</td>
<td>7446</td>
</tr>
<tr>
<td>T37</td>
<td>−8.99</td>
<td>−27.43 to 9.56</td>
<td>0.83</td>
<td>12.42</td>
<td>65.23</td>
<td>1.00</td>
<td>7425</td>
</tr>
<tr>
<td>T38</td>
<td>16.81</td>
<td>−7.39 to 40.61</td>
<td>0.92</td>
<td>15.58</td>
<td>45.69</td>
<td>1.00</td>
<td>7448</td>
</tr>
<tr>
<td>T39</td>
<td>14.22</td>
<td>−14.82 to 42.98</td>
<td>0.84</td>
<td>14.99</td>
<td>52.11</td>
<td>1.00</td>
<td>7901</td>
</tr>
<tr>
<td>T40</td>
<td>−3.62</td>
<td>−25.53 to 18.36</td>
<td>0.63</td>
<td>10.25</td>
<td>64.99</td>
<td>1.00</td>
<td>7758</td>
</tr>
</tbody>
</table>
Higher Subjective Stress Is Associated With Shorter Objective Sleep Duration in Most Target Participants, but the Direction of the Temporal Association Varies

The slope estimate for the association between stress and sleep duration was negative in 86% of all iLMs fitted such that an increase in stress was associated with a decrease in sleep duration (Tables 5 and 6; Figure 6A). This association between stress and sleep duration reached statistical significance in 19 (39%) of the 49 participants (pd>0.975); of these 19 participants, 18 (95%) showed a negative association. Of note, the only participant who showed a significant positive association (T15, for the association between sleep duration and stress the day after) also had no usable actigraphy data over the full spring semester due to a technical issue with their wristband, and their estimated positive slope should therefore be interpreted with caution considering the structured missingness in their data. No other participant in the target sample showed this kind of systematic missingness in either the actigraphy or survey data streams (refer to Figure S1 in Multimedia Appendix 1).

Of the 19 participants who showed a statistically significant relationship between stress and sleep duration, 8 (42%) showed only the stress-then-sleep phenotype, that is, days with shorter sleep durations were preceded by higher stress the day before (Figure 6, in green) but not vice versa; 5 (26%) showed only the sleep-then-stress phenotype, that is, nights with shorter sleep duration were followed by increased stress the day after (Figure 6, in purple); and 6 (32%) showed bidirectional effects such that nights with shorter sleep duration were preceded by increased stress the day before as well as followed by increased stress the day after. Among participants who showed a significant association between today’s stress levels and sleep duration later that night (14/49, 29%), a 1-point increase in a 5-point perceived stress scale was associated with shorter subsequent sleep duration of a median 17 (range 11-33) minutes across participants. Among participants who showed a significant association between today’s stress levels and last night’s sleep duration (11/49, 22%), a 1-point increase in a 5-point perceived stress scale was associated with shorter previous sleep duration of a median 18 (range 10-38) minutes across participants. These effects would compound to more substantial reductions in sleep duration with greater increases in daily stress.

Individual slope estimates of the association between sleep duration and stress showed no clear pattern of association with individuals’ mean sleep duration or the number of daily observations that went into the model (see Figure 6B for mean sleep duration and number of daily observations). Individuals with higher mean stress tended to have a larger absolute slope estimate (see Figure 6B for mean stress), perhaps because participants with very low stress levels have little variance to be modeled.

<table>
<thead>
<tr>
<th>Target participant</th>
<th>Slope (median of the posterior distribution)</th>
<th>95% UIa</th>
<th>pb</th>
<th>ROPEc (±/−)</th>
<th>% UI in ROPE</th>
<th>R-hat</th>
<th>ESSd</th>
</tr>
</thead>
<tbody>
<tr>
<td>T41</td>
<td>−19.81</td>
<td>−43.76 to 4.64</td>
<td>0.94</td>
<td>13.72</td>
<td>30.36</td>
<td>1.00</td>
<td>8205</td>
</tr>
<tr>
<td>T42</td>
<td>4.94</td>
<td>−12.29 to 21.91</td>
<td>0.72</td>
<td>11.89</td>
<td>80.34</td>
<td>1.00</td>
<td>6956</td>
</tr>
<tr>
<td>T43</td>
<td>6.17</td>
<td>−42.60 to 56.27</td>
<td>0.60</td>
<td>33.52</td>
<td>84.63</td>
<td>1.00</td>
<td>7107</td>
</tr>
<tr>
<td>T44</td>
<td>4.48</td>
<td>−7.61 to 16.09</td>
<td>0.76</td>
<td>9.10</td>
<td>78.73</td>
<td>1.00</td>
<td>7931</td>
</tr>
<tr>
<td>T45</td>
<td>−10.23</td>
<td>−23.89 to 3.28</td>
<td>0.93</td>
<td>10.78</td>
<td>53.33</td>
<td>1.00</td>
<td>8686</td>
</tr>
<tr>
<td>T46</td>
<td>10.73</td>
<td>−4.71 to 26.80</td>
<td>0.91</td>
<td>12.00</td>
<td>56.45</td>
<td>1.00</td>
<td>8086</td>
</tr>
<tr>
<td>T47</td>
<td>−20.67c</td>
<td>−35.59 to −5.83</td>
<td>1.00</td>
<td>10.92</td>
<td>8.02</td>
<td>1.00</td>
<td>9068</td>
</tr>
<tr>
<td>T48</td>
<td>−17.71c</td>
<td>−32.53 to −3.60</td>
<td>0.99</td>
<td>10.36</td>
<td>13.20</td>
<td>1.00</td>
<td>8895</td>
</tr>
<tr>
<td>T49</td>
<td>4.96</td>
<td>−9.25 to 18.87</td>
<td>0.76</td>
<td>10.18</td>
<td>78.56</td>
<td>1.00</td>
<td>8084</td>
</tr>
</tbody>
</table>

aUI: uncertainty interval.
bpd: probability of direction.
cROPE: region of practical equivalence.
dESS: effective sample size.
eStatistically significant result.

https://formative.jmir.org/2024/1/e53441 JMIR Form Res 2024 | vol. 8 | e53441 | p.2004 (page number not for citation purposes)
Table 5. Target sample results of individual-level linear models assessing sleep duration associated with stress the day before (all individual-level models included the day of the week as a covariate, but the table only shows model parameters for the main predictor, stress the day before).

<table>
<thead>
<tr>
<th>Target participant</th>
<th>Slope (median of the posterior distribution)</th>
<th>95% UI</th>
<th>pd</th>
<th>ROPE (+/-)</th>
<th>% UI in ROPE</th>
<th>R-hat</th>
<th>ESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>−17.39&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−30.84 to −4.23</td>
<td>0.99</td>
<td>0.90</td>
<td>9.64</td>
<td>1.00</td>
<td>7920</td>
</tr>
<tr>
<td>T2</td>
<td>−25.47&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−38.71 to −12.52</td>
<td>1.00</td>
<td>1.00</td>
<td>9.44</td>
<td>1.00</td>
<td>7532</td>
</tr>
<tr>
<td>T3</td>
<td>−11.75</td>
<td>−36.19 to 12.93</td>
<td>0.83</td>
<td>0.83</td>
<td>56.25</td>
<td>1.00</td>
<td>7497</td>
</tr>
<tr>
<td>T4</td>
<td>−17.88&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−31.13 to −4.48</td>
<td>0.99</td>
<td>0.99</td>
<td>5.45</td>
<td>1.00</td>
<td>7531</td>
</tr>
<tr>
<td>T5</td>
<td>−6.59</td>
<td>−17.46 to 4.17</td>
<td>0.89</td>
<td>0.89</td>
<td>55.09</td>
<td>1.00</td>
<td>8189</td>
</tr>
<tr>
<td>T6</td>
<td>−12.27</td>
<td>−26.09 to 1.72</td>
<td>0.96</td>
<td>0.96</td>
<td>36.82</td>
<td>1.00</td>
<td>8310</td>
</tr>
<tr>
<td>T7</td>
<td>0.37</td>
<td>−7.96 to 8.40</td>
<td>0.53</td>
<td>0.53</td>
<td>78.46</td>
<td>1.00</td>
<td>7568</td>
</tr>
<tr>
<td>T8</td>
<td>−3.97</td>
<td>−31.88 to 24.48</td>
<td>0.60</td>
<td>0.60</td>
<td>78.29</td>
<td>1.00</td>
<td>9005</td>
</tr>
<tr>
<td>T9</td>
<td>−12.75</td>
<td>−39.38 to 13.84</td>
<td>0.83</td>
<td>0.83</td>
<td>56.53</td>
<td>1.00</td>
<td>6773</td>
</tr>
<tr>
<td>T10</td>
<td>−16.42&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−29.99 to −2.96</td>
<td>0.99</td>
<td>0.99</td>
<td>16.55</td>
<td>1.00</td>
<td>8429</td>
</tr>
<tr>
<td>T11</td>
<td>−16.79</td>
<td>−34.71 to 1.64</td>
<td>0.96</td>
<td>0.96</td>
<td>29.51</td>
<td>1.00</td>
<td>8051</td>
</tr>
<tr>
<td>T12</td>
<td>−4.08</td>
<td>−16.42 to 8.22</td>
<td>0.74</td>
<td>0.74</td>
<td>65.81</td>
<td>1.00</td>
<td>7427</td>
</tr>
<tr>
<td>T13</td>
<td>−9.18</td>
<td>−22.63 to 4.30</td>
<td>0.91</td>
<td>0.91</td>
<td>57.36</td>
<td>1.00</td>
<td>8887</td>
</tr>
<tr>
<td>T14</td>
<td>−10.01</td>
<td>−34.82 to 15.03</td>
<td>0.79</td>
<td>0.79</td>
<td>58.66</td>
<td>1.00</td>
<td>6569</td>
</tr>
<tr>
<td>T15</td>
<td>6.40</td>
<td>−8.58 to 21.04</td>
<td>0.80</td>
<td>0.80</td>
<td>54.53</td>
<td>1.00</td>
<td>8457</td>
</tr>
<tr>
<td>T16</td>
<td>−4.81</td>
<td>−28.89 to 19.49</td>
<td>0.65</td>
<td>0.65</td>
<td>78.60</td>
<td>1.00</td>
<td>7154</td>
</tr>
<tr>
<td>T17</td>
<td>0.47</td>
<td>−15.28 to 15.70</td>
<td>0.52</td>
<td>0.52</td>
<td>88.72</td>
<td>1.00</td>
<td>7480</td>
</tr>
<tr>
<td>T18</td>
<td>2.55</td>
<td>−17.09 to 22.65</td>
<td>0.60</td>
<td>0.60</td>
<td>78.35</td>
<td>1.00</td>
<td>8255</td>
</tr>
<tr>
<td>T19</td>
<td>−33.47&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−50.42 to −16.44</td>
<td>1.00</td>
<td>1.00</td>
<td>78.85</td>
<td>1.00</td>
<td>7885</td>
</tr>
<tr>
<td>T20</td>
<td>0.66</td>
<td>−12.06 to 13.07</td>
<td>0.55</td>
<td>0.55</td>
<td>84.76</td>
<td>1.00</td>
<td>7970</td>
</tr>
<tr>
<td>T21</td>
<td>−16.43&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−27.28 to −5.51</td>
<td>1.00</td>
<td>1.00</td>
<td>3.93</td>
<td>1.00</td>
<td>8275</td>
</tr>
<tr>
<td>T22</td>
<td>−8.13</td>
<td>−22.65 to 6.56</td>
<td>0.86</td>
<td>0.86</td>
<td>49.73</td>
<td>1.00</td>
<td>8495</td>
</tr>
<tr>
<td>T23</td>
<td>−16.85</td>
<td>−36.46 to 2.27</td>
<td>0.96</td>
<td>0.96</td>
<td>19.97</td>
<td>1.00</td>
<td>6432</td>
</tr>
<tr>
<td>T24</td>
<td>−20.04&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−35.95 to −4.06</td>
<td>0.99</td>
<td>0.99</td>
<td>11.06</td>
<td>1.00</td>
<td>8168</td>
</tr>
<tr>
<td>T25</td>
<td>−25.95&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−39.11 to −13.25</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>7895</td>
</tr>
<tr>
<td>T26</td>
<td>−7.07</td>
<td>−14.59 to 0.61</td>
<td>0.97</td>
<td>0.97</td>
<td>34.56</td>
<td>1.00</td>
<td>8812</td>
</tr>
<tr>
<td>T27</td>
<td>−7.98</td>
<td>−28.47 to 12.74</td>
<td>0.78</td>
<td>0.78</td>
<td>76.27</td>
<td>1.00</td>
<td>8484</td>
</tr>
<tr>
<td>T28</td>
<td>0.77</td>
<td>−8.97 to 10.58</td>
<td>0.56</td>
<td>0.56</td>
<td>89.44</td>
<td>1.00</td>
<td>8063</td>
</tr>
<tr>
<td>T29</td>
<td>−16.06&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−27.87 to −3.96</td>
<td>1.00</td>
<td>1.00</td>
<td>8.74</td>
<td>1.00</td>
<td>7850</td>
</tr>
<tr>
<td>T30</td>
<td>−11.06&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−19.45 to −2.63</td>
<td>1.00</td>
<td>1.00</td>
<td>9.27</td>
<td>1.00</td>
<td>8544</td>
</tr>
<tr>
<td>T31</td>
<td>7.99</td>
<td>−9.54 to 25.47</td>
<td>0.81</td>
<td>0.81</td>
<td>74.60</td>
<td>1.00</td>
<td>8993</td>
</tr>
<tr>
<td>T32</td>
<td>−15.49</td>
<td>−33.63 to 3.45</td>
<td>0.95</td>
<td>0.95</td>
<td>30.23</td>
<td>1.00</td>
<td>9197</td>
</tr>
<tr>
<td>T33</td>
<td>−15.60</td>
<td>−33.42 to 2.04</td>
<td>0.96</td>
<td>0.96</td>
<td>28.00</td>
<td>1.00</td>
<td>7358</td>
</tr>
<tr>
<td>T34</td>
<td>−20.70&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−32.87 to −8.37</td>
<td>1.00</td>
<td>1.00</td>
<td>0.79</td>
<td>1.00</td>
<td>7973</td>
</tr>
<tr>
<td>T35</td>
<td>0.24</td>
<td>−7.97 to 8.63</td>
<td>0.52</td>
<td>0.52</td>
<td>87.05</td>
<td>1.00</td>
<td>8332</td>
</tr>
<tr>
<td>T36</td>
<td>−11.30&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−22.46 to −0.28</td>
<td>0.98</td>
<td>0.98</td>
<td>25.16</td>
<td>1.00</td>
<td>6861</td>
</tr>
<tr>
<td>T37</td>
<td>−14.13&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−28.07 to −0.10</td>
<td>0.98</td>
<td>0.98</td>
<td>23.84</td>
<td>1.00</td>
<td>7889</td>
</tr>
<tr>
<td>T38</td>
<td>−26.90&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−48.84 to −5.17</td>
<td>0.99</td>
<td>0.99</td>
<td>14.28</td>
<td>1.00</td>
<td>7520</td>
</tr>
<tr>
<td>T39</td>
<td>−13.52</td>
<td>−31.21 to 4.22</td>
<td>0.93</td>
<td>0.93</td>
<td>29.79</td>
<td>1.00</td>
<td>7380</td>
</tr>
<tr>
<td>Target participant</td>
<td>Slope (median of the posterior distribution)</td>
<td>95% UI(^b)</td>
<td>pd(^b)</td>
<td>ROPE(^c) (+/-)</td>
<td>% UI in ROPE</td>
<td>R-hat</td>
<td>ESS(^d)</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------------</td>
<td>-------------</td>
<td>--------</td>
<td>------------------</td>
<td>-------------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>T40</td>
<td>-3.02</td>
<td>-12.06 to 5.95</td>
<td>0.75</td>
<td>4.39</td>
<td>59.48</td>
<td>1.00</td>
<td>7465</td>
</tr>
<tr>
<td>T41</td>
<td>15.91</td>
<td>-5.36 to 37.06</td>
<td>0.93</td>
<td>12.45</td>
<td>35.81</td>
<td>1.00</td>
<td>7703</td>
</tr>
<tr>
<td>T42</td>
<td>-13.10</td>
<td>-41.31 to 15.65</td>
<td>0.82</td>
<td>20.77</td>
<td>71.60</td>
<td>1.00</td>
<td>8825</td>
</tr>
<tr>
<td>T43</td>
<td>-14.40</td>
<td>-35.39 to 5.80</td>
<td>0.92</td>
<td>14.06</td>
<td>48.56</td>
<td>1.00</td>
<td>7357</td>
</tr>
<tr>
<td>T44</td>
<td>-6.76</td>
<td>-15.22 to 1.43</td>
<td>0.94</td>
<td>6.47</td>
<td>46.99</td>
<td>1.00</td>
<td>8141</td>
</tr>
<tr>
<td>T45</td>
<td>-5.39</td>
<td>-20.66 to 10.17</td>
<td>0.75</td>
<td>12.13</td>
<td>81.79</td>
<td>1.00</td>
<td>8410</td>
</tr>
<tr>
<td>T46</td>
<td>-7.56</td>
<td>-17.10 to 2.05</td>
<td>0.94</td>
<td>6.98</td>
<td>44.73</td>
<td>1.00</td>
<td>7513</td>
</tr>
<tr>
<td>T47</td>
<td>-2.84</td>
<td>-14.49 to 8.84</td>
<td>0.69</td>
<td>8.30</td>
<td>83.36</td>
<td>1.00</td>
<td>8012</td>
</tr>
<tr>
<td>T48</td>
<td>-9.00</td>
<td>-20.33 to 2.12</td>
<td>0.95</td>
<td>7.59</td>
<td>39.17</td>
<td>1.00</td>
<td>8326</td>
</tr>
<tr>
<td>T49</td>
<td>-7.55</td>
<td>-18.29 to 3.36</td>
<td>0.92</td>
<td>7.85</td>
<td>52.52</td>
<td>1.00</td>
<td>7613</td>
</tr>
</tbody>
</table>

\(^a\)UI: uncertainty interval.
\(^b\)pd: probability of direction.
\(^c\)ROPE: region of practical equivalence.
\(^d\)ESS: effective sample size.
\(^e\)Statistically significant result.
Table 6. Target sample results of individual-level linear models assessing sleep duration associated with stress the day after (all individual-level models included the day of the week as a covariate, but the table only shows model parameters for the main predictor, stress the day after).

<table>
<thead>
<tr>
<th>Target participant</th>
<th>Slope (median of the posterior distribution)</th>
<th>95% UI</th>
<th>p&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ROPE&lt;sup&gt;c&lt;/sup&gt; (+/−)</th>
<th>% UI in ROPE</th>
<th>R-hat</th>
<th>ESS&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>−13.23</td>
<td>−29.23 to 2.91</td>
<td>0.95</td>
<td>9.09</td>
<td>29.61</td>
<td>1.00</td>
<td>5797</td>
</tr>
<tr>
<td>T2</td>
<td>−5.24</td>
<td>−18.80 to 8.68</td>
<td>0.77</td>
<td>9.94</td>
<td>76.74</td>
<td>1.00</td>
<td>8875</td>
</tr>
<tr>
<td>T3</td>
<td>−11.77</td>
<td>−34.81 to 11.56</td>
<td>0.85</td>
<td>13.68</td>
<td>56.47</td>
<td>1.00</td>
<td>7338</td>
</tr>
<tr>
<td>T4</td>
<td>−16.22&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−32.32 to −0.57</td>
<td>0.98</td>
<td>8.25</td>
<td>15.11</td>
<td>1.00</td>
<td>8021</td>
</tr>
<tr>
<td>T5</td>
<td>−8.21</td>
<td>−18.99 to 2.43</td>
<td>0.94</td>
<td>7.29</td>
<td>42.56</td>
<td>1.00</td>
<td>7377</td>
</tr>
<tr>
<td>T6</td>
<td>−11.94</td>
<td>−25.88 to 1.90</td>
<td>0.96</td>
<td>9.98</td>
<td>37.93</td>
<td>1.00</td>
<td>8926</td>
</tr>
<tr>
<td>T7</td>
<td>4.12</td>
<td>−4.28 to 12.64</td>
<td>0.84</td>
<td>4.82</td>
<td>56.73</td>
<td>1.00</td>
<td>8095</td>
</tr>
<tr>
<td>T8</td>
<td>−7.79</td>
<td>−33.38 to 17.75</td>
<td>0.73</td>
<td>16.89</td>
<td>77.04</td>
<td>1.00</td>
<td>8068</td>
</tr>
<tr>
<td>T9</td>
<td>−11.98</td>
<td>−39.16 to 15.54</td>
<td>0.80</td>
<td>14.68</td>
<td>57.80</td>
<td>1.00</td>
<td>8224</td>
</tr>
<tr>
<td>T10</td>
<td>−19.06&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−32.23 to −5.39</td>
<td>1.00</td>
<td>10.19</td>
<td>7.72</td>
<td>1.00</td>
<td>7881</td>
</tr>
<tr>
<td>T11</td>
<td>−0.62</td>
<td>−18.60 to 17.18</td>
<td>0.53</td>
<td>11.91</td>
<td>85.57</td>
<td>1.00</td>
<td>7375</td>
</tr>
<tr>
<td>T12</td>
<td>−12.49</td>
<td>−26.89 to 1.31</td>
<td>0.96</td>
<td>6.73</td>
<td>18.85</td>
<td>1.00</td>
<td>7661</td>
</tr>
<tr>
<td>T13</td>
<td>−13.95&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−27.24 to −0.47</td>
<td>0.98</td>
<td>10.44</td>
<td>30.02</td>
<td>1.00</td>
<td>7455</td>
</tr>
<tr>
<td>T14</td>
<td>−9.24</td>
<td>−37.81 to 19.76</td>
<td>0.73</td>
<td>12.84</td>
<td>55.80</td>
<td>1.00</td>
<td>5908</td>
</tr>
<tr>
<td>T15</td>
<td>15.64&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1.01 to 30.35</td>
<td>0.98</td>
<td>7.33</td>
<td>11.11</td>
<td>1.00</td>
<td>7561</td>
</tr>
<tr>
<td>T16</td>
<td>−37.82&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−62.78 to −12.78</td>
<td>1.00</td>
<td>15.46</td>
<td>1.60</td>
<td>1.00</td>
<td>7345</td>
</tr>
<tr>
<td>T17</td>
<td>−6.43</td>
<td>−23.61 to 10.03</td>
<td>0.78</td>
<td>11.19</td>
<td>72.94</td>
<td>1.00</td>
<td>7712</td>
</tr>
<tr>
<td>T18</td>
<td>−4.56</td>
<td>−28.81 to 19.13</td>
<td>0.65</td>
<td>11.74</td>
<td>66.61</td>
<td>1.00</td>
<td>8545</td>
</tr>
<tr>
<td>T19</td>
<td>−3.52</td>
<td>−20.73 to 14.26</td>
<td>0.66</td>
<td>12.28</td>
<td>83.96</td>
<td>1.00</td>
<td>8844</td>
</tr>
<tr>
<td>T20</td>
<td>1.55</td>
<td>−10.79 to 13.91</td>
<td>0.60</td>
<td>8.29</td>
<td>84.36</td>
<td>1.00</td>
<td>7304</td>
</tr>
<tr>
<td>T21</td>
<td>−14.76&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−25.94 to −3.31</td>
<td>0.99</td>
<td>7.88</td>
<td>9.48</td>
<td>1.00</td>
<td>7913</td>
</tr>
<tr>
<td>T22</td>
<td>−8.48</td>
<td>−23.16 to 6.68</td>
<td>0.87</td>
<td>8.08</td>
<td>47.68</td>
<td>1.00</td>
<td>8106</td>
</tr>
<tr>
<td>T23</td>
<td>−10.36</td>
<td>−31.70 to 10.72</td>
<td>0.83</td>
<td>9.30</td>
<td>44.67</td>
<td>1.00</td>
<td>7312</td>
</tr>
<tr>
<td>T24</td>
<td>−19.63&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−36.73 to −2.13</td>
<td>0.99</td>
<td>10.92</td>
<td>14.05</td>
<td>1.00</td>
<td>7158</td>
</tr>
<tr>
<td>T25</td>
<td>−11.10</td>
<td>−24.15 to 2.27</td>
<td>0.95</td>
<td>9.44</td>
<td>39.41</td>
<td>1.00</td>
<td>8495</td>
</tr>
<tr>
<td>T26</td>
<td>−10.11&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−17.83 to −2.28</td>
<td>1.00</td>
<td>5.63</td>
<td>10.24</td>
<td>1.00</td>
<td>8767</td>
</tr>
<tr>
<td>T27</td>
<td>−21.40&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−42.43 to −0.97</td>
<td>0.98</td>
<td>14.84</td>
<td>25.22</td>
<td>1.00</td>
<td>8373</td>
</tr>
<tr>
<td>T28</td>
<td>2.60</td>
<td>−7.35 to 12.04</td>
<td>0.70</td>
<td>7.28</td>
<td>84.95</td>
<td>1.00</td>
<td>8551</td>
</tr>
<tr>
<td>T29</td>
<td>−17.68&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−29.37 to −5.97</td>
<td>1.00</td>
<td>8.52</td>
<td>3.66</td>
<td>1.00</td>
<td>8521</td>
</tr>
<tr>
<td>T30</td>
<td>−6.95</td>
<td>−15.37 to 1.40</td>
<td>0.95</td>
<td>5.92</td>
<td>40.05</td>
<td>1.00</td>
<td>8895</td>
</tr>
<tr>
<td>T31</td>
<td>−2.47</td>
<td>−20.99 to 15.76</td>
<td>0.60</td>
<td>13.61</td>
<td>88.88</td>
<td>1.00</td>
<td>9079</td>
</tr>
<tr>
<td>T32</td>
<td>−0.28</td>
<td>−20.33 to 19.80</td>
<td>0.51</td>
<td>10.88</td>
<td>76.02</td>
<td>1.00</td>
<td>8377</td>
</tr>
<tr>
<td>T33</td>
<td>−3.11</td>
<td>−22.10 to 16.05</td>
<td>0.62</td>
<td>10.69</td>
<td>73.59</td>
<td>1.00</td>
<td>7235</td>
</tr>
<tr>
<td>T34</td>
<td>−17.71&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−29.99 to −5.33</td>
<td>1.00</td>
<td>9.07</td>
<td>6.43</td>
<td>1.00</td>
<td>8398</td>
</tr>
<tr>
<td>T35</td>
<td>−7.43</td>
<td>−15.50 to 0.75</td>
<td>0.96</td>
<td>5.75</td>
<td>32.79</td>
<td>1.00</td>
<td>8137</td>
</tr>
<tr>
<td>T36</td>
<td>−8.27</td>
<td>−20.14 to 3.43</td>
<td>0.92</td>
<td>7.70</td>
<td>45.84</td>
<td>1.00</td>
<td>6671</td>
</tr>
<tr>
<td>T37</td>
<td>−10.86</td>
<td>−25.05 to 3.13</td>
<td>0.94</td>
<td>9.32</td>
<td>41.21</td>
<td>1.00</td>
<td>7458</td>
</tr>
<tr>
<td>T38</td>
<td>−17.78</td>
<td>−42.09 to 7.02</td>
<td>0.93</td>
<td>15.98</td>
<td>43.95</td>
<td>1.00</td>
<td>8022</td>
</tr>
<tr>
<td>T39</td>
<td>−15.79</td>
<td>−33.41 to 2.06</td>
<td>0.96</td>
<td>9.20</td>
<td>21.49</td>
<td>1.00</td>
<td>7141</td>
</tr>
<tr>
<td>Target participant</td>
<td>Slope (median of the posterior distribution)</td>
<td>95% UIa</td>
<td>pdb</td>
<td>ROPEc (+/-)</td>
<td>% UI in ROPE</td>
<td>R-hat</td>
<td>ESSd</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------------------</td>
<td>--------</td>
<td>------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>T40</td>
<td>-7.11</td>
<td>-15.56 to 1.26</td>
<td>0.95</td>
<td>4.39</td>
<td>24.88</td>
<td>1.00</td>
<td>7108</td>
</tr>
<tr>
<td>T41</td>
<td>-15.15</td>
<td>-36.52 to 5.96</td>
<td>0.92</td>
<td>12.45</td>
<td>39.48</td>
<td>1.00</td>
<td>8726</td>
</tr>
<tr>
<td>T42</td>
<td>9.05</td>
<td>-20.92 to 38.93</td>
<td>0.72</td>
<td>20.77</td>
<td>79.25</td>
<td>1.00</td>
<td>8862</td>
</tr>
<tr>
<td>T43</td>
<td>-7.80</td>
<td>-29.05 to 14.03</td>
<td>0.76</td>
<td>14.06</td>
<td>72.40</td>
<td>1.00</td>
<td>7545</td>
</tr>
<tr>
<td>T44</td>
<td>-5.63</td>
<td>-14.57 to 3.44</td>
<td>0.89</td>
<td>6.47</td>
<td>57.66</td>
<td>1.00</td>
<td>8151</td>
</tr>
<tr>
<td>T45</td>
<td>-1.63</td>
<td>-19.16 to 16.16</td>
<td>0.58</td>
<td>12.13</td>
<td>86.41</td>
<td>1.00</td>
<td>8594</td>
</tr>
<tr>
<td>T46</td>
<td>-3.81</td>
<td>-13.64 to 5.68</td>
<td>0.79</td>
<td>6.98</td>
<td>75.80</td>
<td>1.00</td>
<td>8400</td>
</tr>
<tr>
<td>T47</td>
<td>-8.61</td>
<td>-21.25 to 3.95</td>
<td>0.91</td>
<td>8.30</td>
<td>47.73</td>
<td>1.00</td>
<td>7590</td>
</tr>
<tr>
<td>T48</td>
<td>-6.65</td>
<td>-17.83 to 4.77</td>
<td>0.87</td>
<td>7.59</td>
<td>56.89</td>
<td>1.00</td>
<td>7442</td>
</tr>
<tr>
<td>T49</td>
<td>10.06</td>
<td>-1.66 to 22.09</td>
<td>0.96</td>
<td>7.85</td>
<td>34.47</td>
<td>1.00</td>
<td>8521</td>
</tr>
</tbody>
</table>

aUI: uncertainty interval.  
bpd: probability of direction.  
cROPE: region of practical equivalence.  
dESS: effective sample size.  
eStatistically significant result.

**Figure 6.** Higher subjective stress associates with shorter objective sleep duration in most individuals. (A) Estimated slopes (corresponding to the median of the posterior distribution) from the individual-level models that assess sleep duration associated with stress the day before (triangles) and with stress the day after (circles) are plotted along the y-axis ordered by participant (x-axis). Statistically significant slope estimates are shaded green (day before) or purple (concurrent). Error bars show 95% uncertainty intervals, and shaded density plots show the full posterior distributions of the slopes. (B) Slope estimates from subpart A are plotted against each participant-level estimate across the study period.

**Person-Specific Estimates Get Attenuated in Group-Based Modeling**

A comparison of the estimates obtained through the iLM and those of an MLM demonstrates that individual-level estimates get systematically attenuated in a group-based approach when there is between-person heterogeneity in the effects (Figure 7). When there was a strong effect and small between-person variability in the tested associations, as with the association between sleep duration and concurrent sleep quality, the iLMs provided slope estimates nearly identical to those estimated through an MLM (Figure 7A). However, for associations that showed a weaker effect and greater degree of between-person variability, group-level approaches systematically flatten individually tailored effect sizes or even reverse the sign of the association (Figure 7B). We observed this in the lead-lag associations between stress and sleep duration (as highlighted with asterisks in Figure 7) and to some degree in the lagged association between sleep quality and sleep duration.
Examining Raw Within-Individual Data Informs the iLM Results

A closer look at individual participants’ data reveals important considerations for the application and interpretation of the iLMs (Figure 8). As with other statistical models, the iLM is unlikely to detect stable associations for participants with too little variability in their data; for instance, participant T45 had consistently low stress levels over the course of the year, with occasional, small increases in stress tied to periods with increased academic demands (eg, ahead of midterm and final examinations periods). This participant’s individual linear models showed null results for both directions of stress-sleep associations (pd<0.975). Meanwhile, participant T29 also had low baseline stress levels, but they presented more frequent and substantial rises in stress throughout the year. Participant T29’s iLMs showed bidirectional negative associations between stress and sleep duration (pd>0.975).

While variability in daily observations over time is necessary to detect linear associations within the iLM, variability alone is not sufficient. Participant T10 showed substantial fluctuations in stress and sleep duration throughout the year, and their iLMs found bidirectional negative associations between stress and sleep duration (pds>0.975). Meanwhile, participant T16 also presented substantial variability in stress and sleep duration, and while their iLM detected a significant association between sleep duration and stress the day after (pd=1.00), no association was detected in the opposite direction (pd=0.65).
Figure 8. Intensive within-individual longitudinal data reveal differences among individuals. Results are shown for individual-level models testing the association between sleep duration and stress the day before (green) and between sleep duration and stress the day after (purple). Column 1 shows the models’ estimated slopes (computed as the median of the posterior distribution) and uncertainty metrics. Symbol shading signifies statistically significant slopes. Error bars show 95% uncertainty intervals (UIs). Shaded density plots show the full posterior distributions of the slopes; gray shaded areas show the regions of practical equivalence (ROPEs). Column 2 shows the models’ predicted sleep duration as a function of stress; shading around the lines indicates 95% UIs. Shaded density plots show the full posterior distributions of the slopes; gray shaded areas show the regions of practical equivalence (ROPEs). Column 3 shows daily observations of actigraphy-derived sleep duration and survey-based perceived stress. Gray dashed lines show the participant’s mean value across the study period. Vertical lines indicate landmark events (labeled) in the academic calendar. Gray shading indicates missing data during the school terms or the winter break, which were excluded from the individual-level linear models. Dur.: duration; E: examinations period; n.s.: not statistically significant; pd: probability of direction; R: reading period; SB: spring break; sig.: statistically significant; TB: Thanksgiving break; WB: winter break.

Discussion

Principal Findings

Stress levels and sleep duration interact in an individual’s daily life, but research has yielded mixed findings regarding the temporal directionality of their associations, with changes in stress preceding changes in sleep, changes in sleep preceding changes in stress, or both. Here, we leveraged a novel individual-level linear regression modeling iLM framework to obtain precision estimates of day-to-day associations between self-reported stress levels and actigraphy-derived sleep duration in a sample of first-year college students studied continuously for a full academic year. While most of the participants (45/49, 92%) showed a negative association between daily stress levels and sleep duration, the temporal direction of the association varied, with all types of lead-lag association previously reported at the group level present within distinct individuals in our sample.

In agreement with prior literature, our results within individuals confirm that stress levels and sleep duration are closely and inversely related in daily life [10-19]. Nearly four-tenths (19/49, 39%) of the participants—each considered an independent test of the association—showed a negative association between daily stress levels and sleep duration, the temporal direction of the association varied, with all types of lead-lag association previously reported at the group level present within distinct individuals in our sample.

Our precision approach further revealed that the temporal directionality of the association between stress and sleep duration varied from person to person, representing all patterns of results reported by prior, group-level studies. For some of the participants (8/49, 16%), heightened stress during the day associated with shorter sleep later that night but not vice versa, in agreement with group results reported in the studies by Marcusson-Clavertz et al [15] and Slavish et al [16]; for others
(5/49, 10%), shorter sleep associated with heightened stress the next day but not vice versa, in agreement with group results in the study by Sin et al [17]; and yet others (6/49, 12%) showed both directions of association, in agreement with group results in the studies by Doane and Thurston [18] and Yap et al [19]. Daily psychological and behavioral experiences such as perceived stress and sleep duration thus seem to interact in person-specific ways, rather than being uniform across the population.

For individuals showing the stress–then–reduced-sleep phenotype, experiences of heightened stress (eg, due to an impending final examination or social conflict) might elicit hyperarousal and rumination [13,51-53], as well as behaviors aimed at mitigating the source of stress (eg, studying or socializing with friends late into the night), all of which can delay sleep and reduce its overall duration. Moreover, for those showing the reduced-sleep–then–stress phenotype, shortened sleep durations might enhance their sensitivity to (and undermine their ability to cope with) academic, interpersonal, or other stressors and thus make them more likely to experience heightened stress levels [10,54,55]. For some individuals, both patterns of effect might occur, with changes in stress levels and sleep duration reinforcing one another and resulting in chains of days with heightened stress and nights of short sleep that succeed one another.

Critically, our results suggest that group-level studies might report inconsistent findings, at least partly, because the dynamic interaction between daily stress levels and sleep duration varies from person to person. When data are aggregated at the group level, individual phenotypes might be obscured, and the resulting group-level estimates are suggestive of generalized effects when in fact they might only apply to a fraction of the sample. Even when hierarchical group models allow for fitting individual estimates, these are biased by the group and tend to attenuate (or shrink) the estimation of individual effects [26]. A comparison of individual slope estimates of stress-sleep associations derived from our iLMs and those derived from an MLM with random intercepts and slopes starkly demonstrated this group bias: the group MLM estimated individual slopes that were systematically attenuated or even reversed sign compared to those estimated by our iLMs.

These results demonstrate the utility of individual-level modeling for characterizing real-world behavioral and psychological dynamics [27-29,34]. Our approach leverages mobile and wearable technology and provides a fit-for-purpose methodology that can turn these devices’ large-scale longitudinal measurements into meaningful insights. The iLM’s model specification is parsimonious by design and easy to interpret, including a single linear term for the main predictor of interest and a day-of-the-week covariate to account for the weekly structure in the outcome variable. Diagnostic checks confirmed that this simple model specification was powerful enough to capture real-world, stable linear associations between psychological phenomena and objective behaviors, while accounting for the time-related dependencies in the daily observations. Individual tailoring is achieved by fitting only 1 person’s data within a model, but the model specification remains identical across individuals, allowing for direct between-person comparisons of results, including estimating the relative prevalence of different phenotypes in the group.

The iLM framework is readily applied to a single individual’s data, making it useful for multiple real-world purposes beyond fundamental research that are increasingly gaining interest in the fields of consumer health informatics and digital health [21,31,32]. The iLM can be applied to data collected through personal devices for self-monitoring as well as for precision approaches in health care settings, where clinicians might use a patient’s data to triage intervention plans. Taking these results as an example, stress management interventions might be first prioritized among individuals for whom heightened stress precedes shortened sleep, while sleep interventions might be prioritized among individuals for whom shortened sleep precedes heightened stress. Moreover, this work might inform a growing body of research and products combining multiple data streams from wearables and smartphones along with machine learning techniques to predict experiences of stress [36,57]. Although prediction was not the focus of our work, understanding the person-specific association between stress and sleep duration, as well as the weekly behavioral patterns revealed by our actigraphy and survey data and individualized approach, could potentially contribute to the identification of periods when individuals are more likely to experience stress so that timely interventions can be offered.

The iLM offers a simple yet powerful precision framework for the estimation of real-world psychological and behavioral associations within the individual, but the results should be interpreted carefully in light of its assumptions and limitations. First, the iLM’s assumptions of linear, stable associations between the predictor and outcome variables are a deliberate attempt to simplify real-world behavioral and psychological dynamics that are highly complex. In the context of this study, these features allowed us to estimate college students’ day-to-day stress-sleep associations that are stable across the fluctuating demands on students within the school semesters as well as straightforward to interpret. However, it is possible that the association between stress and sleep duration is context dependent, varying as a function of the specific source of stress experienced by the individual (eg, academic vs interpersonal) and the broader seasonal demands (eg, whether school is in session). In fact, given the possibility of the latter, we decided a priori to exclude data collected during the winter break from our models. Future research could examine how these contextual demands influence stress-sleep associations, as well as explore nonlinear associations or cumulative effects over time.

Our analyses leveraged leading and lagging patterns in each individual’s time series of stress and sleep duration measures to ascertain the temporal directionality of their association (eg, stress during the day as a predictor of subsequent sleep duration later that night), but we did not implement a controlled experimental manipulation and cannot establish a causal relationship. While our use of intensive longitudinal data collected in the real world grants our results ecological validity, it also exposes them to confounders; for example, it is possible that a significant association between short sleep and higher stress the next day could be explained by the anticipated demands of the next day, such as an examination. Rather than
short sleep duration causing higher stress the next day, the examination might be the primary cause behind both the short sleep (staying up late to prepare) and the stress reported the next day (heightened stress during test taking).

While our current intensive longitudinal data set and individual-level modeling framework passed sanity checks and diagnostics that confirmed data quality and adequate model specification, researchers applying our framework to other data sets and research questions should scrutinize the appropriateness of the data and the model before interpreting the results. Mobile and wearable technologies enable the collection of large behavioral data sets over time, but quantity does not guarantee quality, and long study periods require extra vigilance to ensure that participants remain compliant over time. Quality checks should confirm that the data collected are capturing expected real-world behavior (eg, as suggested by the intuitive changes we observed in students’ behavior between school terms and breaks as well as between weekdays and weekends). Even if the available data are substantial, and participant compliance is high, small variability in the metrics under study could still impede obtaining meaningful estimates of their associations, as demonstrated by participants with minimal fluctuations in stress levels over the course of the academic year. Moreover, investigators should be careful to identify appropriate uses, sanity checks, and interpretations of the iLM for their population of study; for example, part of our iLM validation process included testing the expected positive association between objective sleep duration (measured via the actigraphy wristband) and the participant’s subjective rating for the same sleep event (reported via a daily smartphone survey). While we expect that generally healthy participants will tend to rate nights of shorter-than-usual sleep duration as lower quality, it might not always be advisable to assume a simple linear association between objective sleep duration and subjective sleep quality, especially when modeling data from patients with sleep and psychiatric disorders.

**Conclusions**

Our novel iLM framework leveraged intensive longitudinal data from mobile and wearable devices to obtain individually tailored estimates of day-to-day associations between subjective stress levels and objective sleep duration. While stress and sleep duration were inversely related in most of the participants (45/49, 92%), the iLM revealed that the temporal direction of these associations is person specific, identifying a variety of individual phenotypes that may account for the diverse group-level findings reported in prior literature. Our results demonstrate the utility of individual-level modeling approaches for the assessment of behavioral and psychological associations in the real world. An individualized approach offers a foundation for the characterization of life dynamics at both the individual and group levels, as well as for the development of precision health and well-being interventions tailored to the individual.

**Acknowledgments**

The authors thank Laura Farfel, Marisa Marotta, Erin Phlegar, Lauren DiNicola, and Arpi Youssoufian for their help collecting data. Timothy O’Keefe, Harris Hoke, and Lily Jeong provided valuable assistance in neuroinformatics support. Kenzie W Carlson provided support with Beiwe. Katherine Miclau, Amira Song, Emily Iannazzi, and Hannah Becker helped with the actigraphy quality control procedure. This work was supported by a generous gift from Kent and Liz Dauten, National Institute of Mental Health grants (U01MH116925 and DP2MH103909), a grant from the National Institutes of Health (T90DA022759), and the Sackler Scholar Program in Psychobiology.

**Data Availability**

The data sets generated during this study are not publicly available due to concerns related to participant identifiability but may be available from the corresponding author on reasonable request.

**Authors’ Contributions**

GCIII, J-PO, JTB, and RLB designed the research. GCIII, J-PO, and RLB performed the research. CMVB, GCIII, HR-E, and RLB analyzed the data. PM provided statistical support. CMVB and RLB wrote the paper, and all other authors reviewed the final manuscript.

**Conflicts of Interest**

J-PO is a cofounder and board member of Phebe Health, a commercial entity that operates in digital phenotyping. JTB has received consulting fees from Verily Life Sciences as well as consulting fees and equity from Mindstrong for work unrelated to this study. RLB has received consulting fees from Pfizer, Roche, Alkermes, and Cognito for work unrelated to this study. All other authors declare no other conflicts of interest.

Multimedia Appendix 1

Additional figures show participant-level missing data, comparison between Bayesian and frequentist individual-level linear model estimates, and group-level sleep and stress metrics in the target sample.

[DOCX File: 679 KB - formative_v8i1e53441_app1.docx]
References


33. Scala JJ, Ganz AB, Snyder MP. Precision medicine approaches to mental health care. Physiology (Bethesda) 2023 Mar 01;38(2) [FREE Full text] [doi: 10.1152/physiol.00013.2022] [Medline: 36999270]


Abbreviations

- DPSleep: deep phenotyping of sleep
- ESS: effective sample size
- iLM: individual-level linear model
- MCMC: Markov chain Monte Carlo
- MLM: multilevel linear model
- pd: probability of direction
- ROPE: region of practical equivalence
- UI: uncertainty interval

©Constanza M Vidal Bustamante, Garth Coombs III, Habiballah Rahimi-Eichi, Patrick Mair, Jukka-Pekka Onnela, Justin T Baker, Randy L Buckner. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Effectiveness of a Smartphone App (Heia Meg) in Improving Decisions About Nutrition and Physical Activity: Prospective Longitudinal Study

Christine Olsen, MSc; Daniel Adrian Lungu, PhD
Department of Quality and Health Technology, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway
*All authors contributed equally

Corresponding Author:
Daniel Adrian Lungu, PhD
Department of Quality and Health Technology
Faculty of Health Sciences
University of Stavanger
Kjell Arholms gate 41
Stavanger, 4021
Norway
Phone: 47 51 83 10 00
Email: daniel.a.lungu@uis.no

Abstract

Background: Obesity is a prevalent and serious chronic condition associated with abnormal or excessive fat buildup that poses significant health risks. The rates of overweight and obesity in adults and children continue to rise, with global rates of children with overweight or obesity aged 5-19 years growing from 4% to 18% between 1975 and 2016. Furthermore, in 2017, nearly 4 million people died due to complications arising from being overweight or obese.

Objective: This study aims to investigate the potential impact of the mobile app Heia Meg on promoting healthier lifestyle choices regarding nutrition and physical activity.

Methods: A prospective longitudinal study was conducted in collaboration with the Norwegian Directorate of Health. Participants were recruited through the Heia Meg app and were asked to complete a questionnaire before and after using the app. A total of 199 responses were included in the first (preintervention) questionnaire, while 99 valid responses were obtained in the second (postintervention) questionnaire.

Results: The majority (159/199, 79.9%) of participants were female, and their age ranged from 18 years to 70 years and older. The results show a reduction in BMI after the digital intervention. However, some variables influence the BMI reduction effect: sex, age, education, and smoking. The group that obtained the most benefit from the intervention consisted of those who were male, aged 30-39 years, highly educated, and nonsmokers. Although positive, some of the findings are slightly above the statistical significance threshold and therefore should be interpreted carefully.

Conclusions: Our study found weak evidence to support the effectiveness of the Heia Meg app in promoting healthier lifestyle choices. However, limitations and confounding factors suggest that further research in different populations with larger sample sizes is needed to confirm or disprove our findings.

(JMIR Form Res 2024;8:e48185) doi:10.2196/48185

KEYWORDS
app; BMI; diet; exercise; health; Heia Meg; lifestyle change; longitudinal; mHealth; mobile health; motivation; nutrition; obese; obesity; overweight; physical activity; smartphone apps; weight
Introduction

Overview

Consumers have benefited from the internet’s transformation in information access about a decade ago [1]. Smartphones allow people to access information with a single touch. Smartphones offer a vast variety of health apps that users can download (for free or not). Many offer reference, monitoring, and calculator tools and address a wide range of health-related concerns. Apps count calories and nutrition, calculate BMI, monitor diabetes, handle emergencies, and improve workouts.

Mobile health (mHealth) technology and consumer health informatics research have increased in recent years [1]. A survey by Sarcona et al [2] indicated that mHealth app users had significantly better eating behavior inventory scores or reported more positive eating behavior than nonusers. Fry and Neff [3] found that periodic reminders can aid in behavior improvement. These findings can be used to improve periodic quick interventions, leading to increased effectiveness, positive behavior change, and better health [3].

Health psychology models were used to design the experiment. Most social-cognitive theories presume that the desire to change predicts actual change, yet people rarely act on their intentions [4]. Heia Meg, which in Norwegian means the people’s conviction in their own capacities to produce predetermined levels of performance, influences the circumstances that affect their conduct. Belief in mastery influences people’s thoughts, feelings, motivation, and conduct. Cognitive, motivational, emotional, and selection processes generate this belief.

The health belief model (HBM) was devised in the 1950s to explain why so many people do not participate in disease prevention and detection programs [5]. The HBM is one of the most often used conceptual frameworks in health behavior research since 1950, “both to explain change and maintenance of health-related behaviors and as a guiding framework for health behavior treatments” [5,6]. If individuals regard themselves as susceptible to a condition, believing it could have serious consequences, while believing a course of action could reduce their susceptibility to or severity of the condition, and believing that the anticipated benefits of taking action outweigh the costs, this could increase the likelihood to act [5].

Perceived susceptibility refers to people’s health risks. Perceived severity refers to feelings about the seriousness of developing a disease or leaving it untreated, including medical and clinical implications (such as disability, death, and pain) and possible social consequences (such as effects on work, family life, and social relations) [5,7,8]. The individual’s thoughts regarding the perceived advantage of the potential activities for reducing the threat of illness can influence whether this perception leads to behavior change [5,7]. Other nonhealth factors, such as money savings from quitting smoking or pleasing a family member, may also influence behavior. Individuals with ideal views on susceptibility and severity are unlikely to choose a health intervention unless they believe it can lessen the threat. Perceived barriers are anticipated unfavorable outcomes of a health action. They may prevent people from engaging in the suggested activities. Individuals balance the action’s projected advantages against perceived barriers in an unconscious cost-benefit analysis. Consequently, “susceptibility and severity offer the energy or force to act, and benefits (without barriers) provide a preferred course of action” [5]. Cues to action are inputs that can trigger behaviors and were incorporated in previous HBM formulations. Hochbaum [9] argued that other aspects, such as physical occurrences or media attention, may only increase readiness to act (perceived susceptibility and perceived rewards).

Self-efficacy refers to having faith in one’s own strengths when presented with challenging activities and situations [10]. To attain mastery, one must learn from fresh hurdle situations and practice. A self-efficacy approach was implemented in the app messages as well. Through Heia Meg, messages are delivered as little challenges such as “Try something new, what about a walk before bedtime? Bring a headlamp, use the stairs, and take a 20-minute walk today” or sentences such as “You don’t like hills, yet they have something great to offer. Hills increase heart rate and health. So, imagine a top goal line and go for it.” These messages contain encouraging sentiments that may help people complete the challenge. The second most important source of self-efficacy is what we observe others do or accomplish.

Positive role models can help develop positive self-beliefs and can include family, friends, teachers, coaches, or employers. The third most important source of self-efficacy is social persuasion—receiving positive feedback while executing a difficult task might persuade a person, as they might start considering that they have the abilities and potential to succeed. The Heia Meg app’s daily messages have a positive tone, with statements such as “something is better than nothing.” or “The best session is the one you finish:)” or “A week has passed. Many people find it hard to start, so keep going. Cheers!” Since emotional and physiological factors might affect how someone feels about their ability, the app’s motivational sentences include “Good company helps when motivation fades.” “Know somebody who wants to get in shape? You can request a group activity.” “Before turning, take a 5-minute walk.” or “The first obstacle is often the hardest, but you can continue.”

The I-Change model was developed by Hein de Vries to explain health behavior and motivation [11]. I-Change combines aspects of the theory of planned behavior, the transtheoretical model for health behavior change, the social cognitive theory, the goal-setting theory, and the HBM to create a motivation and behavior change model. Motivation or intention determines behavior. Attitudes, social pressures, and self-efficacy affect motivation. Attitudes are the cognitive and emotional beliefs and costs of an activity. Social modeling, social norms, and social support from others are examples of social influences on a person. New research suggests multiple types of self-efficacy, including stress-, social-, routine-, and skills-based self-efficacy. The I-Change model believes information and antecedent circumstances affect communication results—motivation, awareness, action, and behavior (Figure 1). Preparing and executing detailed plans to achieve the intended behavior increases the likelihood of intentions becoming actions, while barriers decrease these chances [11,12]. Attitudes, social pressures, and self-efficacy expectations all affect a person’s motivation.

https://formative.jmir.org/2024/1/e48185
Overweight and Obesity

According to the World Health Organization, the number of 5- to 19-year-olds with overweight or obesity quadrupled between 1975 and 2016, and overweight and obesity cause more deaths than underweight [13]. We are also seeing a rise in overweight and obesity in Norway, largely due to an energy intake or consumption imbalance [14]. According to the Norwegian Directorate of Health, this increase is linked to increased inactivity, which can lead to various diseases over time, including increased mortality and increased risk of heart attack, stroke, high blood pressure, type 2 diabetes, several cancers, musculoskeletal disorders, and mental disorders. Another US study indicated that physical inactivity is the primary cause of increased mortality and morbidity in adults who are overweight (with a BMI up to 35 kg/m$^2$), but that good physical shape can minimize the risk of weight-related diseases [15]. Increasing weight and obesity have health and economic effects, such as weight-related health problems accounting for 2% to 6% of overall health costs [16].

mHealth

mHealth is a new sector with the potential to reach a large portion of the population cost-effectively. It is a branch of eHealth that uses technology to improve people’s health. Research shows that 2.9% of Google Play apps and 8.8% of Apple App Store apps support healthy behaviors [17]. Cost-effective technologies that record a user’s behavior are gaining prominence. Despite apps’ potential to alter health behavior, there is no proof of their health theory foundation [17]. According to Antezana et al [17], health apps have low levels of theory of behavior change strategies, and improved implementation could lead to more user engagement and better interventions. Mateo et al [18] compared a mobile app to other weight-loss and exercise strategies. In their review of 12 studies, using a mobile app positively affected body weight, but not physical activity.

Heia Meg

The Heia Meg app was developed by the Norwegian Directorate of Health with inspiration from the United Kingdom’s “One You” and a previous Norwegian app called “Slutta” or “Quit” [19]. The project is nondirectly supportive and based on the idea that motivation changes behavior. The “Quit” app is a popular tool for Norwegians trying to quit smoking. Due to the “Quit” app’s success, self-efficacy and positive psychology were also used in Heia Meg.

Heia Meg self-efficacy examples include inspiring health alternatives, push-notification support, and push warnings preventing temptation. The pursuit of happiness is one of the humanities’ most persistent movements [20]. Positive psychology aims to answer the question “What is happiness?” The Heia Meg app uses positive psychology by focusing on stress management and by pushing alerts encouraging users.

Push alerts are regular and scheduled. After downloading the app, users must agree to its terms before choosing 2 of 5 themes to focus on. Users can choose between (1) exercise, (2) mental health, (3) alcohol, (4) sleep habits, and (5) dietary intake. After choosing the 2 themes, the app will send notifications on a regular basis. The notifications include encouragements, theme facts, challenges, and recommendations. The texts are short and written in Norwegian, with no emojis or abbreviations.

The National Institute of Public Health (NIPH) national public health survey presented a report of results on diet, self-reported weight, and weight development in the Norwegian population in 2020 [21]. The report showed that over two-thirds of people wished to reduce weight or had tried to maintain their weight, as well as a large proportion of the population being overweight or obese. According to the findings of the public health survey on body weight and development, male individuals’ average weight, height, and BMI were 86.6 kg, 180.7 cm, and 26.5 kg/m$^2$, respectively [21]. The percentage of male individuals who were overweight or obese (BMI ≥25 kg/m$^2$) was 59%, while the percentage of male individuals who were underweight (BMI <18.5 kg/m$^2$) was 0.8%. On the other hand, the average weight, height, and BMI for female individuals were 86.6 kg, 180.7 cm, and 26.5 kg/m$^2$, respectively [21]. The number of female individuals who were overweight or obese (BMI ≥25 kg/m$^2$) was 47%, while the proportion of female individuals who were underweight was 2.7%. Overall, 16% of female and male individuals had a BMI ≥30 kg/m$^2$. The project is nondirectly supportive and based on the idea that motivation changes behavior. The “Quit” app is a popular tool for Norwegians trying to quit smoking. Due to the “Quit” app’s success, self-efficacy and positive psychology were also used in Heia Meg.

Heia Meg self-efficacy examples include inspiring health alternatives, push-notification support, and push warnings preventing temptation. The pursuit of happiness is one of the humanities’ most persistent movements [20]. Positive psychology aims to answer the question “What is happiness?” The Heia Meg app uses positive psychology by focusing on stress management and by pushing alerts encouraging users.

Push alerts are regular and scheduled. After downloading the app, users must agree to its terms before choosing 2 of 5 themes to focus on. Users can choose between (1) exercise, (2) mental health, (3) alcohol, (4) sleep habits, and (5) dietary intake. After choosing the 2 themes, the app will send notifications on a regular basis. The notifications include encouragements, theme facts, challenges, and recommendations. The texts are short and written in Norwegian, with no emojis or abbreviations.

The National Institute of Public Health (NIPH) national public health survey presented a report of results on diet, self-reported weight, and weight development in the Norwegian population in 2020 [21]. The report showed that over two-thirds of people wished to reduce weight or had tried to maintain their weight, as well as a large proportion of the population being overweight or obese. According to the findings of the public health survey on body weight and development, male individuals’ average weight, height, and BMI were 86.6 kg, 180.7 cm, and 26.5 kg/m$^2$, respectively [21]. The percentage of male individuals who were overweight or obese (BMI ≥25 kg/m$^2$) was 59%, while the percentage of male individuals who were underweight (BMI <18.5 kg/m$^2$) was 0.8%. On the other hand, the average weight, height, and BMI for female individuals were 86.6 kg, 180.7 cm, and 26.5 kg/m$^2$, respectively [21]. The number of female individuals who were overweight or obese (BMI ≥25 kg/m$^2$) was 47%, while the proportion of female individuals who were underweight was 2.7%. Overall, 16% of female and male individuals had a BMI ≥30 kg/m$^2$.
of 30 kg/m$^2$ or higher, which indicates obesity [21]. The average person consumed 2 servings of fruits and vegetables each day, including juice, and according to the survey, only 2.3% reported that they consume at least 5 servings each day. Furthermore, the report stated that 43% of the participants ate fish 2-3 times per week, whereas 7% rarely or never ate fish. Approximately 30% said they ate sweets frequently (at least 3 times per week), while 10% reported eating snacks and 11% reported eating sweet pastries at least 3 times per week [21]. A large percentage of people reported using soft fat or a combination of soft and hard fat types or none for bread, and 61% reported using soft fat, oil, or nonfat for frying. Further, when looking at sugary soft drinks, juices, and soda, 13% reported drinking them 3 times a week or more, with the proportion being highest among younger male individuals and among the least educated. Dietary changes were also mentioned, with 5.4% following a low-carb diet, 3.8% following a calorie-reduced diet, and 3.3% fasting on a regular basis [21].

The survey revealed that Norway’s population needs and wants to improve its lifestyle regarding nutrition and weight loss. Therefore, the aim of this study was to test whether the Heia Meg app can help achieve this goal. Therefore, the following research question was formulated: “Is Heia Meg, a smartphone app, effective in helping users to make healthier lifestyle decisions about nutrition and physical activity?”

**Methods**

**Overview**

Based on past observations using the app and its structure and functioning, we anticipated that it could improve the physical activity behavior and eating habits of users. To test the assumption, we used a prospective longitudinal design. All statistical analyses were conducted using SPSS Statistics (version 26; IBM Corp).

**Recruitment Procedure**

In collaboration with the Norwegian Directorate of Health, we added a link to the questionnaire within the Heia Meg app. This was only accessible to Heia Meg app users with internet connectivity enabled. Besides the in-app advertisement, posters were put on Facebook, Instagram, and Snapchat to foster recruitment. A poster was affixed to the wall of a neighboring fitness center, which contained complete information about the study and how to participate. Because the app and survey were written in Norwegian, so were the posters.

**Participants**

The Heia Meg app was the only channel for the registration of study participants. The sample consisted of users who downloaded the app and agreed to take part in the study.

Everyone who downloaded the app between October 4, 2021, and November 5, 2021, received a message the day after downloading it asking whether they were interested in joining a research study. Participants had to read and sign a consent form confirming that they were 18 years of age or older and that they consented to participate. Participants were asked to provide their email address because that is what was used to send the follow-up survey.

**Materials**

This project’s questionnaire was based on Harris’s [22] NIPH study, “Social circumstances and health: a twin study.” The form was modified for this investigation, and it contains 12 health-related items.

The 12 items response alternative were all defined as a Likert scale, so they could be graded from 1 to 4, with 1 being the worst and 4 being the best. The replies to the 12 items would be used to compute a health score. The score is the sum of the participants’ responses, with a minimum of 12 points (scoring 1 on all 12 questions) and a maximum of 48 points (scoring 4 on all 12 questions). This study took place during the COVID-19 pandemic, which may have affected people’s health, lifestyle, eating, drinking, and mental health.

**Ethical Considerations**

An application, containing the project plan and the questionnaire, was submitted to the Norwegian Center for Research Data (NSD) and the regional ethics committee. As no health or sensitive data were going to be processed, the regional ethics committee confirmed that the project is approved without a formal assessment (application 284804).

Participants had to read and sign a consent form confirming that they were 18 years of age or older and that they consented to participate.

Most of the ethical considerations concern the gathering of personal data and how to organize, implement, and complete it in line with legislative requirements and ethical norms. The Health Research Act, the Health Register Act, laws on population-based health assessments, and the Privacy Ordinance and Personal Data Act offer a comprehensive foundation for medical and health research [23]. For this consent-based research project, the duty to disclose information is accomplished through Netskjem. All personal information was to be anonymized for data processing; hence, a participant identification number was required. Due to participants using their email, their answers were not anonymous; nonetheless, the data were secured, anonymized, and held until the manuscript was submitted, following which they were deleted.

NSD approved the study in relation to the processing of personal data (approval 385157).

No compensation was given to the participants; they volunteered to take part in the study.

**Results**

**Overview**

We received 365 responses, but some did not meet the criteria and were eliminated. In the preintervention questionnaire, 5 people were removed for being 17 years of age or younger, 9 for refusing informed consent, and 20 for answering the questionnaire twice or more. Two people, aged younger than 18 years and 31 years, respectively, had repeated responses, which were eliminated from the postintervention questionnaire.
This study included 298 replies in total: 66.78% (n=199) in the preintervention group and 33.22% (n=99) in the postintervention group.

**Descriptive Statistics**

As depicted in Table 1, female individuals participated more than male individuals. Participants ranged in age from 18 years to 70 years and older, with the majority between 40-49 years (54/199, 27.1%) and 18-29 years of age (43/199, 21.6%). Most participants (63/199, 31.7%) had 4 or more years of college or university. Many (54/199, 27.1%) possessed high school diplomas or apprenticeship certificates. Male individuals have a higher BMI; the mean BMI for female individuals (29.33 kg/m²) falls under the overweight category, while male individuals (31.22 kg/m²) fall under the grade I obesity category. Furthermore, when looking at health score by sex, education, smoking, lifestyle change reason, and BMI, male individuals had a higher health score than female, while female individuals had a higher SD. The group of 30- to 39-year-olds had the highest health score, followed by those over 70 years old. The group of 60- to 69-year-olds had the lowest health score. Higher-educated people got better health scores than individuals with less education. Smokers had a poorer health score than nonsmokers. Comparing BMI and health score, those with the lowest BMI had the highest health score.

Table 2 displays the characteristics of respondents to the postintervention questionnaire. Again, female individuals were more represented than male individuals (76/99, 76% vs 23/99, 23%), and the age groups of 40- to 49-year-olds (22/99, 22%) and 18- to 29-year-olds (21/99, 21%) had the most respondents. We observed that some participants belonging to the 70 years or older age group were still using the app and continued to take part in the study. Male and female individuals’ mean BMI after 1 month with the app show that male individuals have a marginally higher BMI than female, while female individuals have a higher SD. People who were 70 years of age or older had the highest health score; however, there were only 5 participants in this group, and therefore precaution is needed in interpreting this result. Interestingly, the participant group of 18- to 29-year-olds had the lowest health score. People with higher education had a better health score than those with lower education. Nonsmokers had better health score than smokers.

The health score scale’s Cronbach α was .92, indicating good reliability.

For the following phase, we focus on pre-post analyses. We matched respondents from both the pre- and postintervention surveys, leaving us with a sample of 74 participants.

A paired sample 2-tailed t test examined BMI and health score pre-post differences. Results are presented in Table 3 below. The data indicate that the BMI and the health score before intervention were higher than after intervention. More specifically, the differences between the pre- and postintervention groups were 0.284 for the BMI and 0.797 for the health score, as displayed in Table 4 below.

However, the results of the paired sample t test of the BMI and health score before and after intervention were not statistically significant (P=.07 for the BMI and P=.24 for the health score).
Table 1. Preintervention descriptive statistics of the participants recruited through Heia Meg: sex, age group, education, BMI, smoking status, reason for lifestyle change.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants (n=199), n (%)</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40 (20.1)</td>
<td>N/A*</td>
</tr>
<tr>
<td>Female</td>
<td>159 (79.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>43 (21.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>30-39</td>
<td>32 (16.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>40-49</td>
<td>54 (27.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>50-59</td>
<td>36 (18.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>60-69</td>
<td>26 (13.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>70 or older</td>
<td>8 (4)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>32 (16.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>High school or certificate of apprenticeship</td>
<td>54 (27.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>University (less than 4 years)</td>
<td>50 (25.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>University (more than 4 years)</td>
<td>63 (31.7)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>BMI (kg/m²) by Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40 (20.1)</td>
<td>31.22 (6.34)</td>
</tr>
<tr>
<td>Female</td>
<td>159 (79.9)</td>
<td>29.33 (5.72)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>199 (100)</td>
<td>29.71 (5.91)</td>
</tr>
<tr>
<td><strong>BMI groups (kg/m²)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5 (underweight)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>18.5-24.9 (normal weight)</td>
<td>66 (33.2)</td>
<td>22.75 (1.60)</td>
</tr>
<tr>
<td>25.0-29.9 (overweight)</td>
<td>61 (30.7)</td>
<td>27.57 (1.53)</td>
</tr>
<tr>
<td>30.0-34.9 (obesity grade I)</td>
<td>48 (24.1)</td>
<td>32.24 (1.28)</td>
</tr>
<tr>
<td>35.0-39.9 (obesity grade II)</td>
<td>12 (6)</td>
<td>36.45 (2.53)</td>
</tr>
<tr>
<td>≥40 (obesity grade III)</td>
<td>12 (6)</td>
<td>43.34 (3.35)</td>
</tr>
<tr>
<td><strong>Health score by Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40 (20.1)</td>
<td>36.90 (6.56)</td>
</tr>
<tr>
<td>Female</td>
<td>159 (79.9)</td>
<td>33.11 (7.69)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>199 (100)</td>
<td>33.87 (7.62)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>43 (21.6)</td>
<td>33.77 (8.61)</td>
</tr>
<tr>
<td>30-39</td>
<td>32 (16.1)</td>
<td>36.25 (7.60)</td>
</tr>
<tr>
<td>40-49</td>
<td>54 (27.1)</td>
<td>32.96 (7.65)</td>
</tr>
<tr>
<td>50-59</td>
<td>36 (18.1)</td>
<td>33.56 (6.71)</td>
</tr>
<tr>
<td>60-69</td>
<td>26 (13.1)</td>
<td>32.77 (6.40)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Participants (n=199), n (%)</td>
<td>Value, mean (SD)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>70 or older</td>
<td>8 (4)</td>
<td>36.00 (9.07)</td>
</tr>
<tr>
<td>Total</td>
<td>199 (100)</td>
<td>33.87 (7.62)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>32 (16.1)</td>
<td>31.25 (6.05)</td>
</tr>
<tr>
<td>High school or certificate of apprenticeship</td>
<td>54 (27.1)</td>
<td>32.07 (7.67)</td>
</tr>
<tr>
<td>University (less than 4 years)</td>
<td>50 (25.1)</td>
<td>34.08 (7.10)</td>
</tr>
<tr>
<td>University (more than 4 years)</td>
<td>63 (31.7)</td>
<td>36.57 (7.93)</td>
</tr>
<tr>
<td>Total</td>
<td>199 (100)</td>
<td>33.87 (7.62)</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18 (9)</td>
<td>31.56 (4.93)</td>
</tr>
<tr>
<td>No</td>
<td>181 (91)</td>
<td>34.10 (7.80)</td>
</tr>
<tr>
<td>Total</td>
<td>199 (100)</td>
<td>33.87 (7.62)</td>
</tr>
<tr>
<td><strong>Reason for lifestyle change</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better sleeping habits</td>
<td>44 (12)</td>
<td>33.55 (8.37)</td>
</tr>
<tr>
<td>Drink less alcohol</td>
<td>17 (4.6)</td>
<td>35.53 (7.33)</td>
</tr>
<tr>
<td>Get in better shape</td>
<td>110 (30.1)</td>
<td>32.80 (7.85)</td>
</tr>
<tr>
<td>Make better nutrition choices</td>
<td>114 (31.1)</td>
<td>33.72 (7.09)</td>
</tr>
<tr>
<td>Mental health</td>
<td>81 (22.2)</td>
<td>34.22 (7.64)</td>
</tr>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5 (underweight)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>18.5-24.9 (normal weight)</td>
<td>66 (33.2)</td>
<td>38.18 (7.60)</td>
</tr>
<tr>
<td>25.00-29.9 (overweight)</td>
<td>61 (30.7)</td>
<td>34.30 (7.02)</td>
</tr>
<tr>
<td>30.00-4.9 (obesity grade I)</td>
<td>48 (24.1)</td>
<td>30.67 (5.04)</td>
</tr>
<tr>
<td>35.0-39.9 (obesity grade II)</td>
<td>12 (6)</td>
<td>26.67 (4.92)</td>
</tr>
<tr>
<td>≥40 (obesity grade III)</td>
<td>12 (6)</td>
<td>28.00 (7.03)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Table 2. Postintervention descriptive statistics of the participants recruited through Heia Meg: sex, age group, education, BMI, smoking status, reason for lifestyle change.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants (n=99), n (%)</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23 (23)</td>
<td>N/Aa</td>
</tr>
<tr>
<td>Female</td>
<td>76 (77)</td>
<td>N/A</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>21 (21)</td>
<td>N/A</td>
</tr>
<tr>
<td>30-39</td>
<td>17 (17)</td>
<td>N/A</td>
</tr>
<tr>
<td>40-49</td>
<td>22 (22)</td>
<td>N/A</td>
</tr>
<tr>
<td>50-59</td>
<td>17 (17)</td>
<td>N/A</td>
</tr>
<tr>
<td>60-69</td>
<td>17 (17)</td>
<td>N/A</td>
</tr>
<tr>
<td>70 or older</td>
<td>5 (5)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>8 (8)</td>
<td>N/A</td>
</tr>
<tr>
<td>High school or certificate of apprenticeship</td>
<td>28 (28)</td>
<td>N/A</td>
</tr>
<tr>
<td>University (less than 4 years)</td>
<td>24 (24)</td>
<td>N/A</td>
</tr>
<tr>
<td>University (more than 4 years)</td>
<td>39 (39)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>BMI (kg/m²) by</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23 (23)</td>
<td>28.62 (5.25)</td>
</tr>
<tr>
<td>Female</td>
<td>76 (77)</td>
<td>28.52 (5.98)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>99 (100)</td>
<td>28.54 (5.79)</td>
</tr>
<tr>
<td><strong>BMI groups (kg/m²)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5 (underweight)</td>
<td>0 (0)</td>
<td>23.24 (1.45)</td>
</tr>
<tr>
<td>18.5-24.9 (normal weight)</td>
<td>35 (35)</td>
<td>27.22 (1.68)</td>
</tr>
<tr>
<td>25.0-29.9 (overweight)</td>
<td>29 (29)</td>
<td>32.19 (1.28)</td>
</tr>
<tr>
<td>30.0-34.9 (obesity grade I)</td>
<td>23 (23)</td>
<td>36.7 (1.38)</td>
</tr>
<tr>
<td>35.0-39.9 (obesity grade II)</td>
<td>5 (5)</td>
<td>42.7 (1.67)</td>
</tr>
<tr>
<td>≥40 (obesity grade III)</td>
<td>7 (7)</td>
<td>44.34 (1.55)</td>
</tr>
<tr>
<td><strong>Health score by</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23 (23)</td>
<td>35.09 (7.37)</td>
</tr>
<tr>
<td>Female</td>
<td>76 (77)</td>
<td>32.25 (7.90)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>99 (100)</td>
<td>32.91 (7.84)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>21 (21)</td>
<td>30.71 (8.69)</td>
</tr>
<tr>
<td>30-39</td>
<td>17 (17)</td>
<td>32.65 (7.57)</td>
</tr>
<tr>
<td>40-49</td>
<td>22 (22)</td>
<td>33.41 (8.99)</td>
</tr>
<tr>
<td>50-59</td>
<td>17 (17)</td>
<td>33.35 (7.57)</td>
</tr>
<tr>
<td>60-69</td>
<td>17 (17)</td>
<td>32.29 (4.81)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Participants (n=99), n (%)</td>
<td>Value, mean (SD)</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 (5)</td>
</tr>
<tr>
<td>70 or older</td>
<td></td>
<td>41.40 (5.77)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>99 (100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.91 (7.84)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary school</td>
<td>8 (8)</td>
</tr>
<tr>
<td></td>
<td>High school or certificate of apprenticeship</td>
<td>28 (28)</td>
</tr>
<tr>
<td></td>
<td>University (less than 4 years)</td>
<td>24 (24)</td>
</tr>
<tr>
<td></td>
<td>University (more than 4 years)</td>
<td>39 (39)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>99 (100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.91 (7.84)</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td>Yes</td>
<td>8 (8)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>91 (92)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>99 (100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.91 (7.84)</td>
</tr>
<tr>
<td><strong>Reason for lifestyle change</strong></td>
<td>Better sleeping habits</td>
<td>25 (14)</td>
</tr>
<tr>
<td></td>
<td>Drink less alcohol</td>
<td>10 (6)</td>
</tr>
<tr>
<td></td>
<td>Get in better shape</td>
<td>54 (31)</td>
</tr>
<tr>
<td></td>
<td>Make better nutrition choices</td>
<td>47 (27)</td>
</tr>
<tr>
<td></td>
<td>Mental health</td>
<td>40 (23)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>99 (100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.91 (7.84)</td>
</tr>
<tr>
<td><strong>BMI (kg/m^2)</strong></td>
<td>&lt;18.5 (underweight)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>18.5-24.9 (normal weight)</td>
<td>35 (35)</td>
</tr>
<tr>
<td></td>
<td>25.0-29.9 (overweight)</td>
<td>29 (29)</td>
</tr>
<tr>
<td></td>
<td>30.0-4.9 (obesity grade I)</td>
<td>23 (23)</td>
</tr>
<tr>
<td></td>
<td>35.0-39.9 (obesity grade II)</td>
<td>5 (5)</td>
</tr>
<tr>
<td></td>
<td>≥40 (obesity grade III)</td>
<td>7 (7)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>99 (100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.91 (7.84)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Table 3. Paired sample 2-tailed t test statistics of BMI and health score.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value, mean (SD)</th>
<th>Value, SE mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BMI (n=74)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>29.05 (5.45)</td>
<td>0.634</td>
</tr>
<tr>
<td>Postintervention</td>
<td>28.77 (5.38)</td>
<td>0.626</td>
</tr>
<tr>
<td><strong>Health score (n=74)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preintervention</td>
<td>34.65 (7.11)</td>
<td>0.826</td>
</tr>
<tr>
<td>Postintervention</td>
<td>33.85 (6.96)</td>
<td>0.809</td>
</tr>
</tbody>
</table>

Table 4. Paired sample 2-tailed t test and paired differences of BMI and health score.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value, mean (SD)</th>
<th>Value, SE mean</th>
<th>Lower-upper, 95% CIs</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>0.284 (1.32)</td>
<td>0.154</td>
<td>-0.023 to 0.590</td>
<td>1.84 (73)</td>
<td>.07</td>
</tr>
<tr>
<td>Health score</td>
<td>0.797 (5.82)</td>
<td>0.676</td>
<td>-0.551 to 2.145</td>
<td>1.179 (73)</td>
<td>.24</td>
</tr>
</tbody>
</table>
Discussion

Participants

The majority of study participants were female, 79.9% (159/199) before intervention and 77% (76/99) after intervention. Is it because female individuals are more interested in improving their health or have more faith in health-related apps or the app’s design? This is a question that deserves further investigation. Another intriguing observation was the high percentage of participants aged 40-69 years (116/199, 58.3% for the preintervention questionnaire and 56/99, 57% for the postintervention questionnaire), as well as the participation of people aged 70 years and older (8/199, 4% and 5/99, 5%, respectively). When the questionnaire was delivered through Heia Meg, it was assumed that younger people would use it. Instead, we believe that for older individuals, the fact that the app was produced and sponsored by the Norwegian Directorate of Health was reassuring and trustworthy. The relatively small sample size in our study does not allow us to generalize these findings.

Most of the study participants had a higher education (113/199, 56.8% in the preintervention group and 63/99, 64% in the postintervention group). The link between the degree of education and health literacy and the engagement of individuals in their health, including participation in research studies such as ours, is worth discussing further. Furnée et al [24] conducted a meta-analysis to determine the marginal impact of education on self-reported health, while Nummela et al [25] examined the relationship between self-reported health and 3 variables of social economic position (disposable household income, self-reported education, and adequacy of income) and 3 categories of communities (rural areas, highly or sparsely papillated areas, and urban areas) among male and female individuals in southern Finland. The adequacy of income had the strongest positive connection with self-reported health in metropolitan locations among all age groups, demonstrating that while real income is a powerful predictor of health, the adequacy of income is even greater [25]. Furnée et al [24] found that a year of education added 0.036 quality-adjusted life-years. Preliminary calculations show that investing in education improves health.

Individuals with high health literacy may make better decisions regarding their health and well-being [26]. Environmental needs and available resources affect health literacy, highlighting the need to enhance it in schools and communities [27]. Individually, health literacy efforts in schools should focus on teaching meta-cognitive abilities such as critical thinking, self-awareness, and citizenship. Health literacy education should include socioeconomic determinants of health and societal processes that lead to health inequities [27].

Most of our study participants wished to improve their diet and fitness. Male individuals’ mean BMI was higher than female individual’s (31.2 kg/m² vs 28.6 kg/m² preintervention and 29.3 kg/m² vs 28.5 kg/m² postintervention). Respondents in the preintervention group had mostly type I obesity, while female individuals were overweight. The intervention proved beneficial in lowering the mean BMI of study participants. Male individuals’ mean BMI decreased more than female individual’s, and in the postintervention survey, both sexes were overweight. After intervention, both sexes ended up in the overweight category as a mean BMI; several participants were within the BMI category of normal weight; none of the participants were underweight; 10 participants were overweight; and 7 had obesity (group III).

Male individuals had a higher health score than female individuals, both in the preintervention observation (36.9 vs 35.09) and in the postintervention questionnaire (33.1 vs 32.25). Interestingly, both sexes experienced a reduction in their health scores between the pre- and postintervention questionnaires. One possible interpretation of this result is that participants gained self-awareness between the 2 questionnaires, and their replies in the postintervention phase were more realistic than preintervention. Moreover, this study was conducted during the COVID-19 pandemic, and society was locked down and reopened numerous times. We foresee the possibility that the fact that Norwegian society was more open when the study began (October 4, 2021) than when the postintervention questionnaire was sent out (November 25, 2021) could have impacted our results. We all experienced changes in our behavior during the lockdown times of the COVID-19 pandemic, and the same could be true for the participants in this study. In a study conducted during the COVID-19 pandemic, Flanagan et al [28] found that eating behavior altered significantly. Cooking at home rose from 4.49 to 5.18 times per week (P<0.001), while eating out fell from 1.98 to 1.08 (P<0.001). Fast-eating assessments improved from 0.04 to 0.81 (P<0.001), indicating a healthier diet. A total of 35.6% of the participants ate unhealthier food in general, and 43.5% ate more unhealthy snacks [28]. Those who ate poorly had a more sedentary lifestyle, less physical exercise, did not go to bed or sleep later, and reported almost doubling in anxiety [28].

When looking at the health score stratified by age, the group of 30- to 39-year-olds had the highest health score before intervention but one of the lowest after interventions. For those 70 years of age and older, their preintervention health score was 36 and 41.40 after intervention. This age group has more stability in their lives, which influences their perceived health, diet, and exercise habits. However, we must highlight that this age group accounted for less than 5% (13/298) of the study sample; therefore, the findings are not generalizable, and further research is needed to confirm these findings or not.

As anticipated within our research group based on existing evidence, we found that those with the lowest level of education also had the lowest health score, while those with the highest level of education had the highest health score. Smokers have poorer health scores than nonsmokers, and this finding is not unexpected considering available evidence. Tobacco use has long been linked to life-threatening diseases [29], and it is one of the top 10 practices that cause global sickness [5,29].

In most low-income nations, socioeconomic status creates health disparities between persons of different income levels. Chronic diseases and behavioral risk factors are more widespread in low-income and low-educated people [30]. Tobacco, sedentary...
habits, poor diets, and alcohol cause 1 million deaths per year in the United States alone, and changing health behaviors is our best hope for reducing global disease and death [5]. Socioeconomic causes and reducing health disparities should be prioritized by public health officers and policy makers [30].

We presented in the Results section the difference in BMI and health score before and after using the app. We saw that the mean preintervention BMI was higher than the mean postintervention BMI, even though the difference was not statistically significant (P = .07). Given the positive change observed in the BMI, while being aware that there are multiple factors beyond the app that could have influenced this change (eg, motivation), we encourage further research extended to larger samples to investigate the effectiveness of similar apps on the BMI.

This project’s goal was to contribute to digital health app research. Over 200 people joined the study through the app, and a response rate of 49.7% (99/199) was observed in the postintervention questionnaire. Participants were given the possibility of making suggestions about the design and functionality of the app. Some of their feedback messages are similar to the following: “Should have been more specific tips”; “I only get push alerts from Heia Meg, not alerts with a visible note, that’s what I need to get motivated from the app”; “Make it clear, what extreme progression you get when first starting a lifestyle change as physical activity, as well as how these progression values for everyday life”; “It’s motivating to motivate others, and some counsel is too simplistic”; or “It’s great, simple, and uncomplicated.” We used the feedback and reported it to the Norwegian Directorate of Health, which used it as a quality improvement tool.

Although the positive effects we observed were not statistically significant, we cannot exclude that this was due to an insufficient statistical power obtained from our relatively small sample and therefore encourage more research from other scholars who have access to other pools of participants.

When a program or intervention is motivated by a health behavior theory, individuals and communities benefit [5]. Never before have health education and behavior change workers had so many possibilities. Mobile phones, laptops, and smartphone apps have the potential to help people improve their lifestyles. Digital interventions can improve awareness of dangers or benefits, provide cues or reminders for healthy habits, and provide encouragement or training to boost confidence. Digital interventions can help increase self-efficacy through mastery and vicarious experiences, as well as social persuasion, emotional, and physiological factors with sending prompts and challenges for engaging in healthy behaviors, as well as the opportunity to follow friends, family, or a role model.

**Limitations**

When recruiting participants, the survey’s inadequacies harmed the data’s validity and dependability. Several participants modified their behavior aim between the first and second questionnaires, reducing the number of participants for a 2-tailed paired t test. In other cases, people filled out the questionnaire multiple times, giving different ages, education levels, and themes. The questionnaire was not sent out until the day after the app was downloaded, but the recruitment of participants pulled in individuals who were truly using the app, as they had to read the message and click the link to the questionnaire. Our recruitment strategy yielded more female than male individuals, while a more balanced sample could have been more generalizable to the general public. Our sample seems skewed toward higher education, and as for sex, a more balanced sample would make it easier to extend our finding to the general public.

The study’s shortcomings include a lack of a control group to determine the cause and effect of the Heia Meg app, as several confounding factors may have influenced our results (eg, the motivation of each participant and all the exogenous factors). Voluntary participation and a restricted number of participants may risk the study’s external validity, as volunteers may have different opinions than the general public.

**Conclusions**

This research project attempted to investigate the effectiveness of the lifestyle app Heia Meg, developed by the Norwegian Directorate of Health, in helping individuals make healthier lifestyle decisions about nutrition and physical activity. Although we did observe some positive effects, the differences were not statistically significant, and therefore we conclude that from our study, it does not appear that the app intervention leads to better nutrition and physical activity choices. This result could be affected by the limitations and confounding factors described above, and further research is needed to confirm, or not confirm, this conclusion.

**Data Availability**

The data sets generated and analyzed during this study are not publicly available due to the requirements of the Norwegian Center for Research Data to delete the data as soon as the processing is over. However, an anonymized version of the data set is available from the corresponding author on reasonable request.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

HBM: health belief model
mHealth: mobile health
NIPH: National Institute of Public Health
NSD: Norwegian Center for Research Data

©Christine Olsen, Daniel Adrian Lungu. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Community-Dwelling Older Adults’ Readiness for Adopting Digital Health Technologies: Cross-Sectional Survey Study

Dietmar Ausserhofer1,2, PhD; Giuliano Piccoliori1, MD; Adolf Engl1, MD; Angelika Mahlknecht1, MD; Barbara Plagg2, PhD; Verena Barbieri1, PhD; Nicoletta Colletti3; Stefano Lombardo3; Timon Gärtner3; Waltraud Tappeiner2, PhD; Heike Wieser2, PhD; Christian Josef Wiedermann1, MD

1Institute of General Medicine and Public Health, Bolzano, Italy
2Claudiana Research, College of Healthcare Professions Claudiana, Bolzano, Italy
3Provincial Institute of Statistics (ASTAT), Bolzano, Italy

Corresponding Author:
Dietmar Ausserhofer, PhD
Institute of General Medicine and Public Health
Lorenz-Böhler-Str. 13
Bolzano, 39100
Italy
Phone: 39 0471 067290
Email: dietmar.ausserhofer@claudiana.bz.it

Abstract

Background: Digital health technologies offer the potential to improve the daily lives of older adults, maintain their health efficiently, and allow aging in place. Despite increasing evidence of benefits and advantages, readiness for adopting digital interventions among older people remains underexplored.

Objective: This study aims to explore the relationships between sociodemographic-, health-, and lifestyle-related factors and technology use in everyday life and community-dwelling older adults’ readiness to adopt telemedicine, smartphones with texting apps, wearables, and robotics.

Methods: This was a cross-sectional, population-based survey study with a stratified probabilistic sample of adults aged 75 years or older living in South Tyrol (autonomous province of Bolzano/Bozen, Italy). A random sample of 3600 community-dwelling older adults living at home was invited to complete a questionnaire including single items (older adults’ readiness to use health technology) and scales (PRISMA-7; Program of Research on Integration of Services for the Maintenance of Autonomy). Descriptive and logistic regression analyses were performed to analyze the data.

Results: In total, 1695 community-dwelling older adults completed the survey (for a response rate of 47%). In terms of potential digital health technology adoption, wearable devices were favored by 33.7% (n=571), telemedicine by 30.1% (n=510), smartphones and texting apps by 24.5% (n=416), and assistant robots by 13.7% (n=232). Sociodemographic-, health- and lifestyle-related factors, as well as the use of technology in everyday life, played a significant role in explaining readiness to adopt digital health technologies. For telemedicine, age ≥85 years (odds ratio [OR] 0.74, 95% CI 0.56-0.96), financial constraints (OR 0.68, 95% CI 0.49-0.95), and less than 2 hours of physical activity per week (OR 0.75, 95% CI 0.58-0.98) were associated with nonreadiness, while Italian-speaking participants (OR 1.54, 95% CI 1.16-2.05) and those regularly using computers (OR 1.74, 95% CI 1.16-2.60), smartphones (OR 1.69, 95% CI 1.22-2.35), and the internet (OR 2.26, 95% CI 1.47-3.49) reported readiness for adoption.

Conclusions: Community-dwelling older adults display varied readiness toward the adoption of digital health technologies, influenced by age, mother tongue, living situation, financial resources, physical activity, and current use of technology. The findings underscore the need for tailored interventions and educational programs to boost digital health technology adoption among community-dwelling older adults.

(JMIR Form Res 2024;8:e54120) doi:10.2196/54120

KEYWORDS
frail older adults; Italy; Italian; Europe; European; digital health; health technologies; telemedicine; telehealth; eHealth; e-health; adoption; readiness; usage; survey; surveys; questionnaire; questionnaires; robotics; readiness; adoption;
Introduction

As global populations age, health care systems worldwide grapple with the challenge of providing personalized, efficient, and integrated care for an increasing proportion of older adults, with the ultimate goal of facilitating "healthy ageing-in-place" [1]. Digital health technologies offer potential advancements to improve various older adults' health outcomes and access to health care services [2]. There is increasing evidence of the benefits of digital health interventions for community-dwelling older adults in terms of physical function, cognitive performance, depression, behavioral and psychological symptoms of dementia, and overall quality of life [3-6]. Digital health technologies, such as smartphone-based mobile apps or wearables, allow the measurement and tracking of clinical parameters, such as pain, fatigue, fever, arrhythmias, slower walking speed, and insufficient physical activity to prevent or improve the management of chronic diseases, frailty, morbidity, and mortality [7]. For instance, in older adults with type 2 diabetes mellitus, digital health interventions (i.e., mHealth) have been shown to be effective in improving cardiometabolic outcomes [8].

Despite technological advancements and the benefits of digital interventions, the adoption of digital health technologies remains low and inconsistent among older persons, suggesting that underlying factors influence their acceptance, adoption, and use [9]. For example, older adults’ adoption of new digital health technologies, such as social robots to combat social isolation and loneliness [10], depends on multiple factors related to technology, psychological, social and personal aspects, costs, and the environment [11]. Personal factors, including life satisfaction, social relationships, self-perception of health, and everyday activities, are integral facets of an individual’s life, especially in their senior years. These aspects not only shape their overall well-being but also their openness to embracing and readiness to adopt and use potentially beneficial digital health technologies [12]. As the intersections of these domains with digital health adoption have been underexplored, particular attention needs to be given to studying the antecedents of digital health technology adoption, including older adults’ readiness [11]. In Italy, a potential digital health gap among older adults due to infrastructural issues and the lack of digital skills have been described, with differences between age groups and educational levels [13]. In a cross-sectional survey study, less than half of the 1002 respondents were aware of telemedicine services in their region [14]. According to health care professionals, some groups of patients experience difficulties in accessing and using digital health technologies due to sociocultural factors, technological and linguistic challenges, and the absence of caregivers [15]. Yet, little is known about the adoption and use of digital health technologies by community-dwelling older adults. In this study, we aimed to explore the relationships between sociodemographics, health- and lifestyle-related factors, and technology use in their everyday life and community-dwelling older adults’ readiness to adopt digital health technologies, that is, telemedicine, smartphones with texting apps, wearables, and robotics. A deeper understanding of these intersections can inform the design, implementation, and evaluation of digital health technology interventions that resonate closely with the needs and preferences of older adults. Such insights can catalyze the development of more target group-oriented digital health strategies, ensuring not only better health outcomes but also improved quality of life for community-dwelling older adults in the digital age.

Methods

Study Design

This cross-sectional, population-based survey study was conducted jointly by the Provincial Institute of Statistics-ASTAT (Istituto Provinciale di Statistica – Landesinstitut für Statistik) and the Institute of General Medicine and Public Health in the Autonomous Province of Bolzano, South Tyrol between March 1 and May 30, 2023.

Setting and Sample

South Tyrol, the autonomous province of Bolzano, is part of the Trentino–Alto Adige region in Italy, next to Austria (total population: 534,912), with approximately 70% German-speaking, 25% Italian-speaking, and 5% other languages. The target population of the survey comprised approximately 51,000 individuals residing in South Tyrol aged 75 years and older. A stratified probabilistic sampling method was used in this study. The ASTAT randomly selected 3600 community-dwelling adults aged ≥75 years, stratified by age (75-84 years, 85 years and older), sex (male and female), and residency (municipalities), from the register of the current resident population in the whole province. To ensure an adequate level of precision, the sampling of the 3600 individuals considered the distribution of and variation between the strata.

Excluded from the survey were individuals permanently residing in senior living facilities and those who, due to health reasons, were unable to complete the questionnaire independently or with the aid of a family member.

Participant Survey

The participant survey was designed collaboratively by the ASTAT and the Institute of General Medicine and Public Health, based on a similar survey study conducted in 2013 [16] and the INSPIRE population survey applied in Switzerland [17]. The German and Italian language versions, translated from the ASTAT, were reviewed for language equity by the research group at the Institute for General Medicine and Public Health. The final versions were checked and approved by the local health and social authorities. As the survey questionnaire included instruments and items previously used (in 2013 edition of the survey and INSPIRE population survey), we conducted pretesting with 10 older adults aged 75 years and older. These participants were invited to complete the questionnaire and to provide oral feedback regarding the clarity of questions,
difficulty of completion, and the technical functionality of the web-based survey to the research group by phone.

Older adults’ readiness for adoption of digital health technologies was assessed using single items, asking to rate on a 5-point Likert scale (yes/maybe/no/I do not understand what this means/I am already using it) the following four technologies: (1) telemedicine, that is, communicating with your doctor via video or smartphone; (2) smartphone with text messaging apps, that is, reminding you of your clinical condition or taking medications, providing information on how to manage your condition; (3) wearable devices, that is, monitoring your heart rate, blood glucose, physical activity, or SOS device; and (4) assistive robots, that is, supporting you at home for taking medications, or recognizing emergencies.

Participants’ sociodemographics included age (birth year), sex (male/female), native tongue (German/Italian/Ladin/Others), citizenship (Italy/other country), educational level (Below Highschool/Highschool or higher), community and region of origin (rural/urban), living situation (alone/with spouse or family member), children (yes/no), financial resources (excellent or good/adequate/insufficient or low), and overall optimism (yes/no).

Health- and lifestyle-related factors included self-reported health status (poor or moderate/good or very good), frailty (no=0-3 points/yes=4-7 points on PRISMA-7 [Program of Research on Integration of Services for the Maintenance of Autonomy], [18]), physical activity (2 hours or more a week/ less than 2 hours a week/ never), use of home care assistance (eg, from family, nursing team, or private family assistant; yes/no).

Technology use in everyday life was assessed by asking participants if they had already used computers, tablets (yes/no), smartphones (yes/no), or the internet (yes/no).

**Data Collection**

Letters were mailed from the ASTAT to the randomly sampled participants to inform them about the study and to invite them to voluntarily participate by completing the survey alone or with the aid of a family member. Completion of the survey was possible by one of the following: (1) web-based self-completion, (2) paper-based self-completion, or (3) telephonic interviews with collaborators from the ASTAT. One month after the first letter, a second letter was sent to inform them about the study and invite them to participate. The web-based survey was created using LimeSurvey [19].

**Ethical Considerations**

Ethical approval was obtained from the institutional board of the Institute of General Practice and Public Health, Bolzano, Italy (reference number: 03/2023). All study procedures were in accordance with the 1964 Helsinki Declaration and its amendments, European Union General Data Protection Regulation (679/2016), and Italian Data Protection Law (196/2003). Before filling out the online questionnaire, participants were explicitly asked to provide informed consent. Filling out the paper questionnaire and sending it back by post was considered as participants’ informed consent. Participation was voluntary. All data of the study participants were anonymized to protect their identities.

**Statistical Analysis**

Only fully completed questionnaires were included in the statistical analysis, resulting in the exclusion of 73 partially filled-out questionnaires. Descriptive statistics (eg, frequency) were calculated to describe the measured variables. To explore differences between the characteristics of older adults and their readiness to adopt digital health technologies (yes and maybe vs no), we used Fisher exact test. Four binary logistic regression models were used to explore the association between each digital health technology, that is, telemedicine, smartphone and texting apps, wearables and robotics (dependent variables), sociodemographics, health- and lifestyle-related factors, and use of technology in everyday life (independent variable). All analyses were performed using R (version 4.3.1; R Foundation for Statistical Computing) and RStudio (version 2023.6.2.561) with the packages tidyr [20] and lme4 [21]. A P value of less than .05 was considered significant.

**Results**

**Sample Characteristics**

In total, 1695 community-dwelling older adults completed the survey, reflecting a response rate of 47%. As described in *Table 1*, the majority of participants were female (880/1695, 51.9%), aged between 75 and 84 years (1005/1695, 59.3%), living with partners or family (1179/1695, 69.6%), had at least 1 child (79.3%), and had an educational level below high school (1319/1695, 77.8%). Health status was reported in more than half as poor or moderate (1012/1695, 59.7%), not frail (999/1695, 58.9%), yet receiving home care assistance from family, nursing team, or a private family assistant (1121/1695, 66.1%). Approximately 1 in 3 older adults mentioned using a computer or tablet (524/1695, 30.9%), smartphone (678/1695, 40%), and the internet (656/1695, 38.7%) in their everyday lives.
Table 1. Characteristics of study participants (n=1695).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>880 (51.9)</td>
</tr>
<tr>
<td>Male</td>
<td>815 (48.1)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>75-84</td>
<td>1005 (59.3)</td>
</tr>
<tr>
<td>≥85</td>
<td>690 (40.7)</td>
</tr>
<tr>
<td><strong>Native tongue</strong></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>902 (53.2)</td>
</tr>
<tr>
<td>Italian</td>
<td>713 (42.1)</td>
</tr>
<tr>
<td>Ladin</td>
<td>67 (4)</td>
</tr>
<tr>
<td>Other</td>
<td>13 (0.8)</td>
</tr>
<tr>
<td><strong>Citizenship</strong></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>1673 (98.7)</td>
</tr>
<tr>
<td>Other</td>
<td>22 (1.3)</td>
</tr>
<tr>
<td><strong>Community</strong></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>773 (45.6)</td>
</tr>
<tr>
<td>Urban</td>
<td>992 (54.4)</td>
</tr>
<tr>
<td><strong>Living situation</strong></td>
<td></td>
</tr>
<tr>
<td>Living alone</td>
<td>516 (30.4)</td>
</tr>
<tr>
<td>Living with partner or family</td>
<td>1179 (69.6)</td>
</tr>
<tr>
<td><strong>Children</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1344 (79.3)</td>
</tr>
<tr>
<td>No</td>
<td>351 (20.7)</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
</tr>
<tr>
<td>Below highschool</td>
<td>1319 (77.8)</td>
</tr>
<tr>
<td>Highschool or higher</td>
<td>376 (22.2)</td>
</tr>
<tr>
<td><strong>Financial resources</strong></td>
<td></td>
</tr>
<tr>
<td>Excellent or good</td>
<td>483 (28.5)</td>
</tr>
<tr>
<td>Adequate</td>
<td>837 (49.4)</td>
</tr>
<tr>
<td>Insufficient or low</td>
<td>375 (22.1)</td>
</tr>
<tr>
<td><strong>Overall optimism</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1433 (84.5)</td>
</tr>
<tr>
<td>No</td>
<td>262 (16.5)</td>
</tr>
<tr>
<td><strong>Health status</strong></td>
<td></td>
</tr>
<tr>
<td>Poor or moderate</td>
<td>1012 (59.7)</td>
</tr>
<tr>
<td>Good or very good</td>
<td>683 (40.3)</td>
</tr>
<tr>
<td><strong>Frailty (PRISMA-7)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>574 (33.9)</td>
</tr>
<tr>
<td>No</td>
<td>1121 (66.1)</td>
</tr>
<tr>
<td><strong>Physical activity</strong></td>
<td></td>
</tr>
<tr>
<td>2 hours or more a week</td>
<td>733 (43.2)</td>
</tr>
<tr>
<td>Less than 2 hours a week</td>
<td>679 (40.1)</td>
</tr>
</tbody>
</table>
Table 2. Older adults’ readiness for health technology adoption (N=1695).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Yes</th>
<th>Maybe</th>
<th>No</th>
<th>I do not understand what this means</th>
<th>Already using it</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telemedicine, n (%)</td>
<td>510 (30.1)</td>
<td>342 (20.2)</td>
<td>669 (39.5)</td>
<td>83 (4.9)</td>
<td>91 (5.4)</td>
</tr>
<tr>
<td>Smartphone and texting apps, n (%)</td>
<td>416 (24.5)</td>
<td>393 (23.2)</td>
<td>757 (44.7)</td>
<td>87 (5.1)</td>
<td>42 (2.5)</td>
</tr>
<tr>
<td>Wearable devices, n (%)</td>
<td>571 (33.7)</td>
<td>541 (31.9)</td>
<td>489 (28.8)</td>
<td>60 (3.5)</td>
<td>34 (2)</td>
</tr>
<tr>
<td>Assistant robots, n (%)</td>
<td>232 (13.7)</td>
<td>459 (27.1)</td>
<td>888 (52.4)</td>
<td>108 (6.4)</td>
<td>8 (0.5)</td>
</tr>
</tbody>
</table>

Group Differences in Older Adults’ Readiness for Adopting Digital Health Technologies

The tables in Multimedia Appendix 1 present the sociodemographics of the participants and highlight group differences between those who are ready to adopt digital health technology (either responded “Yes” or “Maybe”) and those who responded “No”. In summary, age, educational level, living situation, financial resources, physical activity, and technology use in everyday life significantly influenced older adults’ readiness to integrate various digital health technologies into their lives.

Age plays a key role in digital health technology adoption readiness. In particular, 72.4% (n=697) of older adults between the ages of 75 and 84 years showed readiness to adopt wearables. Living arrangements also had an impact on digital health technology adoption, with individuals living with a partner or family showing higher readiness to integrate all forms of digital health technologies than their counterparts living alone. Similarly, older adults with an educational level of high school or higher; those reporting having adequate financial resources; being physically active; and older adults already using computers, tablets, smartphones, and the internet reported higher readiness to adopt all 4 digital health technologies. In terms of gender differences, except for robotics, more than half of the men reported being ready to adopt digital health technologies and higher readiness than women to use telemedicine and smartphones. Italian-speaking adults living in urban areas who were more optimistic about the future, with good or very good health status, and were not frail, were more likely to adopt 2 out of 4 digital health technologies, namely telemedicine and smartphones with texting apps. Regarding frailty as assessed by PRISMA-7, we observed that nonfrail individuals reported higher readiness to adopt telemedicine, as well as smartphone and texting apps compared with frail individuals.
Factors Influencing Older Adults’ Readiness for Adopting Digital Health Technologies

Tables 3 and 4 present the results of multiple logistic regression analyses examining the association between older people’s readiness to use 4 digital health technologies (telemedicine, smartphone with texting app, wearables, and assistant robot) and various sociodemographic, health- and lifestyle-related factors, and use of technology in everyday life.

Table 3. Multiple logistic regression analyses between older people’s readiness to use telemedicine, smartphones with texting apps and sociodemographics, health- and lifestyle-related factors, and use of technology in everyday life (n=1521).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Telemedicine ORa (95% CI)</th>
<th>P value</th>
<th>Smartphone with texting app OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociodemographic factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (reference group: male)</td>
<td>1.11 (0.87-1.43)</td>
<td>.40</td>
<td>1.00 (0.78-1.28)</td>
<td>.90</td>
</tr>
<tr>
<td>Age ≥85 years (reference group: age 75-84 years)</td>
<td>0.74 (0.56-0.96)</td>
<td>.03</td>
<td>0.75 (0.57-0.98)</td>
<td>.03</td>
</tr>
<tr>
<td>Native tongue (reference group: German)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>1.54 (1.16-2.05)</td>
<td>.003</td>
<td>1.41 (1.07-1.87)</td>
<td>.02</td>
</tr>
<tr>
<td>Ladin and others</td>
<td>1.55 (0.90-2.69)</td>
<td>.11</td>
<td>1.16 (0.67-2.00)</td>
<td>.60</td>
</tr>
<tr>
<td>Rural community (reference group: urban community)</td>
<td>1.03 (0.78-1.36)</td>
<td>.80</td>
<td>0.99 (0.76-1.30)</td>
<td>.90</td>
</tr>
<tr>
<td>Living alone (reference group: living with family)</td>
<td>0.95 (0.74-1.23)</td>
<td>.70</td>
<td>0.95 (0.74-1.23)</td>
<td>.70</td>
</tr>
<tr>
<td>Educational level (reference group: below high school)</td>
<td>1.18 (0.86-1.63)</td>
<td>.30</td>
<td>1.06 (0.78-1.45)</td>
<td>.70</td>
</tr>
<tr>
<td>Financial resources (reference group: excellent or good)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate</td>
<td>0.84 (0.64-1.10)</td>
<td>.20</td>
<td>0.95 (0.73-1.25)</td>
<td>.70</td>
</tr>
<tr>
<td>Insufficient or low</td>
<td>0.68 (0.49-0.95)</td>
<td>.02</td>
<td>0.83 (0.60-1.16)</td>
<td>.30</td>
</tr>
<tr>
<td>Health-related factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall optimism (reference: no)</td>
<td>1.22 (0.88-1.68)</td>
<td>.20</td>
<td>1.21 (0.88-1.66)</td>
<td>.30</td>
</tr>
<tr>
<td>Good or very good health status (reference: poor or moderate)</td>
<td>0.98 (0.75-1.28)</td>
<td>.90</td>
<td>0.96 (0.74-1.24)</td>
<td>.70</td>
</tr>
<tr>
<td>No frailty (reference: frailty)</td>
<td>0.82 (0.59-1.15)</td>
<td>.30</td>
<td>0.88 (0.63-1.23)</td>
<td>.50</td>
</tr>
<tr>
<td>Physical activity (reference group: 2 hours or more a week)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 2 hours a week</td>
<td>0.75 (0.58-0.98)</td>
<td>.034</td>
<td>0.81 (0.63-1.05)</td>
<td>.12</td>
</tr>
<tr>
<td>Never</td>
<td>0.72 (0.50-1.03)</td>
<td>.071</td>
<td>0.59 (0.41-0.85)</td>
<td>.004</td>
</tr>
<tr>
<td>Home care assistance (reference group: no assistance)</td>
<td>0.77 (0.58-1.02)</td>
<td>.065</td>
<td>0.90 (0.68-1.18)</td>
<td>.40</td>
</tr>
<tr>
<td>Technology-related factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of computer, tablet (reference: no use)</td>
<td>1.74 (1.16-2.60)</td>
<td>.008</td>
<td>1.55 (1.05-2.28)</td>
<td>.03</td>
</tr>
<tr>
<td>Use of smartphone (reference: no use)</td>
<td>1.69 (1.22-2.35)</td>
<td>.002</td>
<td>1.68 (1.23-2.29)</td>
<td>.001</td>
</tr>
<tr>
<td>Use of internet (reference: no use)</td>
<td>2.26 (1.47-3.49)</td>
<td>&lt;.001</td>
<td>2.66 (1.77-4.03)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aOR: odds ratio.
Table 4. Multiple logistic regression analyses between older people’s readiness to use wearables, assistant robots and sociodemographics, health- and lifestyle-related factors, and use of technology in everyday life (n=1521).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Wearables OR (95% CI)</th>
<th>P value</th>
<th>Assistant robot OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographic factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female sex (reference group: male)</td>
<td>1.25 (0.98-1.60)</td>
<td>.07</td>
<td>1.02 (0.82-1.28)</td>
<td>.80</td>
</tr>
<tr>
<td>Age ≥85 years (reference group: 75-84 years)</td>
<td>0.78 (0.59-1.01)</td>
<td>.06</td>
<td>0.71 (0.55-0.92)</td>
<td>.009</td>
</tr>
<tr>
<td><strong>Native tongue (reference group: German)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>1.05 (0.80-1.39)</td>
<td>.70</td>
<td>0.93 (0.72-1.20)</td>
<td>.60</td>
</tr>
<tr>
<td>Ladin and others</td>
<td>0.99 (0.59-1.68)</td>
<td>.90</td>
<td>1.40 (0.86-2.28)</td>
<td>.20</td>
</tr>
<tr>
<td>Rural community (reference group: urban community)</td>
<td>1.03 (0.79-1.35)</td>
<td>.80</td>
<td>1.05 (0.82-1.35)</td>
<td>.70</td>
</tr>
<tr>
<td>Living alone (reference group: living with family)</td>
<td>0.75 (0.59-0.96)</td>
<td>.02</td>
<td>0.84 (0.66-1.06)</td>
<td>.15</td>
</tr>
<tr>
<td>Educational level (reference group: below high school)</td>
<td>1.08 (0.79-1.48)</td>
<td>.60</td>
<td>0.95 (0.72-1.26)</td>
<td>.70</td>
</tr>
<tr>
<td><strong>Financial resources (reference group: excellent or good)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate</td>
<td>0.99 (0.76-1.30)</td>
<td>.90</td>
<td>0.84 (0.65-1.07)</td>
<td>.20</td>
</tr>
<tr>
<td>Insufficient or low</td>
<td>0.76 (0.55-1.04)</td>
<td>.09</td>
<td>0.85 (0.62-1.15)</td>
<td>.30</td>
</tr>
<tr>
<td><strong>Health-related factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall optimism (reference: no)</td>
<td>0.98 (0.71-1.34)</td>
<td>.90</td>
<td>1.08 (0.80-1.45)</td>
<td>.60</td>
</tr>
<tr>
<td>Good or very good health status (reference: poor or moderate)</td>
<td>0.88 (0.68-1.14)</td>
<td>.30</td>
<td>1.01 (0.79-1.27)</td>
<td>.90</td>
</tr>
<tr>
<td>No frailty (reference: frailty)</td>
<td>0.79 (0.57-1.10)</td>
<td>.20</td>
<td>0.94 (0.69-1.28)</td>
<td>.70</td>
</tr>
<tr>
<td><strong>Physical activity (reference group: 2 hours or more a week)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 2 hours a week</td>
<td>0.79 (0.61-1.02)</td>
<td>.08</td>
<td>0.87 (0.69-1.11)</td>
<td>.30</td>
</tr>
<tr>
<td>Never</td>
<td>0.46 (0.33-0.65)</td>
<td>&lt;.001</td>
<td>0.76 (0.54-1.06)</td>
<td>.11</td>
</tr>
<tr>
<td>Home care assistance (reference group: no assistance)</td>
<td>0.64 (0.48-0.85)</td>
<td>.002</td>
<td>0.63 (0.48-0.81)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Technology-related factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of computer, tablet (reference: no use)</td>
<td>1.27 (0.85-1.88)</td>
<td>.20</td>
<td>1.18 (0.84-1.68)</td>
<td>.30</td>
</tr>
<tr>
<td>Use of smartphone (reference: no use)</td>
<td>1.36 (0.98-1.89)</td>
<td>.07</td>
<td>1.31 (0.97-1.78)</td>
<td>.08</td>
</tr>
<tr>
<td>Use of internet (reference: no use)</td>
<td>1.19 (0.77-1.83)</td>
<td>.40</td>
<td>1.30 (0.88-1.93)</td>
<td>.20</td>
</tr>
</tbody>
</table>

^OR: odds ratio.

As reported in Table 3, higher age and insufficient financial resources were associated with lower readiness to adopt telemedicine, as well as being active for less than 2 hours a week. Italian-speaking older adults and everyday users of computers or tablets, smartphones, and internet reported significantly higher readiness to adopt telemedicine.

For smartphones with texting apps, adults aged ≥85 years with no physical activity at all were reported to be less likely to adopt this technology, while users of computers, tablets, smartphones, and the internet reported significantly higher readiness to adopt telemedicine.

Living alone, never performing physical activity, and being in need of home care assistance were all associated with lower readiness to adopt wearables. Regarding assistant robots, older adults aged aged ≥85 years and those already receiving home care assistance were less likely to adopt this digital health technology.

Discussion

Principal Findings

The integration of digital health technologies in health care has gained increasing prominence in recent years to improve access to and delivery of services, especially among older adults, who often face unique health care needs. With this cross-sectional, population-based survey study, we aimed to investigate the readiness of community-dwelling older adults in South Tyrol, Italy, to adopt 4 types of digital health technologies. Our study revealed that more than half of the older adults reported readiness to adopt telemedicine and wearable devices. However, a large portion responded with nonreadiness to adopt digital health technologies, that is, smartphones with texting apps or assistant robots. Sociodemographic, health- and lifestyle-related factors, and the use of technology in everyday life played a significant role in explaining older adults’ readiness to adopt digital health technologies.
Our finding on community-dwelling older adults’ readiness to adopt digital health technologies in South Tyrol aligns with a growing body of literature that highlights the increasing acceptance and willingness of older adults to embrace digital health care solutions [22]. In this context, it is evident that readiness to adopt digital health technologies, as observed in South Tyrol’s older population, is part of a more extensive global shift toward digital health technology use. This shift in attitude may be attributed to various factors, including the greater accessibility and user-friendliness of digital devices and health care apps as well as the increasing emphasis on and necessity to use technological solutions during the COVID-19 pandemic [23]. As the use of technology has increasingly become a crucial element of everyday life, older adults seem to embrace the potential benefits of digital health technologies to meet their health care needs, such as improved access to services, enhanced convenience, and better management of chronic conditions, thus playing a crucial role in fostering healthy aging, enhancing social inclusion, and facilitating independent living [24].

Our study revealed a higher adoption readiness for wearables and telemedicine compared with smartphones, texting apps, and assistant robots. One possible explanation for this is the perceived direct health benefits and simplicity associated with wearables and telemedicine. Telemedicine directly addresses the challenges of physical mobility, frequent clinic visits, and waiting times, thereby offering a convenient alternative [23]. Wearables, often designed with user-friendly interfaces, provide real-time health monitoring and a sense of security to older people. In contrast, texting apps may seem less intuitive to a generation that is less accustomed to digital communication, with the added layer of concerns about miscommunication or misunderstanding medical advice. Moreover, older adults are cautious about sharing health-related information online because of privacy and security concerns [25]. Assistant robots, being relatively newer innovations, might invoke apprehensions regarding complexity, safety, and potential dependency. While older adults acknowledge the future potential, readiness to adopt assistant robots is still low because of perceived barriers, such as mismatch between needs and solutions offered by the robots, usability factors, and lack of experience with technology [26].

Our results demonstrate that sociodemographics, health- and lifestyle-related factors, and the use of technology in everyday life play a significant role in shaping older adults’ readiness to adopt digital health technologies. While several sociodemographic, health, and lifestyle-related factors showed significant relationships in simple regression models, their associations were often attenuated or eliminated in multiple regression models. For instance, lower age (below 85 years) was the most consistent sociodemographic factor explaining higher readiness to adopt 3 out of 4 digital health technologies, while higher education or financial resources played a less important role. In our study, Italian-speaking older adults were more inclined to adopt telemedicine, which might be related to higher awareness or potentially more available Italian-language information and campaigns, for example, in television, promoting digital health technology. As language mirrors cultural, societal, and regional differences within South Tyrol, it may play an important role in the acceptance and use of digital health technologies. Recognizing the influence of linguistic and cultural differences on digital health technology adoption can guide future interventions by emphasizing multilingual support and culturally tailored educational and outreach programs.

The association between technology use in everyday life (ie, computer/tablet, smartphone, and internet) and older adults’ readiness to adopt digital health technologies emphasizes the importance of technology literacy, that is, the ability to use, comprehend, manage, and analyze technology safely, effectively, and responsibly. The pronounced readiness among “tech users” in the older age underscores the importance of technological familiarity. Recent research from China has revealed that overcoming technology anxiety is essential in enhancing the adoption and continued use of wearable health technologies among older adults [27]. While developing credible digital health apps that require minimal internet navigation skills, patient education, and collaborative efforts to address access and affordability are urgently warranted [28], tailored training sessions or workshops on basic technology use could substantially bridge the readiness gap. Similarly, health care professionals need to be prepared for the adoption and use of digital health technologies. A recent survey among clinical and nonclinical staff in general practice in South England revealed a significant difference between the self-reported competence in using digital health technologies, with the lowest readiness for using clinical apps and wearables [29].

**Future Implications**

Collaboration between all stakeholders, that is, health care providers, technology developers, policymakers, and the older population, is essential to promote the successful integration of digital health technologies into the health care system for older adults in South Tyrol and beyond. Policymakers and health care providers should consider disparities when designing and implementing digital health interventions, and ensure that technologies, including telemedicine, are accessible and affordable to all older adults to avoid health inequities with inequality in the distribution of health care resources among different populations [30]. Moreover, the use of sensors and wearables for remote monitoring as a source of information for chronic disease management needs to be integrated into primary health care processes so that information with clinical value is not lost along the way and by ensuring data protection [31]. Design adaptations and well-thought-out blended alternatives (ie, combining telemedicine with face-to-face support) may be potential solutions to improve older adults’ user engagement with digital health technologies [32].

While our study aimed to describe and explore more in-depth the experiences of individuals aged 75 years and above, future research should focus on evaluating readiness, actual adoption, and use of digital health technologies among older adults of different ages (eg, 55 years and above) over time. Qualitative research is needed to gain an in-depth understanding of the potential facilitators and barriers to adopting and using digital health technologies, which can guide health care providers, technology developers, and policymakers in creating more targeted and user-friendly digital health technology solutions.
for older adults, ensuring higher adoption rates and beneficial health outcomes. Education and awareness campaigns should address the concerns of older adults, such as safeguards in place to protect their data. Addressing these concerns is crucial for building trust in digital health technologies.

**Limitations**
This study has some limitations that need to be considered. First, its cross-sectional design limits our ability to establish causality or track changes in readiness over time. Longitudinal studies are needed to better understand how attitudes toward digital health technologies evolve among older adults. Additionally, the survey was conducted in South Tyrol, Italy, which may limit the generalizability of our findings to other regions or countries with different health care systems and sociocultural contexts. The participants primarily came from a specific age group, predominantly aged 75-84 years. This narrow age range does not offer a comprehensive picture of the general population’s attitudes across a broader age spectrum. We excluded older adults living in residential care facilities as they are more care-dependent and receive 24/7 institutional care, which might limit their readiness, ability, and need to use digital health technologies. Although the response rate (47%) can be considered excellent, this implies that over half of the potential participants did not respond. There could be systematic reasons for nonresponses, which might have introduced bias. Although we aimed to capture a wide range of community-dwelling older adults, we acknowledge that the findings may not be fully representative, as older adults less inclined or unable to embrace technology might not have participated in this study. Finally, the results are quantitative, which, while providing clear metrics, might not explore deeper reasons or barriers behind older adults’ hesitations or willingness to adopt digital health technologies.

**Conclusion**
This cross-sectional survey study sheds light on the readiness of community-dwelling older adults aged 75 years or older in South Tyrol, Italy, to adopt digital health technologies. While older adults reported being open to embracing these technologies, readiness to adopt telemedicine or wearables was higher than readiness to adopt smartphones with texting apps or assistant robots. Readiness to adopt digital health technologies among older adults was explained by sociodemographic, health-related, and lifestyle factors, with higher age consistently associated with less willingness to adopt digital health technologies, particularly among individuals aged 85 years and older. The use of computers, smartphones, and the internet in everyday life appears to be the most significant driver of readiness to adopt digital health technologies. Key stakeholders, including health care providers, technology developers, and policymakers, require a nuanced understanding of the readiness of older adults for digital health technologies. Among older people, tailored approaches addressing the needs of diverse demographic segments could further enhance the adoption rates. Collaborative initiatives, blending technological innovation with elder-friendly interfaces, and protective policies are vital for the seamless integration of digital health technologies into the lives of older people. Further, longitudinal studies are needed to explore how attitudes toward and the use of digital health technologies evolve among older adults over time. Qualitative research is necessary to gain a deeper understanding of the differential adoption readiness of older adults, including their language differences.

**Data Availability**
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

**Authors’ Contributions**
DA, GP, AE, NC, SL, and TG developed the idea for the study. DA, GP, AE, NC, SL, TG, AM, BP, VB, HW, WT, and CJW contributed to the concept, design, and data collection. DA, GP, AE, NC, SL, and TG contributed to the analysis and interpretation of the data. DA, CJW, HW, and WT contributed to the drafting of the manuscript. All authors contributed to the critical revision of the manuscript and approved the final version.

**Conflicts of Interest**
None declared.

Multimedia Appendix 1
Supplementary tables.

[PDF File (Adobe PDF File), 120 KB - formative_v8i1e54120_app1.pdf ]

**References**
1. WHO’s work on the UN decade of healthy ageing (2021–2030). World Health Organization. URL: [https://www.who.int/initiatives/decade-of-healthy-ageing] [accessed 2024-03-29]

https://formative.jmir.org/2024/1/e54120


https://formative.jmir.org/2024/1/e54120 JMIR Form Res 2024 | vol. 8 | e54120 | p.2038 (page number not for citation purposes)


Abbreviations

ASTAT: The Provincial Institute of Statistics (Istituto Provinciale di Statistica – Landesinstitut für Statistik)
OR: odds ratio
PRISMA-7: Program of Research on Integration of Services for the Maintenance of Autonomy
An Exploration of the Goodness of Fit of Web-Based Tools for Māori: Qualitative Study Using Interviews and Focus Groups

Liesje Donkin1*, BSc, MSc, PGDip (Health Psychology), PGDipArtsDist (Psychology), PGDip (Clinical Psychology), PhD; Marie-Claire Bidois-Putt2*, BSocSci (Hons), MSc; Penelope Hayward1, BA, MA, PGDipArts (Psychology), PGDip (Psychological Practice); Amy Hai Yan Chan2, BPharm (Hons), PhD

1Department of Psychology and Neuroscience, Auckland University of Technology, Auckland, New Zealand
2School of Pharmacy, The University of Auckland, Auckland, New Zealand
*these authors contributed equally

Corresponding Author:
Liesje Donkin, BSc, MSc, PGDip (Health Psychology), PGDipArtsDist (Psychology), PGDip (Clinical Psychology), PhD
Department of Psychology and Neuroscience
Auckland University of Technology
Akoranga Drive
Northcote
Auckland, 0627
New Zealand
Phone: 64 21847886
Email: liesje.donkin@aut.ac.nz

Abstract

Background: Indigenous communities often have poorer health outcomes and services under traditional models of care. In New Zealand, this holds true for Māori people who are tūanga whenua (the indigenous people). Several barriers exist that decrease the likelihood of indigenous communities often have poorer health outcomes and poor service fit under traditional models of care, including access issues, systemic and provider racism, and a lack of culturally safe and responsive services. Web-based interventions (WBIs) have been shown to be effective in supporting mental health and well-being and can overcome some of these barriers. Despite the large number of WBIs developed, more investigation is needed to know how well WBIs fit with an indigenous worldview and how they meet the needs of indigenous communities so that a digitally based future does not drive social and health inequities.

Objective: This study aims to explore the goodness-of-fit of WBIs of Māori individuals, the indigenous people of Aotearoa/New Zealand.

Methods: We used interviews (n=3) and focus groups (n=5) with 30 Māori participants to explore their views about WBIs. Interviews were analyzed using reflexive thematic analysis by members of the research team.

Results: Overall, there was a perception that the design of WBIs did not align with the Māori worldview, which centers around people, relationships, spirituality, and holistic views of well-being. A total of 4 key themes and several subthemes emerged, indicating that WBIs were generally considered a poor fit for Māori. Specifically, the themes were as follows: (1) WBIs are disconnected from the core values of te ao Māori (the Māori worldview), (2) WBIs could be helpful in the right context, (3) there are significant barriers that may make it harder for Māori to use WBIs than other groups, and (4) ways to improve WBIs to help engagement with Māori.

Conclusions: While WBIs are often considered a way to reduce barriers to care, they may not meet the needs of Māori when used as a stand-alone intervention. If WBIs are continued to be offered, developers and researchers need to consider how to develop WBIs that are responsive and engaging to the needs of indigenous communities rather than driving inequities. Ideally, WBIs should be developed by the people they are intended for to fit with those populations’ world views.

(JMIR Form Res 2024;8:e50385) doi:10.2196/50385

KEYWORDS
Indigenous people; Māori; eHealth; mental health; web-based intervention; digital intervention
Introduction

Background

Web-based interventions (WBIs) are therapeutic interventions delivered over technological means such as mobile phones or the internet to facilitate behavior change to improve health or well-being. Over the past 2 decades, the explosion in the number of WBIs has paralleled the growth and increased access to technology worldwide. This growth is driven by the potential of WBIs to reach more people at a lower cost than comparable face-to-face therapies. Several reviews now report that WBIs are as effective as face-to-face interventions [1-4], making WBIs a viable alternative to conventional therapies.

Indigenous communities often face more significant barriers to accessing health care than the dominant cultural group [5-9]. The barriers include systemic racism [10-14], negative attitudes held by health care professionals about indigenous people and traditional methods of healing [14-17], a lack of culturally responsive and safe services [18-20], and significant psychosocial barriers that make even getting to appointments difficult [21]. Due to these barriers and the social inequities created by colonization [22-24], indigenous people often experience worse health and social, psychological, and mortality outcomes. Given the potential benefits of WBIs to overcome some of these barriers, WBIs are increasingly being developed for the hard to reach or engage groups [25-27], such as indigenous communities.

However, one of the reasons that traditional health care underserves indigenous communities is that the care delivery model is often driven by a Western or dominant cultural paradigm [17], which often does not fit with the values and worldview of the indigenous community [28,29]. Thus, existing Western and biomedical models may further perpetuate health inequities [30]. For example, many indigenous communities such as Māori (the indigenous people of New Zealand) [31] hold a collectivist worldview rather than the individualized view prevalent in Western medicine [31,32]. In te ao Māori (the Māori worldview), the relationship and generational history between 2 people are paramount and influence engagement with others. Similarly, wairua (spirituality), Mauri (life force), whakapapa (history), and tikanga (ways of doing things) are paramount in all interactions. These factors are often deprioritized in the medical model, where an adequate patient-clinician relationship is assumed. Similarly, there is often a belief that this relationship will progress within what is largely superficial, impersonal and not responsive to the patient’s cultural needs nor recognizes the history and impact of colonization [33].

As the uptake and persistence of WBIs are often considered low, particularly on a long-term basis [34,35], it is essential to understand the views of potential users to inform the design and determine if WBIs are the best form of intervention. This has been frequently explored with nonindigenous populations, which found that time constraints and lack of perceived worth of the WBI were key reasons for stopping using WBIs [35]. In indigenous communities, uptake of WBIs has been found to be variable and can be affected by intervention characteristics [36,37]. In addition to this, many indigenous communities have poor access to technology and technological infrastructure [38], which can create further barriers to accessing WBIs. Understanding barriers to uptake can help researchers, developers, and policy makers consider how to improve WBIs for indigenous populations or if WBIs are even an appropriate intervention strategy for indigenous communities—because without this understanding, investment in technology may further drive health inequities.

Objective

Currently, there is limited research exploring the views of Māori adults about WBIs and the fit of WBIs with te ao Māori. Given this, this study sought to understand the views of Māori adults about WBIs using a qualitative design and reflexive thematic analysis of interviews and focus groups.

Methods

Overview

This study used a qualitative methodology using a mix of web-based videoconferencing interviews, face-to-face interviews, and focus groups. All interviews were facilitated by MCBP, who also recruited participants using snowball recruitment and social media advertising.

Participants

People were recruited into the study if they identified as Māori, were aged >18 years, and could consent to participation either orally or in a written format. There were no exclusion criteria for this study.

Recruitment

Participants were recruited through a range of recruitment methods using convenience sampling. Participants for 1 focus group (n=8) were recruited through Te Kete Pounamu, a nationally based organization for Māori with lived experience of mental distress and addiction. The remainder of the participants were recruited via the researchers’ professional and personal networks, web-based advertisements, and through relationships formed in the community by research team members.

Procedure

The participants were offered a choice of participating in an individual interview or a focus group discussion. During the initial stages of the meeting with the participants, the interviewer (MBP) opened the session with karakia (prayer) if the participants wished for this to happen. The interviewer then engaged with participants in whanaungatanga (building relationships through shared connection). The participants were then reoriented to the study, and consent was obtained in a written or recorded verbal format. The participants then confirmed that they were happy to be recorded, and the interview or focus group commenced.

Interviews initially started by providing the participants with a definition of WBIs and then asking participants about their views of WBIs. If needed, the participants were prompted to talk about why they felt WBIs did or did not fit with their...
worldview. A specific example of an existing WBI was used as a talking stimulus by demonstrating the WBI if required. At the end of the session, participants were thanked for their time, and the session was closed with karakia if it was opened with one.

Data Gathering

Given that Māori have a strong oral history and prefer to engage kanohi-ki-te-kanohi (face-to-face), interviews and focus groups were used. Combining both methods meant that participants could be part of a group to share their views or talk individually. Consultation with Te Kete Pounamu indicated that their tāngata whaiora (people seeing health) would likely prefer to engage in a focus group. Thus, focus groups were offered to allow this. Individual interviews were offered to enable flexibility in interview times at a time and place that suited individuals. Group interviews were complete kanohi-ki-te-kanohi, while individual interviews were a mix of kanohi-ki-te-kanohi and on the web. Both interviews and group interviews used the same semistructured interview guide with questions to facilitate reflective discussion. Examples of questions used were as follows: “What are your views on digital interventions to support mental health and well-being?” “What do you think has led to you developing these views?” “How does the use of digital interventions fit your background and culture?”

A total of 8 transcripts were produced, consisting of 3 individual interviews and 5 interviews with >2 participants. A total of 30 people were interviewed as part of the study. The interviews were manually transcribed (by HW); checked (by LD); and coded using inductive reflexive thematic analysis by Clarke et al [39] by 2 members (LD and MCBP) of the research team individually and following a 6-step process of familiarization, coding, generating themes, reviewing themes, defining and naming themes, and writing up. Reflexive thematic analysis was chosen as an appropriate methodology for this study as it is aligned with recent projects seeking to understand the views of Māori [40-43] as tāngata whaiora (those seeking health).

The transcripts were not returned to the participants for review. Any disagreements in coding between LD and MCBP were resolved by discussion between LD, MCBP, and PH. Codes, subthemes, and themes were combined on the web for review. The wider team (LD, MCBP, HW, and PH) was chosen as an appropriate methodology for this study as it is aligned with recent projects seeking to understand the views of Māori [40-43] as tāngata whaiora (those seeking health).

The participants had the option to provide scanned copies of the signed consent form to the interviewer if they preferred this method over video consent.

Ethical Considerations

Ethics approval for the study was received from the Auckland Health Research Ethics Committee (AHREC AH23110; expiry October 18, 2024).

Written consent was provided by all participants who completed face-to-face interviews or focus groups following a review of the participant information sheet. In web-based interviews, an oral consent protocol was followed where participants were video recorded giving their consent. The researchers then completed the consent forms based on participant responses, and consents were electronically provided to the participants. The participants had the option to provide scanned copies of the signed consent form to the interviewer if they preferred this method over video consent.

One participant completed their interview or focus group, the recordings were transcribed (completed by MCBP or HW), and all identifying information was removed from the transcript. The deidentified transcripts were provided to the wider research team for review. LD led the review process with MCBP. Codes and themes were discussed with HW, AHYC, and PH once coding was complete. The original recordings were stored separately from the transcripts in a secure manner on a password-protected university-managed research drive aligned with the data management plan.

During the research process, Koha (a gift acknowledging the time spent by participants and sharing their knowledge) in the form of NZ $20 (US $11.87) supermarket vouchers and kai (food) were provided to participants.

Results

Participants

A total of 30 participants participated in this research across 5 group interviews and 3 individual interviews (Table 1).

In terms of participants, ages ranged from 18 to 74 years, with the mean participant age being 41.3 (SD 19.6) years. Of the 30 participants, 11 (37%) identified as a man, 12 (40%) as a woman, 6 (20%) did not specify their gender, and 1 (3%) identified as nonbinary. The participants lived in a mix of main urban centers, such as Auckland and Hamilton, and in more urban areas, such as Northland and Rotorua.

<table>
<thead>
<tr>
<th>Transcript number</th>
<th>Format</th>
<th>Label</th>
<th>Participants, n (%)</th>
<th>Duration (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Group</td>
<td>Group 1</td>
<td>12 (40)</td>
<td>184</td>
</tr>
<tr>
<td>2</td>
<td>Individual</td>
<td>Interview 1</td>
<td>1 (3)</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>Group</td>
<td>Group 2</td>
<td>8 (27)</td>
<td>180</td>
</tr>
<tr>
<td>4</td>
<td>Group</td>
<td>Group 3</td>
<td>3 (10)</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>Individual</td>
<td>Interview 2</td>
<td>1 (3)</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>Individual</td>
<td>Interview 3</td>
<td>1 (3)</td>
<td>53</td>
</tr>
<tr>
<td>7</td>
<td>Group</td>
<td>Group 4</td>
<td>2 (7)</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>Group</td>
<td>Group 5</td>
<td>2 (7)</td>
<td>68</td>
</tr>
</tbody>
</table>
Findings

The participants were largely aligned with their views on WBIs. Although they often offered a critical perspective on WBIs, they also quickly indicated that others might find WBIs helpful even though they did not believe that WBIs would be useful for themselves. Through the analysis of the transcripts, 4 themes emerged, as listed in Table 2.

These themes and subthemes are discussed in more detail in the subsequent sections.

Table 2. Themes and subthemes generated from the data.

<table>
<thead>
<tr>
<th>Theme number</th>
<th>Theme</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WBIs(^a) are disconnected from the core values of te ao Māori (the Māori worldview)</td>
<td>• Te ao Māori is about wairua, and WBIs cannot replicate this.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Māori models of health are holistic, whereas WBIs are singular in focus.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• WBIs could be a strand in the weaving of a well-being kete (basket).</td>
</tr>
<tr>
<td>2</td>
<td>WBIs could be helpful in the right context</td>
<td>• WBIs could be useful for people who “moved with the times.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Apps could be a tool but not the solution—a blended care approach is needed.</td>
</tr>
<tr>
<td>3</td>
<td>Barriers to using WBIs for Māori</td>
<td>• WBIs come with an upfront cost that may drive inequitable access.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Technical issues put people off.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Literacy and language may make engagement difficult.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mistrust due to years of systemic racism and broken promises.</td>
</tr>
<tr>
<td>4</td>
<td>Ways to improve WBIs to help engagement with Māori</td>
<td>• Māori imagery is key for Māori to connect.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Māori models of well-being should be at the heart of all interventions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Improved integration of the te reo Māori language would make WBIs more appealing.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• By Māori for Māori.</td>
</tr>
</tbody>
</table>

\(^a\)WBI: web-based intervention.

WBIs Are Disconnected From the Core Values of Te Ao Māori

Overview

Te ao Māori (the world view of Māori) and Māori culture are built around relationships, collectivism, and a shared connection. Considering the collective is vital, and a treatment model focused on the individual’s pathology alone without connection or relationships is at odds with Māori beliefs and values. The participants felt that WBIs may not understand and replicate the critical relational aspect, which is crucial in all aspects of life for Māori. Therefore, most participants did not consider WBIs as something that would fit with Māori culture or worldview. A participant said the following:

*Because you are not face to face...and as Māori’s, we are...like, unless we know you and love you, we’ll never be as open though. No way! Ah, it’s, it’s, I don’t know...it just is* [Group 5]

Another participant stated the following:

*I definitely think doing something online definitely lacks whanaungatanga [relationships/the process of forming relationships] and that kind of personal relationships with your doctor or your medical provider* [Group 5]

Many of the older participants struggled to conceptualize what an automated program such as a chatbot would look like and how this could be used to support well-being, as this was removed from the concept of relationships, which are key to wellness and healing. Even when considering web-based therapy delivery, this had significant barriers to developing whanaungatanga. For some, there was a perception that without a real in-person connection with others, it would be easy to mask true feelings and intentions, which would further limit the benefits of the WBI. Others felt that people might inadvertently disclose more than they wanted to due to the WBI not seeming like a relationship (and more like a diary), and this could be harmful due to mistrust in how this information would be used. The overall message is that a person’s absence meant the absence of connection and healing. A participant said the following:

*For healing purposes, people are necessary in our culture* [Group 1]

Another participant stated the following:

*I know people, they would like the online thing because they don’t feel like they are disclosing much about themselves. They are, but they don’t feel like they are* [Interview 2]

Another participant said the following:

*At the same time, when its online you can kind of feel some like you’re wearing a mask. So, it’s not really you* [Interview 3]

The perceived lack of connection for users with WBIs hindered honesty and information sharing. This ultimately limited the benefits and healing that could be obtained when exclusively using WBIs. For some participants, there was the belief that WBIs and the digital world contributed to poor health as people became more disconnected from others; lost their ability to communicate; and tended to live in a digital world that was disconnected from their whakapapa (family history or genealogy), whenua (land), atua (gods or spiritual beings), and...
values. Some perceived turning to a digital world for help as a lack of personal responsibility for healing, which would further exacerbate long-term difficulties and have implications for future generations. This was particularly emphasized for automated interventions with a lack of relationship and accountability. A participant stated the following:

I do I think there is a whole generation, like the younger generation, who will know nothing else but digital stuff and like babies who now are growing up seeing nothing but masks on people’s faces. It’s the same kind of thing they won’t know how to relate unless it’s seen, or read it, or hear it. You know, tap into it online. You know? Their connections might be doing the de-de-de finger scrolling online, whatever online that might be their connection. But for me, I’m old school. I like people [Group 1]

Te Ao Māori Is About Wairua and WBIs Cannot Replicate This

When Māori connect, they do so with an energy transfer between 2 people, which conveys many things, including meaning to nuanced words. As such, te ao Māori is about wairua (spirituality) and Mauri (energy or life force), which respondents felt could not be conveyed by a digital tool. The 1D nature of the WBIs often felt empty to participants, robotic, and therefore not healing. For some, there was a perception that turning to WBIs would further drive participants away from te ao Māori and may worsen the underlying illness mechanism. A participant in stated the following:

We need that human connection. But it’s not just physical connection. It’s the frequency, it's the energy, it’s the Wairua [Group 1]

A participant stated the following:

Well-being for Māori comes from connection. I think well-being, particularly in terms of Māori well-being, you know? The energy and frequency that comes with healing...um...is can be misinterpreted; can be absent [Group 2]

Another participant stated the following:

The last thing you want is for people to be dependent on digital apps for their well-being. That would be disastrous when people don’t know how to go to other people in a group and don’t be empowered enough to heal themselves [Group 1]

Māori Models of Health Are Holistic, Whereas WBIs Are Singular in Focus

Many participants reflected on Māori models of healing as being holistic and considering all areas of well-being. Specifically, one was unlikely to become well or heal from illness by just considering 1 facet of treatment (eg, thinking styles), when being well also included good physical health, strong connections to wānau (family) and whenua (the land), whanaungatanga (relationships), a connection to a higher purpose and meaning, and spirituality, all of which were considered essential for people to flourish. The approach of WBIs was largely seen to be singular in focus and did not fit with te ao Māori views of what was needed for healing. It was recognized that the perceived approach of WBIs was not aligned with Māori models of health and was more aligned with traditional Western models of treatment, such as psychological therapy or a medical model.

The common themes demonstrating the desire for a holistic approach included being able to meet the needs of wānau through accessing food, forming relationships, and recognizing that conversations about Mauri and spirituality were absent. Even when WBIs incorporate te reo Māori (the Māori language), the use of kupu (words) are often perceived as tokenistic, with poor translation that fails to capture the nuances of te reo Māori.

WBIs Could Be a Strand in the Weaving of a Well-Being Kete

Despite the belief that WBIs were unlikely to meet the ideal delivery of an intervention to support the well-being of Māori, many participants were open to considering using a WBI. Specifically, there was a perception that WBIs could be used as an adjunct to a treatment or therapeutic intervention already underway, particularly when there was already a strong relationship with the therapist or the care team. A participant stated the following:

I feel like I could get the information I need to do more. Cause, obviously, an app can only go so far, and your mental health is really...you know? You’re paddling that waka [boat]...People can help you, but you’re the main navigator of that boat...I think it will help start it will help you build the waka and get it onto the water, um, but then yeah...I am guessing that you would have to put in the work yourself to go find other resources [Interview 3]

Another participant stated the following:

Am just not sure we would be able to actually heal people properly, but it could help at least getting them toward the resources they need [Group 3]

WBIs Could Be Helpful in the Right Context

Overview

The participants noted several strengths of WBIs and were open to WBIs being a tool that might benefit some people more than others. Context was important, as was the support around the WBI. Specifically, there was a view that while WBIs might be good for some people, using them in isolation, without connection to people, might actually exacerbate issues.

WBIs Could Be Useful for People Who “Moved With the Times”

There was a perception that WBIs might suit those that have moved away from traditional modes of healing, such as the younger generations, but those groups did not always support this view. Many participants reported that they felt “older,” meaning they were more likely to struggle with WBIs than younger generations. There was a strong sense that rangatahi (young people) would be more inclined to use WBIs than kūmātua (elders), who were less confident with technology. This was observed in this study by the difficulty many kūmātua...
experienced accessing the stimulus WBI and the confusion around what the application was asking. The participants often indicated that they did not believe they could learn new technology and tended to avoid using it. As such, the perceived stress of using the WBIs outweighed the potential benefits. A participant stated the following:

I think the young Māori people might have embraced it, but the older ones, we say, “It’s pākehā! [non-Māori].” They won’t; they won’t switch onto it. Could you imagine your mother doing it? [Interview 2]

Younger participants tended to voice more concerns about functionality and technological issues with WBIs and less about the values disconnection expressed by the older participants. While open to using WBIs, the younger participants still had reluctance to do so. A low tolerance for poor functionality was indicated, with participants highlighting that the complicated log-in processes or errors would lead to them feeling frustrated and uninstalling an application. This indicated that significant barriers to use were complicated sign-up or sign-in processes and outdated or faulty technology. A participant stated the following:

I feel like I would probably uninstall it. I mean, I don’t really have the patience for broken web pages and things...I don’t think most people do. Especially on phones and your meant to be doing a million things at once [Group 3]

The younger participants also supported the views of the older participants around holistic care and the desire for a face-to-face connection. Thus, the assumption that WBIs were embraced by rangatahi was not necessarily true. The younger participants wanted interpersonal connection and did not think WBIs could meet this. WBIs were instead seen as a source of information but unlikely to be considered for therapeutic intervention or connection with others.

**Apps Could Be a Tool but Not the Solution: a Blended Care Approach Is Needed**

Many participants felt that they would be more open to using a WBI when integrated in the context of an existing, trusted relationship. The perception was that the role of the relationship would be to facilitate healing, while the WBI could serve for monitoring, support, providing information, promptly answering questions, and potentially facilitating connections to others beyond the primary relationship. The relationship would center around wairua, whanaungatanga, and whānau, while the WBI might provide exercises, reminders, and ways to keep track of important things. Relaxation and mindfulness recordings were specifically noted as something that could be provided or used by participants in WBIs. Contacting a trusted support or therapist through the WBI was also seen as a way to increase whanaungatanga and healing if used in a timely and appropriate manner.

**Barriers to Using WBIs for Māori**

There were several barriers to using WBIs for participants, and many participants shared that they believed other Māori would also experience some of these challenges.
or that the person using the WBIs would not be a priority. A participant stated the following:

I share something you share something you know that kind of kōrero or that I don’t know if they’re listening. I mean, while I’m talking, what are they doing? Emailing somebody? [Group 1]

The participants raised further concerns about the confidentiality of information shared with WBIs and how this would be used, particularly with regard to information potentially being shared with government agencies and the potential negative impact this could have on the participant and their whānau. This mistrust was not limited to WBI but to the researchers, the institutions, and the groups that supported it. A participant stated the following:

What puts me off is that there is nothing that comes across as Māori. This is a pākehā app for pākehā, pākehā solutions, pākehā reference pākehā, you know? And I think that’s one of the, um, big issues of today, um, well-being solutions is that pākehā are trying to form solutions using methodologies that aren’t Māori, you know? [Group 1]

Many participants reported struggling with interventions that Pākehā developed as they lacked the depth of understanding about the historical impact of colonization on Māori and how this may impact the person is seeking help. Many felt that interventions and Pākehā therapists and WBIs could not address this as they did not have these lived experiences. A participant stated the following:

I would want someone who can understand what lived experience is... [Group 2]

Another participant stated the following:

I prefer to speak with people, real people. And when it comes to, um, some of my historical stuff, then I would want someone who can understand what lived experience is. You know, email doesn’t cut it either, you know, again who is this person that I’m emailing? [Group 1]

Mistrust was further exacerbated by the perception that Māori did not design many WBIs for that were reportedly for Māori. This was not only due to the WBI being perceived as a poor fit for te ao Māori but also due to Pākehā ultimately designing many WBIs for Pākehā (or the dominant paradigm) and then adapting (often poorly) these for Māori. Therefore, very few interventions were specifically designed for Māori, and even fewer were created by Māori. A participant said the following:

There’s no Māori kupa in here. Not even a “kia ora,” welcome [Group 1]

Another participant said the following:

They had a real photo, but, um...whoever created didn’t look at the photo properly and it had somebody poking their tongue. So what time was given to that? I felt it was lacking big time cause if I’m not in a great space and I see someone poking their tongues, saying, you know? I’m feel challenged! What’s up with that? [Group 1]

The participants were also wary of research teams that used a Māori interviewer but did not involve Māori in key decision-making. Questions were raised around the motivations of the teams and whether they came from genuine caring and a desire to reduce inequities (it was noted that the interviewer was challenged about this on >1 occasion). There was a belief that these teams that did not have genuine motivation would create solutions that were unlikely to resonate with Māori. A participant stated the following:

You know, they’re not using Māori or mātauranga Māori [knowledge] or Māori methodologies to create Māori solutions or do the research. You know they have got a brown face here, but I mean, who’s behind it and why they behind it? You know? [Group 4]

Ways to Improve WBIs to Help Engagement With Māori

Despite reservations about WBI, the participants provided several recommendations that could make WBIs more engaging for Māori. These were specifically about features of WBIs that would make them more engaging for Māori who may be interested in using WBIs.

Māori Imagery Is Key for Māori to Connect

The most common subtheme was that the participants wanted to see themselves reflected in the WBI. Imagery was important, and the participants wanted to see a mix of images, including pictures of Māori people, whānau (families), and hapu (communities). This imagery also included other key things such as bodies of water, native bush, maunga (mountains), Māori art and carvings, and buildings easily identified as Māori (eg, wharenui—meeting houses). This range of imagery tied back to a sense of holistic approaches to well-being, including the connection to people, to whenua (land), and to key important spiritual sites.

The participants also highlighted that Māori are diverse people and that not just 1 image connects with all Māori. Images depicting kapa haka or individuals performing a pukana (facial expression) were frequently viewed as oversimplified and occasionally offensive, simplifying the complexity of Māori culture. Instead, the participants would be drawn to WBIs that used a range of imagery of people showing the diversity of Māori. A participant stated the following:

A diverse range of Māori women. Um...sizes, ages...um, you know, skin tone. You know? Just like everything [Interview 3]

This was considered important by almost all participants, particularly as some participants identified that Māori are a strongly oral and visual culture with images of kowhaiwhai (weaving), whakairo (carving), and taonga (treasured items) drawing the eye and creating connection. For participants, many WBIs still mimicked clinical rooms in terms of the colors used and simple pages, which further created a sense of depersonalization and disconnection. A participant stated the following:

It seems very like clinical in a way, um, because of the coloring. Like, it’s all blue, like, it kind of just
makes me think of, like, the doctors, um, like, it’s very formal I think [Group 3]

Māori Models of Well-Being Should Be at the Heart of All Interventions

Several participants noted that if WBIs interventions could be constructed around Māori models of well-being, WBIs would be likely be more engaging for Māori. The key model suggested was Te Whare Tapa Whā [45]. While other participants did not explicitly outline a model, they articulated the need for a holistic approach to care that supported all aspects of wellness, including connection to culture as a path for healing. A participant stated the following:

Inclusive of the mental effort, you know the spiritual and all that stuff and the physical but so is this are they talking about holistic well-being or are they just talking about mental well-being because you can’t have one it’s, like, Te Whare Tapa Whā thing, ah, you can’t just focus on the one when you are expecting others to fall into place you have to work on the whole lot [Group 1]

Another participant said the following:

There needs to be some sort of cultural, um, tool that helps...that grounds people. I guess whakapapa was one of those things [Group 1]

Improved Integration of the Te Reo Māori Language Would Make WBIs More Appealing

The participants agreed that correct and appropriate use of te reo Māori was important if WBIs were to engage Māori. Recognizing that the fluency of te reo varied, the participants wanted WBIs to be able to be modified based on the user’s fluency (such as entirely in te reo Māori or with only a few words). Even for those only beginning their te reo journey, keywords such as a greeting should be used meaningfully. The participants also noted that, based on past experiences, simply translating a few words into te reo lost the nuances of the language and could lead to misunderstanding and, at times, felt tokenistic. There was an emphasis on the correct use of Māori words, which often carry contextual meanings. Tokenistic inclusion of these words can lead to misunderstandings or, at worst, be offensive. Therefore, people developing WBIs for Māori needed someone fluent in te reo Māori working on the content rather than using simplified translation tools.

By Māori for Māori

A strong theme that came through was that WBIs, interventions, and tools that were developed by Māori for Māori resonated more strongly with Māori participants and were more likely to be engaging and used. Trusted institutions and sites tended to result in resources that were more readily used. A participant said the following:

I have searched through...through Te Ora [website] you know to find out stuff I think because it’s a Māori organization there’s a sense of connection being a Māori organization [Group 1]

A clear understanding of the whakapapa (history and origin) of the research project and the research team was deemed crucial, ideally with the project being designed and led by Māori for Māori. In addition, there was discussion about incorporating key models of health and acknowledging the influence of significant Māori figures in the project’s development.

Discussion

Principal Findings

This study is one of the first to explore the views of Māori, the indigenous people population of New Zealand, about WBIs. Although participants could see the potential benefits of WBIs when used in the context of a strong existing relationship, there were concerns that a digital tool would not be able to facilitate healing due to the perception of WBIs being 1D in their focus rather than holistic, the potential to drive people further away from te ao Māori, and the lack of genuine connection that could be made on the web. Significant barriers to using WBIs were highlighted, including the impact of social inequities, which hindered access to the technology needed to engage with WBIs. Educational disadvantage also contributed to difficulties, particularly with text-heavy platforms. In addition, concerns about confidentiality and mistrust in the motivations of researchers (and government) due to experiences of colonization were clear.

This study makes a unique contribution to understanding WBIs, how they fit with the view of Māori as the indigenous people of New Zealand, and how indigenous communities may perceive and respond to WBIs. Our findings are in contrast with previous studies that explored uptake and engagement with WBIs in indigenous populations and ethnic minority groups. A recent scoping review exploring the use and uptake of web-based therapeutic interventions among indigenous populations in Australia, New Zealand, the United States, and Canada found moderate uptake of WBIs and potentially improved health outcomes associated with them [36]. Of the 31 studies, 9 (29%) were from New Zealand, with 3 (10%) relating to 1 web-based cognitive behavior therapy–based intervention (Smart, Positive, Active, Realistic, X-factor thoughts), 1 (3%) relating to minimizing risky alcohol use, and others (n=5, 16%) on specific noncommunicable diseases. The review findings reported a more positive experience with WBIs, possibly because of the younger age of participants and the way the WBIs were implemented, for example, delivered in school-based settings or with significant input and coproduction with Māori. Similar to our findings, the review highlighted the importance of tailoring content and presentation formats to ensure cultural relevance; appropriateness; and a customizable, easy-to-use interface. Another systematic review looking at the use of digital technologies to improve the mental health and well-being of indigenous people reported 27 studies that generally support the effectiveness of digital technologies in aiding the provision of mental health services but acknowledged that decolonizing and culturally appropriate approaches are needed [46].

The studies in these reviews generally do not examine the user experience of using the WBIs in depth. The review was limited in the degree to which it could explore barriers to accessing
WBIs resulting from cultural and linguistic diversity, low health literacy, limited digital capabilities, and infrastructural and resource limitations for individuals and communities in different geographic locations—concepts that our participants expressed as potential barriers to WBI use. Similarly, a recent rapid review examining the use of web-based care in indigenous populations highlighted several barriers to engagement with web-based care—cost, accessibility, digital literacy, and language [47]—which align with our findings. The review emphasized the importance of building relationships and trust and ensuring the infrastructure is present to support technology navigation with indigenous populations, echoing our findings. Our study builds on these reviews by being one of a few studies to examine the views of Māori using qualitative methodology, which may explain the more in-depth findings articulated by respondents.

Limitations

Several limitations for this study exist. First, although the sample size was large for a qualitative study, Māori, like all indigenous people, are diverse; therefore, the views reflected in this study may not apply to all Māori. These views can likely be beneficial in shaping the development of WBIs. However, an effective WBI should be developed by Māori with the understanding that one intervention will not fit everyone. Instead, WBIs need to be developed with Māori models of well-being being central for use alongside a strong kanohi-ki-te-kanohi (face-to-face) therapeutic relationship. Previous reviews [36] recognize that the definitions of health often used are less holistic and relational than indigenous models of health and well-being, which may affect the interpretation of published studies in this area or result in studies appearing to be more effective than what would be perceived from a more holistic framework.

Second, this study explored the views of WBIs and used an existing intervention as an example for participants. This means that interventions that may have a different development process (eg, Māori developed) may have resulted in different views by the participants around the acceptability and usability of WBIs. The views of the participants may also be affected by the stimulus chosen as the example.

Finally, while most of the research team were Māori, the researchers varied in the strength of their connection to te ao Māori and their knowledge of this. This means that the analysis of the transcripts and subsequent findings may have varied if other people with different understandings and connections were to analyze the transcripts.

Reflections

One challenge for the research team was to avoid replicating some of the concerns and barriers regarding WBIs that were raised by the participants. Specifically, one aspect was the composition of the research team, where the individual conducting participant interviews was Māori, while the senior authors of the paper were not Māori. For context, this project is 1 part of a 2-part study designed from the onset with a diverse team in consultation with a Kaupapa Māori nongovernment organization. MCBP was involved in early discussions with the research team about the project and was brought into the team in recognition of the mātauranga Māori (knowledge unique to Māori) that she brings. MCBP was supported to be an active and equal member of the research, including being provided with support to upskill in qualitative methodologies. PH and HW were engaged in reviewing the themes to ensure interpretation and understanding were correct and to add to the richness of the interpretation, as the research team did not want to assume understanding and acknowledged their limitations around the lived experiences of Māori. Author order was decided among the team (MCBP, HW, PH, LD, and AYHC) based on roles in the project, with the decision to make MCBP and LD joint lead authors in recognition of the different roles in the project. We recognize that Māori should ideally lead projects exploring the view of Māori, and the experience that MCBP has within this project has meant that she has been able to colead a subsequent project and is beginning to develop her own research pathway as a Māori researcher with lived experience.

Conclusions

Through in-depth qualitative interviews and focus groups with Māori (indigenous people of New Zealand), WBIs were found to be generally considered a poor fit for Māori as the design of WBIs did not align with the Māori worldview or concepts of well-being. This contrasts with previous findings, where WBIs have been shown to be effective in supporting mental health and well-being and can overcome some of the traditional barriers to help seeking. With the large number of WBIs being developed, these findings are important in highlighting key considerations for WBIs to promote engagement with Māori, particularly considering how well WBIs fit with the indigenous worldview and how they meet the needs of indigenous communities in a culturally appropriate manner. While WBIs may have a place in supporting the well-being of Māori, WBIs alone are unlikely to achieve the same benefits expected for a non-Māori population and may further drive health inequities if not properly implemented and supported.

Acknowledgments

The authors would like to acknowledge the participants who contributed to this study and Te Kete Pounamu for their support, advice, and help with recruitment for this project. In addition, funding was provided by Chorus and Oakley Mental Health Foundation.
**Data Availability**

The data sets generated and analyzed during this study are not publicly available due to participants not consenting to the sharing of data beyond the research team with consent given for publication.

**Authors' Contributions**

LD is a health and clinical psychologist and researcher of New Zealand European descent who works in the area of digital interventions. LD was involved in the design of the project, ethics application, recruitment, overseeing interviews and supporting MCBP to complete the interviews, checking the transcripts, overseeing data analysis including supporting MCBP to upskill in thematic analysis, coding and generation of themes, and writing the manuscript. MCBP is a Mōari (Ngāti Whakaue and Te Arawa) and a research assistant for this project. MCBP was involved in the recruitment, facilitated interviews, and focus groups as well as the checking of transcripts, coding of interview data, and generation of themes. HW is a Mōari (Ngāti Tahu) and a PhD candidate at the University of Auckland. HW was involved in the project as a transcriber of the interviews. HW was also involved in discussions with the wider research team about the themes uncovered in the data. PH is a Mōari (Ngāti Porou, Ngapuhi, Te Whānau-a-Apanui) psychologist and an academic who is completing her PhD. PH was involved in the discussion where there was disagreement in the coding of the transcripts and was involved in the wider team discussion about the generated codes. AHYC is an academic clinical pharmacist with Asian heritage. She has particular expertise in digital health, big data, and behavioral medicine. AHYC was involved in the project development, reviewing the themes from the interviews, and contributing to the manuscript.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

WBI: web-based intervention
Abstract

Background: Poor sleep quality can elevate stress levels and diminish overall well-being. Japanese individuals often experience sleep deprivation, and workers have high levels of stress. Nevertheless, research examining the connection between objective sleep assessments and stress levels, as well as overall well-being, among Japanese workers is lacking.

Objective: This study aims to investigate the correlation between physiological data, including sleep duration and heart rate variability (HRV), objectively measured through wearable devices, and 3 states (sleepiness, mood, and energy) assessed through ecological momentary assessment (EMA) and use of rating scales for stress and well-being.

Methods: A total of 40 office workers (female, 20/40, 50%; mean age 40.4 years, SD 11.8 years) participated in the study. Participants were asked to wear a wearable wristband device for 8 consecutive weeks. EMA regarding sleepiness, mood, and energy levels was conducted via email messages sent by participants 4 times daily, with each session spaced 3 hours apart. This assessment occurred on 8 designated days within the 8-week timeframe. Participants’ stress levels and perception of well-being were assessed using respective self-rating questionnaires. Subsequently, participants were categorized into quartiles based on their stress and well-being scores, and the sleep patterns and HRV indices recorded by the Fitbit Inspire 2 were compared among these groups. The Mann-Whitney U test was used to assess differences between the quartiles, with adjustments made for multiple comparisons using the Bonferroni correction. Furthermore, EMA results and the sleep and HRV indices were subjected to multilevel analysis for a comprehensive evaluation.
Results: The EMA achieved a total response rate of 87.3%, while the Fitbit Inspire 2 wear rate reached 88.0%. When participants were grouped based on quartiles of well-being and stress-related scores, significant differences emerged. Specifically, individuals in the lowest stress quartile or highest subjective satisfaction quartile retired to bed earlier (β<0.001 and P=0.01, respectively), whereas those in the highest stress quartile exhibited greater variation in the midpoint of sleep (P<0.001). A multilevel analysis unveiled notable relationships: intraindividual variability analysis indicated that higher energy levels were associated with lower deviation of heart rate during sleep on the preceding day (β=-0.12, P<0.001), and decreased sleepiness was observed on days following longer sleep durations (β=-0.10, P<0.001). Furthermore, interindividual variability analysis revealed that individuals with earlier midpoints of sleep tended to exhibit higher energy levels (β=-0.26, P=0.04).

Conclusions: Increased sleep variabilities, characterized by unstable bedtime or midpoint of sleep, were correlated with elevated stress levels and diminished well-being. Conversely, improved sleep indices (eg, lower heart rate during sleep and earlier average bedtime) were associated with heightened daytime energy levels. Further research with a larger sample size using these methodologies, particularly focusing on specific phenomena such as social jet lag, has the potential to yield valuable insights.

Trial Registration: UMIN-CTR UMIN000046858; https://center6.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000053392

KEYWORDS
wearable device; sleep feedback; well-being; stress; ecological momentary assessment; feasibility study

Introduction

Sleep is one of the most important health behaviors for human beings, and a lack of it affects both physical and mental health [1,2]. Furthermore, subjective stress and well-being are closely related to sleep quality [3-5]. Subjective methods for evaluating sleep quality are the Consensus Sleep Diary [6] and Pittsburgh Sleep Quality Index [7]; however, their results sometimes differ from those of objective assessments [7,8]. Therefore, objective sleep evaluation methods have been developed, and research using these methods is ongoing.

Polysomnography is the gold standard for objective sleep assessment methods [9], but it is expensive and physically demanding. Hence, wearable devices have emerged as recent alternatives for sleep assessment, offering less burdensome options compared with polysomnography [10]. For example, commercially available devices such as Fitbit (Fitbit Inc.) can measure motion, heart rate variability (HRV), and respiratory rate, and based on these measurements, sleep quality can be measured with reasonable accuracy [11]. Furthermore, HRV, a quantitative biomarker of autonomic activity [12], has been reported to be useful for the quantitative assessment of stress [13] and is highly associated with subjective well-being [14,15]. In addition, self-administered questionnaires such as the Epworth Sleepiness Scale have been used [16]; however, they do not capture diurnal variations in sleepiness. One method to capture symptom variability is ecological momentary assessment (EMA), which is a prospective, repeated sampling of symptoms in real time in the participants’ natural environment, with minimal burden on respondents [17]. EMA has been widely used to accurately assess respondents’ mood and stress levels during the day [18]. Moreover, several studies have examined sleep conditions and their relationship to daytime sleepiness using EMA [19-23]. In recent years, studies have combined objective sleep assessments, such as those obtained by wearables, with EMA-based assessments [24-26]. It has long been reported that, on average, Japanese people go to bed later and sleep for shorter durations than people in other countries [27,28]. A survey of 13 countries in 2020 and 2021 also reported that Japan had the lowest subjective satisfaction with sleep for 2 consecutive years [29]. In Japan, the number of individuals with work-related mental health problems continues to increase annually [30]; furthermore, in a 2023 survey on workers’ health and well-being, Japan ranked last among 30 countries [31]. However, only a few studies have evaluated the relationship between sleep and stress and well-being among Japanese workers using a combination of objective sleep assessments and EMA. This study investigated the relationship between sleep and stress and well-being among Japanese workers. We evaluated the relationship between sleep quality/daytime sleepiness and stress levels/perception of well-being (hereinafter, stress/well-being) by combining the assessment of sleep quality using wearable devices with real-time monitoring of daytime sleepiness and mental status using EMA. In particular, we examined the relationships between objectively measured sleep indices and HRV using Fitbit Inspire 2 (Fitbit Inc.) and the participants’ subjective conditions or feelings to verify the following hypotheses: (1) people with short total sleep time and late bedtime have high stress levels and low perceived well-being; (2) people with high sleep quality have high energy and good mood during the day.

Methods

Study Design

The study was designed to explore the correlations between physiological and behavioral data (eg, sleep, daytime activity, HRV, and blood glucose) objectively measured by wearable devices, and the participants’ perception of well-being, stress levels, daily mood, energy, and sleepiness assessed through questionnaires and EMA. The participants were asked to wear 2 wearable devices, namely, the Fitbit Inspire 2 and the FreeStyle Libre (Abbott Diabetes Care) for acquiring the physiological/behavioral data and 24-hour blood glucose data.
respectively. In this study, our focus was on examining the relationships between sleep and HRV indices measured using the Fitbit Inspire 2, alongside the concurrent assessment of participants’ psychological status through EMA and stress/well-being questionnaires. Other aspects of the study will be reported separately.

The detailed study design consists of recruiting a total of 40 participants, comprising individuals with impaired glucose tolerance (not meeting the diagnostic criteria for diabetes mellitus) and healthy participants. Initially, participants completed a questionnaire aimed at evaluating their eating behaviors and stress/well-being. Following this, they were instructed to wear the Fitbit Inspire 2 continuously for 8 consecutive weeks. In addition, during the initial 2 weeks of the study, participants were instructed to wear the FreeStyle Libre continuously during the day. After the completion of this 2-week period, participants were asked to once again complete the questionnaire assessing their eating behaviors. Furthermore, they received guidance on eating behaviors from a dietitian. Following the initial 4 weeks of wearing the Fitbit Inspire 2 device, participants were requested to wear the FreeStyle Libre once more for a duration of 2 weeks. This period coincided with the final 2 weeks of the 8-week Fitbit Inspire 2 wearing period. Subsequently, upon concluding the 8-week duration of wearing both devices, participants were once again asked to complete the questionnaires assessing their eating behaviors, as well as another questionnaire aimed at evaluating their stress/well-being. Furthermore, EMA was conducted to evaluate sleepiness, mood, and energy levels. This assessment involved participants sending emails 4 times a day, with each email sent 3 hours apart. This process occurred on 8 designated days throughout the 8-week study period.

Participants

The inclusion and exclusion criteria for participants are detailed in Textbox 1. Eligible participants were adults, defined as individuals at least 20 years old (the age of adulthood in Japan at the time of recruitment), who possessed a smartphone (either Android or iOS). The study aimed to investigate the correlation between sleep, HRV, 24-hour blood glucose levels, and various psychological factors including sleepiness, energy, mood, and stress/well-being. Therefore, inclusion criteria encompassed individuals expected to exhibit high blood glucose variability, alongside normal participants. Specifically, participants with abnormal glucose tolerance (hemoglobin A\textsubscript{1c} (HbA\textsubscript{1c}) levels ranging from 5.8 to <6.5) who did not meet the criteria for diabetes mellitus were considered eligible. Psychiatric disorders were neither inclusion nor exclusion criteria for participant selection. Recruitment for the study occurred in February 2022, with participants recruited at a 1:1 male-to-female ratio.

Textbox 1. Inclusion and exclusion criteria.

1. **Inclusion criteria**
   - Patients with hemoglobin A\textsubscript{1c} (HbA\textsubscript{1c}) values of 5.8 ≤ HbA\textsubscript{1c} <6.5 at the most recent physical examination undergone within the previous year, or healthy participants with no glucose intolerance.
   - Adults aged 20 years or older at the time of consent acquisition.
   - Smartphone users who owned and used a smartphone that was compatible with the app used in this study.

2. **Exclusion criteria**
   - Patients who had been diagnosed and treated for diabetes mellitus in the past.
   - Patients with comorbidities that could affect the data measured by the wearable wristband device, such as paralysis of the upper limbs.
   - Others who were deemed as being inappropriate participants for the study by the principal researcher or coresearchers.

Data Collection

The data collected in this study are outlined comprehensively in Textbox 2. Further, the data collection schedule is presented in Table 1. Data collection occurred between February 2022 and April 2022. The Perceived Stress Scale (PSS) [32] was used for the assessment of subjective stress. For the assessment of subjective well-being, we used the Satisfaction With Life Scale (SWLS) [33], the Japanese version of the Scale of Positive and Negative Experience (SPANe-J) [34], and the Japanese version of the Flourishing Scale (FS-J) [34]. To evaluate sleepiness, the Japanese version of the Epworth Sleepiness Scale (JESS) [35] was used.
Textbox 2. Data collected in this study.

1. Background information (results of company medical checkup)
   - Gender, age, and preexisting medical conditions
   - Results of the most recent medical checkup undergone within the previous year

2. Questionnaire
   - Brief-Type Self-Administered Diet History Questionnaire (BDHQ) [36]
   - Eating Behavior Questionnaire (EBQ) [37]
   - Perceived Stress Scale (PSS) [32]
   - Satisfaction With Life Scale (SWLS) [33]
   - The Japanese version of the Scale of Positive and Negative Experience (SPANE-J) [34]
   - The Japanese version of the Flourishing Scale (FS-J) [34]
   - The Japanese version of the Epworth Sleepiness Scale (JESS) [35]

3. Ecological momentary assessment
   - Sleepiness (Visual Analog Scale [VAS] format): scores range from 0=not sleepy at all to 100=very sleepy (maximum sleepiness imaginable).
   - Mood (VAS format): scores range from 0=negative mood (lowest mood imaginable) to 100=positive mood (best mood imaginable).
   - Energy (VAS format): scores range from 0=no energy at all to 100=full of energy (greatest state of energy imaginable).

4. Wearable Devices
   - Fitbit Inspire 2 data (eg, sleep, heart rate variability, and body movements)
   - FreeStyle Libre data (24-hour blood glucose)

Table 1. Schedule for data collection.

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Baseline</th>
<th>2 weeks</th>
<th>Intermediate evaluation</th>
<th>Interval (4 weeks)</th>
<th>2 weeks</th>
<th>Final evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background information</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical checkup</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDHQ(^a)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBQ(^b)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSS(^c)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWLS(^d)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPANE-J(^e)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FS-J(^f)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JESS(^g)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) BDHQ: Brief-Type Self-Administered Diet History Questionnaire.
\(^b\) EBQ: Eating Behavior Questionnaire.
\(^c\) PSS: Perceived Stress Scale.
\(^d\) SWLS: Satisfaction With Life Scale.
\(^e\) SPANE-J: The Japanese version of the Scale of Positive and Negative Experience.
\(^f\) FS-J: The Japanese version of the Flourishing Scale.
\(^g\) JESS: The Japanese version of the Epworth Sleepiness Scale.

Fitbit Inspire 2 data were collected continuously throughout the 8-week study period. FreeStyle Libre data and EMA data were collected 2 weeks after baseline and 2 weeks before the final evaluation. Data collection was facilitated through electronic patient-reported outcomes, a self-administered instrument completed by participants independently. Electronic patient-reported outcome response links were disseminated via
email and accessed by participants through their smartphones. An identical method was used for EMA.

As previously noted, not all data from this study are presented within this paper. The results of the assessment of eating behaviors and blood glucose–related data, along with the corresponding analyses, were reported elsewhere [38,39].

**EMA**

In the EMA, participants were prompted to respond to questions regarding 3 states: sleepiness (ranging from “not sleepy at all” to “very sleepy [maximum sleepiness imaginable]”), mood (ranging from “negative mood [lowest mood imaginable]” to “positive mood [maximum mood imaginable]”), and energy (ranging from “no energy at all” to “full of energy [greatest state of energy imaginable]”). These responses were provided using the Visual Analog Scale (VAS) format. The design of these VAS items was informed by previous studies that combined EMA with the VAS [40,41].

Throughout the study period, the questions were administered 2 times a week on weekdays, occurring on 2 consecutive days (such as Tuesday and Thursday or Wednesday and Friday). These questions were delivered at defined times of the day, spaced 3 hours apart. Specifically, the questions were presented at 9 AM, noon, 3 PM, and 6 PM. Participants were then requested to respond to the questions during the following time intervals: 9 AM to noon, noon to 3 PM, 3 PM to 6 PM, and after 6 PM, respectively.

The EMA score data collected on the same day as the email was sent were used in the analysis. Fitbit Inspire 2 activity indicators, including heart rate, number of steps, and metabolic equivalents (METs), were aggregated within the 1-hour time window immediately preceding the EMA response time. Each EMA response is collected as time-stamped data, and the analysis is conducted retrospectively from the time stamp, specifically 1 hour before the EMA response time. Sleep indices were compiled from Fitbit Inspire 2 sleep records where the waking time fell within the time window from 6 PM on the day preceding the EMA response date to 9 AM on the EMA response date. In instances where multiple sleep records were present within this timeframe, the sleep periods were consolidated, and sleep indices were recalculated accordingly.

**Fitbit Inspire 2**

Fitbit Inspire 2 is a commercially available device equipped with a triaxial acceleration sensor, which allows for the calculation of body movement, sleep patterns, and other related metrics. In addition, it features a pulse sensor that collects data for each heartbeat using the photoelectric pulse wave method. Participants were instructed to wear the Fitbit Inspire 2 device, distributed at the outset of the study, continuously for a designated period (8 weeks). Participants were permitted to temporarily remove the Fitbit Inspire 2 device for short durations, such as during bathing or when recharging the device. As the Fitbit Inspire 2 has a data storage capacity of approximately 7 days, participants were requested to periodically transmit the collected data to the cloud using their smartphones.

Fitbit Inspire 2 records were considered valid record days if at least 80% of the day was recorded. When matching against EMA data, only participants with a wear rate exceeding 80% in the hour preceding the most recent hour of EMA activity were included. For investigations into the relationship between EMA results and sleep, only participants with a wearing rate of 80% or greater between 6 PM on the day before the response date and 9 AM on the day of the response were included in the analysis.

**Statistical Analysis**

All data collected during the study were incorporated into the analysis. We adopted an exploratory analysis approach, thoroughly examining the various relationships among the data obtained from the wearable devices, EMA, and assessments of stress/well-being.

The Fitbit Inspire 2 device provides data on time in bed (TIB), mid-wake time, and sleep duration. Using this information, we computed bedtime, midpoint of sleep, and sleep efficiency (calculated as 100 × sleep duration/TIB) for the analyses conducted in this study. However, we chose not to include sleep stage estimates provided by the device in our analyses because of concerns regarding their validation. Therefore, this parameter was excluded from our analysis. Based on the pulse data obtained from the Fitbit Inspire 2, we calculated the SD of the NN intervals (SDNN) to assess HRV. In analyzing the SDNN, we used a 15-minute time window. This decision was made considering that the pulse was measured once every 5 seconds. Using a shorter time window, such as 5 minutes, might lead to less reliable values because of the limited amount of data available. Although SDNN is typically calculated based on each pulse interval, the Fitbit Inspire 2 reports pulse rate data every 5 seconds. Therefore, in this study, the SDNN was calculated based on the pulse rate provided at 5-second intervals.

The participants were stratified into quartiles based on their scores on the stress/well-being questionnaires. We then compared the sleep and HRV indices measured by the Fitbit Inspire 2 between the groups with scores in the lowest and highest quartiles. To assess the differences between these 2 groups, we used the Mann–Whitney U test. Given the multiple comparisons conducted, we applied the Bonferroni correction to adjust for potential type I errors.

The study also investigated the relationships between the sleep and HRV indices measured by the Fitbit Inspire 2 and the daily mood, energy levels, and sleepiness assessed by EMA.

In the multilevel analysis, 3 types of multilevel models were constructed, with the EMA scores for energy, mood, and drowsiness serving as the dependent variables. The independent variables were selected through forward stepwise selection using the Akaike information criterion as the evaluation index. The sleep index, sleep heart rate, and sleep HRV index measured by the Fitbit Inspire 2 were considered potential candidates for the independent variables in the models. In addition to the main predictors, control variables such as age, sex, response time, and MET time were included in the analysis. However, to mitigate the influence of activity immediately preceding the response on the EMA score, the aggregate value of the MET...
time from 1 hour before the response to the time of the response was used. Sleep indices were separated into 2 components: the mean value across all participants (interindividual), such as average bedtime among participants, and within-participant deviations (intraindividual), such as bedtime deviations for each participant. These components were then used as explanatory variables in the analysis. The statistical analyses described above were performed using either Python (Python Foundation) or R (R Foundation) programming languages.

**Ethical Consideration**
This study was carried out with the approval of the Keio University School of Medicine Ethics Committee (ID 20211103). Before enrollment, all participants provided written informed consent. All collected information was anonymized and used solely for the purpose of the study. Participants did not receive any monetary compensation; however, they were provided with feedback on their own data collected during the study.

**Results**

**Participant Characteristics**
A cohort of 40 eligible participants, comprising office workers from a real estate company, was enrolled in this study in February 2022. All participants successfully completed the entire study protocol. It is noteworthy that participants were required to work from their respective offices throughout the study duration, and none took extended vacations or engaged in remote work for prolonged periods during the study period. The demographic characteristics of the participants are summarized in Table 2. The mean age of the participants was 40.4 (SD 11.8) years, and the male-to-female ratio was 1:1. No individuals were excluded from participation due to meeting the exclusion criteria at the time of recruitment. Responses to the background information questionnaire revealed that none of the participants were currently undergoing treatment for psychiatric disorders. Furthermore, no serious adverse events, including instances of COVID-19 infection, were reported throughout the study duration. Histograms displaying the distribution of scores on the stress/well-being questionnaire within each quartile group are provided in Figures S1-S6 in Multimedia Appendix 1.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>40.4 (11.8)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>20 (50)</td>
</tr>
<tr>
<td>Perceived Stress Scale (PSS), mean (SD)</td>
<td>19.2 (6.4)</td>
</tr>
<tr>
<td>Satisfaction With Life Scale (SWLS), mean (SD)</td>
<td>22.1 (6.1)</td>
</tr>
<tr>
<td>SPANE-J positive experience (SPANE-P), mean (SD)</td>
<td>20.7 (5.0)</td>
</tr>
<tr>
<td>SPANE-J negative experience (SPANE-N), mean (SD)</td>
<td>16.9 (4.5)</td>
</tr>
<tr>
<td>The Japanese version of the Flourishing Scale (FS-J), mean (SD)</td>
<td>39.8 (7.1)</td>
</tr>
<tr>
<td>The Japanese version of the Epworth Sleepiness Scale (JESS), mean (SD)</td>
<td>8.1 (4.0)</td>
</tr>
</tbody>
</table>

**EMA and Fitbit Inspire 2 Adherence**
All participants successfully completed the program without dropping out. The overall response rate for EMA was 87.34% (1118/1280). Regarding the difference between EMA transmission time and response time, the median (IQR) was 11.8 (3.2-42.4) minutes. However, 1 participant was unaware of receiving emails until midway through the study period because the emails had been inadvertently sorted into the spam folder. The mean daily wear time for all Fitbit Inspire 2 users was 1214 (SD 374) minutes. In addition, the Fitbit Inspire 2 data recovery rate during the study period was 87.98% (2,837,881 minutes/3,225,600 minutes).

**Bedtime, TIB, Midpoint of Sleep, and Psychological State**
In terms of the relationship with JESS scores, compared with the group with high daytime sleepiness (third quartile to maximum; high-score group), the group with low daytime sleepiness (minimum to first quartile; low-score group) exhibited significantly earlier bedtime and midpoint of sleep (r=0.24, P<.001 and r=0.25, P<.001, respectively) (Figures 1 and 2). In the low JESS group, the median bedtime and midpoint of sleep were 11:21 PM and 2:36 AM, respectively. By contrast, in the high JESS group, the median bedtime and midpoint of sleep were 12:12 AM and 3:54 AM, respectively.
Figure 1. Box plot showing the distribution of bedtime in the low- and high-score groups of the JESS score. The dashed line in the box plot refers to the mean value and the whiskers refer to the minimum and maximum values within the IQR 1.5 range (same for subsequent figures). JESS: The Japanese version of the Epworth Sleepiness Scale.

Figure 2. Box plot showing the distribution of the midpoint of sleep in the low- and high-score groups of the JESS score. JESS: The Japanese version of the Epworth Sleepiness Scale.

Regarding subjective stress, the median bedtime was 11:30 PM in the low PSS group and 12:45 AM in the high PSS group. Notably, the median bedtime was significantly earlier in the low PSS group ($r=0.226$, $P<.001$; Figure 3).
In terms of well-being scores, the group with high SWLS scores and low SPANE-J negative experience (SPANE-N) scores exhibited significantly earlier bedtime ($r = -0.23$, $P = 0.01$ and $r = 0.23$, $P < .001$, respectively; Figures 4 and 5). Specifically, the median bedtime was 11:48 PM and 11:24 PM in the high SWLS group and the low SPANE-N group, respectively. Conversely, in the low SWLS group and the high SPANE-N group, the median bedtime was 12:51 AM and 12:54 AM, respectively.

**Figure 3.** Box plot showing the distribution of bedtime in the low and high score groups of the PSS score. PSS: Perceived Stress Scale.

**Figure 4.** Box plot showing the distribution of bedtime in the low- and high-score groups of the SWLS. SWLS: Satisfaction With Life Scale.
Variability in the Midpoint of Sleep and Psychological State

The SD of the midpoint of sleep per week was significantly larger (|Hedge g|>0.5) in the group with low scores on the well-being scales, namely, the FS-J and the SPANE-J positive experience (SPANE-P), compared with the group with high scores (r=−0.11, P=.006 and r=−0.3, P<.001, respectively). Concerning the FS-J, the average SD of the midpoint of sleep per week was 75 minutes in the low-score group compared with 44 minutes in the high-score group. Similarly, based on the SPANE-P, the average SD was 67 minutes in the low-scoring group versus 37 minutes in the high-scoring group (Figures 6 and 7).
Similarly, the SD of the midpoint of sleep was significantly greater in the group with high subjective stress as measured by the PSS compared with the group with low subjective stress ($r=0.3$, $P<.001$). Specifically, on the PSS, the median SD of the midpoint of sleep was 33 minutes in the low-score group versus 66 minutes in the high-score group (Figure 8).

**Figure 6.** Box plot showing the distribution of variability in the midpoint of sleep in the low and high score groups of the FS-J. FS-J: The Japanese version of the Flourishing Scale.

**Figure 7.** Box plot showing the distribution of variability in the midpoint of sleep in the low- and high-score groups of the SPANE-P. SPANE-J: The Japanese version of the Scale of Positive and Negative Experience; SPANE-P: SPANE-J positive experience.
Sleep Duration, Sleep Efficiency, and Psychological State

There was no significant relationship between sleep duration and psychological state. However, the sleep efficiency was higher in the group with high scores on the PSS compared with the group with low scores ($r=0.25, P=.006$).

HRV

HRV was assessed separately during daytime and bedtime. HRV during sleep, specifically the SDNN, was significantly higher in the high SWLS group compared with the low SWLS group ($P=.003$).

EMA Response and Data Obtained From Fitbit Inspire 2

The intraclass correlation coefficients for the EMAs performed in this study were as follows: energy, 0.407; mood, 0.4075; and sleepiness, 0.3286.

The results of the multilevel analysis for the EMA response and data obtained from the Fitbit Inspire 2 are presented in Tables 3-5. Regarding intraindividual variability, energy levels tended to be higher when the deviation of heart rate during sleep on the previous day was lower ($\beta=-.12, P<.001$; Table 3). Furthermore, sleepiness levels tended to be lower on days with a higher deviation of time in bed (TIB; $\beta=-.10, P<.001$; Table 5).

Based on the interindividual evaluation of the relationships, participants with a lower deviation of the midpoint of sleep tended to have relatively higher energy ($\beta=-0.26, P=.04$; Table 3).
Table 3. Multilevel analysis of energy.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Energy Estimates</th>
<th>CI</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-0.12$</td>
<td>$-0.40$ to $0.17$</td>
<td>$.42$</td>
</tr>
<tr>
<td>Age</td>
<td>$-0.09$</td>
<td>$-0.33$ to $0.15$</td>
<td>$.48$</td>
</tr>
<tr>
<td>Gender</td>
<td>$0.22$</td>
<td>$-0.18$ to $0.62$</td>
<td>$.29$</td>
</tr>
<tr>
<td>Metabolic equivalents</td>
<td>$0.02$</td>
<td>$-0.04$ to $0.07$</td>
<td>$.54$</td>
</tr>
<tr>
<td>Response time</td>
<td>$0.03$</td>
<td>$-0.04$ to $0.09$</td>
<td>$.40$</td>
</tr>
<tr>
<td>Deviation of heart rate during sleep (intraindividual)</td>
<td>$-0.12$</td>
<td>$-0.19$ to $-0.05$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Midpoint of sleep (interindividual)</td>
<td>$-0.26$</td>
<td>$-0.50$ to $-0.02$</td>
<td>$.04$</td>
</tr>
</tbody>
</table>

\textsuperscript{a}In addition to a random intercept, random effects were added to the coefficient of response time, metabolic equivalents, and deviation of heart rate during sleep for each user.

Table 4. Multilevel analysis of mood.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Mood Estimates</th>
<th>CI</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-0.08$</td>
<td>$-0.36$ to $0.21$</td>
<td>$.60$</td>
</tr>
<tr>
<td>Age</td>
<td>$-0.02$</td>
<td>$-0.20$ to $0.23$</td>
<td>$.87$</td>
</tr>
<tr>
<td>Gender</td>
<td>$0.13$</td>
<td>$-0.28$ to $0.53$</td>
<td>$.54$</td>
</tr>
<tr>
<td>Metabolic equivalents</td>
<td>$0.04$</td>
<td>$-0.02$ to $0.09$</td>
<td>$.19$</td>
</tr>
<tr>
<td>Response time</td>
<td>$0.08$</td>
<td>$0.03$ to $0.13$</td>
<td>$.002$</td>
</tr>
<tr>
<td>Deviation of the midpoint of sleep (intraindividual)</td>
<td>$-0.04$</td>
<td>$-0.13$ to $-0.05$</td>
<td>$.38$</td>
</tr>
<tr>
<td>Bedtime (interindividual)</td>
<td>$-0.19$</td>
<td>$-0.41$ to $0.03$</td>
<td>$.09$</td>
</tr>
</tbody>
</table>

\textsuperscript{a}In addition to a random intercept, random effects were added to the coefficient of response time, metabolic equivalents, and deviation of the midpoint of sleep for each user.

Table 5. Multilevel analysis of sleepiness.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Sleepiness Estimates</th>
<th>CI</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-0.16$</td>
<td>$-0.09$ to $0.42$</td>
<td>$.21$</td>
</tr>
<tr>
<td>Age</td>
<td>$-0.09$</td>
<td>$-0.27$ to $0.09$</td>
<td>$.33$</td>
</tr>
<tr>
<td>Gender</td>
<td>$-0.38$</td>
<td>$-0.74$ to $0.02$</td>
<td>$.04$</td>
</tr>
<tr>
<td>Metabolic equivalents</td>
<td>$-0.09$</td>
<td>$-0.15$ to $0.03$</td>
<td>$.005$</td>
</tr>
<tr>
<td>Response time</td>
<td>$0.07$</td>
<td>$-0.14$ to $0.00$</td>
<td>$.04$</td>
</tr>
<tr>
<td>Deviation of time in bed (intraindividual)</td>
<td>$-0.10$</td>
<td>$-0.15$ to $-0.04$</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\textsuperscript{a}In addition to a random intercept, random effects were added to the coefficient of response time, metabolic equivalents, and deviation of time in bed for each user.

Discussion

Principal Findings

In this study, we investigated the relationships between sleep indices and HRV, measured through the use of wearable devices, and the daily sleepiness, energy, and mood status assessed by EMA, along with stress and well-being questionnaires, in a cohort of 40 participants. The response rate of the participants was relatively high, and several significant associations were observed: greater variability of the midpoint of sleep was linked to lower well-being and higher stress levels, later bedtimes were associated with increased sleepiness and higher stress levels, and greater HRV during sleep was correlated with higher well-being.

The intraclass correlation coefficients of the EMA performed in this study exceeded 0.1 (energy, 0.4071; mood, 0.4075; and sleepiness, 0.3286). Hence, it can be concluded that the EMA VAS scores in this study captured not only within-individual (within-participant) variation but also psychological differences.
between individuals. The findings regarding the relationship between psychological state and sleep were consistent with previous studies, suggesting that wearable devices and EMA are valuable methods for quantifying both sleep and psychological state.

The results regarding mood and sleepiness were impacted by the level of activity around the response time, highlighting the importance of considering real-time information, including waking time, collected by both EMA and wearable devices when investigating the relationship between psychological state and sleep.

Comparison With Prior Work

Regarding participation rates in studies using EMA, a meta-analysis on EMA compliance reported a typical response rate of 78%. Specifically, the average response rate was 92% for prompts 2-3 times per day for nonclinical participants, and 74% for prompts 3-4 times per day [42]. Therefore, we consider the EMA response rate in this study (1118/1280, 87.34%) to be within a reasonable range. Although not extensively detailed in this paper, it is noteworthy that this study also involved participants wearing the FreeStyle Libre in parallel. Therefore, it was significant that responses were collected without dropouts, despite participants likely having additional tasks compared with a typical study.

The results of this study revealed significantly higher HRV during sleep in individuals with higher scores on the SWLS. It is established that HRV is greater during high-quality sleep and that sleep-related HRV is compromised in individuals with sleep disorders and mental health conditions [43,44]. The findings of this study align with prior research indicating a relationship between sleep quality and well-being [3-5]. Moreover, the results demonstrated that individuals with earlier bedtimes exhibited lower levels of sleepiness, as measured by the JESS, and lower stress levels, as measured by the PSS. Previous studies have suggested that late bedtime is associated with daytime sleepiness and that sleep deprivation can impact daytime stress [27,45], which is consistent with our findings.

Furthermore, our study revealed that the SD of the midpoint of sleep was correlated with stress and well-being. In recent years, the issue of social jetlag—referring to the misalignment between an individual’s internal circadian rhythm and external social demands, often due to work and other societal factors—has gained recognition, particularly among urban office workers. This phenomenon has been noted in Japan as well [46,47]. Social jetlag has been increasingly associated with stress and well-being [48,49]. Our study findings among urban office workers similarly demonstrated a link between disruptions in sleep rhythm and stress/well-being, consistent with previous research on social jetlag. To our knowledge, there have been no studies investigating social jetlag in Japanese individuals using wearable devices or similar tools for objective assessment of sleep. This presents a promising area for future research, offering the opportunity to explore the relationship between social jetlag, sleep patterns, and various health outcomes in this population.

In this study, sleep efficiency was higher among participants reporting higher subjective stress. Typically, it is understood that when subjective stress is high, sleep efficiency and quality are lower [50,51]. However, the reasons for these results in our study remain unclear and warrant further investigation. It is plausible that the group reporting higher subjective stress may have experienced fatigue from work, leading them to fall asleep earlier. However, this is a hypothesis that requires further investigation in future studies.

Regarding the results of the EMA, later bedtimes were correlated with lower energy levels and mood. Previous research has indicated that poor sleep quality can impact positive emotions throughout the day [52], and the findings of our EMA analysis corroborate this trend [23]. Thus, the results of this study are consistent with prior findings in the literature.

Limitations

This study had several limitations. First, as this was an exploratory study, formal power calculations were not conducted, and a practical sample size was adopted. A larger sample size might have enabled the detection of additional findings. Second, the study recruited a small number of participants from a single company in Japan. Given the presumably similar working environment of the participants, the generalizability of the results may be limited. For instance, sleep habits are likely to vary by occupation, as job stress and working hours are known to influence habitual sleep duration and quality [53]. To enhance the generalizability of the study results, it would be necessary to recruit participants from a diverse range of occupations with various lifestyles. Further, the study relied on the Fitbit Inspire 2 for the measurement using wearable devices, which may have limitations compared with more advanced monitoring technologies. Despite providing thorough instructions to participants before the study and conducting regular checks for defective products throughout the study period, intrinsic limitations such as device performance were unavoidable. Moreover, in this study, HRV was calculated from heart rate measurements with a sampling time of 5 seconds, which may be less accurate compared with using methods such as electrocardiography. Finally, the EMA in this study entailed participants answering questions 2 times a week, 4 times a day, to avoid placing excessive burdens on them. However, for instance, if evaluating social jetlag, it would be important to compare weekends and weekdays. Including additional days of the week could potentially yield more meaningful results. It is worth noting that Japanese individuals often tend to select the week could potentially yield more meaningful results. It is worth noting that Japanese individuals often tend to select answers that fall in the midpoint of the range in surveys [54]. This conservative response tendency might have limited their ability to reflect subtle differences in their condition from one day to the next in their EMA responses. Given the limited number of studies using EMA in healthy Japanese adults, it would be beneficial in future research to investigate the potential presence of a response tendency in this population. Furthermore, our EMA design had its limitations. Recognizing that all participants in this study were full-time workers who might often have difficulty responding immediately, we set the EMA notification time as a fixed time but provided a range of response deadlines. Indeed, a randomly scheduled EMA is preferable for evaluating highly variable psychological states. However, the
accuracy of the EMA in this study is limited because of the use of scheduled EMAs, which can lead to participant learning and anticipation. In our study, analysis was conducted based on response time rather than the sending time of the EMA. This approach ensured that regardless of any significant discrepancy between the sending and response times, the answers were elicited at the time of the response. However, it is important to note that this approach can introduce bias by only collecting data when participants choose to respond. Nevertheless, considering that the median difference between the sending and response times was 11.8 minutes, we believe that this issue was not significant in our study.

Conclusions
In this study, the combination of objective assessment of sleep using a wrist-worn device worn by participants, along with subjective symptom assessment through EMA, provided valuable insights into the relationship between sleep and stress/well-being, particularly considering daily variations in sleep patterns. For example, increased sleep variability, characterized by unstable bedtime or midpoint of sleep, was linked to elevated stress levels and diminished well-being. Furthermore, improved sleep metrics, such as lower heart rate during sleep and earlier average bedtime, correlated with enhanced daytime energy levels. Conducting broader-scale studies, such as those on social jetlag, using a methodology akin to that of this study and encompassing a wider array of occupations, could provide valuable insights.

Acknowledgments
We thank all the participants who made this study possible and the members of the Wellness Promotion Department at Mori Building Co., Ltd who helped with the recruitment of the study participants.

Data Availability
All data generated and analyzed are included in this article and a previously published article [38,39]. Raw data sets analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
SK, KK, and TK belong to endowed chairs funded by Mori Building Co. The others have no conflicts of interest.

Multimedia Appendix 1
Histograms depicting the groups formed within each quartile by the population at each score.

References


49. Kashiwagi et al. JMIR FORMATIVE RESEARCH


52. Luckhaupt SE, Tak S, Calvert GM. The prevalence of short sleep duration by industry and occupation in the National Health Interview Survey. Sleep 2010 Feb;33(2):149-159 [FREE Full text] [doi: 10.1093/sleep/33.2.149] [Medline: 20175398]

Abbreviations

BHQ: Brief-Type Self-Administered Diet History Questionnaire
EBQ: Eating Behavior Questionnaire
EMA: ecological momentary assessment
FS-J: The Japanese version of the Flourishing Scale
HRV: heart rate variability
JESS: The Japanese version of the Epworth Sleepiness Scale
METs: metabolic equivalents
PSS: Perceived Stress Scale
SDNN: SD of the NN intervals
SPANE-J: The Japanese version of the Scale of Positive and Negative Experience
SPANE-N: SPANE-J negative experience
SPANE-P: SPANE-J positive experience
SWLS: Satisfaction With Life Scale
TIB: time in bed
VAS: Visual Analog Scale

Please cite as:

Assessment of Stress and Well-Being of Japanese Employees Using Wearable Devices for Sleep Monitoring Combined With Ecological Momentary Assessment: Pilot Observational Study
JMIR Form Res 2024;8:e49396
URL: https://formative.jmir.org/2024/1/e49396
doi:10.2196/49396
PMID:38696237

©Shotaro Kinoshita, Sayaka Hanashiro, Shiori Tsutsumi, Kiko Shiga, Momoko Kitazawa, Yasuyo Wada, Jun Inaishi, Kazuhiro Kashiwagi, Toshikazu Fukami, Yasumasa Mashimo, Kazumichi Minato, Taishiro Kishimoto. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Group Cohesion and Necessary Adaptations in Online Hearing Voices Peer Support Groups: Qualitative Study With Group Facilitators

Alison Branitsky1,2,3,4, BA, MRes; Eleanor Longden1,2,3,4, BSc, MSc, PhD; Sandra Bucciv1,3,4, BSc (Hons), ClinPsyD; Anthony P Morrison1,2,4, BSc (Hons), ClinPsyD; Filippo Varese1,3,4, BSc, MRes, PhD, ClinPsyD

1Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
2Psychosis Research Unit, Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom
3Complex Trauma and Resilience Research Unit, Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom
4Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Alison Branitsky, BA, MRes
Division of Psychology and Mental Health, School of Health Sciences
The University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 07936813441
Email: alison.branitsky@postgrad.manchester.ac.uk

Abstract

Background: Face-to-face hearing voices peer support groups (HVGs), a survivor-led initiative that enables individuals who hear voices to engage with the support of peers, have a long-standing history in community settings. HVGs are premised on the notion that forming authentic, mutual relationships enables the exploration of one’s voice hearing experiences and, in turn, reduces subjective distress. As such, group cohesion is assumed to be a central mechanism of change in HVGs. The rise of digital mental health support, coupled with the COVID-19 pandemic, has resulted in many HVGs adapting to online delivery. However, to date no studies have examined the implementation of these online groups and the adaptations necessary to foster cohesion.

Objective: This study aims to understand the experience of group cohesion among HVG facilitators in online groups compared with face-to-face groups. Specifically, we examined the ways in which the medium through which groups run (online or face-to-face) impacts group cohesion and how facilitators adapted HVGs to foster group cohesion online.

Methods: Semistructured qualitative interviews were conducted with 11 facilitators with varied experience of facilitating online and face-to-face HVGs. Data were analyzed using reflexive thematic analysis.

Results: The findings are organized into 3 themes and associated subthemes: nonverbal challenges to cohesion (lack of differentiation, transitional space, inability to see the whole picture, and expressions of empathy); discursive challenges to cohesion (topic-based conversation and depth of disclosure); and necessary adaptations for online groups (fostering shared experience and using the unique context to demonstrate investment in others). Despite challenges in both the setting and content of online groups, facilitators felt that group cohesion was still possible to achieve online but that it had to be facilitated intentionally.

Conclusions: This study is the first to specifically investigate group cohesion in online HVGs. Participants noted numerous challenges to group cohesion when adapting groups to run online, including the unnaturally linear narrative flow of dialogue in online settings; lack of transitional spaces, and associated small talk before and after the session; ease of disengagement online; inhibited sharing; and absence of shared physical presence online. Although these challenges were significant, facilitators nevertheless emphasized that the benefits provided by the accessibility of online groups outweighed these challenges. Necessary adaptations for cultivating group cohesion online are outlined and include capitalizing on moments of humor and spontaneity, using group activities, encouraging information sharing between participants using the chat and screen-sharing features, and using objects from participants’ environments to gain deeper insight into their subjective worlds.
Introduction

Background

Hearing voices peer support groups (HVGs) have a long-standing history within the community [1]. The establishment of HVGs is one of the primary objectives of the Hearing Voices Movement (HVM), an international, survivor-led coalition of voice hearers and their allies, which strives to shift professional and public attitudes toward voice hearing away from biogenetic, pathology-based models toward understandings that locate voice hearing as a complex psychological experience imbued with subjective meaning (eg, psychosocial, cultural, or spiritual) [1]. HVGs are premised on the notion that forming genuine, mutual relationships between group members can support individuals to safely explore their voice hearing experiences and, as such, may result in positive psychosocial change [2-11]. As a result, group cohesion is posited to be of central importance in HVGs [7,9].

The concept of group cohesion arises from group psychotherapy literature and refers to participants’ feelings of connectedness to one another and to the group as a whole [12]. Group cohesion is believed to be the foundational feature upon which all other therapeutic work takes place [12]. Although their distinct ideologies mean theories of group psychotherapy cannot be uncritically applied to peer support groups [13], it may nevertheless be instructive to further investigate the role of group cohesion in HVGs.

Although peer support is an established aspect of mental health provision, a more recent development concerns their online delivery; a process largely expedited by the COVID-19 pandemic, wherein many forms of support, including HVGs, were forced to be held remotely. As a result, there has been growing interest in understanding if and how group cohesion can form in online groups [14-16]. The literature on online group cohesion has been largely theoretical and anecdotal but suggests that although group cohesion does occur, there are various challenges to its achievement, including the ease of disengagement and the lack of embodied presence of both the participants and therapist [14]. Given that online peer support represents a potentially accessible and scalable form of support that can be implemented in multiple contexts globally, it is necessary to further investigate the development of group cohesion in online spaces. This type of investigation is particularly important to HVGs as they are an existing resource that exist internationally and can provide vital support to individuals who may otherwise be insufficiently supported by mental health services [17,18].

Study Objectives

This study aims to understand HVG facilitators’ experience of group cohesion in online groups compared to face-to-face groups. Specifically, the following issues were examined: (1) How does the medium through which groups run (online vs face-to-face) impact group cohesion? and (2) How do facilitators adapt HVGs to foster group cohesion online?

Methods

Design

Semistructured qualitative interviews were conducted with 11 HVG facilitators. Data were analyzed using a hybrid of deductive and inductive reflexive thematic analysis (RTA), a flexible approach that recognizes the centrality of the researcher’s subjectivity as an analytical tool and emphasizes reflective engagement with theory, data, and interpretation [19]. The study was further underpinned by a critical realist epistemological position, which acknowledges the existence of an objective social world while recognizing that one’s understanding of that social world is shaped by one’s experiences within it [20]. The subjective realities of both participants and researchers were acknowledged while still aiming to understand and explain experiences that exist beyond the study sample [21].

Topic guides were developed by AB in consultation with patient and public involvement and engagement (PPIE) experts with relevant experience of HVG facilitation or membership. Specifically, topic guides were informed by the supposition by Yalom [12] that group cohesion is the foundation for effective groups; as such, the interviews focused on facilitators’ experiences of if and how group cohesion was cultivated in both face-to-face and online settings. Topic guides were further informed by previous literature on the proposed mechanisms of action in HVGs [6] as well as the lived experience of AB and PPIE representatives as HVG facilitators. All questions were piloted with PPIE experts before commencing recruitment, with topic guides revised based on PPIE feedback. Questions were open ended and sequencing was determined based on the flow of conversation. Probes were used throughout to elaborate on relevant topics. Topic guides were iterative, with specific questions being added or emphasized as preliminary themes began to be constructed. Reflective logs were kept throughout data collection and analysis. The COREQ (Consolidated Criteria for Reporting Qualitative Research) were consulted in reporting this research (Multimedia Appendix 1) [22].

Participants

Participants in this study comprised an international sample of HVG facilitators, who were recruited using convenience and snowball sampling methods through the HVM (Figure 1). Adults who spoke English and had at least 3 months of experience facilitating HVGs were eligible for interview. The sample size was determined through information power [23] and thematic sufficiency [19], with the former referring to the degree in which participants hold information to address the research question and the latter referring to the point at which data collection is sufficiently complete as to meaningfully answer the research questions. This approach to data completeness was chosen over
the more common idea of data saturation, which is subjective in nature and theoretically incompatible with RTA [24].

**Ethical Considerations**

Ethics approval was received from the University of Manchester Research Ethics Committee (2022-13944-22907), and all participants provided written informed consent. Participants received a participant information sheet outlining the parameters of study involvement and were given a minimum of 24 hours to decide whether they wanted to take part. The voluntary nature of the research was emphasized throughout, including during the qualitative interviews. All interviews were confidential, and transcripts were pseudonymized with any other identifiable information removed. Due to resource constraints, no compensation was offered for study participation. Furthermore, due to the potentially identifiable nature of the data, interview transcripts were not made available via public data repositories, and the data were not shared outside the research team.

**Procedure**

Potential participants contacted AB via email to express their interest in the study. Interviews were conducted by AB, a researcher with lived experience of voice hearing and HVG group facilitation, between April and June 2022. Participants were aware of AB’s background as an HVG facilitator. All interviews were conducted online via Zoom (Zoom Video Communications, Inc) [25] and were audio recorded using Zoom’s built-in encrypted recording software. Interviews were then transcribed verbatim. Participants were assigned a pseudonym, and identifying information was removed from the transcripts. The median interview length was 58 minutes 42 seconds (range 38 min 7 s to 89 min 20 s).

**Analysis**

A predominantly inductive approach to data analysis was adopted. Data were open coded, and preliminary codes were inductively generated based on participants’ responses. However, given that the topic guide was framed on the theoretical assumption that group cohesion was crucial for both face-to-face and online HVGs, deductive analysis was also used to allow for the identification of codes that could meaningfully answer the research question. As such, both latent and semantic codes were generated. The iterative approach to data analysis by Braun and Clarke [19] was followed: (1) data familiarization through repeated reading of transcripts, (2) generating initial codes, (3) generating initial themes, (4) reviewing initial themes, (5) defining and naming themes, and (6) producing the report.

The final thematic structure was derived by AB in consultation with EL, SB, APM, and FV. Data analysis took place using NVivo (version 12; QSR International) [26]. No attempt was made to establish interrater reliability as it is antithetical to the philosophical position of RTA [27]; however, sense checking among the research team, the use of reflective notes and memos, and the 20-point guide by Braun and Clarke [27] for the assessment of RTA research quality were used to enhance the trustworthiness of the data. Furthermore, validation of the thematic structure was sought from all study participants, all of whom consented to member checking, as well as from PPIE representatives.

**Reflexivity**

All the researchers in this study hold perspectives on voice hearing that align with those of the HVM; specifically, that voice hearing is an inherently meaningful psychological experience that is worthy of ongoing exploration. AB is a PhD-level researcher; EL is an academic psychologist; and SB, APM, and FV are academic clinical psychologists; all of whom have experience supporting and developing interventions, including digital interventions, for individuals who hear voices. All researchers had extensive experience conducting and analyzing qualitative data. AB and EL have similarly been members and facilitators of HVGs (AB has experience with online and face-to-face HVGs, and EL has experience with face-to-face HVGs). The background of the research team, particularly their commitment to developing various forms of psychosocial support for voice hearers, influenced the generation and interpretation of the data. To maintain transparency around how the researchers’ backgrounds were interacting with the data, AB kept reflective logs to record impressions about interviews and emergent themes and document ways in which findings paralleled or differed from her lived experience as a facilitator. Emphasis was placed on ensuring that divergent views were represented within both analysis and reporting.

**Results**

**Overview**

A total of 11 participants consented to take part in the study. Facilitators ranged in age from 25 to 52 (mean 41.27, SD 7.40) years with most participants identifying as male (7/11, 64%) and White (10/11, 91%). Participants had between 1 and 10 (mean 5.45, SD 3.06) years of experience facilitating HVGs.
Key participant characteristics are presented in Table 1, with a table of full participant characteristics available in Multimedia Appendix 2.

Results were organized into three themes and associated subthemes: (1) nonverbal barriers to cohesion (lack of differentiation, transitional space, inability to see the whole picture, and expressions of empathy), (2) discursive barriers to cohesion (topic-based conversation and depth of disclosure), and (3) necessary adaptations for online groups (fostering shared experiences and using the unique context to demonstrate investment in others). The findings from the first 2 themes informed the necessary adaptations presented in the third theme. A thematic map of the findings is presented in Figure 2.

Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Name</th>
<th>Age (y)</th>
<th>Gender</th>
<th>Geographical location</th>
<th>Group medium</th>
<th>Group moved online during the COVID-19 pandemic?</th>
<th>Length of time facilitating (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sean</td>
<td>41</td>
<td>Man</td>
<td>Western Europe</td>
<td>Face-to-face</td>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Annika</td>
<td>34</td>
<td>Woman</td>
<td>Northern Europe</td>
<td>Both</td>
<td>Yes</td>
<td>4</td>
</tr>
<tr>
<td>Patrick</td>
<td>46</td>
<td>Man</td>
<td>Western Europe</td>
<td>Online</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>Rachel</td>
<td>25</td>
<td>Woman</td>
<td>North America</td>
<td>Face-to-face</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Arjun</td>
<td>36</td>
<td>Man</td>
<td>Western Europe</td>
<td>Both</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>Callum</td>
<td>52</td>
<td>Woman</td>
<td>Western Europe</td>
<td>Both</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>Sabina</td>
<td>36</td>
<td>Woman  or gender fluid</td>
<td>Western Europe</td>
<td>Both</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Michail</td>
<td>46</td>
<td>Man</td>
<td>North America</td>
<td>Both</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>Lex</td>
<td>46</td>
<td>Man</td>
<td>Western Europe</td>
<td>Both</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>Noah</td>
<td>47</td>
<td>Man</td>
<td>Western Europe</td>
<td>Face-to-face</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Isabella</td>
<td>45</td>
<td>Woman</td>
<td>Western Europe</td>
<td>Both</td>
<td>Yes</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 2. Thematic map depicting the relationships between themes and subthemes.

Nonverbal Barriers to Cohesion

Groups that took place online lacked many of the nonverbal elements of communication that were previously relied upon to create a cohesive space. These nonverbal elements included both the group setting and cues between HVG participants.

Lack of Differentiation

HVGs were often described as “unique” (Lex) and “different” (Rachel) spaces, whose ethos of curiosity, openness, and nonjudgmental approaches toward voice hearing were intentionally cultivated and led to connections that could not be fostered elsewhere:

It wasn’t ever said...but we all picked up that describing these sorts of [voice hearing] experiences wasn’t welcomed in other groups. So we participated in other groups on another level. But have these [voice hearing] experiences close to our chest. So when you have this space to talk about it and then realize “oh I’m so far from alone in this” that you have all these different individuals... and all of a sudden it’s kind of like, “oh you share this with me as well.” And it feels very opposite of being alone. Like, it was really connected. [Annika]

In turn, many facilitators took great care to create a soothing environment that reflected the distinctness of the groups:
We had always put a lot of emphasis on making a nice atmosphere. Much more than in our [other] groups. We turned off the industrial lights; we had candles; we sat in a circle. You don’t even need to sit on a chair if you don’t want to. We had like yoga meditation chairs. And some people sat on the floor. And it was so different... We were really thinking of the atmosphere. And then when we moved online it’s just, uh, the screen and the face and you couldn’t really make this atmosphere in the same way. [Annika]

In contrast, online meetings tended to feel indistinct from both other mental health groups, as well as other online gatherings, wherein participants were often sitting in the same place regardless of whether they were joining the HVG, a family event, or a work meeting. In this regard, this lack of differentiation resulted in grief as it felt some HVGs lost the distinctiveness they had in the face-to-face setting. Groups such as Annika’s, which adhered very strongly to the ethos of the HVM, particularly struggled with the transition online because the truly distinct nature of the space was largely lost.

Transitional Space

Facilitators spoke to the importance of transitional space, or the “before and after” (Lex) of the group, where participants could “walk together to the bus stop” (Lex) or meet in the doorway over tea and coffee to enjoy small talk and forge “potential relationships outside of this room” (Annika). The social energy likewise made its way into the meeting, as described by Sabina: “Often there’s coffee or cookies or something like that and so often a conversation...it starts with...’the coffee is stronger this time than last time’.” As such, entering a group without a transitional space could potentially feel lonelier and more jarring, especially for new members:

Whereas with online groups, particularly if you’ve never been before, it really is, it’s that all or nothing: “I’m in the waiting room by myself and now I’m with all of these people that I don’t know. And there’s not that, there’s no clear etiquette about how I’m supposed to behave because every group is completely different.” [Annika]

Similarly, facilitators also reported how a lack of intentional time for small talk could inhibit members’ subsequent willingness to share, given that they no longer had initial opportunities to connect on less vulnerable topics.

Inability to See the Whole Picture

As described by facilitators, online groups posed a unique visual context: that rather than sitting in a circle, as is customary with face-to-face groups, participants were instead confined to small, 2D screens online, if their cameras were even on at all. As such, the information and input participants received from and about one another was diminished, which could make individuals feel less connected. As Lex described, “some people get their camera off. Only audio...that makes a whole different dynamic,” whereas Michail noted how “on Zoom it’s just so much easier to hide.” Even in cases where participants had their cameras on, engagement could still be difficult to assess, as it was impossible to tell, for example, if “you’re busy playing with your phone” offscreen (Arjun).

In this regard, one clear impact of meeting online was the ease with which participants could disengage with the group. As Michail reported, “it’s just so much easier to be distracted on Zoom.” Although seeing members in their own home enabled them to join the group from an environment that felt safe and comfortable, at times this could also be taken to the extreme when participants visibly engaged in other activities during the meeting:

[Group member] sat there and poured himself a glass of beer. And drank it during the group. And it’s like, that would have never happened if we were sitting together in a circle because it’s just, it doesn’t work like that. [Isabella]

Not all facilitators felt it was problematic if group members engaged in other activities during the group, especially if doing so made it easier for them to participate. However, it did highlight the importance of discussing expectations about engagement with their respective members in a way that was unnecessary in face-to-face groups. Annika described having to create group agreements online that had not previously existed: “we had to have rules like, ‘you have to be dressed, you’re going to a meeting’ and ‘please do not be doing something else’.”

The ability to disengage could also pose unique challenges within the online context. Disagreement between members could be particularly troublesome in this regard, as abruptly leaving could prevent disagreements from being resolved within the group space: “One time there was sort of a heated discussion, one of the guys just hung up...but that was hard and kind of a bummer. So it’s easier to check out in that way” (Michail).

Facilitators further noted that at times it could be difficult to navigate confidentiality online, not least because many participants did not have access to truly private spaces. As Annika described, “Like they had the computer in the kitchen and a parent could just walk by whenever. It didn’t really feel as confidential.” This, in turn, could impact the trust that group members were able to develop with one another. Annika noted that this was particularly pronounced for group members who had not been a part of her group before it transitioned to being held online, stating, “I don’t remember somebody starting to trust online rather than in person.”

Expressions of Empathy

One of the greatest challenges posed by online groups was the lack of proximity between group participants. This resulted in participants’ needs not always being responded to in a way that felt appropriately compassionate:

I had a couple instances where I’d get on and somebody looks like they’re in trouble...they’re way over there. I can’t help them way over there. But if they’re in front of me, you can try and do something you know...If they’re definitely in front of you, you can try to give them a hug if it comes to it. [Patrick]
Other facilitators spoke of how being in the same space and seeing and hearing the same things in one’s surroundings naturally fostered a sense of connection. As described by Lex, “there’s also the nonverbal part and being in the same place and seeing the same sun or rain outside when you’re having a break… I think that’s really powerful.”

The importance of physical contact, such as “you shake each other’s hand or a sometimes a hug” (Lex) was important not just for expressing social bonds, but for conveying empathy and understanding toward participants who were distressed. Isabella compared the experiences in her online and face-to-face group in the following way:

Whereas online we had as well that people were like distressed or crying and there isn’t that same level of connection or compassion if that makes sense. Somehow, because there is this distance. There is this distance of not being in the, in the same room together. [Isabella]

Facilitators described it as “natural” (Patrick) to interact in person, and that one could achieve a richer, more embodied experience wherein nonverbal cues provided important communicative information: “You know, just being in person, being in the presence of other people, the body language and just you know, you forget how much you get out of sitting with someone” (Michail).

**Discursive Barriers to Engagement**

Without the nonverbal cues present in face-to-face meetings, facilitators noted that there were changes in both the style and content of verbal communication in online groups.

**Topic-Based Conversation**

Facilitators noted how the group medium could have a distinct impact on the flow of conversation. For example, a defining feature of face-to-face groups was topic-based conversations, or conversations that were centered on a particular issue to which everyone contributed, regardless of whether it was particularly relevant to them at the time. These conversations were marked by a “free flow” (Annika) of “collaborative interrupting” (Noah), where “you go to something and it leads to something else” (Arjun). “There’s no uncomfortable silences [and] everybody gets to speak” (Sean) and “[participants] really bounced off each other and related to each other on experiences that you wouldn’t imagine that anybody could relate on” (Rachel):

If it was running well, we really didn’t need to facilitate much…And maybe someone came in wanting to start with a, with a topic or, that they kind of added to another person’s story, but really like, it flowed. [Annika]

These kinds of interactions reduced the reliance on the facilitator to keep the discussion going, with participants instead empowered to carry on the conversation themselves:

It seemed like it was definitely a running machine when everybody did kind of play a role. It definitely lessened any outlook on me as some kind of authority figure for everyone to see…And I think it made everybody feel a bit more of a sense of connection. Like it was a little community and that they all were really relating to each other. [Rachel]

This “natural” (Patrick) conversation allowed discussions to deepen in a way that felt appropriate and engaging: “they respond to each other. Sometimes asking questions, deepening questions to each other” (Noah).

In contrast to face-to-face meetings, the physical distance between participants in online groups made topic-based conversation more difficult. In the absence of nonverbal cues typically present in face-to-face meeting, Isabella noted that it was difficult for participants to know when to speak next:

They have to lift the hand and say, “can I say something”…because sometimes they don’t know like, “is it my turn to speak?” “Can I speak?” Because you don’t have that connection.

This was true even in groups that transitioned to being online and where the participants had previously connected face-to-face. The resultant flow was markedly more sequential, with individuals tending to bring up a topic that was relevant to them, and the topic then shifting when the next person spoke: “Yeah, in the online setting there was, ok this topic was talked about. Now I bring this topic for new conversation” (Sabina). Thus, when stories were shared as isolated silos, participants were not connecting with each other as intimately as they would when face-to-face:

It’s more like one says something and then the other one then tells her story, the other one then tells his story and then the next one. There’s not really like interaction as much as when we were in a room together. [Isabella]

**Depth of Disclosure**

Some facilitators felt that conversation could be “flattened” (Michail) online, that there was more hesitation sharing “intimate” (Sabina) details of one’s life, and that, in general, “you don’t go too heavy in the group” (Patrick). As a result, conversations sometimes lacked the depth that was often felt in face-to-face meetings. This, in turn, could hinder the development of trusting relationships because participants felt neither open enough to talk about vulnerable subjects nor secure enough to connect over day-to-day experiences. Sabina described one such situation:

In [in-person groups, the conversation is deeper but also could often go to more lighter topics, like their favorite TV series or something like that. And in online setting it wasn’t as deep, but it wasn’t also as light. And I had the feeling like [the participant[s], they didn’t feel secure enough to put this light things. So, they feel “ok this topic about TV series is too dumb or too, um, superficial or too, yeah whatever;” And in [in-person group[s] they, after deep topic, they went after the connecting together; they feel secure enough to put also this more superficial [topic]. And it brings, brings a little bit of lightness and cosiness in the group setting.
Necessary Adaptations for Online Groups

Despite the challenges posed by the online medium, many facilitators were still able to run cohesive and successful groups. This theme outlines the specific ways in which cohesion between members was cultivated online.

Fostering Shared Experience

Creating and building on shared experiences was a central feature of forming cohesion within online groups and was achieved in 2 primary ways: the first being capitalizing on spontaneity, and the second through group activities. In terms of the former, moments of spontaneity served as the online corollary to infectious laughter in face-to-face spaces; indeed, the fact that participants were not in a shared space could, at times, enhance the experience as they had access to objects that would not have been available within in-person groups:

A couple of times this spontaneous thing happened where one group member was wearing a new hat and he was telling us about his hat. And everybody in the group went off and got a hat. And it was so funny and fun and he still will talk about that day. And we laughed so much that day...And I remember feeling like, “it feels so good to laugh like this.” And we’ll laugh about that day still. And he wears the hat all the time and sometimes he’ll say, “c’mon guys, where are your hats? You know it’s our group. Where are your hats guys? C’mon?” You know this sort of spontaneous thing that couldn’t really happen in person. [Michail]

In this respect, spontaneous, shared moments provided a sense of continuity between groups and served as a crucial moment that could be repeatedly referred to, ultimately enhancing the sense of group identity. Following this experience, Michail wondered if similar occurrences should be fostered more deliberately: “Boy I wonder if it’s nice to have an ice breaker activity where you actually share something from your space.” Similarly, such moments also enabled group members to connect with one another over other topics of mutual interest, experience, or “personal things like music” (Lex), which left facilitators feeling like they were having “a co-human kind of experience” (Lex).

Although valuable, facilitators of online HVGs also noted that moments of spontaneous humor could be few and far between. However, when this was the case, the second means of fostering shared experience could be used, namely group activities. These could range from voice-related work, such as introducing members to the Maastricht interview [28], a technique used to explore voices’ characteristics, origins, and emotional impacts; practices such as reiki; or “playing music” (Patrick) and “watching videos together” (Lex). These activities could provide a starting point for conversation: “we’ll play [Beyond Possible: How the Hearing Voices Movement Transforms Lives [29]] and it really brings up a lot of stuff to talk about” (Michail), whereas at other times, participating in the activity was a shared experience in and of itself:

Sometimes...we’ll have something like...something more structured [co-facilitator will] bring somebody in to talk about reiki or have a guest speaker for meditation or something. Um, or tapping [as described by the emotional freedom technique] and stuff like that. [Callum]

The ease with which information could be shared online was an asset that offered individuals a novel way of participating in the group and sharing their experiences or interests. The chat and screen-sharing feature allowed ideas to be shared in real time, which enabled the entire group to have access to the same information. In turn, such information could be informative (“maybe if someone was sharing a visual aid or like ‘I was listening to something. Here is the link.’ It would go in the chat” [Annika]) or humorous (“if there was a funny video or something that was much easier to share in the online group” [Sabina]).

Using the Unique Context to Demonstrate Investment in Others

Facilitators emphasized the fact that group members were equally able to foster intimacy online as they did in face-to-face settings, but that the manner of achieving this was necessarily different. For example, seeing people in their own homes provided “an extra possibility to show an interest in their world” (Lex). At times, facilitators could gain a more contextualized sense of who a person was and what their life was like outside the group when they were able to see them in their home environments:

We had someone who wasn’t comfortable speaking, but he would attend, and he would write in the chat. And we would ask “is it okay if we read from the chat to the group” um, and, and he was okay with that. But he was doing that because he was concerned about who around him could hear. [Michail]

Online groups also offered an additional level of accessibility, thus enabling individuals to take part who either did not feel comfortable participating in or did not have access to face-to-face groups. For some members, joining a group online was “a lower bar to participation” wherein “it makes it a little easier for folks who are reticent to just kind of try it” (Michail). The online medium also provided voice hearers with the opportunity to engage with the group in a way that worked best for them. For some members, this opened up new channels of communication: “if you’re not confident to speak...you can type. You’ll be able to type [in the chat]” (Callum). For others, it enabled them to engage in a way that felt safe to them:

We’ve had [participants] say “can I join and not show my face? Is that ok?” And we say...”sure, as long as it’s ok with the group, as long as the group doesn’t feel weird about it but we want you to participate as fully as you can. Or in a way that feels right for you.”...It gives a different set of opportunities for modulating engagement. [Michail]

The distance and anonymity of online groups was additionally advantageous for some members, making it easier to disclose personal experiences precisely because they could talk about them in a “safe space” (Lex) that felt very distant and removed from the rest of their lives: “I would say you can very [easily]
communicate stuff on Zoom you probably wouldn’t do it the person was living down the road...[online] it’s like, it’s not like it’s gonna walk out of the room” (Patrick).

Discussion

Principal Findings

This study is the first to explore how the medium through which HVGs are delivered impacts cohesion within the group. It is similarly the first to analyze the specific adaptations necessary to cultivate group cohesion in online peer support settings. In terms of building group cohesion, facilitators identified several nonverbal and discursive barriers to running groups online. The primary nonverbal barriers included the lack of differentiation between HVGs and other online spaces; the lack of transitional space online; the reduced visual input online, which prevented individuals from fully seeing what was going on in one another’s environment; and the inability to express and receive empathy nonverbally to the same extent as in face-to-face settings. These barriers, in turn, impacted discursive communication within the group, with online groups having an overprescriptive flow of dialogue where only one person could share at a time, thus resulting in discussions that tended to build less naturally on a single topic and where participants were discouraged from sharing as deeply. However, despite these challenges, facilitators were able to adapt to running cohesive groups online by (1) capitalizing on moments of spontaneity and using group activities to build a shared experience and (2) bringing members’ subjective environments into the group space.

These findings largely correspond with previous literature outlining how nonverbal elements of the group, both in terms of group setting and nonverbal communication between members, strongly influence group cohesion. For example, Weinberg [15] described how one of the tasks of group therapists was creating a “holding space” in which therapeutic work can take place, including deliberately choosing and arranging furniture, lighting, and objects in the room in such a way as to further therapeutic aims. Indeed, Payne et al [9] described one of the main benefits of HVGs as creating a “containing” environment in which anomalous experiences and intense emotions can be shared and processed. Similarly, facilitators in this study often took great care to create a soothing and inviting atmosphere that reflected the distinct values of HVGs; however, in the absence of this control, they could subsequently struggle to create an environment that felt sufficiently safe and containing to encourage disclosure and cohesion between members.

In turn, Weinberg [15] further noted how the opportunity for small talk before, during, and after meetings was necessary for establishing group cohesion. HVG facilitators similarly described how a lack of transitional space into and out of the group inhibited this small talk and acted as a barrier to connection. This is perhaps surprising given previous research on HVGs, which suggests that members primarily connect over the shared experience of voice hearing [4,7,30]. Future online HVGs, and online peer support groups more broadly, may therefore benefit from prioritizing time for these more informal, “co-human” (Lex) connections. This may be achieved by intentionally building in transitional space: for example, by opening the online room early and allowing participants to stay for several minutes after the group ends, encouraging small talk between participants, and having refreshment breaks during the group.

The flattening of visual cues online has similarly been demonstrated to be a barrier to cohesion. Within individual therapeutic settings, Grondin et al [31] argued that empathy and, by extension, alliance (the individual corollary to cohesion), are established through an iterative cycle of producing and perceiving cues that enable the interpretation of others’ emotional states. Although these cues can be verbal [32], they are very often nonverbal and include signals such as facial expression [33], body posture [34], and eye contact [35]. In online settings, these cues tend to be diminished [15], especially in instances where participants have their cameras off and thus, visual cues are neither sent nor received. This lack of visual cues can result in individuals feeling as though they are not being sufficiently empathically received by others [31,35], as well as having the potential to undermine trust and disclosure through reduced awareness as to whether other members are in a truly confidential space [15]. This study is largely consistent with these findings, with facilitators additionally noting the importance of overt nonverbal interactions, such as handshakes or hugs, in the establishment of cohesion. Interestingly, these barriers were still present in HVGs that transitioned from face-to-face during the COVID-19 pandemic, with facilitators often describing highly cohesive groups that subsequently struggled online, despite being composed of largely the same people. This can perhaps be explained by a sense of loss that accompanied the online transition [15], particularly for groups that adhered strongly to the ethos of the HVM and placed great emphasis on creating a distinct, user-led environment. Although not a replacement for nonverbal cues, future online groups may benefit from verbalizing the nonverbal; for example, by emphasizing the distinct values at the start of each group, acknowledging when a nongroup member appears in the background, or expressing that one wishes they could give another member a hug.

The distinct discursive elements of online groups were similarly consistent with previous research. For example, Weinberg [15] contended that there can be an unnaturally linear flow of dialogue in online psychotherapy groups because participants are in their own “boxes” and less able to pick up on nonverbal cues that indicate when the next person can speak. Similarly, facilitators in this study found that not only did discussions tend to focus on one person at a time rather than flow naturally between participants but disclosure was also not always as personal as in the face-to-face groups. This is a potentially important barrier to cohesion, and previous research suggests that facilitators can overcome this challenge by verbalizing statements and emotions, which previously would have remained nonverbal as a means of enhancing the continuity of dialogue between participants [35].

Although facilitators often noted significant challenges to group cohesion, many were nonetheless able to run successful and cohesive HVGs online. While previous research recommends acknowledging members’ subjective environment in moments...
where the space is disturbed (eg, a parent walks into the room during the group [15]). HVG facilitators went further and highlighted the utility of using each member’s individual space as a way of proactively showing interest and investment in their world. Given their open structure [30] and nonmanualized approach, HVGs may be particularly well-suited for this endeavor, as aspects of a member’s home environment can be incorporated more seamlessly and spontaneously into group discussion. Furthermore, previous literature recommends the use of “ice breaker” activities at the beginning of the group to help individuals ease into the space and connect with one another [36]. Sharing items from one’s home may serve as way of having less vulnerable conversations while still enabling members to connect with one another and gain unique insight into their subjective worlds. Similarly, HVG facilitators note that either planned (eg, sharing a video on voice hearing) or spontaneous activities (eg, members making music together) can be a useful tool in establishing cohesion. In this respect, facilitators in this study also emphasized the importance of capitalizing on moments of humor and spontaneity as a means of fostering shared experience between members. Although previous research has supported an association between humor and therapeutic alliance [37], this is the first study to specifically outline its utility in peer support groups and highlight how the unshared environment may lend itself more easily to these spontaneous moments. When moments of humor or spontaneity do arise, future groups should therefore purposefully use these moments to strengthen connections between members.

Mental health services may face particular challenges in providing adequate support to individuals who hear voices [17,18]. Therefore, despite the challenges outlined, caution should be taken in either interpreting online HVGs as inferior forms of support or in minimizing the value of their accessibility. Specifically, online groups allow members to engage more easily in ways that felt comfortable for them, as well as offering support to those who cannot access face-to-face groups. Furthermore, it was notable that facilitators did not mention voices and associated experiences (eg, paranoid beliefs, particularly those about being surveilled by technology) as being a major barrier to either group cohesion or to individuals successfully accessing the online space.

A significant strength of this study is its international scope, being the first of its kind to recruit participants throughout Europe and North America in a single group. Furthermore, none of the findings were region specific, meaning facilitators from across the global north may potentially benefit from implementing the strategies recommended below. Although HVGs are distinct in their ethos, the findings and recommendations may also be applicable to other forms of online support (although in non-HVG spaces, care should be taken to emphasize how the latter may be distinct from other types of mental health support groups). However, it should also be noted that the findings cannot be uncritically applied to groups facilitated in the global south, where norms and expectations about relational dynamics in peer support spaces may differ [38] and further research would be required.

Finally, the study must be interpreted in view of its limitations. Although previous research has highlighted the centrality of group relationships to the successful running of HVGs [6,7,9], the concept of group cohesion arises from psychotherapy literature rather than the peer support literature, and therefore cannot be uncritically applied [13]. Furthermore, the sample was racially homogeneous, with most participants identifying as White. Given the global spread of HVGs, future studies would be greatly enhanced by recruiting a more ethnically representative sample. Furthermore, 87% (7/8) of the facilitators with online experience had to adapt a face-to-face group to an online platform following the COVID-19 pandemic, and additional research should focus on delineating the development of cohesion in groups that started online compared to those that had to transition into being so. Finally, although efforts were made to member check results with all participants, responses were only received from 2 individuals (who suggested no substantial changes), and future research would be strengthened from incorporating a wider range of reflections into the analysis.

Conclusions

In conclusion, it is possible to overcome the distinct challenges inherent to the medium to deliver cohesive HVGs online. Facilitators should be mindful that there are both nonverbal and discursive barriers to cohesion, many of which originate from the distance and detachment inherent in the online space. However, group cohesion may be enhanced by highlighting the distinct nature of the group, creating transitional space, verbalizing the nonverbal, capitalizing on spontaneity, and using participants’ unique environments to foster intimacy. Recommendations for promoting cohesive groups online can be found in Textbox 1, and it is hoped such forums may continue to prove an accessible and implementable form of support not only for individuals who hear voices but also for anyone who may benefit from peer engagement.
Textbox 1. Recommendations for optimizing group cohesion in online hearing voices peer support groups.

<table>
<thead>
<tr>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Avoid admitting all participants into the meeting at once. If possible, start</td>
</tr>
<tr>
<td>the meeting 10-15 minutes before the designated start time and allow</td>
</tr>
<tr>
<td>participants to arrive early to allow time to settle into the group space.</td>
</tr>
<tr>
<td>• Plan time for small-talk opportunities before the group starts and after it</td>
</tr>
<tr>
<td>ends.</td>
</tr>
<tr>
<td>• Intentionally begin the group by reminding participants of its norms and values.</td>
</tr>
<tr>
<td>This helps to differentiate the group from other online spaces.</td>
</tr>
<tr>
<td>• Use the environment by having participants share objects from their homes.</td>
</tr>
<tr>
<td>Where participants have their cameras off, consider using auditory materials</td>
</tr>
<tr>
<td>such as music.</td>
</tr>
<tr>
<td>• Be mindful of the flow of conversation; encourage participants to respond to</td>
</tr>
<tr>
<td>one another rather than the facilitator.</td>
</tr>
<tr>
<td>• Use group activities, either psychoeducational (eg, sharing information or</td>
</tr>
<tr>
<td>videos) or social (eg, everyone sharing an object from their environment)</td>
</tr>
<tr>
<td>where appropriate.</td>
</tr>
<tr>
<td>• When interruptions occur (eg, a family member appears in the background),</td>
</tr>
<tr>
<td>acknowledge and talk about it with the group.</td>
</tr>
<tr>
<td>• Frequently return to conversations about group norms and expectations (eg, is</td>
</tr>
<tr>
<td>it acceptable to be multitasking while in the group?).</td>
</tr>
<tr>
<td>• Have refreshment breaks and encourage participants to bring food and drink</td>
</tr>
<tr>
<td>back to the group the same way they would in face-to-face meetings.</td>
</tr>
<tr>
<td>• Capitalize on moments of humor and spontaneity by encouraging other members</td>
</tr>
<tr>
<td>to join in and not moving to another topic too quickly.</td>
</tr>
</tbody>
</table>

Acknowledgments
The authors would like to thank all participants for sharing their experiences and the patient and public involvement and engagement representatives for their consultation and support of this study.

Data Availability
All data generated and analyzed during this study are not publicly available due to privacy and ethical restrictions.

Conflicts of Interest
AB and EL have received financial compensation for presentations and trainings on the Hearing Voices Movement and hearing voices peer support group facilitation. SB is director and shareholder of CareLoop Health Ltd, a spin-out from the University of Manchester to develop and market digital solutions for remote monitoring using smartphones for mental health conditions, currently schizophrenia and postnatal depression. SB also reports research funding from the National Institute for Health and Care Research (NIHR) and The Wellcome Trust. All other authors declare no other conflicts of interest relevant to this study.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 416 KB - formative_v8i1e51694_app1.pdf]

Multimedia Appendix 2
Full participant characteristics.
[DOCX File, 16 KB - formative_v8i1e51694_app2.docx]

References


Abbreviations
- **COREQ**: Consolidated Criteria for Reporting Qualitative Research
- **HVG**: hearing voices peer support group
- **HVM**: Hearing Voices Movement
- **PPIE**: patient and public involvement and engagement
- **RTA**: reflexive thematic analysis

*Please cite as:*  
Branitsky A, Longden E, Bucci S, Morrison AP, Varese F  
Group Cohesion and Necessary Adaptations in Online Hearing Voices Peer Support Groups: Qualitative Study With Group Facilitators  
JMIR Form Res 2024;8:e51694  
URL: https://formative.jmir.org/2024/1/e51694  
doi:10.2196/51694  
PMID:38701439

©Alison Branitsky, Eleanor Longden, Sandra Bucci, Anthony P Morrison, Filippo Varese. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Accuracy of the Apple Watch Series 4 and Fitbit Versa for Assessing Energy Expenditure and Heart Rate of Wheelchair Users During Treadmill Wheelchair Propulsion: Cross-sectional Study

Marius Lyng Danielsson1, MSc; Melanie Vergeer1,2, MSc; Guy Plasqui2, PhD; Julia Kathrin Baumgart1, PhD

1Centre for Elite Sports Research, Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway
2Department of Nutrition and Movement Sciences, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Marius Lyng Danielsson, MSc
Centre for Elite Sports Research
Department of Neuromedicine and Movement Science
Norwegian University of Science and Technology
Smistadvegen 11
Trondheim, 7026
Norway
Phone: +47 47863154
Email: mlyngd@gmail.com

Abstract

Background: The Apple Watch (AW) Series 1 provides energy expenditure (EE) for wheelchair users but was found to be inaccurate with an error of approximately 30%, and the corresponding error for heart rate (HR) provided by the Fitbit Charge 2 was approximately 10% to 20%. Improved accuracy of estimated EE and HR is expected with newer editions of these smart watches (SWs).

Objective: This study aims to assess the accuracy of the AW Series 4 (wheelchair-specific setting) and the Fitbit Versa (treadmill running mode) for estimating EE and HR during wheelchair propulsion at different intensities.

Methods: Data from 20 manual wheelchair users (male: n=11, female: n=9; body mass: mean 75, SD 19 kg) and 20 people without a disability (male: n=11, female: n=9; body mass: mean 75, SD 11 kg) were included. Three 4-minute wheelchair propulsion stages at increasing speed were performed on 3 separate test days (0.5%, 2.5%, or 5% incline), while EE and HR were collected by criterion devices and the AW or Fitbit. The mean absolute percentage error (MAPE) was used to indicate the absolute agreement between the criterion device and SWs for EE and HR. Additionally, linear mixed model analyses assessed the effect of exercise intensity, sex, and group on the SW error. Interclass correlation coefficients were used to assess relative agreement between criterion devices and SWs.

Results: The AW underestimated EE with MAPEs of 29.2% (SD 22%) in wheelchair users and 30% (SD 12%) in people without a disability. The Fitbit overestimated EE with MAPEs of 73.9% (SD 7%) in wheelchair users and 44.7% (SD 38%) in people without a disability. Both SWs underestimated HR. The device error for EE and HR increased with intensity for both SWs (all comparisons: P<.001), and the only significant difference between groups was found for HR in the AW (~5.27 beats/min for wheelchair users; P=.02). There was a significant effect of sex on the estimation error in EE, with worse accuracy for the AW (~0.69 kcal/min; P<.001) and better accuracy for the Fitbit (~2.08 kcal/min; P<.001) in female participants. For HR, sex differences were found only for the AW, with a smaller error in female participants (5.23 beats/min; P=.02). Interclass correlation coefficients showed poor to moderate relative agreement for both SWs apart from 2 stage-incline combinations (AW: 0.12-0.57 for EE and 0.11-0.86 for HR; Fitbit: 0.06-0.85 for EE and 0.03-0.29 for HR).

Conclusions: Neither the AW nor Fitbit were sufficiently accurate for estimating EE or HR during wheelchair propulsion. The AW underestimated EE and the Fitbit overestimated EE, and both SWs underestimated HR. Caution is hence required when using SWs as a tool for training intensity regulation and energy balance or imbalance in wheelchair users.
agreement; validity; accuracy; cross sectional; physiology; disability; disabled; upper-body exercise; upper body; exercise; physical activity; ergospirometer; fitness; vital; vitals; energy; expenditure; mHealth; wearable; wearables; mobile health; smartwatch; smartwatches; apple watch; fitbit; digital health; energy expenditure; heart rate; wheelchair; wheelchairs; fitness trackers; tracker; trackers

Introduction

Wheelchair users are generally less active than people without a disability, which increases their risk of developing noncommunicable diseases such as cardiovascular disease, type 2 diabetes, and obesity [1-4]. In fact, the prevalence of obesity is 2.5 times greater in wheelchair users compared to people without a disability [5]. This is related to an energy intake that exceeds their energy expenditure (EE), which is approximately 5% to 40% lower in wheelchair users compared to people without a disability [6-8].

Total EE in people without a disability is approximately 1900 to 2900 kcal/day and comprised of 3 components: 60% to 75% attributed to resting EE (REE), 10% to diet-induced thermogenesis, and 15% to 30% to physical activity EE (PAEE) [9,10]. The lower total EE of wheelchair users is mainly related to reduced REE and PAEE [11-13]. Wheelchair users with a spinal cord injury (SCI) have a 14% to 27% lower REE due to reduced fat-free mass and sympathetic nervous system activity [11]. Furthermore, wheelchair users have lower PAEE due to a smaller amount of active muscle mass during upper-body exercise compared to what ambulatory people without a disability expend during walking or running [14,15]. Also, PAEE typically ranges between 6% and 36% of the total daily EE in wheelchair users with a SCI [16,17] and is the most modifiable of the 3 components. Therefore, PAEE may be particularly useful for obtaining a balance between EE and energy intake.

Criterion devices for measuring EE, such as direct or indirect calorimetry, are restricted to the laboratory setting and expensive to use. Therefore, more accessible devices that accurately estimate EE within the population of wheelchair users are needed. Smart watches (SWs) are widely used to provide feedback on estimated EE and monitor physical activity intensity (eg, through monitoring heart rate [HR]) [14,18-20]. If sufficiently accurate, the feedback provided by SWs may serve as a tool to counteract obesity and promote physical activity in wheelchair users.

Commonly used cutoffs for acceptable accuracy of parameters provided by wearable devices are ±10% in free-living settings and ±3% in standardized settings [18,21]. However, even in standardized settings, SWs often estimate HR and EE values outside of this range in both wheelchair users and people without a disability [18-22]. Additionally, the development of wheelchair user–specific estimation algorithms for EE and HR is especially challenging because of the high heterogeneity and disability-related differences in physiological functioning in this population. Currently, the only study that evaluated the accuracy of estimated HR with a commercially available SW (Fitbit Charge 2 [Fitbit Inc]) found lower accuracy for wheelchair users (mean absolute percentage errors [MAPE] of approximately 10%-20% dependent on level of SCI) compared to people without a disability (approximately 8% MAPE) [23]. Furthermore, the sole study that assessed the accuracy of the estimated EE in the commercially available Apple Watch (AW) Series 1 (Apple Inc) with a wheelchair-specific setting reported a MAPE of 29% [24]. Notably, both studies report that the measurement error for estimating HR and EE increased with higher-intensity exercise [23,24]. While follow-up studies have not yet been conducted, one would expect companies to further improve the accuracy of their HR and EE estimation algorithms.

Therefore, the aim of this study was to assess the accuracy of the AW Series 4 (in the wheelchair-specific setting “outdoor push walking pace”) and the Fitbit Versa (in the treadmill running mode) for estimating EE and HR during wheelchair propulsion at different intensities. We decided to include both wheelchair users and a control group consisting of people without a disability to investigate if the wheelchair setting was specifically adjusted for wheelchair users.

Methods

Participants

A total of 20 wheelchair users and 20 people without a disability were included in the study. Both groups consisted of 11 male participants and 9 female participants and had similar demographic characteristics (Table 1). Participants were included if they were aged between 18 and 60 years and without injury or other health issues that could be aggravated by physical exertion. Included in the wheelchair user group were individuals that used a manual wheelchair as a main form of transport or were ambulatory wheelchair users. The wheelchair user group was comprised of individuals with SCI (n=11), spina bifida (n=2), and cerebral palsy (n=2). A total of 5 participants had other neurological, musculoskeletal, or joint impairments. Participants were recruited from sports associations and organizations for people with disabilities in Norway and social media.

https://formative.jmir.org/2024/1/e52312

doi:10.2196/52312
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Groups and sex</th>
<th>Age (years), mean (SD)</th>
<th>Body mass (kg), mean (SD)</th>
<th>Body height (cm), mean (SD)</th>
<th>BMI (kg/m²), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combined</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>35.3 (11.8)</td>
<td>74.8 (15.2)</td>
<td>174.5 (10.9)</td>
<td>24.5 (4.1)</td>
</tr>
<tr>
<td>Male participants</td>
<td>36.3 (12.2)</td>
<td>81.1 (11.9)</td>
<td>181.9 (7.2)</td>
<td>24.4 (2.8)</td>
</tr>
<tr>
<td>Female participants</td>
<td>34.1 (11.5)</td>
<td>67.1 (15.6)</td>
<td>165.3 (7.1)</td>
<td>24.6 (5.3)</td>
</tr>
<tr>
<td><strong>Wheelchair users</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>37.4 (12.6)</td>
<td>74.5 (18.6)</td>
<td>172.5 (12.2)</td>
<td>24.9 (5.3)</td>
</tr>
<tr>
<td>Male participants</td>
<td>40.0 (12.9)</td>
<td>80.4 (14.3)</td>
<td>180.5 (8.5)</td>
<td>24.5 (3.3)</td>
</tr>
<tr>
<td>Female participants</td>
<td>34.1 (12.1)</td>
<td>67.2 (21.3)</td>
<td>162.7 (8.0)</td>
<td>25.3 (7.2)</td>
</tr>
<tr>
<td><strong>People without a disability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>33.3 (10.8)</td>
<td>75.2 (11.4)</td>
<td>176.2 (9.9)</td>
<td>24.2 (2.4)</td>
</tr>
<tr>
<td>Male participants</td>
<td>32.6 (10.8)</td>
<td>81.9 (9.4)</td>
<td>183.5 (5.9)</td>
<td>24.3 (2.4)</td>
</tr>
<tr>
<td>Female participants</td>
<td>34.0 (11.5)</td>
<td>67.0 (7.8)</td>
<td>167.3 (5.3)</td>
<td>23.9 (2.6)</td>
</tr>
</tbody>
</table>

**Study Protocol**

Three test days of wheelchair propulsion with different treadmill incline-speed combinations were conducted within 2 consecutive weeks. A minimum of 24 hours separated each test day, and sessions occurred at approximately the same time of day to account for diurnal variations. All test days started with a 5-minute warmup at a 0.5% incline at a self-chosen speed that corresponded to a rating of perceived exertion of 7-9 on the Borg scale [25]. Then, 3 standardized 4-minute stages were performed at a predetermined incline for the day (either 0.5%, 2.5%, or 5%) with increasing speed across the stages. The order of the test days was counterbalanced. The speed at each incline was established through pilot testing and determined to be manageable for the participants (Table 2). Anthropometric data (age and sex) were collected before testing, and body mass and height were collected on the first test day. Participants were instructed to avoid high intensity training and alcohol consumption 24 hours before testing, avoid caffeine on the day of testing, and fast for at least 2 hours before testing.

Table 2. Overview of the standardized speeds for the 3 test days (0.5%, 2.5%, or 5% incline) for male participants (without tetraplegia) and female participants or male tetraplegic wheelchair users.

<table>
<thead>
<tr>
<th>Test days and participants</th>
<th>Stage 1, speed (km/h)</th>
<th>Stage 2, speed (km/h)</th>
<th>Stage 3, speed (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5% day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male participants</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Female participants or male participant with tetraplegia</td>
<td>3</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2.5% day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male participants</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Female participants or male participant with tetraplegia</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5% day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male participants</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Female participants or male participant with tetraplegia</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Equipment**

Participants’ body mass was measured using a Kistler force plate (Kistler 9286BA; Kistler Instruments AG) before the first test day. Body mass was determined for participants in the wheelchair user group while seated in their own wheelchair and obtained by subtracting the mass of the individual wheelchair (range 6.5-18.2 kg). All people without a disability were weighed while standing without any equipment. Participants wore a facemask (7450 V2 Series; Hans Rudolph Inc), which was connected to a Vyntus CPX ergospirometer with a mixing chamber (Vyaire, Medical GmbH) to measure gas exchange as 10-second averages, from which the criterion device EE was calculated. The Vyntus CPX was calibrated against a known gas mixture of 15% O₂ and 5% CO₂ before every test. Participants were fitted with a Polar HR monitor (version M400; Polar Electro Oy) and a Polar chest strap (version H10; Polar Electro Oy), which served as the criterion device for HR.

Participants wore 2 SWs on their nondominant wrist: an AW Series 4 software version OS 7.3.3 and a Fitbit Versa (2017) software version OS 5.0. The SWs tracked HR using...
photoplethysmography, and both SWs had a built-in accelerometer and gyroscope. The SW placement (closest to the wrist) was counterbalanced. Participant characteristics were entered in the devices, and the activity settings of “outdoor push walking pace” (AW) and “treadmill running” (Fitbit) were used throughout testing. “Treadmill running” was chosen for the Fitbit in the absence of a wheelchair-specific setting. Wheelchair users used their personal wheelchair, and people without a disability used a standardized wheelchair (Küschall K-Series Attract; Invacare; mass 11.7 kg). The wheelchairs were secured on a motorized 5x3 m treadmill (Forcelink Technology) with a mobile traverse bar attached to side rails (Figure 1). The side rails were equipped with safety stoppers to prevent participants from rolling off the back of the treadmill.

Figure 1. Test setup on the treadmill with the manual wheelchair attached to the traverse safety bar.

Blood lactate concentrations (mmol/L) were measured in a rested state and after every stage from a 20-μl blood sample obtained from the participants earlobe [26]. Concentrations were analyzed using the Biosen C-Line Sport lactate measurement system (EKF Industrial Electronics). Rating of perceived exertion was obtained for muscular, respiratory, and total effort on the 6-20 Borg scale after each stage [25].

Data Analysis
The criterion EE was calculated for every 10-second average VO₂ and VCO₂ values using the Weir formula [27]:

\[
EE \text{ (kcal/min)} = 3.941 \times VO_2 \text{ (L/min)} + 1.106 \times VCO_2 \text{ (L/min)}
\]

An average over the entire 4-minute period (as opposed to a steady-state average) was calculated for EE and HR for each stage of the criterion device. This was done since the SWs displayed an estimated average EE and HR for each entire stage. It was not possible to retrieve data with a higher time resolution from the SWs. The EE values provided by the criterion device and SWs were converted to kcal/min for comparison. Lastly, incomplete 4-minute stages were excluded from the analyses.

Missing Data
In total, 20 of the 360 performed stages were incomplete and excluded from the analyses. Incomplete stages mostly occurred at the highest speeds and often at the 5% incline, which was due to a combination of an age- or disability-related lack of physical fitness or upper-body strength. Additionally, equipment failure or human error contributed to the following missing data: incorrect activity setting (AW, n=2), no HR recorded after activity (AW, n=32), and a lost HR signal (criterion, n=1).

Statistical Analysis
Statistical analyses were conducted, and visualizations created in R Studio (version 4.2.1; R Core Team) [28]. Descriptive statistics were calculated for EE and HR for criterion devices, AW, and Fitbit and visualized with box plots using the R Studio package ggplot2 (Multimedia Appendix 1).

Absolute Agreement
The MAPE was used to establish the difference between criterion devices and SWs for both EE and HR during each stage:

\[
MAPE = \frac{1}{n} \sum \frac{|C_p - SW_p|}{|C_p|} \times 100
\]

where criterion devices are represented by \(C\), smart watches as \(SW\), participants as \(p\), and the total number of participants that completed the respective stage as \(n\). In addition to the separate MAPEs visualized in a figure, an overall MAPE is provided in text, with the mean and SD being calculated across all participants and stages. The MAPE was categorized based on commonly used accuracy cutoffs for measuring EE, HR, and steps with wearable devices, with an acceptable error of ±10% in free living settings and ±3% in standardized settings [18,21].
Our categorization for the MAPE was therefore as follows: poor (>20%), moderate (10.1%-20%), good (3.1%-10%) and excellent (0%-3%).

The results of linear mixed model analyses are presented in our main results, in addition to Bland-Altman plots in the Multimedia Appendix 2. The main reason for choosing linear mixed model analyses as our main analyses was the repeated-measures design of our data collection with corresponding dependency in data [29], since all participants conducted several stages. The linear mixed model analyses were used to assess if there was a significant difference in EE and HR between criterion devices and SWs and to investigate the effect of group and sex as well as the increase in intensity on these device differences. Speed and incline were not adjusted for in the mixed model analyses, as we were interested in the estimation error of the SWs across all intensities and not within each speed-incline combination. As such, we also did not need to adjust for multiple comparisons. Participant ID was included as a random-intercept effect in these models to account for dependency in the data. An α value of .05 was used to indicate statistical significance. There was no deviation of the residuals from normality (checked by visual inspection of Q-Q plots) and no violation of the assumption of homoscedasticity (checked by plotting the fitted values against the residuals; plot_model function, R sjPlot package). The inclusion or exclusion of outliers did not change the results of the mixed model analyses. We therefore decided to include the analyses with outliers, as they represent the actual estimation errors of the AW.

### Results

#### Absolute Agreement

For the EE reported by the AW Series 4, the MAPE (with all inclines and stages combined) was 27.4% (SD 16.7%) in wheelchair users and 32.1% (SD 14.4%) in people without a disability. The EE provided by the Fitbit Versa had a MAPE of 73.9% (SD 57.2%) in wheelchair users and 44.7% (SD 37.8%) in people without a disability. Absolute agreement based on the MAPE for each incline-stage combination was mostly poor (Figure 2).

For HR, the MAPE with all stages and inclines combined was 8.5 (SD 10.4%) in wheelchair users and 8.1 (SD 13.6%) in people without a disability for the AW, and 17.4 (SD 12.4%) in wheelchair users and 14.3 (SD 10.7%) in people without a disability for the Fitbit. The absolute agreement for HR in each incline-stage combination was moderate-good for the AW and poor-moderate for the Fitbit (Figure 2).

The mixed model analyses indicated that EE was underestimated by the AW and overestimated by the Fitbit, while both SWs underestimated HR (Figure 3). Additionally, the differences between criterion and comparison devices increased negatively with higher exercise intensity for EE and HR (all comparisons, P < .001; Figure 3 and Multimedia Appendix 2). This led to reduced accuracy in the AW (larger underestimation) and improved accuracy in the Fitbit (lower overestimation) with increased intensity. EE and HR differences between SWs and criterion devices were not significantly different in wheelchair users compared to people without a disability in most comparisons (P > .06), with the exception of the HR reported by the AW (~5.27 beats/min; P = .02). Furthermore, for EE, the differences were significantly larger in female participants for the AW (~0.69 kcal/min; P = .02). For HR, the only sex difference was found for the AW, with smaller differences between the AW and criterion device in female participants (5.23 beats/min; P = .02).

#### Relative Agreement

Interclass correlation coefficients (ICCs) were used to quantify relative agreement between criterion devices and SWs for both EE and HR. The ICCs were calculated using the 2-way random effects model with absolute agreement by using the icc function from the irr package. ICCs were categorized based on widely used cutoff points into poor (<0.5), moderate (0.5-0.75), good (0.75-0.9), and excellent relative agreement (>0.9) [30,31].

#### Ethical Considerations

The data collection and processing were approved by the Norwegian Centre for Research Data (216680) and conducted in accordance with the Declaration of Helsinki. All participants were informed of the study purpose, design, potential risks, and the possibility to withdraw without penalty before signing the consent form. Participation was voluntary and without financial compensation beyond insight into the individual’s collected data. The collected data were deidentified.
Figure 2. Mean absolute percentage error (MAPE) between criterion devices and smart watches for energy expenditure and heart rate on all 3 test days (0.5%, 2.5%, and 5% incline) and stages with increasing speed. Values are presented separately for wheelchair users and people without a disability. MAPEs were categorized as poor (>20%), moderate (10.1%-20%), good (3.1%-10%), and excellent (0%-3%).
Figure 3. Regression lines (with shaded 95% CIs) separated by groups for the differences in energy expenditure (EE) and heart rate (HR) between smart watches (SWs) and criterion devices based on linear mixed model analyses: (A) EE: Apple-Vyntus; (B) EE: Fitbit-Vyntus; (C) HR: Apple-Polar; and (D): HR: Fitbit-Polar. The x-axis shows the criterion device values, while the y-axis shows the absolute difference between SWs and criterion devices. A regression line below zero indicates underestimation of SWs compared to criterion devices, while 1 above zero indicates overestimation.

Relative Agreement
Apart from 1 EE (Figure 4; a 5% incline for stage 3) and 1 HR ICC (Figure 4; a 0.5% incline for stage 2), all remaining ICCs indicate poor to moderate relative agreement between the criterion devices and SWs (Figure 4). The ICCs for each incline-stage combination for the AW had a range from 0.12 to 0.57 for EE and from 0.11 to 0.86 for HR. For the Fitbit, the corresponding ranges were from 0.06 to 0.85 for EE and from 0.03 to 0.29 for HR.
Figure 4. Interclass correlation coefficients (ICCs) between criterion devices and smart watches (SWs) for wheelchair users and people without a disability on all 3 test days (0.5%, 2.5%, and 5% incline) and stages with increasing speed. ICCs were categorized as poor (<0.5), moderate (0.5-0.75), good (0.76-0.9), and excellent (>0.9) for energy expenditure and heart rate.

Discussion

Summary

The aim of this study was to assess the accuracy of the AW Series 4 and the Fitbit Versa for estimating EE and HR during wheelchair propulsion at different intensities. The AW underestimated EE and the Fitbit overestimated EE, suggesting that neither of the SWs are accurate enough for estimating EE in wheelchair users. Furthermore, both the AW and Fitbit underestimated HR. Lastly, the differences in HR and EE between SWs and criterion devices increased with increasing intensity, and they mostly did not differ between groups.
EE Findings

The MAPE of approximately 30% and poor to moderate ICCs of the AW Series 4 are similar to the ones reported by Moreno et al [24] for the AW Series 1. Since the participant characteristics were similar those in this study, there seems to be no improvement in the AW’s EE estimation algorithms. The SW algorithms are proprietary technology, and it cannot therefore be determined why the AW underestimates EE for all participants. A plausible explanation is that Apple intentionally chose to report lower values for the sake of obesity prevention. Another possible explanation is that the data for developing the AW algorithms were collected from wheelchair users with lower training status or higher levels of SCI, who have a lower EE compared to the wheelchair users tested in this study. If the latter is the case, injury and fitness levels are important factors to consider when estimating the EE of wheelchair propulsion.

In this study, the AW was found to be equally inaccurate when estimating EE for wheelchair users and people without a disability, a finding that also aligns with Moreno et al [24]. Since the AW does not request user information on, for example, impairment levels, this finding indicates that the watch’s software is not capable of identifying these types of individual characteristics from other factors such as movement patterns or HR responses. In contrast, sex could be inputted into the AW settings, and we found a significantly better estimated EE for male participants. Preliminary evidence [32] indicates that algorithms for wearable devices are developed mostly based on reference data from male participants and may therefore be less accurate for female participants.

The AW’s underestimation of EE increased with intensity, which contradicts findings from previous AW studies [24,33]. While Pope et al [33] reported lower accuracy at moderate compared to low and high intensity during running, Moreno et al [24] reported consistent accuracy across wheelchair propulsion stages with an increasing stroke rate. The test protocol with the standardized stroke rate increases in Moreno et al [24] might more closely resemble the way the AW estimation algorithm works. Possibly, the AW uses the accelerometry data to determine how much exercise intensity, and thereby EE, has increased. In contrast, in this study, intensity increased with higher speeds at a given incline, with the steeper incline days being more physiologically taxing with a larger anaerobic contribution than the flatter incline days. Furthermore, it is possible that the AW EE estimation algorithms were developed mostly based on low-intensity data or without a physiological intensity measure (eg, HR) since wheelchair users spend most of their day at or below low-intensity exercise levels. In line with this, preliminary findings of our research group [34] indicate that estimation algorithms developed for wheelchair users perform less well on high-intensity data if they are only developed based on low- to moderate-intensity data.

The Fitbit did, in contrast to the AW and other SW studies on EE [24,33], show a systematic decrease in the error with higher intensities for both groups, which resulted in lower overestimations. This finding was most likely related to using the “treadmill running mode” in the absence of a wheelchair-specific setting. This setting leads the estimation algorithm to expect a weight-bearing and leg-dependent activity with higher muscle activation. However, wheelchair propulsion is a non–weight-bearing activity that allows longer rest between cycles, especially at lower speeds and inclines. The reduction of the Fitbit error at higher intensities might therefore be a result of more active muscle mass, faster cycle rates, and longer cycle lengths with increased incline or speed. As such, the physiological effort of wheelchair propulsion may be more similar to running at higher intensities. Although the Fitbit displayed improved accuracy at higher intensities, the MAPEs were far greater than the ±3% acceptable accuracy cutoff, and the Fitbit should therefore not be used to estimate EE in wheelchair users.

HR Findings

Both SWs showed better accuracy for HR compared to EE. However, only the AW had a MAPE below the arbitrary cutoff of ±10% that is commonly used for acceptable accuracy in free-living activities, with no values below the ±3% cutoff for standardized settings [18,21]. Both SWs additionally showed reduced accuracy at the highest intensities, which is in agreement with previous findings of HR measured from wrist-worn devices during running [22]. Furthermore, wheelchair users were found to have a larger underestimation of HR (ie, −5 beats/min) compared to people without a disability. While the reasons for this are somewhat unclear, it seems like this is due to more negative outliers in wheelchair users. Overall, the high MAPE variance and mostly poor to moderate (AW) and poor (Fitbit) ICCs indicate a high risk of individual inaccuracy when monitoring HR from these wrist-worn SWs during wheelchair propulsion.

Methodological and Future Considerations

Two main factors need to be addressed for better EE and HR estimation algorithms in wrist-worn SWs: (1) the sensor hardware (eg, Apple Watch) and (2) the sensor software (estimation algorithm).

The sensor hardware of current wrist-worn technology is not capable of reporting precise or consistent HR signals during activity, which was partly highlighted by the missing data in this study. With regard to further improving the estimation algorithms, an assessment is needed on the impact and relative importance of factors such as personal characteristics (sex, age, body mass, training status, etc) or more wheelchair users–specific aspects such as type and level of impairment. These additional factors may increase estimation accuracy and reduce variation within the highly heterogeneous wheelchair user group, even without hardware improvements.

Furthermore, investigation into appropriate cutoffs for acceptable accuracy of EE and HR provided by SWs is needed. The commonly used cutoffs (±3% in standardized and ±10% in free-living settings) are based on previous research assessing step count during walking [35-38] or pushes during wheelchair propulsion [39]. It may be plausible to establish higher cutoffs for EE or HR, especially when testing a heterogeneous wheelchair user group during an upper-body activity. However, a standardized low cutoff is essential to avoid differentiating acceptable accuracy between parameters in both controlled and free-living settings. Accurate estimates are also crucial for various populations, for example, to address an imbalance...
between energy intake and expenditure in athletes trying to regulate nutrition for performance purposes or in wheelchair users attempting to prevent or counteract obesity. Therefore, it is likely the ±10% error cutoff for free-living activities is too high.

Lastly, the effect of filtering done by the AW and Fitbit on the EE or HR data deserves mentioning. The AW removed many average HR values, which is likely attributed to a low number of data points. Comparatively, Fitbit reported data for all activities, although with lower accuracy. For now, we advise the use of HR belts for increased accuracy of the parameters investigated.

**Conclusion**

Neither the wrist-worn AW nor Fitbit were sufficiently accurate for estimating EE or HR during wheelchair propulsion. The AW underestimated EE while the Fitbit overestimated the EE across all incline-stage combinations. The underestimation of the AW increased and the overestimation of the Fitbit decreased with higher intensities, suggesting that neither watch sufficiently adjusts for the change in intensity. Additionally, both SWs underestimated HR. High MAPEs were found for both SWs and parameters (ie, EE and HR), in addition to the poor relative agreement indicated by low ICCs. Furthermore, neither the wheelchair-specific algorithm for estimating EE nor its ability to differentiate between wheelchair users and people without a disability have been improved for the AW Series 4 as compared to the previously investigated AW Series 1. Overall, our findings suggest that caution is required when using SWs as a tool for training intensity regulation and energy balance or imbalance in wheelchair users.

**Acknowledgments**

Testing was performed with the laboratory facilities and equipment from NeXt Move, the Norwegian University of Science and Technology, the Faculty of Medicine at the Norwegian University of Science and Technology, and the Central Norway Regional Health Authority. We appreciate the work of everyone involved in “the digital wheelchair” project, in particular Cecilia Severin and Roy Mulder for contributions previously and during data collection, in addition to all the colleagues and master students assisting in the laboratory. This study was financed by strategic funding from the Faculty of Medicine and Health Science at the Norwegian University of Science and Technology, an innovation scholarship, and Discovery funding at the same university. The funding bodies had no role in the design of this study, its execution, analyses, interpretation of the data, or decision to submit results.

**Data Availability**

The data sets generated during this study and corresponding analyses are available from the corresponding author on reasonable request. If used for further studies, the request includes a data application form and a data transfer agreement.

**Conflicts of Interest**

None declared.

---

Descriptive box plots comparing absolute values for criterion energy expenditure from Vyntus and hear rate from Polar to comparison devices Apple Watch and Fitbit. Data were split between wheelchair users and people without a disability, and individual-case differences were added in (shaded gray).

**Multimedia Appendix 1**

Multimedia Appendix 2

Bland-Altman plots visualizing absolute differences between criterion values and Apple or Fitbit values. The data were split between wheelchair users and people without disability and color coded for sex (male: blue circles, female: orange triangles). Furthermore, in accordance with Krouwer, criterion values were used on the x-axis rather than the average value of criterion and comparison device.

**References**


Abbreviations

AW: Apple Watch
EE: energy expenditure
HR: heart rate
ICC: interclass correlation coefficient
MAPE: mean absolute percentage error
PAEE: physical activity energy expenditure
REE: resting energy expenditure
SCI: spinal cord injury
SW: smart watch
Original Paper

Needs for Successful Engagement in Telemedicine Among Rural Older US Veterans and Their Caregivers: Qualitative Study

Jacqueline Hannah Boudreau\textsuperscript{1}, MPH; Lauren R Moo\textsuperscript{2,3}, MD; Meaghan A Kennedy\textsuperscript{2,4}, MD, MPH; Jennifer Conti\textsuperscript{1}, MPH; Chitra Anwar\textsuperscript{1}, MA; Camilla B Pimentel\textsuperscript{1,2,5}, MPH, PhD; Kathryn A Nearing\textsuperscript{6,7}, MA, PhD; William W Hung\textsuperscript{8,9}, MD, MPH; Eileen M Dryden\textsuperscript{1}, PhD

\textsuperscript{1}Center for Healthcare Organization and Implementation Research, VA Bedford Healthcare System, US Department of Veterans Affairs, Bedford, MA, United States
\textsuperscript{2}New England Geriatric Research, Education, and Clinical Center, VA Bedford Healthcare System, US Department of Veterans Affairs, Bedford, MA, United States
\textsuperscript{3}Department of Neurology, Harvard Medical School, Boston, MA, United States
\textsuperscript{4}Department of Family Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, United States
\textsuperscript{5}Department of Public Health, Zuckerberg College of Health Sciences, University of Massachusetts Lowell, Lowell, MA, United States
\textsuperscript{6}Eastern Colorado VA Geriatric Research Education and Clinical Center, Aurora, CO, United States
\textsuperscript{7}Division of Geriatric Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
\textsuperscript{8}Bronx Geriatric Research Education and Clinical Center, James J Peters VA Medical Center, Bronx, New York, NY, United States
\textsuperscript{9}Department of Geriatrics and Palliative Medicine, Icahn School of Medicine, New York, NY, United States

Corresponding Author:
Jacqueline Hannah Boudreau, MPH
Center for Healthcare Organization and Implementation Research
VA Bedford Healthcare System
US Department of Veterans Affairs
200 Springs Rd.
Bedford, MA, 01730
United States
Phone: 1 6033059605
Email: jacqueline.boudreau@va.gov

Abstract

Background: Telemedicine is an important option for rural older adults who often must travel far distances to clinics or forgo essential care. In 2014, the Geriatric Research, Education, and Clinical Centers (GRECC) of the US Veterans Health Administration (VA) established a national telemedicine network called GRECC Connect. This network increased access to geriatric specialty care for the 1.4 million rural VA-enrolled veterans aged 65 years or older. The use of telemedicine skyrocketed during the COVID-19 pandemic, which disproportionately impacted older adults, exacerbating disparities in specialty care access as overburdened systems shut down in-person services. This surge presented a unique opportunity to study the supports necessary for those who would forgo telemedicine if in-person care were available.

Objective: In spring 2021, we interviewed veterans and their informal caregivers to (1) elicit their experiences attempting to prepare for a video visit with a GRECC Connect geriatric specialist and (2) explore facilitators and barriers to successful engagement in a telemedicine visit.

Methods: We conducted a cross-sectional qualitative evaluation with patients and their caregivers who agreed to participate in at least 1 GRECC Connect telemedicine visit in the previous 3 months. A total of 30 participants from 6 geographically diverse GRECC Connect hub sites agreed to participate. Semistructured interviews were conducted through telephone or the VA’s videoconference platform for home telemedicine visits (VA Video Connect) per participant preference. We observed challenges and, when needed, provided real-time technical support to facilitate VA Video Connect use for interviews. All interviews were recorded with permission and professionally transcribed. A team of 5 researchers experienced in qualitative research analyzed interview transcripts using rapid qualitative analysis.

Results: From 30 participant interviews, we identified the following 4 categories of supports participants described regarding successful engagement in telemedicine, as defined by visit completion, satisfaction, and willingness to engage in telemedicine
in the future: (1) caregiver presence to facilitate technology setup and communication; (2) flexibility in visit modality (eg, video from home or a clinic or telephone); (3) technology support (eg, determining device compatibility or providing instruction and on-demand assistance); and (4) assurance of comfort with web-based communication, including orientation to features like closed captioning. Supports were needed at multiple points before the visit, and participants stressed the importance of eliciting the varying needs and preferences of each patient-caregiver dyad. Though many initially agreed to a telemedicine visit because of pandemic-related clinic closures, participants were satisfied with telemedicine and willing to use it for other types of health care visits.

Conclusions: To close gaps in telemedicine use among rural older adults, supports must be tailored to individuals, accounting for technology availability and comfort, as well as availability of and need for caregiver involvement. Comprehensive scaffolding of support starts well before the first telemedicine visit.

(JMIR Form Res 2024;8:e50507) doi:10.2196/50507

KEYWORDS
caregivers; geriatrics; older adults; rural veterans; rural; specialty care; telehealth; telemedicine; veterans

Introduction

Not all older adults have the means or willingness to successfully participate in telemedicine visits. In a survey of medically high-need, high-risk older US veterans, approximately half of respondents were unwilling to engage in video visits. A quarter of those willing lacked the needed technology, and these individuals were more likely to be older, less health literate, or living in more socioeconomically disadvantaged areas [1]. Concerns persist about technology access and use among older adults for telemedicine, such as those who lack broadband access, and common age-associated communication challenges related to sight, hearing, and cognitive limitations [2,3].

However, when telemedicine is successful, its advantages may be especially critical for older, rural adults and their caregivers [4]. Compared to in-person appointments, telemedicine has been shown to have similar, and sometimes better, clinical outcomes [5-8]. It has similar quality, including for visits related to cognitive decline [9], has high levels of patient satisfaction, and is associated with timely access to specialty care [10] which can help support patients with mobility or transportation issues [11,12]. Telemedicine cuts down on travel complications [13], bridging far distances to clinics for older adults living in rural areas [14] and reduces patient and caregiver stress [15]. Further, telemedicine improves care continuity and decreases missed care opportunities [16], including in dementia clinic settings [10].

Those who are older, lower-income, or living in rural areas have historically been less likely to engage in telemedicine visits [17], disparities that worsened during the COVID-19 pandemic. The pandemic disproportionately impacted older adults (World Health Organization) [18], exacerbating disparities in specialty care access as overburdened care systems shut down nonessential in-person services. Even as the overall use of telemedicine soared, gaps among rural patients and those in the oldest age groups widened [19,20]. Yet, the forced pivot to telemedicine provided a unique opportunity to study the support needs of rural, older adults who may not have otherwise used these services.

The US Veterans Health Administration (VA) serves as an optimal setting in which to explore factors influencing successful engagement in telemedicine for rural older patients. More than half of the 2.8 million rural veterans receiving VA services are aged 65 years or older [21]. The VA has been a leading force in telemedicine for 20 years [22,23] in its efforts to expand access to health care services to all veterans, recently celebrating the milestone of 20 years of providing telemedicine services. In 2014, the VA's Office of Rural Health funded GRECC Connect to use telemedicine to expand access to geriatric specialty care for rural, older veterans with multiple chronic conditions and complex care needs. Geriatric specialists from urban centers of excellence called Geriatric Research, Education, and Clinical Centers (GRECCs) focused on aging worked to build partnerships with rural clinicians to serve older adults living in rural areas through clinic-to-clinic or clinic-to-home visits [24,25].

In spring 2021, we interviewed older adults and their caregivers as part of an evaluation of GRECC Connect’s national network of geriatric telemedicine specialty services. We conducted interviews using the same VA videoconferencing platform used for clinical visits, enabling us to observe and interact with participants as they navigated technological and communication challenges and, if needed, provide necessary technical support or instruction. While other studies have examined satisfaction with telemedicine in older adults [9,10] and competencies of providers, including previsit preparation [26], few studies have looked in-depth at the experiences of older patients inexperienced in telemedicine as they prepare to engage in visits. We sought to (1) elicit patient and caregiver experiences attempting to participate in a video visit with a GRECC Connect geriatric specialist, and (2) explore facilitators and barriers to successful engagement in a telemedicine visit among this group. In this paper, we relay lessons learned from the experiences of these rural older adults and their caregivers, as well as the study team, highlighting the kinds of supports needed for successful engagement in telemedicine among older rural adults and caregivers.

Methods

Study Design

We conducted a cross-sectional qualitative evaluation with patients and, when appropriate, their informal caregivers, who...
had agreed to participate in at least 1 VA GRECC Connect telemedicine visit in the previous 3 months. The study team, led by an anthropologist and health services researcher, was experienced in qualitative research. Throughout the evaluation, the team consulted with an advisory group composed of GRECC Connect leaders, clinicians with expertise in geriatrics, neurology, and primary care, health services researchers, and a veteran consultant. The VA Bedford Healthcare System’s Institutional Review Board determined this evaluation of GRECC Connect services to be nonresearch.

Setting

In December 2020, we sent letters to clinical leads at 7 of the 15 GRECC Connect hub sites, interdisciplinary geriatric specialty teams who partner with rural community-based outpatient clinics (CBOCs), to assess their willingness to serve as recruitment sites. Potential sites were chosen based on their diverse geographic location across regions of the United States and high volume of telemedicine visits performed during fiscal year 2020.

Participants

Interviewees were recruited from lists of patients who agreed to participate in a GRECC Connect telemedicine visit at 1 of the 7 hub sites from December 2020 to March 2021. Inclusion criteria included age of ≥65 years and residential rurality (rural-urban commuting area code >1, in line with VA practice). We defined telemedicine as (1) video visits conducted between a remote geriatric specialist and a patient located at home (VA Video Connect; VVC) or at a rural VA CBOC (clinical video telemedicine; CVT) or (2) telephone visits conducted between a remote geriatrics specialist and a patient. Some participants had agreed to a video visit but, for various reasons, only completed a phone call with the clinician. We reviewed the electronic health record to exclude potential participants who passed away or were currently hospitalized or in hospice.

Recruitment

We sent patients a recruitment letter detailing the goals of the evaluation and inviting them to participate in an interview. Overall, 1 of 3 team members (CA, JHB, and JC) conducted a brief phone screening with the patients and, in some instances, caregivers who agreed to participate to confirm recall of the GRECC Connect index appointment—a recent GRECC Connect telemedicine appointment around which to ground the interview. During the screening call, a staff member administered a technology questionnaire to gauge participant comfort with and access to technology (Multimedia Appendix 1).

Data Collection

We asked participants about their preference for completing the interview through VVC videoconference or telephone. Members of the study team (CA, JHB, JC, and EMD) conducted semistructured qualitative interviews using participants’ preferred modality. In cases where patient participants had some degree of cognitive impairment (eg, dementia) or communication challenges (uncorrected hearing loss), we interviewed the patient-caregiver dyad. Based on needs identified through the technology questionnaire administered during initial screening phone calls, interviewers sometimes began the interview appointments early to provide instruction for initiating and connecting through the videoconference application. We documented technical support needs in interview notes, which offered additional context for the experiences participants shared. Interview questions were drafted by a subset of the study team, shared with the advisory group for feedback, piloted with 2 different older patients, and then finalized. The guide focused on the support received to engage the dyad in telemedicine, what worked and did not, preferences for medical visit modality, satisfaction, and recommendations for improving the telemedicine experience and available support.

Analysis

Interviews were audio-recorded and professionally transcribed verbatim. The study team (CA, JHB, JC, EMD, and MAK) analyzed interviews using rapid qualitative analysis [27,28]. Analysts summarized individual interviews using a structured template organized by key conceptual categories or domains. We met regularly to develop consensus about the template domains, which included a priori and emergent domains from the interview guide and content of interviews, respectively. To achieve consensus in the content and application of the template used to organize salient aspects of interview transcripts, 2 evaluators summarized 2 initial transcripts. Other members of the evaluation team reviewed this work. We then summarized 10 additional interviews in pairs, or triads, to maintain consistency in applying the template. The remaining transcripts were summarized individually. The team met to resolve uncertainties and refine content domains as needed. Summary templates were condensed into a single matrix where each row contained data for an individual transcript and each column represented a domain. This matrix allowed us to discern and distill patterns in the data within and across domains. We shared key themes and illustrative quotes with the advisory group, who used their expertise to help interpret the findings.

Ethical Considerations

The VA Bedford Healthcare System Institutional Review Board determined this work was undertaken to inform VA operations as part of program evaluation and quality improvement activities and was not human subjects research.

Results

Participants

We interviewed 30 patients and 26 caregivers who had agreed to participate in at least 1 telemedicine visit in the previous 3 months at 7 geographically diverse GRECC Connect hubs (2 in the Midwest, 1 in the Northeast, 2 in the West, and 2 in the South). Table 1 provides a summary of participant and index-visit characteristics as determined from electronic health record data and interview responses. Patient participants were all male, 93% (28/30) non-Hispanic and White, with a mean age of 76 years. Interviewed caregivers (some had additional caregivers) were most often the patient’s spouse (20/26, 77%) and thus were often older adults themselves. A total of 13% (4/30) of the patients we interviewed did not have a known caregiver. During interviews, sometimes the caregiver was the main respondent in cases where the patient experienced

https://formative.jmir.org/2024/1/e50507

JMIR Form Res 2024 | vol. 8 | e50507 | p.2096
(page number not for citation purposes)
cognitive or communication challenges. Indeed, 87% (26/30) of index telemedicine visits with geriatrics specialists were related to the patient’s cognitive challenges.

**Modality of Index Visit**

At least two-thirds of participants’ index appointments (specific, recent GRECC Connect appointments used to identify participants and ground interviews) were conducted using a video modality (Table 1). Some had agreed to a video appointment but either missed it or were unable to initiate it through the home videoconference application. Of these, some visits were performed by phone instead. A total of 90% (27/30) of index visits were not the patient’s first telemedicine encounter, though many had only begun telemedicine visits during the COVID-19 pandemic.
Table 1. Characteristics of older, rural veteran- and caregiver-interview participants and their index Veterans Health Administration telemedicine visits between December 2020 and March 2021.

<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients (n=30)</strong></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>76 (6.24; 66-87)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>30 (100)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>White and non-Hispanic</td>
<td>28 (93)</td>
</tr>
<tr>
<td>White and Hispanic</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Caregivers (n=26), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Unpaid caregivers</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>4 (13)</td>
</tr>
<tr>
<td>1 known caregiver</td>
<td>22 (73)</td>
</tr>
<tr>
<td>&gt;1 known caregivers</td>
<td>4 (13)</td>
</tr>
<tr>
<td><strong>Interviewed caregiver relationship to patient</strong></td>
<td></td>
</tr>
<tr>
<td>Spouse</td>
<td>20 (77)</td>
</tr>
<tr>
<td>Nonspousal significant other</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Child or child’s spouse</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Other family member</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>GRECC(^a) connect visit characteristics among participants (n=30), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Reason for visit</strong></td>
<td></td>
</tr>
<tr>
<td>Initial cognitive impairment</td>
<td>16 (53)</td>
</tr>
<tr>
<td>Follow-up cognitive impairment</td>
<td>10 (33)</td>
</tr>
<tr>
<td>Other(^b)</td>
<td>4 (13)</td>
</tr>
<tr>
<td><strong>Visit modality</strong></td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td>7 (23)</td>
</tr>
<tr>
<td>VA(^c) video connect (VVC(^d)—video to home)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>Clinical video telehealth (CVT(^e)—video to rural clinic)</td>
<td>12 (40)</td>
</tr>
<tr>
<td>CVT per medical record; patient described phone visit</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Combination of phone and VVC</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>First experience with telemedicine</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2 (7)</td>
</tr>
<tr>
<td>No</td>
<td>27 (90)</td>
</tr>
<tr>
<td>Unknown or no data</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Present in visit</strong></td>
<td></td>
</tr>
<tr>
<td>Patient only</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Patient and caregiver</td>
<td>22 (73)</td>
</tr>
<tr>
<td>Caregiver only</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

\(^a\)GRECC: Geriatric Research, Education, and Clinical Centers.
\(^b\)Falls, physical therapy, medication consultation after stroke, evaluation of sleep hygiene.
\(^c\)VA: US Veterans Health Administration.
\(^d\)VVC: VA Video Connect.
\(^e\)CVT: clinical video telemedicine.
needed or desired supports
Participants described supports they experienced or would recommend before a video visit. They noted that such support should be provided at multiple points between appointment scheduling and the time of the appointment to ensure success. Supports (available, desired, or recommended) fell into the following 4 categories: (1) presence of a caregiver; (2) choice and flexibility in visit modality (eg, video visit or phone); (3) technology support, including assurance of compatible devices, instruction, and provision of on-demand assistance; and (4) assurance of comfort with web-based communication, orientation to features of the videoconferencing platform.

presence of a caregiver
Patients relied on caregivers for assistance with technology and communication, as needed, to successfully participate in telemedicine visits. Among those who successfully initiated the video visit, caregivers were frequently responsible for setting up in-home technology. Ease of connecting, thanks to the presence of a family caregiver, led to favorable views on telemedicine. As per 1 caregiver,

I know that older people that aren’t computer savvy or want to be on the computer. They can get a little intimidated and I know that’s a challenge for some.... I go in and I make sure it’s set up and make sure that he’s connected like we did today with you.... It’s been a very easy process to work with him to work with the telehealth and set it up and to go with it.

However, not all caregivers were comfortable with the technology. One said, “I’ve been known to push the wrong buttons sometimes and cause errors, so. Other than my hesitation or lack of confidence in electronics, there really wasn’t [any issues].”

Further, caregivers often supported the patient in communication, liaising between the patient and the clinician, especially for patients with hearing challenges or cognitive impairment. One patient told us during his video interview, “This is my memory here, over my shoulder,” referring to his spouse. However, caregivers cautioned that the clinician should be cognizant of including the patient in communication rather than just relying on the caregiver.

choice and flexibility in visit modality
Participants seldom felt that the telemedicine modality had been a choice. Rather, sometimes telemedicine was the only option available due to pandemic-related clinic closures, and the patient or dyad was informed of visit modality at or after appointment scheduling. Other reasons for telemedicine use and location (home or clinic) included patient or caregiver concern about COVID-19, distance from the specialty clinic, and other logistical concerns. One caregiver said, “We were told it was to be done over the telephone because of COVID-19. [The patient] wasn’t vaccinated then and he had a lot of health issues... Especially breathing issues and stuff so, I don’t think they even wanted like to try to take that attempt with him.”

Some said that having options, when possible, would make them more comfortable with telemedicine. For example, a caregiver recalled that the staff member scheduling the patient’s appointment assured her that telemedicine was appropriate for her father’s evaluation but that alternatives were available if the dyad was unsatisfied with the visit’s quality:

I thought [telemedicine] seemed like a good idea, but I was a little bit uncertain as to how effective it would be for the type of evaluation that needed to be done.... I mentioned it to the person that I spoke to when I scheduled the visit for Dad and she was very clear that if it did not seem to go well, that if it didn’t seem to work well that they would, of course, schedule an in-person visit, so I was fine to go ahead and start with the telehealth visit after she said that.

A patient shared the experience of repeatedly being scheduled for CVT appointments from the rural CBOC, even though he would have preferred to do telemedicine appointments from his home. After several visits, he talked to his doctor to have it changed: “I’d rather do it at home here and talk to [the clinician]. And, you know, finally I just told Dr X, I said, you know, I don’t think I need to have all these nurses standin’ around here listening to our call.”

The telephone enabled additional flexibility for telemedicine encounters. Having the ability to switch from video to the telephone if technical issues arose allowed patients and caregivers to continue the appointments and participants to continue interviews. The phone also afforded caregivers the opportunity to join patients’ visits when they were unable to attend in-person. One caregiver, the patient’s child, who lives in another state, explained, “I wasn’t there, but they call me, or my dad calls me on his cell phone so I can overhear the conversation. They put me on speaker, and we’ve done this twice now and it has been wonderful.”

technology support
Participants stressed that having a positive experience with the technology is critical to successful telemedicine engagement. The ability to connect to video appointments varied, with many experiencing roadblocks such as poor internet connectivity and uncertainty about how to connect to the video software, especially from home. One patient who agreed to try but was unable to complete his visit over video said, “They told me I could try [VVC] and I tried it a couple days before and I still couldn’t do it, so I just got frustrated with it.”

Some participants experienced technology challenges due to incompatibility between their equipment and the VA’s videoconferencing platform. While the VA has the capability to offer free, internet-connected tablets to patients who don’t have their own devices, we found participants at 5 out of the 6 VA hub sites had not been offered a device by anyone in the VA. When the availability of these VA-owned devices was raised in interviews, many showed interest in a VA-issued device. One caregiver remarked, “With the tablet, we could walk outside if it gets, like I can’t [go] outside with my [landline] phone so I just thought like a tablet would be great because we could go sit on the bench.... That would be perfect for him.”
Participants desired more formal instruction before the video appointment date, such as test calls or group education for older adults and caregivers. Some preferred printed and mailed materials, while others found it easier to keep track of digital resources and were overwhelmed by receiving too much written information. Appointment reminders sent through electronic mail were 1 channel participants identified for receiving such instructions. A few recalled receiving in-depth instructions to test audio and video before the appointment. Our experience preparing participants for video interviews corroborated this need. As I participant said, “(For) telemedicine, if I know how to get on there and if I know how to, you know, set everything up...like today, my wife was able to set it up because you gave her the instructions, and that was great.” One patient who had struggled to connect to the VA videoconferencing platform before the interview said he wished he had received instruction sooner:

I would just say [to VA employees], “Hey, if you don’t get it to work, tell them to scroll down and hit ‘Start.’” Cause, you know, if I would’ve known that. Of course, they probably figure, “Well, you can see that,” and I mentioned earlier, I’m sure that that’s why nothing ever went through because, you know, I never seen the “Start” ’til [I received live instruction] today.

We observed that some dyads needed more than 1 round of instruction to support successful engagement. Those who performed test calls typically found it helpful and stressed its importance, but some, like the patient previously mentioned, had challenges with test calls but did not know how to reach anyone for live technical support. One participant recommended, “If [the patient is] not connected with the internet and knows a little bit...there could be somebody provided through the VA that would help them.”

Those who participated in CVT visits from a local rural outpatient clinic benefited from additional technology support from a nurse or technician who was present during setup. Participants who experienced this additional help found it valuable. One patient received hearing aids from the telemedicine nurse. When asked if he would recommend CVT to others, another patient said, “Oh heck yeah! Even the lady setting up stuff like this [is] helpful and everything.”

Comfort With Web-Based Communication

Participants described and displayed variability in comfort with and preferences for telemedicine visits. For CVT appointments from rural community outpatient clinics, in particular, preferences for having staff in the room during the appointment session varied. In a previous example highlighted earlier in this paper, a patient disliked having extra staff in the room for privacy reasons. Another patient we interviewed had positive feedback about a CVT call where the staff member left after setting up the technology and ensuring that the patient and caregiver were comfortable: “Whoever the person was that, you know, got the doctor on the TV and, you know, made us comfortable. Yeah, it worked very well.... I really liked the fact that she left us—the person that set it up left us alone with the doctor so we had privacy. I liked that part of it.” However, another participant liked having a nurse in the room who could ask the clinician questions on his behalf if he needed help with web-based communication. Another participant was unaware of the optional closed captioning feature of the VA’s videoconferencing platform and recommended that this should be promoted for patients with hearing difficulties.

“More Than Just COVID-19”

Participants were typically satisfied with their telemedicine experience, and many said they were interested in continuing using telemedicine for at least some appointments. One participant said, “I totally agreed to [the telemedicine visit] and the only reason was because of COVID. But then, after we’d done it a couple of times, it was—it’s so much more than COVID. I mean, I think [the VA] should continue doing it all the time.” Some still preferred or would like to at least keep some in-person appointments.

“They May Surprise You With What They Know”; A Note on Tailoring Supports

Overall, patients and their caregivers desired options tailored to their preferences and needs. Participants noted that not all older adults have the same technological abilities. One participant acknowledged that “[technology is] a great mystery to some [older adults],” but said they personally had “[not] really hesitated to use what they [the VA] offer tech-wise” and that other older patients “may surprise you with what they know.” Still, not all patients or their caregivers were tech-savvy. Some who formerly felt confident in their technological literacy had lost abilities due to aging or cognitive decline, leading to feelings of frustration or shame. One patient, who engineered military jets before his retirement, explained how he went from being among the top in his technical field to relying on his spouse to set up his computer and the effect it had on his confidence:

I’ve worked all over the world. I’ve worked on the F-18s, the design of it, and worked on all the missiles and all this and I was considered one of the top techs or engineers or directors.... Now, when I had a problem, I would invite my wife on my computer. I could probably work on it, but she’s faster and she does it, you know, better or she’ll say, “Remember, you had to do this,” and I didn’t remember that, so I kind of shied away from it, you know?

Discussion

Principal Results

Patient and caregiver participants identified supports critical to the successful initiation of video visits. These included an explanation of options (from local clinic vs from home), instruction and test calls, on-demand technology support, and orientation to video communication norms and features, including sensitivity to privacy preferences. Supports desired spanned multiple points in time between when a patient is offered a telemedicine appointment and the appointment itself, suggesting that ongoing support may be needed to adequately prepare patients and their caregivers.
Our work contributes to the research on telemedicine in several ways. Unlike other studies, we provided an in-depth look at the preappointment preparation experiences of a group of older adults inexperienced in telemedicine. Our methods in this study allowed a uniquely intimate perspective on the telemedicine experiences of our participants. By conducting qualitative interviews using the same technology used in these appointments, we were able to solicit participants’ nuanced accounts of their experiences and recommendations, observe some of the challenges they described, and embody a support role to facilitate engagement in real time. Assisting patients in connecting to the video visit through phone or troubleshooting issues using the chat feature, a previously identified support [29], was often helpful and valued by participants. Table 2 shows recommended strategies to support older telemedicine participants and their caregivers based on patient and caregiver experiences, interviewer observations, and input from our advisory group stakeholders.

Proactively assessing the needs of older adults and their caregivers and discussing available supports is needed. While the VA provides some of the supports participants desired, such as on-demand assistance, closed captioning, and 4G-enabled tablets, many participants were not aware of these resources. System-level interventions may be needed to extend tablet services to more VA patients. A recent evaluation of the tablet program to date found that the VA distributed tablets to more than 7000 patients with access needs in more than 850 inpatient settings, mostly for mental health but increasingly for specialty appointments, though broadband has been a challenge [30]. As in other studies [1,31], we identified a segment of older adults who are interested in participating in telemedicine but who lack the appropriate technology. Offering alternatives to these patients, such as conducting the visit from a local clinic, receiving a clinic-issued device, or suggesting borrowing a device from a family member or friend, may enable engagement in video visits [31]. Patients and their caregivers desired more tailored training for video visit technology use. A study done by Hawley et al [31] showed that trainings for older adults tailored to telemedicine interest level and capability and informed by an initial needs assessment reduce barriers to engagement in video visits.

Patients’ physical and psychological comfort should be addressed to ensure a positive experience that respects patient beneficence and autonomy. Other studies have identified previsit preparation as an important clinician competency domain for video telemedicine with older adult patients, including optimizing the clinician environment for audibility and visibility and identifying who should be included in the visit and their roles in the patient’s care [26]. Our results echo privacy concerns in the telemedicine community regarding the sharing of sensitive information and diagnoses with care partners or others who happen to be in the appointment environment, which in some situations can pose distress to patients or caregivers [32,33]. Mishkin and colleagues [32] recommend multilevel strategies to address these challenges, including system-level patient reminders that encourage patients to ensure privacy before joining appointments, provider discretion in the appropriateness of telemedicine for sensitive appointments, and partnering with local clinics to provide private spaces to see distant providers, as the VA does in its GRECC Connect clinical video visits. Greater attention may be needed to respect patient autonomy by eliciting preferences about the presence of staff and care partners during these appointments.

Social and technological needs are not one-size-fits-all and should be approached with compassion and supportive solutions. This may be especially pertinent for appointments with patients being seen for memory loss and other cognitive challenges whose needs and abilities may be rapidly changing as disease progresses, but is important to consider with all older adults, who have a range of technical aptitude. Studies have shown that clinicians anticipate that older adults will have trouble with remote technology [34]. Yet even within what may appear to be a homogeneous group, we saw variation in what supports were wanted and needed.

Many participants did need help successfully engaging in video visits. Caregivers and staff at CBOCs were key supports for setting up the technology, troubleshooting issues, and getting logged into appointments. Numerous studies have found that the presence of a caregiver facilitates engagement in video visits for older adults, particularly for setting up technology and for patients with dementia [13,14,29,35]. Additionally, we found that participants with hearing or cognitive challenges relied on caregivers for communication with clinicians.
Table 2. Recommended strategies to support older telemedicine participants and their caregivers based on patient and caregiver experiences from telemedicine appointments between December 2020 and March 2021, interviewer observations from phone and videoconference interviews, and input from our advisory group stakeholders.

<table>
<thead>
<tr>
<th>Support</th>
<th>Time period</th>
<th>Prior to appointment date</th>
<th>Day of appointment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of a caregiver</td>
<td>Referral or scheduling</td>
<td>• Identified individuals, including caregiver, provide or help test technology</td>
<td>• If in-home visit, caregiver initiates videoconference technology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinician or staff include both patient and caregiver in communication, where possible</td>
<td>• Clinician or staff include both patient and caregiver in communication, where possible</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinician or staff identify person(s) who will provide or assist with technology</td>
<td></td>
</tr>
<tr>
<td>Choice and flexibility in visit modality</td>
<td>Prior to appointment date</td>
<td>• If in-home visit, clinic staff provides contact instructions in case of technology challenges</td>
<td>• Clinician or staff provides backup contact in case modality fails</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinician or staff explain appointment options</td>
<td>• If in-home visit, clinic staff solicit patient or caregiver privacy preferences, when appropriate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinician or staff solicitor patient or caregiver modality preferences</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If in-home visit, clinic staff provides contact instructions in case of technology challenges</td>
<td></td>
</tr>
<tr>
<td>Technology support</td>
<td>Day of appointment</td>
<td>• Hospital-administered individual or group education</td>
<td>• Clinic staff assist patient or caregiver with connecting to video platform</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinic staff provide internet-connected devices</td>
<td>• Clinic staff provide contact for troubleshooting technology challenges</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinic staff ensure test call was successful</td>
<td></td>
</tr>
<tr>
<td>Assurance of comfort with virtual communication</td>
<td>Prior to appointment date</td>
<td>• Appointment reminders include information about videoconference features</td>
<td>• If in-home visit, staff or clinician explains videoconference features</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinician or staff consult with patient or caregiver about technology comfortability and needs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinic staff provide information for technology assistance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinic staff or encourage test call</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Identified individuals, including caregiver, provide or help test technology</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinician or staff include both patient and caregiver in communication, where possible</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinician or staff identify person(s) who will provide or assist with technology</td>
<td></td>
</tr>
</tbody>
</table>

Limitations

We interviewed patients who agreed to a telemedicine visit and to an interview through phone or video, so there may be unexplored prohibitive barriers for this population among those who declined either. However, many of our participants were not offered in-person visit alternatives because of pandemic closures. Further, while we gained insights into the VVC experience through observation through our interview method, we were not able to observe the CVT experience from a VA clinic. Participants reported far fewer technical challenges in these instances, likely because they had on-site technicians who were responsible for establishing the video connection to the hub site specialists. While our study population reflects the demographics of older, rural veterans and their caregivers, it is important to note that our participants were predominantly White and male and typically had female caregivers who were spouses. Future qualitative studies about the experiences of older, rural adults should explore the social and technological needs of those with different ethnic, racial, gender, or sexual identities, including identity-discordant patient-caregiver dyads. Several studies conducted over the same time frame found that patients who are Black and living in lower-income areas are more likely to engage in telephone visits while patients who are White and living in higher-income areas are more likely to engage in video visits [36-38].

Conclusions

Telemedicine has the potential to expand access to specialty health care services for rural older adults. Yet widening gaps in use following the COVID-19 pandemic surge show that this population needs more support to engage in telemedicine successfully. Older adults and their caregivers may need ongoing support over various touch points to ensure successful engagement in video telemedicine. Supports are not one-size-fits-all and should instead be tailored to individual older patients, considering their access to and comfort with videoconferencing technology, comfort with the clinician, and availability of and need for caregiver involvement. Technology continues to be a barrier for many, in part due to gaps in broadband access in rural areas. If we do not address these barriers, we are increasing the inequity for this most vulnerable population.

Acknowledgments

The authors wish to thank Robert Bruno, a veteran consultant on our project team, for providing valuable input and feedback on our work. This work was supported by funding from the US Department of Veterans Affairs, Office of Rural Health, and Veteran
Rural Health Resource Center-Salt Lake City for portfolio project FY21-16035. MAK is supported by the Career Development Award (IK2 RX003930) from the US Veterans Health Administration (VA) Rehabilitation Research and Development Service. CBP is supported by the Career Development Award (IK2 HX003184) from the VA Health Services Research and Development Service. The funders had no role in the preparation, review, or approval of the paper or decision to submit it for publication. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the US Department of Veterans Affairs or the US government.

Data Availability
Data sharing is not applicable to this article as the qualitative interview data contains participants’ sensitive information. More information is available from the corresponding author.

Conflicts of Interest
None declared.

Multimedia Appendix 1
During screening calls for interviews with veterans and caregivers who participated in a telemedicine visit between October 2020 and March 2021, qualitative research team members administered a technology questionnaire to gauge participant comfort with and access to technology. This questionnaire was used to help determine interview modality (telephone or videoconference) and anticipate interview technology challenges.

References


Abbreviations
- CBOC: community-based outpatient clinics
- CVT: clinical video telemedicine
- GRECC: Geriatric Research, Education, and Clinical Centers
- VA: US Veterans Health Administration
- VVC: VA Video Connect

©Jacqueline Hannah Boudreau, Lauren R Moo, Meaghan A Kennedy, Jennifer Conti, Chitra Anwar, Camilla B Pimentel, Kathryn A Nearing, William W Hung, Eileen M Dryden. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Patient and Provider Perspectives About the Use of Patient-Generated Health Data During Pregnancy: Qualitative Exploratory Study

Sarah R MacEwan1,2, PhD; Ramona G Olvera2, PhD; Pallavi Jonnalagadda2, DrPH; Naleef Fareed2,1, PhD; Ann Scheck McAlearney2,3,4, ScD

1Division of General Internal Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States
2Center for the Advancement of Team Science, Analytics, and Systems Thinking in Health Services and Implementation Science Research, College of Medicine, The Ohio State University, Columbus, OH, United States
3Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, United States
4Department of Family and Community Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Sarah R MacEwan, PhD
Division of General Internal Medicine
College of Medicine
The Ohio State University
700 Ackerman Rd
Suite 4100
Columbus, OH, 43202
United States
Phone: 1 614 293 0164
Email: Sarah.MacEwan@osumc.edu

Abstract

Background: There is increasing interest in using patient-generated health data (PGHD) to improve patient-centered care during pregnancy. However, little research has examined the perspectives of patients and providers as they report, collect, and use PGHD to inform obstetric care.

Objective: This study aims to explore the perspectives of patients and providers about the use of PGHD during pregnancy, including the benefits and challenges of reporting, collecting, and using these data, as well as considerations for expanding the use of PGHD to improve obstetric care.

Methods: We conducted one-on-one interviews with 30 pregnant or postpartum patients and 14 health care providers from 2 obstetrics clinics associated with an academic medical center. Semistructured interview guides included questions for patients about their experience and preferences for sharing PGHD and questions for providers about current processes for collecting PGHD, opportunities to improve or expand the collection of PGHD, and challenges faced when collecting and using this information. Interviews were conducted by phone or videoconference and were audio recorded, transcribed verbatim, and deidentified. Interview transcripts were analyzed deductively and inductively to characterize and explore themes in the data.

Results: Patients and providers described how PGHD, including physiologic measurements and experience of symptoms, were currently collected during and between in-person clinic visits for obstetric care. Both patients and providers reported positive perceptions about the collection and use of PGHD during pregnancy. Reported benefits of collecting PGHD included the potential to use data to directly inform patient care (eg, identify issues and adjust medication) and to encourage ongoing patient involvement in their care (eg, increase patient attention to their health). Patients and providers had suggestions for expanding the collection and use of PGHD during pregnancy, and providers also shared considerations about strategies that could be used to expand PGHD collection and use. These strategies included considering the roles of both patients and providers in reporting and interpreting PGHD. Providers also noted the need to consider the unintended consequences of using PGHD that should be anticipated and addressed.

Conclusions: Acknowledging the challenges, suggestions, and considerations voiced by patients and providers can inform the development and implementation of strategies to effectively collect and use PGHD to support patient-centered care during pregnancy.
Introduction

Patient-centered care can improve health outcomes by addressing each patient’s individual needs, increasing patients’ ownership of their health decisions, and improving patient-physician relationships [1-4]. In obstetrics, patient-centered care is recognized as important, yet providing this care during pregnancy remains challenging [5-8] as providers need to gather information from patients about their health and well-being throughout their pregnancies. To facilitate the gathering of person-centered information, there is increasing interest in the role of patients in reporting health-related data to their providers to inform patient-centered care.

Patient-generated health data (PGHD) is any health-related information that is gathered from or recorded directly by patients [9-11]. These data include self-reported symptoms, health behaviors, and physiologic measurements. While PGHD is frequently gathered during in-person appointments, opportunities also exist to collect these data between visits. To date, the collection of PGHD between in-person appointments to inform patient care has largely focused on physiologic measurement data, such as blood glucose levels and blood pressure [12-18]. The collection of these measurements is of particular interest during pregnancy, as diabetes and hypertension present common and serious risks to both mothers and their babies [19-27]. Although other aspects of patient-reported health status have been identified as important to informing patient care during pregnancy, such as quality of life, pain, and mental health status [28], limited evidence exists about the collection of these data between in-person appointments, or about the use of this information to inform patient care.

Patient sharing of PGHD between in-person appointments has been shown to improve patient-provider communication [29], enhance patients’ involvement in their clinical visits [30], and increase insight into patients’ health between clinical visits [31]. While evidence is building to understand if and how the use of PGHD can improve health outcomes [32], it is largely understudied in the context of pregnancy [33,34]. The use of PGHD may be particularly beneficial to improve obstetric care, as pregnancy encompasses a period during which many individuals experience frequent changes in their health status, and there are potentially serious complications that can occur that can impact maternal and infant health outcomes.

While there is growing interest in implementing strategies to collect PGHD, there has been little exploration of the experiences of patients and providers as they share and receive these data [35]. When implementing strategies to collect and use PGHD to inform patient care, technological considerations, such as integration of data into the electronic health record [12-14], have been at the forefront of research efforts. Less attention has been paid to the experiences of patients and providers as they collect and report PGHD, their perceptions of the impact of using PGHD, and the practical considerations that are critical to the implementation of strategies to expand the collection and use of these data. Exploring the perspectives of both patients and providers is, therefore, critical to improving our understanding of their needs and preferences, many of which should be acknowledged and addressed in clinical guidelines, reimbursement strategies, policy, and rules of engagement for the collection and use of PGHD [35].

This study aims to explore the perspectives of pregnant patients about sharing PGHD and the perspectives of obstetric providers about collecting and using PGHD for clinical decision-making during pregnancy. We were interested in understanding the benefits of using PGHD, as perceived by patients and providers, as well as the challenges that each group noted. Finally, we sought to gather patients’ and providers’ suggestions for ways to improve or expand the collection and use of PGHD during pregnancy.

Methods

Study Design and Setting

This exploratory qualitative study was guided by a constructivist and interpretivist research paradigm [36]. Through one-on-one interviews, we sought to explore the perspectives of our study participants within the unique contexts of their roles and experiences. We recognize the reflexivity of our research team in our data collection and the interpretation of our study findings, where research team members included bachelor-trained and doctoral-trained health services researchers. The reporting of our findings is guided by the Standards for Reporting Qualitative Research checklist [37].

We conducted one-on-one interviews with patients and providers from 2 obstetrics and gynecology ambulatory care clinics at The Ohio State University Wexner Medical Center, an academic medical center (AMC), between September 2020 and January 2021. Patients who were 18 years or older, pregnant or up to 90 days postpartum, and spoke English were eligible to participate. Eligible providers included physicians and nurses who worked in either or both clinics. Purposeful sampling was used to recruit patients to the study by telephone call inviting their participation, while providers were recruited by email. Emails sent to providers included a Qualtrics survey link to indicate their interest in participating and their preference for a telephone or videoconference interview.

Data Collection

Separate semistructured interview guides were used to support data collection from patients and providers (Multimedia Appendices 1 and 2). These interview guides were developed by 3 members of the research team (ASM, SRM, and NF) who worked in either or both clinics. Purposeful sampling was used to recruit patients to the study by telephone call inviting their participation, while providers were recruited by email. Emails sent to providers included a Qualtrics survey link to indicate their interest in participating and their preference for a telephone or videoconference interview.
symptoms and physiologic measures. The patient interview guide was tested with 2 individuals—1 pregnant and 1 postpartum—before finalizing the interview questions. Provider interviews asked about current strategies for collecting patient-generated information, opportunities to implement new strategies to collect this information in the future, and potential challenges faced when collecting this information.

Patient interviews were conducted by 2 female members of the research team who were either doctoral-trained or bachelor-trained (SRM and Abigail Petrecca). Patients did not know their interviewer prior to their interview. Patients were contacted by telephone and were given the option to complete the interview at the time of the call or schedule their interview at another time that was convenient for them.

Providers who were interested in participating in the study from the initial recruitment email provided a preferred time to schedule either a phone or videoconference (ie, Zoom) interview. Provider interviews were completed by 2 members of the research team, 1 female and 1 male, who were both doctoral trained (SRM and NF). Most providers did not know their interviewer prior to their interview; however, some providers had met their interviewer in a professional capacity prior to their interview. All interviews were audio-recorded and transcribed verbatim to permit rigorous qualitative analysis.

Data Analysis
We used thematic analysis to code the transcripts from all interviews [38]. First, we created separate preliminary coding dictionaries for the patient and provider interview transcripts based on questions from the semistructured interview guides (Multimedia Appendix 3). Three doctoral-trained or bachelor-trained members of the research team (SRM, NF, and Holly Heffer) then coded 2 patient transcripts and 2 provider transcripts using these preliminary coding dictionaries. The coders met to discuss discrepancies in initial coding and refined the coding dictionary. Two of the initial coders (SRM and Holly Heffer) individually coded the remaining transcripts using the refined coding dictionary, meeting frequently to ensure consistency between the coders. After this deductive coding, the 2 coders met to discuss emergent subthemes they identified during the coding process and inductively developed subcodes that were added to the coding dictionary. One coder (Holly Heffer) then applied these subcodes to all the transcripts. The coding team reached thematic saturation based on consensus among coders that themes fully covered emergent topics across transcripts [39]. ATLAS.ti software (ATLAS.ti Scientific Software Development GmbH) was used to support the coding and analysis process.

Ethical Considerations
The Ohio State University’s institutional review board approved this study (2020B0038). Participation was voluntary and all participants provided verbal informed consent. Interview transcripts were deidentified to protect participant privacy and confidentiality. Patients and nonphysician participants received a US $25 gift card in appreciation for their participation.

Results
Participant Characteristics
Interviews were conducted with 30 patients and 14 providers. Patients had a mean age of 30 years and included 22 pregnant individuals (mean gestational age: 28.6 weeks) and 8 postpartum individuals (mean weeks postpartum: 6.3). Providers included 9 physicians and 5 nurses. Patient interviews lasted an average of 16 minutes, while provider interviews lasted an average of 26 minutes.

Thematic Analysis
Overview
Three themes were identified across patient and provider interviews: (1) current collection of PGHD during pregnancy, (2) suggestions for reporting PGHD during pregnancy, and (3) the impact of reporting PGHD during pregnancy. A fourth theme was identified only from provider comments: (4) considerations for expanding the collection of PGHD during pregnancy. Themes and their respective subthemes are discussed below and summarized in Table 1.
<table>
<thead>
<tr>
<th>Themes and subthemes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current collection of PGHD(^a) during pregnancy, by health concern and measurement</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Diabetes</strong></td>
<td></td>
</tr>
<tr>
<td>Blood glucose</td>
<td>• 4-8 daily blood glucose values reported weekly through email</td>
</tr>
<tr>
<td>Symptoms (eg, indications of ketoacidosis)</td>
<td>• Only reported at in-person appointments</td>
</tr>
<tr>
<td><strong>Hypertension</strong></td>
<td></td>
</tr>
<tr>
<td>Blood pressure</td>
<td>• Out-of-range values reported by phone</td>
</tr>
<tr>
<td>Symptoms (eg, indications of pre-eclampsia)</td>
<td>• Only reported at in-person appointments</td>
</tr>
<tr>
<td><strong>Mental health</strong></td>
<td></td>
</tr>
<tr>
<td>Perceptions of mood, anxiety, and depression</td>
<td>• Only reported at in-person appointments</td>
</tr>
<tr>
<td><strong>Fetal movement</strong></td>
<td></td>
</tr>
<tr>
<td>Kick counts</td>
<td>• Decreases in movement reported by phone</td>
</tr>
<tr>
<td><strong>Other symptoms of pregnancy complications</strong></td>
<td></td>
</tr>
<tr>
<td>Nausea, bleeding, cramping, discharge, vaginal pressure, dizziness, fainting, itching, and pain</td>
<td>• Worsening symptoms reported by phone or patient portal messaging</td>
</tr>
<tr>
<td><strong>Suggestions for reporting PGHD during pregnancy</strong></td>
<td></td>
</tr>
<tr>
<td>Use patient portal for PGHD collection</td>
<td>• Reporting PGHD in the patient portal could allow data to be stored in a way that is accessible to both patients and all members of the health care team.</td>
</tr>
<tr>
<td>Use reminders to collect PGHD</td>
<td>• Reminders to take and record measurements could increase adherence with PGHD reporting.</td>
</tr>
<tr>
<td>Collect information about symptoms</td>
<td>• Collecting symptoms, in addition to physiologic measurements, could facilitate the identification of concerns or complications.</td>
</tr>
<tr>
<td>Expand collection of PGHD to health concerns that are currently only assessed in person</td>
<td>• Collection between in-person appointments could improve clinical awareness of information related to fetal movement and mental health concerns.</td>
</tr>
<tr>
<td><strong>Impact of reporting PGHD during pregnancy between in-person appointments</strong></td>
<td></td>
</tr>
<tr>
<td>Informs patient care</td>
<td>• Collecting PGHD facilitates the identification of concerns and the adjustment of patient care.</td>
</tr>
<tr>
<td>Encourages patient involvement in their care</td>
<td>• Reporting PGHD increases attention to and awareness of health.</td>
</tr>
<tr>
<td><strong>Considerations for expanding the collection of PGHD during pregnancy between in-person appointments</strong></td>
<td></td>
</tr>
<tr>
<td>Considerations for the patient role</td>
<td>• Target and tailor PGHD collection for each patient’s clinical needs.</td>
</tr>
<tr>
<td></td>
<td>• Provide patient guidance for reporting information that may be difficult to quantify (eg, nausea, severity of bleeding).</td>
</tr>
<tr>
<td>Considerations for the provider role</td>
<td>• Present data in a way that conveys necessary information and trends over time (eg, alert to abnormal values that suggest clinical concern).</td>
</tr>
<tr>
<td></td>
<td>• Assign roles and responsibilities for who is reviewing and responding to PGHD as it is collected.</td>
</tr>
<tr>
<td>Considerations for unintended consequences</td>
<td>• Have mechanisms in place to identify and escalate emergencies.</td>
</tr>
<tr>
<td></td>
<td>• Recognize potential increase in nonemergent concerns and subsequent clinical burden and cost.</td>
</tr>
</tbody>
</table>

\(^a\)PGHD: patient-generated health data.
Current Collection of PGHD During Pregnancy

Patients and providers described the current collection, reporting, and use of PGHD during pregnancy. Patients and providers reported that the majority of PGHD were collected or reported during in-person appointments. It was noted to be less common for patients to report PGHD or for providers to collect PGHD between their appointments.

Providers described collecting PGHD related to symptoms, as well as self-reported physiologic measurements including blood pressure and blood glucose values, and noted using these data to direct patient care. Although information about symptoms was predominantly collected in person, providers reported informing patients to call the AMC if they experienced any concerning or worsening symptoms between visits. For example, 1 provider described:

I tell patients to keep a log [of fetal kicks] if they feel like, you know, to get a basis for what their babies do, because some babies are more active and some are less active. So, typically we're not like reviewing those logs. I just tell them to keep that information so if they noticed that it's less frequent, they can, you know, let us know. [provider 3]

Providers commented that they asked some patients to regularly measure their blood pressure at home between appointments and contact their provider if their blood pressure exceeded a certain limit. Providers also reported that they asked patients with diabetes to track their blood sugar levels at home (taking between 4 and 8 daily measurements) and share those values with them each week. Patients most frequently reported sharing blood sugar values with a clinic nurse via email.

Suggestions for Reporting PGHD During Pregnancy

Patients and providers described several opportunities to improve or expand the collection of PGHD between in-person appointments.

Use Patient Portal for PGHD Collection

For physiologic measurements that were already being reported between appointments to monitor hypertension and diabetes (ie, self-reported blood pressure and blood glucose measurements), patients suggested that it might be better to report this information in the AMC’s patient portal, where the data could be stored and available for the whole care team to view. One patient suggested:

I think I’d rather have it in the patient portal because then it’s like recorded. It’s logged, everyone has access to it. So, it’s not just one person has an email, you know, the whole staff team could have availability of it and then you have it. [patient 13]

Use Reminders to Collect PGHD

Patients and providers also commented that reminders might improve adherence to taking and reporting these measurements. As 1 provider suggested:

We probably could be better about being sure that patients are potentially even sent reminders about like how often they should be checking their blood pressures ...whether it be morning and night or just one time during the day. You know, it’s sending a reminder to them to be like, “Hey, can you enter your blood pressure for the day?” [provider 9]

Collect Information About Symptoms

For patients with diabetes or hypertension, providers suggested collecting additional information about symptoms between in-person appointments; for example, asking about complications among patients with diabetes, as 1 provider suggested: “With the diabetic patients, I think it would be helpful to know, are you having anything that could be signs and symptoms of ketoacidosis” (provider 5). Another provider suggested asking about symptoms of pre-eclampsia: “I think certainly if we have a patient who has preeclampsia, asking them for symptoms like headache, change in vision, things like that would be important” (provider 8).

Expand Collection of PGHD to Health Concerns That Are Currently Only Assessed in Person

Patients and providers also proposed that it could be useful to collect PGHD between appointments related to health concerns that were currently only assessed in person, including changes in fetal movement and evaluating mental health status. One provider described the opportunities to collect information about fetal movement:

With like you mentioned earlier, fetal movement, there’s no formalized way that we do that except, you know, talk to people in office visits. But if you want to collect data, that would be a good thing to do if patients could just weekly send them something saying, yes fetal movement has been the same, or fetal movement has been less, or things like that. [provider 5]

With regard to evaluating mental health status, 1 patient recommended:

Maybe the postpartum depression, stuff like that, maybe have them look more into that. Because the only time I’ve done the questionnaire for that was when I took my twins to the pediatrician. [patient 11]

A provider echoed this suggestion:

I think if there could be some sort of system where it tracks people’s mood and then if it was really bad, you know some certain threshold, that it sent a notification to the provider. I think that would be good, especially for postpartum depression. [provider 2]

Impact of Reporting PGHD During Pregnancy Between In-Person Appointments

Patients and providers also noted the positive impacts of reporting PGHD during pregnancy, including that sharing PGHD helped to inform patient care and helped to encourage patients’ involvement in their care.

Informs Patient Care

First, providers commented that reporting PGHD allowed them to identify issues and adjust patient care between clinical visits.
For example, collecting blood sugar values allowed providers to adjust diabetes medication, as 1 patient with pregestational diabetes described, “It [reporting blood sugar values] helped, it definitely helped regulate. It helped them to figure out how to adjust my [insulin] pump to regulate my diabetes in general” (patient 23). One provider described how having information between in-person visits helped them make health care decisions for their patients:

- I think that they empower you to feel like you're making the right decision and care for the patient. I mean, I know when I have, when I have the information documented and I have it to work off of, I think two ways in which I think it’s helpful: one, it gives me the comfort level as the provider that I’m making an informed decision about their care. I think, in particular, about the diabetic patients and the hypertensive patients, it’s good to have that information so that you know that you can make changes to their care that feel appropriate and safe. [provider 3]

Providers also noted that knowing PGHD between in-person appointments could help them adjust patient care in a timely manner:

- The blood pressure, blood sugars, you know, in between visits, if we knew that we were having issues, you know, we would be able to assist them quicker than you know, waiting until the next appointment. Or even with postpartum depression, or, you know, fetal movement counts. You know, those things we would be able to address quicker if we did it in between visits. If we were able to track it in between visits. [provider 15]

**Encourages Patient Involvement in Their Care**

Patients and providers also reflected that reporting PGHD could encourage patients’ involvement in their care. One patient explained their increased attention to their health when asked to report information between in-person appointments:

- I just feel like you’re more aware of things and pay more attention to things if you are like sending that information in between appointments...sometimes you wait till you almost have your appointment and then you start paying attention to things. Where if you’re doing it like every week or throughout, then you’re just more aware, and you’re basically paying more attention to what’s going on. [patient 28]

Another patient described their increased sense of control in their health care with regard to reporting their blood glucose levels: “With what I have done with the insulin and some concerns with the gestational diabetes at the beginning, it does make me feel more in control, and it does make me feel like I have more say in my care” (patient 29). A provider’s comment echoed this perspective: “I do think it will give patients more ownership of their care” (provider 5).

Providers also expressed that collecting PGHD could help them understand patients better and potentially strengthen the patient-provider relationship, which might similarly encourage patients’ involvement in their care: “I think it would just allow for a more like solid or strengthened patient-doctor relationship that you know, hopefully the patient would feel that I'm more involved and checking in on her, and kind of aware of what's going on in between visits” (provider 4).

**Considerations for Expanding the Collection of PGHD During Pregnancy Between In-Person Appointments**

Finally, providers mentioned several considerations when discussing opportunities to expand the collection of PGHD to support obstetric care. These fell into the following categories: considerations for the patient role, considerations for the provider role, and considerations for unintended consequences.

**Considerations for the Patient Role**

With respect to patients’ roles in reporting PGHD, providers noted the importance of targeting and tailoring instruments (eg, questionnaires) to collect this information. Providers stressed that PGHD collection from each patient should be personalized to that patient’s individual clinical needs. As 1 provider described:

- Linking certain diagnoses to it and certain conditions to it, it makes them much more targeted and tailored approaches rather than just kind of a “here’s everything you could possibly answer.” And that would be completely overwhelming to people and completely turn them off. [provider 9]

Providers also commented about the need for guidance when asking patients to report information that may be hard to quantify, including the perceived severity of symptoms that are subjective. One provider described this challenge in the context of nausea:

- I think nausea is hard because I think so many women also experience nausea. And we try to manage those women as an outpatient as best as we can. And I think sometimes it’s patient perspective that helps. You know, you can have someone who’s really nauseous and vomits all the time and they’re able to manage that at home with medications. ...And then you have other women who vomit but they feel like their amount of nausea is something they need to be seen for. Their amount of nausea and their amount of vomiting objectively may be less than someone else who’s controlling symptoms at home. So, I think that would be a little tricky. [provider 2]

In another example, a provider explained this consideration in the context of bleeding:

- You know, bleeding is a little bit harder because that’s a little bit, so subjective. But it can be, I think there’s ways to potentially quantify that to people's satisfaction, such that they would feel okay with it. Like in the sense of you know, if you have spotting on your tissue, okay, you know, did you see it again? Or like, check again with like another wipe or in like another like two hours, if it's still there come in. Or if it’s fully filling the tissue paper or even like a panty liner come in, those types of things. [provider 9]
Considerations for the Provider Role

With regard to the provider’s role in reviewing and responding to PGHD, providers stressed the importance of how the data would need to be presented to them. Providers reflected that, depending on the type of PGHD, they may not need to review every value or data point, but rather the data should be presented in a way to alert them to abnormal values that would be cause for concern. One provider gave the example:

I mean some things we need to know, like blood sugars. We have to know every single blood sugar. But you know, perhaps something like a fetal movement count, we don’t really need to know about that until it’s abnormal. [provider 12]

Providers also commented on the importance of presenting the data in a way that could help them understand trends over time (eg, tables or charts), rather than going through data presented as text alone: “If we have a formalized portal, that would be great, that would come up with a you know, a graph or a table rather than have to wade through a bunch of texts” (provider 5).

Additionally, providers identified the importance of assigning roles and responsibilities among the care team to ensure that PGHD was collected appropriately, promptly reviewed, and acted upon by the care team. One provider explained these considerations:

I’d say there’s the systemic issue of you know, who is responsible for making sure that information is read? When is it supposed to be read? And what ways are we supposed to respond to it? Who’s supposed to be responding to it? How frequently are they supposed to be checking? [provider 13]

Considering roles and responsibilities was critical due to the additional workload these processes presented to the health care team, as 1 provider commented:

During pregnancy, my concern with collecting all the information would be about who is assessing it. So, if we are sending out text messages once a week or every day, making sure that there’s a nurse who’s going to be really thoughtful about looking at it so that the concerns, symptoms, complaints are not automatically pushed to the physician to review and look at. I think there’s already a lot of work being done and physician burnout, and so my concern would be that with all this extra data, is that overwhelming the physician’s already limited time and brain capacity during the day to kind of review all that information? [provider 8]

Considerations for Unintended Consequences

Finally, providers described considerations regarding the potential unintended consequences of collecting PGHD. First, providers noted the need to identify and react to emergencies as indicated by patients’ responses. One provider gave an example in the context of a patient reporting changes in fetal movement:

If there was a way for the system to automatically tell the patient if she said, “I’m not feeling the baby move, I have decreased fetal movement,” then there should be no delay. It should be like an automatic response to come to the hospital and get evaluated. That is the only thing. I feel like there’s some liability there if a rare event or something would happen. [provider 4]

Another provider described how processes might be designed to help identify emergencies in the context of identifying pre-eclampsia:

We have the potential for automation to kind of play a factor in the sense that you know, if you have certain triggering blood pressure thresholds that would, you know, rapidly bring that to the attention of somebody, or more specifically guide a patient through, okay you have this blood pressure, is this a second blood pressure you’ve already checked because it was elevated before? If not, you know, re-check it in 15 minutes after you’ve sat down, relaxed, all that thing. Or if this blood pressure is still severely elevated above this level after 15 minutes, you know, please, you know, report to the like emergency room or, you know, call this number specifically so we can get you in touch with the provider and guide you through the care process. [provider 9]

In addition to identifying and reacting to emergencies, providers also expressed consideration that the collection of PGHD may increase the number of nonemergent concerns being reported during pregnancy. One provider explained their consideration:

I think my biggest concern would just be with, with extra data comes up with extra medical intervention. That would be my biggest concern. I think them reporting it would be fine. But would we be more apt to bring people into the hospital then, to assess these symptoms? More testing? When ultimately maybe they were, you know, most of them were fine. [provider 8]

Some providers questioned whether reporting PGHD would result in patients reporting a greater number of concerns, which could also potentially increase medical costs and cause extra work for clinic staff:

I’d be a little bit worried that we would get a lot of responses that we would be following up on that wouldn’t necessarily amount to anything. And that would be a lot of extra work for the staff in bringing those patients in and tracking those outcomes on so many patients that we have. [provider 2]

Discussion

Principal Results

Our qualitative analysis provides insight into patients’ and providers’ perspectives on the value and challenges of collecting and using PGHD to inform obstetric patient care. While the use of PGHD during pregnancy has most commonly been reported for objective physiologic measurements (eg, blood glucose and blood pressure) [40,41], our patient and provider participants...
also expressed interest in the collection of subjective measurements including anxiety, depression, and symptoms indicative of pregnancy complications. Recognizing the interest in PGHD beyond physiologic measurements is important, as studies have demonstrated that the collection of a range of different types of PGHD can improve understanding of the unique context of each patient’s pregnancy. For example, pregnant women with heart disease identified general well-being, mental health, fatigue, and quality of life as important topics for which PGHD, in addition to clinical outcomes specific to cardiac function, may be critical information to share with their providers [42]. Similarly, in a study of pregnant women with gestational diabetes, the collection of PGHD including patient perspectives of anxiety, self-knowledge about diabetes, and social support has been proposed to provide a more comprehensive picture of the patient that helps explain why some patients experience challenges with blood glucose management [43]. Increasing holistic awareness of the patient through the collection of PGHD may also help providers target interventions (e.g., education) as a means to improve maternal health outcomes [44].

The growing emphasis on patient-centered care during pregnancy has brought to light the potential of PGHD as an important factor that can enhance the delivery of value-based care [23,28,45-47]. Evidence is growing surrounding the use and effectiveness of strategies for using PGHD to inform obstetric care. For example, uploading self-reported measures of blood glucose to a patient portal has been associated with improvements in glycemic control in a sample of pregnant and nonpregnant patients [13], and continuous monitoring of blood glucose has been associated with improved neonatal outcomes [46,48]. Evidence will continue to grow from several ongoing clinical trials on this topic, with the greatest focus of these studies on patient-generated physiologic measurements [49,50].

Despite interest in expanding the use of PGHD during pregnancy, little work has investigated the perspectives of stakeholders in the use of PGHD, including those of pregnant individuals providing this information and the providers receiving and using this information to improve obstetric care. While patients and providers in our study commented about the clinical benefits of using PGHD, for which there is a mixture of evidence in the literature [15,51,52], they also suggested the additional benefit of improving patients’ involvement in their care through the collection of PGHD. In nonpregnant populations, the collection of PGHD has been recognized to create an opportunity for self-reflection among patients that can encourage patient-provider communication, which, in turn, increases provider awareness of patient concerns [53]. Similarly, others have suggested additional benefits of using PGHD beyond their impact on health outcomes, such as improved patient satisfaction with care as a result of improved patient-provider communication [29,45,54].

In imagining the expanded use of PGHD to inform obstetric care, providers in our study explained several considerations in the practical implementation of strategies to collect and use this information. These considerations included questions surrounding the patient’s role in reporting PGHD, the provider’s role in using PGHD, and the potential unintended consequences of collecting PGHD. These considerations bring to light how challenging it can be to introduce PGHD into clinical practice. Furthermore, it will be important to evaluate these considerations within the unique context of the multiple ways in which PGHD can be collected, synthesized, and displayed. Examples of methods in operation or development for these purposes include the use of electronic health records and patient portals [12], automated SMS text messages [55-57], mobile health apps [29], wearable devices [58,59], and remote patient monitoring systems [60,61].

While interest in expanding the use of PGHD during pregnancy was expressed by providers in our study, there is little guidance on the collection of PGHD and a paucity of research on using PGHD in this patient population [33,34]. Furthermore, the infrastructure to measure and synthesize PGHD is not extensive in the United States [45]. For example, challenges noted in the literature include how PGHD can be collected, how to best interpret PGHD, and how meaningful thresholds of PGHD are indicative of an action required by the patient or provider [62-64]. Notably, many of the approaches to address these challenges will be specific to providers, their health care systems, and the patient populations they serve.

One mechanism to increase our collective knowledge surrounding the collection and use of PGHD in pregnancy is through the existing call to action for randomized controlled trials in maternal health to standardize the selection, collection, and reporting of outcomes based on PGHD that reflect the perspectives of study participants [65]. Such an endeavor could facilitate the translation of research into clinical practice for PGHD collection and use while increasing contextual knowledge about patient health and health care experiences that can be used to improve the patient-centeredness of obstetric care. Future research should focus on the evaluation of methods for using PGHD in obstetric care to begin to build evidence and inform implementation strategies for this practice.

**Limitations**

One limitation of our study is that the participant population was drawn from a single AMC. Our findings, therefore, represent perspectives specific to the patient and provider populations at our clinical sites. In addition, while our study population included both patients and providers, our analytical approach was not designed to identify convergence or divergence of perspectives across groups. Future research that explicitly assesses similarities and differences in the perspectives of patients and providers on these topics is important to inform interventions for PGHD collection and reporting that meet the needs and preferences of both stakeholder groups. We conducted this study during the COVID-19 pandemic, a period during which pregnant individuals may have experienced changes in their clinical care due to the many impacts of the pandemic. Due to restrictions on in-person research at this time, study participants were recruited by phone or email, which may have influenced their willingness to participate. The remote nature of the interviews, conducted by phone or videoconference, may have also impacted the sharing of information, as compared with an in-person interview. Furthermore, changes in care experienced due to social distancing and other restrictions may...
have increased interest in remote communications, including those that involved the collection of PGHD. Finally, we did not systematically collect participant clinical characteristics, such as pregnancy complications, that could impact participants’ perspectives about PGHD. Including the collection of these data in future studies may be important to understand the perspectives of individuals based on their specific clinical needs.

**Conclusions**

Our qualitative analysis presents patient and provider perspectives on the collection of PGHD and its use during pregnancy. Participant responses collectively supported the use of PGHD to improve obstetric care. While several opportunities to expand the use of PGHD during pregnancy were mentioned, many considerations for implementing strategies to use PGHD were also noted, shedding light on the challenging nature of collecting and using this information in clinical practice.

**Acknowledgments**

The authors thank Johanna S DeStefano, PhD, Professor Emerita, The Ohio State University, for the generous donation that supported this work. The authors acknowledge Holly Heffer, Abigail Petrecca, Lindsey Sova, and Matthew Rohl for their assistance with this research and express their gratitude to the patients and providers who participated in this study.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Patient semistructured interview guide.
[DOCX File, 27 KB - formative_v8i1e52397_app1.docx ]

Multimedia Appendix 2
Provider semistructured interview guide.
[DOCX File, 29 KB - formative_v8i1e52397_app2.docx ]

Multimedia Appendix 3
Patient and provider coding dictionaries.
[XLSX File (Microsoft Excel File), 27 KB - formative_v8i1e52397_app3.xlsx ]

**References**

2. NEJM Catalyst. What is patient-centered care? NEJM Catal 2017;3(1) [FREE Full text] [doi: 10.1056/CAT.17.0559]


42. Hall C, D’Souza RD. Patients and health care providers identify important outcomes for research on pregnancy and heart disease. CJC Open 2020;2(6):454-461 [FREE Full text] [doi: 10.1016/j.jcoep.2020.05.010] [Medline: 33305204]


Abbreviations

AMC: academic medical center
PGHD: patient-generated health data
Relationship Between Product Features and the Prices of e-Cigarette Devices Sold in Web-Based Vape Shops: Comparison Study Using a Linear Regression Model

Yanyun He¹, PhD; Qian Yang¹, MPH; Yousef Alish², BSc; Shaoying Ma¹, PhD; Zefeng Qiu³, MS; Jian Chen³, PhD; Theodore Wagener¹⁴, PhD; Ce Shang¹⁴, PhD

¹Center for Tobacco Research, The Ohio State University Wexner Medical Center, Columbus, OH, United States
²Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University, Columbus, OH, United States
³Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, United States
⁴Medical Oncology Division, Department of Internal Medicine, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Yanyun He, PhD
Center for Tobacco Research
The Ohio State University Wexner Medical Center
3650 Olentangy River Rd
Columbus, OH, 43214
United States
Phone: 1 9739607969
Email: yanyun.he@osumc.edu

Abstract

Background: Open-system electronic cigarette (EC) product features, such as battery capacity, maximum output wattage, and so forth, are major components that drive product costs and may influence use patterns. Moreover, continued innovation and monitoring of product features and prices will provide critical information for designing appropriate taxation policies and product regulations.

Objective: This study will examine how product features are associated with the prices of devices sold in web-based vape shops.

Methods: We draw samples from 5 popular, US-based, web-based vape shops from April to August 2022 to examine starter kits, device-only products, and e-liquid container–only products. We implemented a linear regression model with a store-fixed effect to examine the association between device attributes and prices.

Results: EC starter kits or devices vary significantly by type, with mod prices being much higher than pod and vape pen prices. The prices of mod starter kits were even lower than those of mod devices, suggesting that mod starter kits are discounted in web-based vape shops. The price of mod kits, mod device–only products, and pod kits increased as the battery capacity and output wattage increased. For vape pens, the price was positively associated with the volume size of the e-liquid container. On the other hand, the price of pod kits was positively associated with the number of containers.

Conclusions: A unit-based specific tax, therefore, will impose a higher tax burden on lower-priced devices such as vape pens or pod systems and a lower tax burden on mod devices. A volume- or capacity-based specific tax on devices will impose a higher tax burden on vape pens with a larger container size. Meanwhile, ad valorem taxes pegged to wholesale or retail prices would apply evenly across device types, meaning those with advanced features such as higher battery capacities and output wattage would face higher rates. Therefore, policy makers could manipulate tax rates by device type to discourage the use of certain device products.

(JMIR Form Res 2024;8:e49276) doi:10.2196/49276

KEYWORDS

e-cigarette devices; price; web-based vape shop; battery capacity; output wattage; e-cigarette; vape; vaping; consumers; costs; prices; pricing; feature; features; device; devices; consumer; cost; smoking products; product safety; consumer safety; smoking; smoker; smokers; tax; taxes; taxation; economic; economics; regression; regression model; health economics
Introduction

The electronic cigarette (EC) market has experienced a surge in growth, especially among adolescents and young adults in the United States, which prompted a series of government regulatory actions such as flavor product restrictions and e-cigarette taxes at various government levels [1]. Although ECs pose addiction risks among adolescents and young adults, they also have the potential to help people who smoke combustible tobacco to quit smoking [2], making EC regulations challenging and controversial. Nonetheless, effective EC regulations require that policy makers weigh the benefits and risks of ECs [3,4].

One important feature distinguishing ECs from cigarettes is the wide range of EC models, ranging from basic disposable models to more complex rechargeable devices, which may affect product appeal, use patterns, and health consequences. The use of certain EC devices such as mods and pods may be associated with more frequent EC use and nicotine dependence symptoms [5-7]. The majority of adolescent and young adult EC users report using rechargeable pods (eg, JUUL) [7-11], while smokers who successfully quit smoking are more likely to use open-tank systems or mods [12,13]. The choices of models are further associated with the frequency of EC use and nicotine dependence, making EC devices or models an important product attribute for policy makers to regulate [5-7].

In addition to regulating product attributes by implementing product standards, policy makers could also impose taxes on ECs, including devices, to change their appeal. To attract new users, many EC retailers offer starter kits at discounted prices [14,15], which typically include a rechargeable device and replaceable e-liquid tanks or pods [16,17]. Therefore, taxes and promotion restrictions may be needed to decrease the affordability or appeal of these starter kits to prevent youth initiation. On the other hand, the perceived cost of the devices may make ECs seem expensive compared to traditional tobacco products, which may prevent people from switching from cigarettes to ECs [18,19]. Therefore, monitoring EC device prices and attributes associated with price differences is crucial for policy makers and public health officials to make informed decisions about EC taxation and other pricing policies [5,6].

Rapid advancements in technology have resulted in an increasing number of EC devices being offered in the market, making it challenging to monitor the market and keep track of pricing trends. The existing surveillance data also have the limitation of only providing information on brands sold in brick-and-mortar stores and missing the products sold in other retail channels such as vape shops and web-based stores. Prior studies examining web-based EC stores or vendors revealed that the prices offered digitally were much cheaper than physical stores and few collected sales taxes in their state of business or based on shipping addresses [20,21]. Web-based retail websites engage in a variety of promotional strategies, such as promo or discounts, customer rewards, loyalty programs, and so forth [20,21]. Moreover, they use multiple marketing techniques such as linking their websites to social media platforms, using celebrity endorsement, posing misleading messages about ECs (eg, ECs are healthier, cleaner, and effective as a cessation aid), and so forth, to attract younger population [22-24]. As the US Food and Drug Administration (FDA) issued denial orders to approximately 1 million EC-flavored products in 2020, how this will shift the marketplace and product features remains unknown. It is therefore critical to conduct comprehensive and rapid surveillance on EC devices.

Finally, a growing literature suggests that EC taxes and prices are effective in reducing EC use, with a 10% increase in prices associated with an 11.5% decrease in sales or purchases [25]. However, it is also shown that the current tax bases for ECs vary significantly by state; some states adopt specific taxes, whereas others adopt ad valorem taxes. Given that ECs have a wide range of configurations and features, different tax bases may lead to different tax burdens on different EC types (eg, vape pens vs disposables) [26]. The taxation policies also differ by state regarding whether to tax devices, which are the more durable components compared to refillable cartridges or e-liquid in open-system ECs. A better understanding of EC devices and their costs is needed to guide taxation policies for devices.

In response to these research needs, this study aims to bridge the gap in the literature by analyzing EC device data from popular web-based retailers to evaluate the distribution of device prices and features and investigate the associations between device characteristics and prices in the marketplace using a hedonic pricing model [27]. The results of this study will provide insights to policy makers considering product standards and taxation policies for ECs.

Methods

Data Sources

After conducting extensive research through a combination of Google searches and Reddit discussions in 2021, we curated a list of 5 popular web-based vaping shops. Using the latest information available, we prioritized the top results from Google searches and Reddit threads. Specifically, we focused on the top 3 results without physical addresses from Google and identified 2 highly regarded shops without physical addresses from a Reddit discussion dated in 2020. This comprehensive approach ensured that our selection process was thorough and reflective of the most current and popular web-based vape shops available.

From April to August 2022, we conducted a study on open-system EC device products sold from these 5 web-based vape shops. In total, we identified 1166 reusable products after charging or changing batteries. These products include starter kits, device-only products, and container-only products. Starter kits refer to products that include both the heating device and containers, such as mod kits, pod kits, and vape pens. Device-only products only include the heating devices, such as mod and pod devices. Container-only products are just replacement tanks or pods and contain no liquid or solutions. To examine the relationship between prices and device attributes, we focused our analysis on starter kits and device-only products as replacement tanks or pods are components instead of devices.
Outcome Variable

The outcome variable in our analysis was the log-transformed effective price (ie, after discounts) in US dollars, extracted from the store web pages.

Explanatory Variables

The following product attributes were selected for regression analysis: the number of containers (ie, tanks or pods or cartridges), container volume size (ie, maximum e-liquid capacity per tank, pod, or cartridge in mL), maximum output wattage (divided into 3 groups with each group comprising approximately one-third of the total observations: 5-39 W, 40-85 W, and more than 85 W), and battery capacity (divided into 3 groups: less than 900 mAh, 900-1499 mAh, and 1500 mAh and more). For the products with missing battery capacity (n=488) or missing wattage information (n=91), we manually checked each product web page. We found that the missing battery capacity in mod kits (n=273) or mod device–only products (n=209) was due to their use of 18,650- or 21,700 mAh–sized batteries and a lack of inclusion in the kits or devices. The missing battery capacity information on other products (n=6) and missing wattage information were purely due to a lack of information on the product web pages. To fully use the sample that we collected for the analysis, a missing category was added for output wattage and battery capacity. In addition to these attributes, the number of coils, rings, cables, chargers, batteries, glasses, and chips were included as control variables.

Statistical Methods

We used a 3-step approach to investigate pricing patterns and product attributes of EC devices sold in web-based vape shops. First, we computed the price distribution for all EC device products and identified the respective brands. Second, we examined the battery capacity and maximum output wattage features of products for starter kits and device-only products. Finally, we analyzed a linear regression model to examine the association between device prices and product attributes using a hedonic pricing model [27], controlling for store-specific unobservable factors using store-fixed effect and stratified by device types (mod kit, mod devices only, pod kits, and vape pens). Pod device–only products were excluded from the analysis because of the small sample size (n=8). SEs were clustered at the store level to account for intertemporal correlations among products sold in the same store. As the outcome variable was the log form of device price, our estimates reflect the percentage change in price due to a 1-unit change in a continuous independent variable. In the case of a categorical independent variable, our estimates indicate the percentage change in price for being in a certain category compared to the comparison category.

Given that our sample comprises a large number of brands (93 brands) and many brands (59/93, 63% brands) have fewer than 5 products, we did not control for brand-fixed effects. Nonetheless, we conducted sensitivity analysis and estimated alternative models where brand effects are controlled using random effects and generalized estimating equations.

Ethical Considerations

In this study, we collected data from 5 US-based, web-based vape shops. Thus, no human subjects were involved, and the determination of no human subjects was approved by the Ohio State University institutional review board (study 2020E1328).

Results

Table 1 shows the summary statistics of EC device products. We identified 1166 products from 93 unique brands, including 427 mod kits, 348 pod kits, 50 vape pens, 229 mod device–only products, 8 pod device–only products, and 104 replacement tanks or pods. Among starter kits, mod kits have a mean price of US $51.46 (SD US $24.23), which is significantly higher (P<.001) compared to pod kits (US $24.72, SD US $8.78) and vape pens (US $29.50, SD US $15.60). The mean price of mod kits is significantly lower (P<.001) than that of mod device–only products (US $58.93, SD US $41.98). In addition, the median price of mod starter kits is about US $6 cheaper than the sum of individual mod device and e-liquid container prices (US $44.99 and US $8.99, respectively), suggesting that mod kit prices are heavily discounted.

Table 1. Summary statistics of EC device products from 5 popular, US-based, web-based vape shops from April to August 2022 (n=1166).

<table>
<thead>
<tr>
<th>Product type</th>
<th>Price (US $), mean (SD)</th>
<th>Price (US $), range</th>
<th>Price (US $), median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Starter kits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mod kit (n=427)</td>
<td>53</td>
<td>51.46 (24.23)</td>
<td>17.99-244.99</td>
</tr>
<tr>
<td>Pod kit (n=348)</td>
<td>58</td>
<td>24.72 (8.78)</td>
<td>4.99-84.99</td>
</tr>
<tr>
<td>Vape pen (n=50)</td>
<td>16</td>
<td>29.50 (15.60)</td>
<td>11.95-99.95</td>
</tr>
<tr>
<td><strong>Device-only products</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mod device–only products (n=229)</td>
<td>48</td>
<td>58.93 (41.98)</td>
<td>17.99-299.99</td>
</tr>
<tr>
<td>Pod device–only products (n=8)</td>
<td>8</td>
<td>11.73 (6.95)</td>
<td>3.99-20.99</td>
</tr>
<tr>
<td><strong>Container-only products</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replacement tanks or pods (n=104)</td>
<td>19</td>
<td>11.80 (7.83)</td>
<td>1.99-33.99</td>
</tr>
</tbody>
</table>

aEC: electronic cigarette.
The price distribution for all devices except pod kits, which follow a normal distribution, is positively skewed (i.e., having high prices; Figures 1-4). Few pod device–only products (n=8) were identified, and they were significantly less expensive than pod kits. This reveals that web-based stores are more likely to sell pod devices in starter kits instead of individual products. The price distributions of vape pens and pod kits are similar: the lower quartile and median prices for vape pens and pod kits are almost identical.

**Figure 1.** Price distribution of mod products from 5 US-based popular vape shops from April to August 2022 (mod device: n=229, mod kit: n=427).

**Figure 2.** Price distribution of pod products from 5 US-based popular vape shops from April to August 2022 (pod device: n=8, pod kit: n=348).
Table 2 displays the distribution of product attributes based on the device type. Pod kits tend to have more replacement tanks or pods included in the kits than mod kits and vape pens. However, the volume size of the containers in mod kits is significantly larger than that in pod kits and vape pens. Most mod kits (396/427, 92.74%) have a battery capacity above 1500 mAh (123/427, 28.81%) or high-end 18,650- and 21,700 mAh–sized batteries (273/427, 63.93%), which are not included in the kits. On the other hand, most pod kits (282/348, 81.04%) have a battery capacity below 900 mAh (152/348, 43.68%) or between 900 mAh and 1500 mAh (130/348, 37.36%). This is probably because these products also typically have much higher concentrations of nicotine so they can still deliver high doses of nicotine with lower wattage and therefore do not need the
same battery power. Of the 50 vape pens, 21 (42%), 14 (28%),
and 12 (24%) have a battery capacity above 1500 mAh, between
900 mAh and 1500 mAh, and below 900 mAh, respectively.
Additionally, 90.63% (387/427) of mod kits have output wattage
between 40 W and 85 W (194/427, 45.43%) or above 85 W
(193/427, 45.2%), while the majority of pod kits (293/348,
84.2%) have output wattage less than 40 W (216/348, 62.07%) or
between 40 W and 85 W (77/348, 22.13%). Most vape pens
(21/50, 42%) have output wattage below 40 W. The majority
of the mod device–only products (209/229, 91.27%) are sold
without including batteries (18,650 or 21,700 mAh) in
web-based vape shops. More than half (154/229, 67.25%) of
the mod device–only products have output wattage greater than
85 W. These findings suggest that mod kits and mod
device–only products are distinct from pod kits and vape pens
in terms of battery capacity, output wattage, and volume size.

Table 3 illustrates the association between device prices and
product attributes, stratified by device types (mod kit, mod
device–only products, pod kits, and vape pens). We found that
the price of mod kits is not significantly associated with the
number of tanks or pods ($P_{.84}$) included or with the volume
size of the container ($P_{.07}$). Mod kits with advanced battery
sizes (18,650 or 21,700 mAh) are priced 21.5% higher than
those with less than 900 mAh batteries, even though the battery
is not included in the kit. Mod kits with output wattage over 85
W are priced 27.3% higher than those with less than 40-W
output. Similarly, the prices of mod device–only products are
higher when they have greater battery capacity or output wattage.

On average, the price of pod kits is 10% higher as the number
of tanks or pods increases. However, the price of pod kits is not
significantly associated with the volume size of tanks or pods
($P_{.60}$). Furthermore, the price of pod kits is positively
associated with battery capacity, with a capacity between 900
mAh and 1500 mAh and greater than 1500 mAh being 16%
and 23.9% higher, respectively, than those with less than 900
mAh battery. The price of pod kits with output wattage between
40 W and 85 W is 12.5% higher compared to those with output
wattage less than 40 W.

Unlike mod and pod devices that can be sold as part of a starter
kit or as individual devices, vape pens are rechargeable devices
sold exclusively as starter kits. The price of vape pens is
positively associated with volume sizes, with every 1-mL
increase in volume size associated with 3.7% higher prices. In
addition, compared to vape pens with a battery capacity below
900 mAh, those with a battery capacity above 1500 mAh are
priced 11.9% higher. Output wattage is not significantly
associated with the prices of vape pens.

The sensitivity analysis using a random effect model or
generalized estimating equations to account for brand effects
is reported in Multimedia Appendix 1. The results are very
similar to our main findings using ordinary least square
regressions.

Table 2. The distribution of product features based on product type from 5 popular, US-based, web-based vape shops from April to August 2022 (n=1062).

<table>
<thead>
<tr>
<th>Product features</th>
<th>Mod kits (n=427)</th>
<th>Pod kits (n=348)</th>
<th>Vape pens (n=50)</th>
<th>Mod device (n=229)</th>
<th>Pod device (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanks or pods, mean (SD)</td>
<td>1.103 (0.327)</td>
<td>1.325 (0.516)</td>
<td>1 (0.000)</td>
<td>N/Aa</td>
<td>N/A</td>
</tr>
<tr>
<td>Tank or pod volume size, mean (SD)</td>
<td>4.695 (1.968)</td>
<td>2.709 (1.162)</td>
<td>3.117 (2.102)b</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Battery capacity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capacity&lt;900 mAh</td>
<td>5 (1.2)</td>
<td>152 (43.7)</td>
<td>12 (24)</td>
<td>4 (1.8)</td>
<td>8 (100)</td>
</tr>
<tr>
<td>900 mAh≤capacity&lt;1500 mAh</td>
<td>26 (6.1)</td>
<td>130 (37.4)</td>
<td>14 (28)</td>
<td>3 (1.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Capacity≥1500 mAh</td>
<td>123 (28.8)</td>
<td>63 (18.1)</td>
<td>21 (42)</td>
<td>13 (5.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>Battery not included</td>
<td>273 (63.9)</td>
<td>N/A</td>
<td>N/A</td>
<td>209 (91.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Battery missing</td>
<td>N/A</td>
<td>3 (0.9)</td>
<td>3 (6)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Output wattage, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output&lt;40 W</td>
<td>28 (6.6)</td>
<td>216 (62.1)</td>
<td>21 (42)</td>
<td>7 (3.7)</td>
<td>4 (50)</td>
</tr>
<tr>
<td>40 W≤output&lt;85 W</td>
<td>194 (45.4)</td>
<td>77 (22.1)</td>
<td>10 (20)</td>
<td>54 (23.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Output≥85 W</td>
<td>193 (45.2)</td>
<td>2 (0.6)</td>
<td>1 (2)</td>
<td>154 (67.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Wattage output missing</td>
<td>12 (2.8)</td>
<td>53 (15.2)</td>
<td>18 (36)</td>
<td>4 (6.1)</td>
<td>4 (50)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

b9 out of 50 vape pens had missing volume size information.
Table 3. The association between product features and the prices of electronic cigarette devices stratified by device type from 5 popular, US-based, web-based vape shops from April to August 2022 (n=1045).

<table>
<thead>
<tr>
<th>Products and features</th>
<th>Coefficient</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mod kits (n=427)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of tanks or pods</td>
<td>0.008</td>
<td>.84</td>
</tr>
<tr>
<td>Tank or pod volume size</td>
<td>0.038</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Battery capacity: &lt;900 mAh as a comparison group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900 mAh&lt;capacity&lt;1500 mAh</td>
<td>0.117</td>
<td>.27</td>
</tr>
<tr>
<td>Capacity≥1500 mAh</td>
<td>0.139</td>
<td>.06</td>
</tr>
<tr>
<td>Battery not included</td>
<td>0.215$^{a,b}$</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Maximum output wattage: &lt;40 W as a comparison group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 W&lt;output&lt;85 W</td>
<td>0.049</td>
<td>.35</td>
</tr>
<tr>
<td>Output≥85 W</td>
<td>0.273$^{a,b}$</td>
<td>.02</td>
</tr>
<tr>
<td>Wattage output missing</td>
<td>−0.098</td>
<td>.55</td>
</tr>
<tr>
<td><strong>Mod device–only products (n=229)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Battery capacity: &lt;900 mAh as a comparison group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900 mAh&lt;capacity&lt;1500 mAh</td>
<td>0.503$^{a,b}$</td>
<td>.03</td>
</tr>
<tr>
<td>Capacity≥1500 mAh</td>
<td>1.034$^{a,b}$</td>
<td>.04</td>
</tr>
<tr>
<td>Battery not included</td>
<td>0.844$^{a,b}$</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Maximum output wattage: &lt;40 W as a comparison group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 W&lt;output&lt;85 W</td>
<td>0.455$^{b,c}$</td>
<td>.01</td>
</tr>
<tr>
<td>Output≥85 W</td>
<td>0.693$^{b,c}$</td>
<td>.003</td>
</tr>
<tr>
<td>Wattage output missing</td>
<td>0.532</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Pod kits (n=348)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of tanks or pods</td>
<td>0.100$^{a,b}$</td>
<td>.04</td>
</tr>
<tr>
<td>Tank or pod volume size</td>
<td>0.011</td>
<td>.60</td>
</tr>
<tr>
<td><strong>Battery capacity: &lt;900 mAh as a comparison group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900 mAh&lt;capacity&lt;1500 mAh</td>
<td>0.160$^{b,d}$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Capacity≥1500 mAh</td>
<td>0.239$^{b,c}$</td>
<td>.008</td>
</tr>
<tr>
<td>Battery missing$^e$</td>
<td>0.506$^{b,c}$</td>
<td>.002</td>
</tr>
<tr>
<td><strong>Maximum output wattage: &lt;40 W as a comparison group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 W&lt;output&lt;85 W</td>
<td>0.125$^{b,h}$</td>
<td>.05</td>
</tr>
<tr>
<td>Output≥85 W$^e$</td>
<td>−0.023</td>
<td>.67</td>
</tr>
<tr>
<td>Wattage output missing</td>
<td>0.046</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Vape pens (n=41)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vape pen volume size</td>
<td>0.037$^{b,d}$</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Battery capacity: &lt;900 mAh as a comparison group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900 mAh&lt;capacity&lt;1500 mAh</td>
<td>0.093</td>
<td>.14</td>
</tr>
<tr>
<td>Capacity≥1500 mAh</td>
<td>0.119$^{b,c}$</td>
<td>.008</td>
</tr>
<tr>
<td>Battery missing</td>
<td>0.765$^{b,d}$</td>
<td>.001</td>
</tr>
</tbody>
</table>

**Note:**
- $^a$: Indicates significance compared to the baseline category.
- $^b$, $^c$, $^d$, $^e$: Indicates significance compared to other categories within the same mod type.
Discussion

Using unique data gathered from 5 web-based stores that sell nationally, this study examines the pricing of rechargeable EC starter kits and devices. The findings reveal that the prices of mod devices and starter kits are on average US $59 to US $51, respectively, approximately twice as high as those of pod starter kits (US $25) and vape pens (US $30). Moreover, all rechargeable devices are much more expensive than disposable devices (cost about US $8) [28]. The average price of mod starter kits (US $51) from the 5 stores is similar to the price reported in a 2016 study (US $56) [20], which evaluated starter kit costs from 44 web-based vendors. Therefore, consumers considering using ECs, especially mod products, can expect relatively high initial costs. This supports previous findings suggesting that consumers tend to choose more affordable disposable options when first experimenting with ECs [29].

Furthermore, we discovered that the price distribution of both mods (starter kits and devices) and vape pens are positively skewed, suggesting that certain products are priced much higher than their average counterparts. This finding is consistent with prior studies based on web-based vendors and studies using the Standardized Tobacco Assessment for Retail Settings: Vape Shops surveillance tool to document prices that found that advanced mod products are priced much higher than regular mods [30,31].

The price analysis of EC devices further illustrates that greater battery capacity and output wattages are associated with higher device or kit prices. This is not surprising given that battery capacity and output wattage are key factors that determine nicotine delivery and user behaviors [21,32,33]. Studies have shown that tank ECs (eg, mod devices) can achieve much higher blood nicotine levels over a longer duration [34]. Survey data also suggest that smokers who successfully quit cigarette smoking using ECs are more likely to use tanks or mod devices [12,13]. It is possible that the higher output wattage and battery capacity, which ensure longer use before needing to be charged and reduce the risk of unexpected power outages, lead to more frequent e-cigarette use and may have assisted in transitions from smoking to vaping. However, greater battery capacity and output wattages could also attract youth and young adult users who report often trying or using multiple devices. Nonetheless, higher output wattage could also expose users to higher toxicant emissions and exposure to higher amounts of particulate matter, which may be harmful to human health [35,36]. The FDA and policy makers may need to take all of the factors (eg, product appeals in youth vs adult smokers) into consideration when setting product standards for batteries and volume sizes.

Our findings provide several key insights about designing EC pricing policies (eg, taxes) for devices. While there is growing literature that increasing EC taxes and prices reduces consumption, there is a lack of evidence that distinguishes between EC types and components, such as devices versus consumables such as e-liquid and cartridges. As a growing number of states start to tax ECs, not all EC-taxing states impose excise taxes on devices. Moreover, there is no clear guidance on how best to tax devices such as choosing tax rates and bases. Our findings suggest that EC starter kits or devices vary significantly by type, with mod prices being much higher than pod and vape pen prices. A unit-based specific tax therefore will impose a higher tax burden on lower-priced devices such as vape pens or pod systems and a lower tax burden on mod devices. A volume- or capacity-based specific tax on devices will impose a higher tax burden on vape pens with large container sizes. On the other hand, ad valorem taxes based on wholesale or retail prices will impose uniform tax burdens across all device types and consequently tax devices with higher battery capacities and output wattage at a higher rate. Therefore, policy makers could manipulate tax rates by device type to discourage the use of certain device products according to the health literature on the relative harms of ECs. For example, if cheap devices are preferred by smokers who are considering ECs, ad valorem tax may be preferred over specific taxes as the former imposed lower taxes on lower-priced products. In contrast, if the goal is to prevent youth from trying ECs and youth are more interested in cheap devices, a specific tax will be more favorable than ad valorem taxes in raising the prices of cheap devices.

In addition to taxation policies, our findings also highlight the importance of promotion restrictions. In web-based stores, the prices of mod starter kits are even lower than those of mod devices, suggesting mod starter kits are discounted. If mod products are mostly used by adult smokers to quit and the initial costs of ECs are a barrier to completely transitioning from cigarettes to ECs, such discounts should be allowed. However, if mods are found to attract youth and young adults, promotion restrictions may be needed to reduce their affordability.
Finally, we used data collected from 5 web-based stores that sell nationally. Although these data are not representative of the US web-based EC marketplace, they provide valuable information on device attributes and costs. It is also important to acknowledge that web-based stores or sales lead to challenges for regulations, including low prices and low compliance with state taxes [21,37]. A prior study further shows that there are international sites that sell ECs to the United States and these sites did not have age verification and detectable health warnings [38]. Future research is needed to understand how international markets in the web-based space may impact use behaviors, price minimization, and policy effectiveness.

There are some limitations of this study. First, we have very limited data on pod devices sold as stand-alone products and therefore do not have sufficient statistical power to conclude the pricing differences between pod starter kits and pod devices. Future studies are needed to address this gap. Second, we did not control for all the factors that affect prices because many factors are either not available or not measurable, such as production costs, consumer preferences, and so forth. Future studies may address this limitation. Nonetheless, we assessed all attributes that are presented on the web-based store web page, which arguably contains all the information that consumers see when they make purchasing decisions. Finally, the US FDA has approved a limited list of EC products. Many e-liquid products have or will become illegal for either failing to submit a premarket approval application or having their applications denied. Therefore, the demand for open-system devices could be significantly reduced as a result. However, given that many e-liquid products remain available in the marketplace and devices could still be used with 0-nicotine e-liquid, we consider monitoring device features continuing to be an important endeavor.

In summary, we provide the first assessment of how product features are associated with device or starter kit prices for the following distinct device types sold in the US digital market: mods, vape pens, and pods. The results can be used to design EC product standards and pricing policies by policy makers.

Acknowledgments
This study was funded by The Ohio State University Comprehensive Cancer Center for Tobacco Research Pilot Grant and by the National Cancer Institute of the National Institutes of Health and Food and Drug Administration U54CA287392 OSU Tobacco Center of Regulatory Science and Center for Tobacco Products at Rutgers U01CA278695. CS is funded by the National Cancer Institute (R21CA249757; Principal Investigator CS; 9/16/2021-8/31/2023). TW is funded by the Center Core Grants (P30) of the National Institute of Health and the National Cancer Institute (2P30CA016058-45). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health or the Food and Drug Administration.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
CS conceived of the idea. YH cleaned the data, conducted the analysis, and drafted the paper. QY drafted part of the paper. YA made plots. ZQ and JC scraped data from web-based vape shops. CS, SM, and TW reviewed the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Sensitivity analysis.

References


Abbreviations

EC: electronic cigarette
FDA: Food and Drug Administration
Decision Support for Managing Common Musculoskeletal Pain Disorders: Development of a Case-Based Reasoning Application

Fredrik Granviken1,2; Ottar Vasseljen1†; Kerstin Bach3; Amar Jaiswal3; Ingebrigt Meisingset1,4

1Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway
2Clinic of Rehabilitation, St Olavs Hospital, Trondheim, Norway
3Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
4Unit for Physiotherapy Services, Trondheim Municipality, Trondheim, Norway
†deceased

Corresponding Author:
Fredrik Granviken
Department of Public Health and Nursing
Norwegian University of Science and Technology
Postboks 8905
Trondheim, 7491
Norway
Phone: 47 93059497
Email: fredrik.granviken@ntnu.no

Abstract

Background: Common interventions for musculoskeletal pain disorders either lack evidence to support their use or have small to modest or short-term effects. Given the heterogeneity of patients with musculoskeletal pain disorders, treatment guidelines and systematic reviews have limited transferability to clinical practice. A problem-solving method in artificial intelligence, case-based reasoning (CBR), where new problems are solved based on experiences from past similar problems, might offer guidance in such situations.

Objective: This study aims to use CBR to build a decision support system for patients with musculoskeletal pain disorders seeking physiotherapy care. This study describes the development of the CBR system SupportPrim PT and demonstrates its ability to identify similar patients.

Methods: Data from physiotherapy patients in primary care in Norway were collected to build a case base for SupportPrim PT. We used the local-global principle in CBR to identify similar patients. The global similarity measures are attributes used to identify similar patients and consisted of prognostic attributes. They were weighted in terms of prognostic importance and choice of treatment, where the weighting represents the relevance of the different attributes. For the local similarity measures, the degree of similarity within each attribute was based on minimal clinically important differences and expert knowledge. The SupportPrim PT’s ability to identify similar patients was assessed by comparing the similarity scores of all patients in the case base with the scores on an established screening tool (the short form Örebro Musculoskeletal Pain Screening Questionnaire [ÖMSPQ]) and an outcome measure (the Musculoskeletal Health Questionnaire [MSK-HQ]) used in musculoskeletal pain. We also assessed the same in a more extensive case base.

Results: The original case base contained 105 patients with musculoskeletal pain (mean age 46, SD 15 years; 77/105, 73.3% women). The SupportPrim PT consisted of 29 weighted attributes with local similarities. When comparing the similarity scores for all patients in the case base, one at a time, with the ÖMSPQ and MSK-HQ, the most similar patients had a mean absolute difference from the query patient of 9.3 (95% CI 8.0–10.6) points on the ÖMSPQ and a mean absolute difference of 5.6 (95% CI 4.6–6.6) points on the MSK-HQ. For both ÖMSPQ and MSK-HQ, the absolute score difference increased as the rank of most similar patients decreased. Patients retrieved from a more extensive case base (N=486) had a higher mean similarity score and were slightly more similar to the query patients in ÖMSPQ and MSK-HQ compared with the original smaller case base.

Conclusions: This study describes the development of a CBR system, SupportPrim PT, for musculoskeletal pain in primary care. The SupportPrim PT identified similar patients according to an established screening tool and an outcome measure for patients with musculoskeletal pain.

(JMIR Form Res 2024;8:e44805) doi:10.2196/44805
**Introduction**

**Background**

Musculoskeletal pain conditions are the leading cause of disability and a major societal burden worldwide [1]. Common interventions for musculoskeletal pain either lack evidence to support their use or, at best, have modest or only short-term effects [2,3]. Treatment guidelines are based on randomized controlled trials considering effects on the group level with little or no consideration for the huge variation in patient stories and individual symptoms, even within more narrowly defined diagnostic entities, for example, low back pain. Thus, applying group-level evidence to individual patients and the relevance of one-size-fits-all treatment guidelines have been questioned [4]. In addition, the highly selected patients in most clinical trials do not match clinical settings where patients often present with comorbidities and large variations in symptoms and clinical history. Thus, clinicians are at unease with and often do not follow evidence-based guidelines [5]. Different attempts at subgrouping patients have been explored [6,7], but most attempts of subgrouping patients according to symptoms and clinical characteristics, and offering matched treatments (stratified care), have yet to demonstrate superior treatment outcomes [8].

It has been argued to focus less on diagnostic classification in musculoskeletal pain and more on prognostic factors to inform treatment decisions and improve treatment outcomes [9]. Factors influencing patients’ course and treatment outcomes are many, making decisions on the best treatment approach challenging for clinicians. In this situation, artificial intelligence (AI) may add decision support [10]. An intriguing AI method relevant to musculoskeletal pain disorders is case-based reasoning (CBR), where experiences from past problems and their solutions are used to solve new problems [11]. CBR may advance decision-making for musculoskeletal pain disorders and improve patient care and outcomes by providing information for tailoring the treatment. CBR has been used in fields such as the oil industry [12]; weather prediction [13]; and different aspects of health care [14], for example, kidney functioning in an intensive unit setting [15]; assessment and diagnosis of depression in palliative care [16], follow-up of patients who underwent stem cell transplantation [17], and diabetes management [18]. More recent studies have used CBR in diagnostics [19] and promoting self-management [20,21]. Yin et al [19] developed a CBR-based decision support system capable of differentiating between 2 types of probable primary headaches challenging for physicians in clinical care. The system was found to have high accuracy and differentiated probable migraine and probable tension-type headache much better than a guideline-based system. In a previous study from our research group, CBR was used to capture patient experiences and find the best treatment advice for patients by evaluating how to carry out a similarity-based retrieval [22]. This was further used for tailoring self-management support for patients with low back pain through a smartphone app. The app was provided as an adjunct to usual care and compared to usual care only in a randomized controlled trial. The patients in the intervention group reported a larger improvement in disability compared to those receiving usual care only [20]. The effect of the app has recently also been tested in a 3-armed randomized controlled trial among patients with neck and back pain in the specialist care. The authors reported no differences in effects among use of the app, a web-based nontailored self-management support tool, or usual care alone [21].

**Objective**

In this study, we used CBR to build a system for decision support in patients with musculoskeletal pain seeking physiotherapy care. This study describes the development of the system and demonstrates the system’s ability to identify similar patients.

**Methods**

The CBR Cycle Versus Physiotherapy Way of Solving a Problem

CBR has been described as a 4-step process, known as the CBR cycle: retrieve, reuse, revise, and retain (Figure 1) [11]. The most similar case or cases are retrieved from the collection of previous cases (stored in the case base), where a case is a set of data that represents a problem with its solution from the past. Knowledge of the CBR model (eg, adaptation rules) is applied to fit a new problem to an existing solution (reuse). The solution for the new case is tested for success and revised if necessary. The system learns as useful experiences of the new case are retained for future problem-solving, such that the case base is continuously updated with new or modified cases. Building and refining the collection of cases, the case base, is an important step in the CBR process [23]. The CBR methodology assumes that similar problems have similar solutions. Translated to medical terms, the problem is defined by a detailed description of the patient’s characteristics, signs, and symptoms, and the solution is defined by the treatment leading to a successful outcome. New patients are matched to previous similar patients (problems) with a successful outcome, and their treatment is used to inform treatment for the new patient (solution).
Figure 1. The case-based reasoning cycle, adapted from the study by Aamodt and Plaza [11].

An important reason for choosing CBR as the AI method of choice in this study (Figure 1) was its logic and resemblance with how physiotherapists approach new patients in clinical care (Figure 2). When a new patient consults a physiotherapist (Figure 2), the physiotherapist collects information about the patients’ symptoms, performs a clinical examination, and then tries to recall his experiences with similar patients from the past (ie, Retrieve in Figure 2). Knowledge and experience with previous similar patients with a successful outcome are used to guide treatment for the new patient (ie, Reuse in Figure 2), and the treatment is adapted and revised if necessary to fit the new patient. The physiotherapist gains experience with the new patient and may thus increase their knowledge of treatment leading to a successful outcome (ie, Retain in Figure 2). This process resembles the structure of the CBR cycle. The main difference in problem-solving between a physiotherapist and a CBR system is that the physiotherapist is limited by his memory and experiences, while a CBR system can use experiences from many different physiotherapists and thus use a much larger case base for decision support.
Development of the CBR System, SupportPrim PT, in Musculoskeletal Pain

In this study, we focus on the retrieval phase of CBR. We demonstrate the CBR system, SupportPrim PT, for musculoskeletal pain in two steps: (1) how similar patients were identified and (2) an evaluation of SupportPrim PT’s ability to identify similar patients. The system also displays solutions (ie, treatment suggestions) for new patients based on previous successful cases, but this part will only be described briefly. The medical community is the target audience for this study, and we have, therefore, used nontechnical language.

Patient Similarity Measures

The SupportPrim PT was built using myCBR (myCBR v3 and its rest API v2), which leverages patient data from the past to identify the most similar patients to advise management [24]. We used the local-global principle in CBR to identify similar patients, where similarity is calculated by a weighted sum function [25]. Global similarity measures are attributes used to identify similar patients, where an example of an attribute is a patient’s age or pain intensity. The weighting of these attributes (ie, global weighting) represents the relevance of the different attributes for the identification of similar patients, in our case, in terms of prognostic importance and choice of treatment, while local measures weight similarities between different values for the same attribute.

For the development of global weights, we first created a baseline CBR system by assigning equal weights to all the attributes, a second system with assigned weights based on a data-driven approach [49], and then a CBR system that used expert knowledge. We decided to use the expert knowledge approach to emphasize evidence of prognostic factors across different musculoskeletal conditions. We validated the weighting of the attributes in an iterative process using a sample of 14 patients representing 5 distinct phenotypes of musculoskeletal complaints [50], ranging from good to poor prognosis for a successful outcome [51]. The validation aimed to retrieve the most similar patient from the same phenotype as the queried patient, which is the new patient. In addition, we weighted attributes we believed were important for choosing adequate physiotherapy treatment higher (eg, mental distress, insomnia, and work ability; Multimedia Appendix 1).

For the local similarity measures, we decided the degree of similarity between values within each attribute from 0 (not similar at all) to 1 (full conformity between scores; Multimedia Appendix 1). To guide this work, we used knowledge about minimal clinically important difference, which is the smallest difference that is clinically important for the patient. For instance, this could be 2 points in the numeric rating scale for
pain intensity [52], which means that values within this range were regarded as completely similar. For attributes where information about minimal clinically important difference were lacking, we determined this by consensus within the study group. We did not always assume a linear relationship between scores on an attribute in the local similarity measure, where the local similarity for the same absolute difference in score could differ if the score was at the upper or lower end of the scale (eg, physical activity, where we defined full conformity—1.0 similar—between 5 and 6 to 7 days, while 0 versus 1 day were defined as only 0.4 similar).

**How to Find a Similar Patient**

The SupportPrim PT calculates a similarity score to find similar patients. A similarity score is the weighted sum of all the local similarity scores divided by the total possible weighting (Table 1), giving a similarity score between 0 and 1. Calculation of the similarity score between a query patient (Q) and the most similar patient in the case base (C) is shown in equation 1, earlier described by Bergman [25], where “w” is the weight of the attribute “i.” For each attribute “i,” the local similarity is defined as “sim_i(q, c),” where “q” is the value of the attribute for the query patient and “c” is the value of that respective attribute for the patient case from the case base. Finding the most similar patient is the result of the retrieval process where all patients are compared to the query patient, and the most similar patients are returned. The patient with the highest similarity score will be the most similar to the query patient.

**Equation 1** shows the calculation of the similarity score between a query patient (Q) and the most similar patient in the case base (C).

**Calculations (Table 1).** A similarity score is the weighted sum of all the local similarity scores divided by the total possible weighting (Table 1), giving a similarity score between 0 and 1. Calculation of the similarity score between a query patient (Q) and the most similar patient in the case base (C) is shown in equation 1, earlier described by Bergman [25].

**Table 1.** A similarity score is the weighted sum of all the local similarity scores divided by the total possible weighting (Table 1), giving a similarity score between 0 and 1. Calculation of the similarity score between a query patient (Q) and the most similar patient in the case base (C) is shown in equation 1, earlier described by Bergman [25].

### Table 1. Example of a calculation of similarity score showing the query patient with the 4 most similar patients in the case-based reasoning system SupportPrim PT for patients with musculoskeletal pain disorders.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Attributes&lt;sup&gt;a&lt;/sup&gt; (weight&lt;sup&gt;b&lt;/sup&gt;)</th>
<th>Score</th>
<th>Local sim&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Score</th>
<th>Local sim&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Score</th>
<th>Local sim&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Score</th>
<th>Local sim&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Score</th>
<th>Local sim&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Total score</th>
<th>Similarity score&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query patient</td>
<td></td>
<td>2</td>
<td>N/A&lt;sup&gt;e&lt;/sup&gt;</td>
<td>8</td>
<td>N/A</td>
<td>3</td>
<td>N/A</td>
<td>Moderate</td>
<td>N/A</td>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Patient A</td>
<td>1.8</td>
<td>0.8</td>
<td>7</td>
<td>1.0</td>
<td>3</td>
<td>1.0</td>
<td>Moderate</td>
<td>1.0</td>
<td>5</td>
<td>1.0</td>
<td>20.4</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Patient B</td>
<td>2.0</td>
<td>1.0</td>
<td>6</td>
<td>0.8</td>
<td>2</td>
<td>0.8</td>
<td>Great</td>
<td>0.6</td>
<td>7</td>
<td>1.0</td>
<td>19.2</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Patient C</td>
<td>1.5</td>
<td>0.6</td>
<td>9</td>
<td>1.0</td>
<td>4</td>
<td>0.8</td>
<td>Slight</td>
<td>0.8</td>
<td>3</td>
<td>0.6</td>
<td>16.0</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Patient D</td>
<td>2.2</td>
<td>0.8</td>
<td>4</td>
<td>0.3</td>
<td>5</td>
<td>0.6</td>
<td>Normal</td>
<td>0.4</td>
<td>4</td>
<td>0.8</td>
<td>13.6</td>
<td>0.62</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> The attributes used to identify similar patients.

<sup>b</sup> The relevance of the different attributes for the identification of similar patients, ranging from 1 to 8, where higher weights represent higher relevance.

<sup>c</sup> Total score Similarity score<sup>c</sup> calculated as: [((attribute 1 weight × attribute 1 local sim)+((attribute 2 weight × attribute 2 local sim))+...+(attribute n weight × attribute n local sim))/attribute 1 weight + attribute 2 weight +...+ attribute n weight] = Similarity score. Patient A: [(8×0.8)+(4×1.0)+(2×1.0)+(4×1.0)+(4×1.0)]/(8+4+2+4+4)=0.93.

<sup>d</sup> The degree of similarity between values within each attribute. Ranges from 0 to 1, where 0 means not similar at all and 1 means full conformity between scores.

<sup>e</sup> N/A: not applicable.

In Table 1, we show a query patient with its 4 most similar patients ranked according to the similarity score (patient A, B, C, and D). We exemplify the calculations of the similarity score by showing 5 of the 29 global measures (ie, attributes) used to represent cases in the CBR system. Patient A will be the most similar, B second most similar, C third most similar, and D the fourth most similar. Furthermore, which is beyond the scope of this study, important patient information and treatment description from the most similar patients having successful outcome will be displayed for the physiotherapist in a clinical dashboard.

To populate the case base for the SupportPrim PT, we systematically collected data from patients and physiotherapists in primary care of Norway. The case base consists of data on patient characteristics, prognostic factors, description of treatments, and outcomes from patients aged ≥18 years with musculoskeletal pain in any of these areas: shoulder, neck, upper or low back, hip, knee, or with complex pain as primary contact reason. Classification of complex pain was at the discretion of the treating physiotherapist based on a combination of the overall severity of symptoms, the number of pain sites, the clinical examination, and the patient history.

**System’s Ability to Identify Similar Patients**

To explore how the SupportPrim PT performs in finding similar patients, we assessed the similarity score of all patients in the case base, ranked from the most similar to the least similar patients, with the scores on the short form Örebro Musculoskeletal Pain Screening Questionnaire (ÖMSPQ) [53] and the Musculoskeletal Health Questionnaire (MSK-HQ) [54].

For similarity scores, each patient was compared with all other patients in the case base and ranked similarly for all patients to attain a rank order of most similar patients for all patients in the case base. Both questionnaires are used across different musculoskeletal pain conditions. ÖMSPQ is an established...
prognostic tool for long-term disability and failure of return to work, with a total score of 0 to 100, where higher scores indicate a worse long-term disability. The ÖMSPQ questionnaire emphasizes biopsychosocial variables related to future disability—similar to the global measures in the CBR system. The ÖMSPQ includes pain, self-perceived function, distress, return to work expectancy, and fear avoidance beliefs. The MSK-HQ is a generic musculoskeletal outcome measure that can be used for different musculoskeletal conditions. It contains 14 key items: severity of pain or stiffness, physical function or activity, work or daily activities, symptoms interference, independence, sleep, fatigue or low-energy levels, emotional well-being, understanding of condition and treatment, confidence in being able to manage symptoms, and overall symptom impact. The total score range is 0 to 56, with higher scores indicating better musculoskeletal health.

Assessing the similarity scores with the 2 established instruments was done with the case base used to build the SupportPrim PT (n=105) and then repeated in a larger case base (n=486). For the latter, we imported additional patients in the case base from another study to assess the performance of the similarity scores in a larger case base.

For patients with musculoskeletal problems, there is rarely 1 ideal solution. Different treatments could lead to a satisfactory outcome and, thus, work as solutions to the problem. In the final CBR system, after identifying similar patients using the local-global principle, similar previous successful patients were filtered, and the description of their treatment was displayed to inform treatment for the new patient (“solution”). The criterion for a successful outcome was a combination of pain intensity and function measured at baseline and at 3-month follow-up, where also the change score on MSK-HQ and the patient’s global perceived effect were included in the combined outcome measure. Details of our definition of a successful outcome are described in Multimedia Appendix 2 [53-56].

Ethical Considerations
The Regional Committee for Medical and Health Research Ethics in mid-Norway approved the study (51566/2019 and 49308/2020). All patients provided written informed consent to participate in the study. Patients did not receive any compensation for participating. Study data are deidentified, and no identification of individual participants is possible.

Results

Descriptive Characteristics of Patients in the Case Base
The original case base consisted of 105 patients, with complete data gathered from 22 physiotherapists in primary care collected from January 2020 to January 2021. The patients’ mean age was 46 (SD 15) years, the majority were women (77/105, 73.3%), and pain duration was >3 months for most patients (87/105, 82.9%; Table 2).
Table 2. Characteristics of patients with musculoskeletal pain disorders included in the case-based reasoning system SupportPrim PT (N=105).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>77 (73.3)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>46.0 (15.2)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>26.9 (6.1)</td>
</tr>
<tr>
<td>Higher educationa, n (%)</td>
<td>59 (56.2)</td>
</tr>
<tr>
<td>Pain durationb, n (%)</td>
<td>87 (82.9)</td>
</tr>
<tr>
<td>Current smoker, n (%)</td>
<td>9 (8.6)</td>
</tr>
<tr>
<td>Pain intensity, mean (SD)</td>
<td>4.7 (2.1)</td>
</tr>
<tr>
<td>PSFSc, mean (SD)</td>
<td>4.0 (2.7)</td>
</tr>
<tr>
<td>Work ability, mean (SD)</td>
<td>6.2 (2.9)</td>
</tr>
<tr>
<td>Musculoskeletal risk groupd (n=101), n (%)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>42 (41.6)</td>
</tr>
<tr>
<td>Medium</td>
<td>47 (46.5)</td>
</tr>
<tr>
<td>High</td>
<td>12 (11.9)</td>
</tr>
<tr>
<td>Main body pain region, n (%)</td>
<td></td>
</tr>
<tr>
<td>Neck</td>
<td>16 (15.2)</td>
</tr>
<tr>
<td>Shoulder</td>
<td>18 (17.1)</td>
</tr>
<tr>
<td>Back</td>
<td>19 (18.1)</td>
</tr>
<tr>
<td>Hip</td>
<td>16 (15.2)</td>
</tr>
<tr>
<td>Knee</td>
<td>13 (12.4)</td>
</tr>
<tr>
<td>Complex</td>
<td>23 (21.9)</td>
</tr>
<tr>
<td>MSK-HQé, mean (SD)</td>
<td>37.7 (8.5)</td>
</tr>
<tr>
<td>ÖMSPQf, mean (SD)</td>
<td>43.1 (15.6)</td>
</tr>
</tbody>
</table>

*a*Education above high school.  
*b*Pain duration >3 months.  
*c*PSFS: Patient Specific Functional Scale; a higher value indicates better function.  
*d*Musculoskeletal risk groups: 0-4 is low risk, 5-8 is medium risk, and 9-12 is high risk.  
*e*MSK-HQ: Musculoskeletal Health Questionnaire; a higher value indicates better musculoskeletal health.  
*f*ÖMSPQ: Örebro Musculoskeletal Pain Screening Questionnaire; higher scores indicate worse long-term disability.

### The System’s Ability to Identify Similar Patients

The SupportPrim PT built to use for decision support in patients with musculoskeletal pain consisted of 29 weighted attributes, each having a defined local similarity measure to identify similar patients. To demonstrate the system’s ability to identify the most similar patients, all patients in the case base were compared to each other, that is, each patient was queried against the rest of the patients, and this was repeated for all patients. Rank 1 thus represents the average similarity score for the most similar patient (“best match”) to the query patient for all patients in the case base. The ranks from the most similar to least similar were then plotted against each rank’s absolute difference on the ÖMSPQ (Figure 3) and MSK-HQ (Figure 4) scores. The most similar patients had a mean absolute difference from the query patient of 9.3 (95% CI 8.0-10.6) points on the ÖMSPQ and a mean absolute difference of 5.6 (95% CI 4.6-6.6) points on the MSK-HQ. For both ÖMSPQ and MSK-HQ, the absolute score difference increased as the rank of most similar patients decreased (Figures 3 and 4).
To assess the performance of the similarity scores in a larger case base, we imported additional patients into the case base, resulting in a case base of 486 patients. When we compare the mean similarity score for the most similar patients for the 2 case bases, we see that the patients retrieved from the larger case base had a slightly higher mean similarity score (Figure 5). We compared the ranks of most similar patients for all patients in both case bases with the ÖMSPQ (Figure 6 and Table 3) and MSK-HQ (Figure 7 and Table 4). The results showed that the larger case base identified slightly more similar patients with a smaller absolute mean difference on ÖMSPQ and MSK-HQ for the ranks compared to the original case base.
**Figure 5.** The mean similarity score of the 15 most similar patients with musculoskeletal pain disorders in the 2 case bases in the case-based reasoning system SupportPrim PT.

**Figure 6.** The absolute mean difference in the short form Örebro Musculoskeletal Pain Screening Questionnaire (ÖMSPQ) between queried patients with musculoskeletal pain disorders and most similar patients in ranked order for the 2 different size case bases in the case-based reasoning system SupportPrim PT.
Table 3. The absolute mean difference in the short Örebro Musculoskeletal Pain Screening Questionnaire (score range 0-100) between queried patients with musculoskeletal pain disorders and the 3 most similar patients for the 2 different case bases in the case-based reasoning system SupportPrim PT.

<table>
<thead>
<tr>
<th>Case base with 105 patients, absolute mean difference (95% CI)</th>
<th>Case base with 486 patients, absolute mean difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most similar patient</td>
<td>9.3 (8.0-10.6)</td>
</tr>
<tr>
<td>Second most similar patient</td>
<td>11.2 (9.5-13.0)</td>
</tr>
<tr>
<td>Third most similar patient</td>
<td>10.2 (8.6-11.7)</td>
</tr>
</tbody>
</table>

Figure 7. The absolute mean difference in the Musculoskeletal Health Questionnaire (MSK-HQ) between the queried patients with musculoskeletal pain disorders and the most similar patients in ranked order for the 2 different size case bases in the case-based reasoning system SupportPrim PT.

Table 4. The absolute mean difference in the Musculoskeletal Health Questionnaire (score range 0-56) between queried patients with musculoskeletal pain disorders and the 3 most similar patients for the 2 different case bases in the case-based reasoning system SupportPrim PT.

<table>
<thead>
<tr>
<th>Case base with 105 patients, absolute mean difference (95% CI)</th>
<th>Case base with 486 patients, absolute mean difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most similar patient</td>
<td>5.6 (4.6-6.6)</td>
</tr>
<tr>
<td>Second most similar patient</td>
<td>7.1 (6.0-8.2)</td>
</tr>
<tr>
<td>Third most similar patient</td>
<td>6.8 (5.8-7.9)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

This study describes the development of the CBR system, SupportPrim PT, using the local-global principle to identify similar patients with musculoskeletal pain disorders. The SupportPrim PT successfully identified similar patients.

When comparing the similarity scores from the SupportPrim PT and their rankings with ÖMSPQ and MSK-HQ scores, we found that the SupportPrim PT successfully identified the most similar patients to the queried patients across the musculoskeletal pain conditions (Figures 3 and 4). The mean score differences on the questionnaires between the queried and the most similar patients increased linearly or curvilinearly with increasing rank order (less similar patients). These results are not unexpected as SupportPrim PT contains 29 mainly prognostic attributes partly overlapping with items on the ÖMSPQ and MSK-HQ. Mean differences between queried patients and best matches of 9.3 points on ÖMSPQ and 5.6 points on the MSK-HQ are also expected. The SupportPrim PT uses a larger number of attributes for patient similarity.
comparisons, possibly making the system more comprehensive in mapping patients’ symptoms and prognostic factors than the shorter comparative questionnaires. Using a larger case base, the SupportPrim PT yielded patients with slightly higher mean similarity scores and lower absolute mean difference between the queried patients and the most similar patients on the ÖMSPQ and MSK-HQ. This may indicate that a larger case base may improve the performance of the system [57]. The case base of 105 patients is nevertheless representative of primary physiotherapy care in Norway, with descriptive data being consistent with data from a large longitudinal observational study [58].

Key challenges when developing a CBR system are definition of the case representation, similarity measure development, and retrieval strategy. Attribute selection and weighting could be dependent on expert knowledge and limited by the number, type, and quality of the attributes included. We selected attributes based on their prognostic value or potential to influence treatment choice. There is good documentation for prognostic factors being similar across musculoskeletal diagnostic groups [42-46]. Classifying patients according to similar prognostic factors rather than diagnosis may be more fruitful in improving care [9]. To base interventions on diagnoses is relevant if a causal pathway between diagnosis and choice of treatment is established. However, a clear understanding of causal pathways is often lacking in musculoskeletal pain complaints. Thus, physiotherapy interventions are commonly directed toward symptom alleviation, for example, advice, reassurance, self-management, exercise, and manual therapy. Therefore, patients may best be treated within a prognostic framework with more emphasis on specific prognostic factors on the individual level [43].

Definition of global weights in CBR is challenging. We used expert knowledge instead of data-driven methods [49]. Among the attributes selected, the sum score of Hopkins Symptom Checklist 10-item, with questions about anxiety, depression, and somatization, was weighted highest, as emotional distress is an important mediating factor for the treatment effect [59] and also potentially modifiable by physiotherapy interventions [60]. Decisions were based on consensus between the authors with different backgrounds and extensive experience both from research and clinical work.

Limitations
Instead of attribute selection based on expert knowledge, automated data-driven attribute selection methods [61] could have been used. We acknowledge that other attributes not included in our system could have improved the process of identifying similar patients. A limitation of the study is that the process of assigning weights of the global measures did not follow a formal consensus method. This could have resulted in different weighting of the attributes. In addition, comparative studies of data-driven and expert-driven approaches to decide the weighting of different attributes should be explored in future work.

Conclusions
Advising treatment for new patients using previous similar patients with successful outcome represents a move toward more individualized treatment rather than relying on the best evidence of average effects in clinical trials. It is important to acknowledge that the SupportPrim PT is not built to replace the clinical expertise and experience of the therapist but to work as a decision support. Furthermore, AI might create an uncomfortable situation for clinicians and patients, not having complete control and being uncertain, not understanding what is in the system, not having the possibility to tell which attributes are used in the model, and thus not trusting the system [10]. We believe CBR can address this uncertainty by being an easy-to-understand and explainable AI method [62], with expert and domain knowledge being an integrated part of the system, which increases the likelihood for clinicians to trust it. This study describes the development of a CBR system, SupportPrim PT, for musculoskeletal pain in primary care. It demonstrates the system’s ability to identify similar patients on an established screening tool and an outcome measure used for patients with musculoskeletal pain disorders. The SupportPrim PT will be integrated into a clinical decision support system and tested in a full-scale randomized controlled trial in primary health care to evaluate its effectiveness among physiotherapists and their patients. The SupportPrim PT was developed for decision support for physiotherapists in managing patients with musculoskeletal pain disorders, but we think such an explainable system could be applicable to other health care personnel for patients where decision support is needed.

Acknowledgments
The authors are grateful to Anita Formo Bones for administration and support, Paola Marin Veites for technical help, Danielle van der Windt for inputs to the manuscript, and all physiotherapists and patients for participating in the data collection. The project was funded by The Norwegian Research Council and The Norwegian Fund for Post-Graduate Training in Physiotherapy.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
All authors contributed to the design of the study. KB and AJ had the main responsibility for the technical parts. FG, IM, and OV completed the data collection and analysis. FG drafted the manuscript. All authors discussed the results, revised the manuscript, and approved the final version.
Conflicts of Interest

None declared.

Multimedia Appendix 1
Attributes in the case-based reasoning system, SupportPrim PT.

Multimedia Appendix 2
Definition of a successful outcome in the case-based reasoning system, SupportPrim PT.

References


Abbreviations

AI: artificial intelligence
CBR: case-based reasoning
MSK-HQ: Musculoskeletal Health Questionnaire
ÖMSPQ: Örebro Musculoskeletal Pain Screening Questionnaire
Medical Students’ Perceptions on Identifying and Addressing Emotional Responses in Emergency Medicine: Pilot Investigation

Anish Kumar Agarwal1,2, MD, MPH, MS; Rachel Gonzales1,2, MPH; Cory Munden1, MD; DaCarla Albright3, MD; Suzana Tsao1, DO

1Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States
2Center for Health Care Transformation and Innovation, Penn Medicine, Philadelphia, PA, United States
3Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Anish Kumar Agarwal, MD, MPH, MS
Department of Emergency Medicine
University of Pennsylvania
423 Guardian Drive
410 Blockley Hall
Philadelphia, PA, 19104
United States
Phone: 1 215 573 6784
Email: anish.agarwal@pennmedicine.upenn.edu

Abstract

Background: Training in acute care, such as emergency medicine (EM), where exposure to critically ill and injured patients is high, impacts the well-being of trainees and contributes to burnout. Investigating how, and if, trainees prepare for these situations is necessary to ensure they are supported.

Objective: This study aimed to evaluate medical students’ perspectives and emotional preparedness for handling acute care and trauma.

Methods: We conducted a pilot investigation using a remote digital survey of medical students during their EM clerkship at a large, urban academic institution. The primary outcome of interest was student-reported preparedness and comfort in handling trauma and critical care patient encounters. Secondary outcomes included awareness of well-being resources and comfort in accessing digital well-being resources.

Results: A total of 57 medical students completed the voluntary digital survey, and half of the students (n=28, 49%) reported having witnessed the care of a critically ill or a penetrating trauma patient (eg, a victim of gun violence). A majority (n=40, 70%) had thought about how these events may impact them, and over half felt unprepared to identify the emotional impact these cases may have on them (n=31, 54%) or address the emotional or mental health impact (n=36, 63%). Less than a quarter (n=14, 25%) were aware of digital mental health resources, and 58% (n=33) did not feel fully comfortable connecting with resources if needed. Students who had previously witnessed critical care were significantly more likely to report feeling well prepared in identifying the emotional impact and addressing this impact.

Conclusions: In this cross-sectional survey, students did not feel fully prepared to identify or address the emotional impact of working in EM. Additionally, they lacked awareness of or comfort with accessing digital institutional resources meant to support their well-being, such as a large web-based platform. These findings can help inform and guide interventions by educational and academic leaders. The aim would be to create and promote environments that empower students with tools to identify their own emotions and connect to well-being resources.

(JMIR Form Res 2024;8:e50827) doi:10.2196/50827

KEYWORDS

well-being; burnout; medical education; coping; student; students; university; college; acute care; trauma; traumatic; emotion; emotional; stress; distress; psychological; cross-sectional; survey; surveys; critical; critically; perception; perspectives; prepared; preparedness
Introduction

Health care–associated burnout persists in medicine and can be identified early in medical training [1-4]. This syndrome has negative impacts at the individual level, and it also affects patient care and the health care system by contributing to more medical errors, lower patient satisfaction, reduced productivity, poor clinical teaching and role modeling for trainees, increased cost, and increased physician attrition and ultimately contributes to physician suicide [5-8]. Physicians have a higher risk for suicide than the general population, and these mental health risks have been labeled as known “occupational hazards” [9,10]. Notably, the rate of US physicians annually completing suicide is estimated to be equivalent to the number of students in 3 graduating medical school classes [11,12].

Studies have demonstrated that the impacts of burnout and mental health symptoms, such as depression and anxiety, begin early in medical education and are found in medical students, residents, and physicians in training [2,13-15]. Studies have also found that rates of depression are higher for medical students than other trainees, suggesting this group may be particularly in need of interventions to support well-being and prevent burnout [11,14]. The strain and emotional toll of working in health care emerges early in training. For students, investigating how one identifies, processes, and copes with feelings of anxiety, sadness, depression, or stress related to their clinical experiences within clerkships remains understudied [16]. Emergency medicine (EM) physicians consistently report some of the highest rates of burnout, with EM often being referred to as the “center” of burnout [6,17]. EM physicians are 3 times more likely to be burned out compared to non-EM physicians [4].

Investing in strategies to help trainees identify their own emotions related to providing care, how to cope, and how to sustain well-being is critical for the future of the workforce [18]. Rather than awaiting burnout to evolve, examining how prepared these students feel is necessary to ensure that health systems and medical schools adequately support and proactively maintain medical students' well-being. This is especially important given the rise of mental health symptoms and burnout in health care within the backdrop of the COVID-19 pandemic and social unrest related to racial injustice and rising gun violence.

The goal of this study was to assess medical students’ preparation during their EM rotation to understand how students self-identify their capacity to deal with emotionally charged clinical settings (eg, critical care cases or trauma).

Methods

Ethical Considerations

This pilot investigation used an electronic voluntary survey administered to second- and third-year medical students at the University of Pennsylvania in Philadelphia during their EM clerkship at a large, urban academic institution. The study was approved by the University of Pennsylvania Institutional Review Board (849318). All research methods, consent, and activities were performed in accordance with the university guidelines and regulations.

Eligible Participants and Study Type

Inclusion criteria consisted of students in their second (preclerkship or M2) or third (clerkship or M3) year of medical school. This study was cross-sectional, and a voluntary response sample was used. There were no specific exclusion criteria, as other students were not invited to the survey.

Recruitment Procedures

Participants were invited via email, completed informed consent, and were not compensated for participating. Data were collected and aggregated for analysis. The students at this medical school are routinely surveyed, and questions from this study were incorporated into the preexisting and ongoing school surveying. In total, the survey was sent to 161 students in their preclerkship (M2) and 153 students in their clerkship (M3) during the final week of classes in December 2021.

Approach and Analysis

This cross-sectional pilot study was developed by the research team, with expertise in medical education (authors ST and DA), qualitative methods (AKA and RG), and clinician well-being (AKA). No previous instrument, to the knowledge of the study team, exists; thus, the instrument was developed and pilot-tested in this study. All answers were anonymous, and no demographic information was collected. The primary outcome of interest was student-reported preparedness and comfort in handling trauma and critical care patient encounters. Secondary outcomes included awareness of well-being resources available to them during their clerkship, feelings of preparedness, and comfort in accessing well-being resources. Comparisons were done using chi-square tests in Stata IC 16.1 (StataCorp), and a P<.05 was considered statistically significant.

Results

A total of 57 medical students completed the voluntary survey; 26 (46%) of them were M2 students, and 31 (54%) were M3 students. Almost half (n=28, 49%) of the students reported having witnessed the care of a critically ill or injured patient (defined as a victim of gun violence). Most (n=40, 70%) students had thought about how these events may impact them, but most did not feel fully prepared to identify the emotional impact these cases may have on them (n=31, 55%) or prepared to address this emotional or mental health impact (n=36, 63%). Although resources are widely available to support students’ well-being at this institution, only 25% (n=14) were aware of these institutional resources to help them cope with the emotions involved with care, and 58% (n=33) did not feel fully comfortable connecting with resources if needed (Table 1).

Differences were identified between those students who had witnessed the care of a critically ill or injured patient and those who had not (Table 2). Students who had witnessed such care were more likely to feel well prepared in identifying the emotional impact of these cases (n=7, 25% vs n=0, 0%; P=.007) and in addressing this impact (n=9, 32% vs n=0, 0%; P=.001). No significant differences were found in student awareness of...
resources and their comfort in connecting with these resources to cope with the emotions involved with care.

Table 1. Medical students’ perspectives on the emotional impact and preparedness of caring for critical patients.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Values (N=57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you witnessed the immediate care of a critically ill or injured patient?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>28 (49)</td>
</tr>
<tr>
<td>No</td>
<td>29 (51)</td>
</tr>
<tr>
<td>Have you thought about how these events may impact you?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>40 (70)</td>
</tr>
<tr>
<td>No</td>
<td>15 (26)</td>
</tr>
<tr>
<td>N/A</td>
<td>2 (4)</td>
</tr>
<tr>
<td>How prepared do you feel in identifying the emotional impact these cases may have?</td>
<td></td>
</tr>
<tr>
<td>Well prepared</td>
<td>7 (12)</td>
</tr>
<tr>
<td>Prepared</td>
<td>18 (32)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>23 (40)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>8 (14)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (2)</td>
</tr>
<tr>
<td>How prepared do you feel in addressing with the emotional impact these cases may have?</td>
<td></td>
</tr>
<tr>
<td>Well prepared</td>
<td>9 (16)</td>
</tr>
<tr>
<td>Prepared</td>
<td>11 (19)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>25 (44)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>11 (19)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Are you aware of university resources to support learners as they cope with the emotions involved with care?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14 (25)</td>
</tr>
<tr>
<td>No</td>
<td>42 (74)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (2)</td>
</tr>
<tr>
<td>How comfortable you would feel in connecting with resources if needed?</td>
<td></td>
</tr>
<tr>
<td>Very comfortable</td>
<td>7 (12)</td>
</tr>
<tr>
<td>Comfortable</td>
<td>16 (28)</td>
</tr>
<tr>
<td>Somewhat comfortable</td>
<td>23 (40)</td>
</tr>
<tr>
<td>Not comfortable</td>
<td>10 (18)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (1.8)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Table 2. Medical students’ perspectives on personal emotional reaction, preparation, and coping skills based on prior exposure.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Ever witnessed the care of a critically ill or injured patient (N=57)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=28)</td>
<td>No (n=29)</td>
</tr>
<tr>
<td>Have you thought about how these events may impact you?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>22 (78.6)</td>
<td>18 (62)</td>
</tr>
<tr>
<td>No</td>
<td>6 (21.4)</td>
<td>9 (31)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>How prepared do you feel in identifying the emotional impact these cases may have?</td>
<td></td>
<td>.007</td>
</tr>
<tr>
<td>Well prepared</td>
<td>7 (25.0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Prepared</td>
<td>10 (35.7)</td>
<td>8 (28)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>10 (35.7)</td>
<td>13 (45)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>1 (3.6)</td>
<td>7 (24)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>How prepared do you feel in addressing with the emotional impact these cases may have?</td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Well prepared</td>
<td>9 (32.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Prepared</td>
<td>6 (21.4)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>12 (42.9)</td>
<td>13 (45)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>1 (3.6)</td>
<td>10 (35)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Are you aware of university resources to support learners as they cope with the emotions involved with care?</td>
<td></td>
<td>.22</td>
</tr>
<tr>
<td>Yes</td>
<td>9 (32.1)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>No</td>
<td>19 (67.9)</td>
<td>23 (79)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>How comfortable you would feel in connecting with resources if needed?</td>
<td></td>
<td>.63</td>
</tr>
<tr>
<td>Very comfortable</td>
<td>5 (17.9)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Comfortable</td>
<td>8 (28.5)</td>
<td>8 (28)</td>
</tr>
<tr>
<td>Somewhat comfortable</td>
<td>11 (39.3)</td>
<td>12 (41)</td>
</tr>
<tr>
<td>Not comfortable</td>
<td>4 (14.3)</td>
<td>6 (21)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
<td>1 (4)</td>
</tr>
</tbody>
</table>

*N/A: not applicable.

Discussion

Principal Findings

Given EM’s high rates of burnout and the vulnerable role medical students hold as trainees, medical students undergoing their EM clerkship are at high risk of emotional strain and stress [15]. The rising focus on physician mental health extends to those in training [8]. There is a gap in understanding how prepared medical students feel in identifying and addressing their emotional response and the impact experiences in EM may have upon them. This study assessed medical students’ preparation to better understand their needs and guide interventions toward key priority areas of focus.

This study found that, regardless of whether students had already witnessed the care of a critically ill patient, most students did not feel prepared to identify or address the emotional impact associated with these situations. This highlights the need to train students early and normalize the emotional impact of working in medicine [19,20]. The stigma associated with mental health in health care has pervaded the classroom and hospitals, and to build structures to prevent burnout, we must begin early to help trainees identify feelings of anxiety, depression, or stress as they experience them [13,21-23]. A proactive approach would provide students with the tools and resources they need to adequately identify their emotions and connect to appropriate resources when needed.

It is essential for institutions not only to have resources available to support students’ well-being and mental health but also to make these resources readily accessible and easy to navigate. Our findings reveal that even in an environment where resources are present, students may be unaware of these resources or may not be comfortable accessing them. The University of Pennsylvania School of Medicine has a robust infrastructure within the medical school, a separate web-based mental health...
and well-being platform, accessible to the entire health system community [24]. However, students in this study remained unaware of their availability or accessibility. Institutions must work to incorporate these resources into the clinical and teaching environments to reduce the stigma that may prevent students from accessing them.

Finally, the significant differences in feelings of preparedness between those students who had and those who had not witnessed the care of critical care patients suggest that students do not feel prepared to identify the emotional impact such experience may have until it has happened. It is important to act proactively to prepare students to experience the care of critical patients. EM provides an ideal environment to do so, as the likelihood that students will be placed in an emotionally charged setting is high [17,25]. Venues such as critical care, pediatrics, obstetrics and gynecology, as well as surgery provide other opportune areas for schools to think about deploying focused interventions to where students may need them most.

Limitations
This study has some limitations. First, to protect confidentiality and privacy, we did not collect demographic information. This prevented us from analyzing how these findings might differ by age, race, or ethnicity. Second, as medical students self-selected to participate, selection bias may play a role. The participants surveyed here may not accurately represent the experiences of all medical students in the clerkship program. Additionally, this study was performed at a single urban, academic program and may not be applicable to all medical students at other various institutions. We are also limited by a sample of 56 students, which may further limit our ability to generalize these findings.

Conclusions
Similar to other roles in health care, medical students do not feel fully prepared to identify or address the emotional impact of working in acute care. Additionally, they lack awareness of or comfort with accessing institutional resources designed to support their well-being. Medical students who have not witnessed the care of a critically ill or injured patient were more likely to feel unprepared in identifying or addressing the emotional impact such an event might have on them. These findings can help inform and guide interventions by educational and academic leaders. The aim would be to create and promote training environments that empower students with tools to identify their own emotions and connect to well-being resources. We need to normalize the conversation around mental health in the health care workforce and reduce stigma early in medicine, beginning with our medical students.

Acknowledgments
We would like to acknowledge the students and staff of the University of Pennsylvania Department of Emergency Medicine.

Data Availability
The data sets generated and analyzed during the study are not publicly available due to the sensitivity of data content in responses, but deidentified data are available by email request to the corresponding author upon reasonable request.

Authors' Contributions
AKA planned, executed, and led this study. RG planned data collection and completed analysis. CM drafted and revised the manuscript. DA executed the study and revised the manuscript. ST planned and supervised the study. All authors contributed to the draft, revisions, and manuscript preparation.

Conflicts of Interest
None declared.

References

https://formative.jmir.org/2024/1/e50827


Abbreviations

EM: emergency medicine
©Anish Kumar Agarwal, Rachel Gonzales, Cory Munden, DaCarla Albright, Suzana Tsao. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Factors Associated With Perception of Stigma Among Parents of Children With Cleft Lip and Palate: Cross-Sectional Study

Yanan Zhang¹, MD; Xinwen Zhang¹, BS; Jinzhuo Jiang², MD; Wanhua Xie³, PhD; Daoman Xiang¹, PhD

¹Ophthalmology and otorhinolaryngology Department, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangdong Provincial Clinical Research Center for Child Health, Guangzhou, China
²College of Engineering, South China Agricultural University, Guangzhou, China
³Outpatient Department, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangdong Provincial Clinical Research Center for Child Health, Guangzhou, China

Corresponding Author:
Wanhua Xie, PhD
Outpatient Department
Guangzhou Women and Children's Medical Center, Guangzhou Medical University
Guangdong Provincial Clinical Research Center for Child Health
9 Jinsui Road
Guangzhou, 510623
China
Phone: 1 13725370379
Fax: 1 2038076020
Email: xiewanhua1@126.com

Abstract

Background: Parents of children with cleft lip with or without cleft palate (CL/P) often face stigmatization, which has a significant impact on their quality of life and mental health. However, to date, there is a lack of comprehensive, multicenter empirical research on parents of children with CL/P in China, particularly those with large-scale samples.

Objective: This study aimed to identify major factors that contribute to the perception of stigma experienced by parents of children with CL/P.

Methods: A cross-sectional survey was conducted. A total of 104 parents of children diagnosed with CL/P in 2 hospitals were selected by convenience sampling. Demographics and disease information, the Chinese Perception of Stigma Questionnaire, the Center for Epidemiological Studies Depression Scale, and the Social Anxiety Scale were used in this study. Descriptive statistics, t-tests, and one-way ANOVA were used to compare the differences between participants’ demographic information and perception of stigma. Multivariable linear regression was performed to assess associations between demographic factors, social anxiety, depression, and perception of stigma.

Results: The mean scores for the dimensions of perception of stigma, depression, and social anxiety were 22.97 (SD 9.21), 38.34 (SD 8.25), and 22.86 (SD 6.69), respectively. Depression and social anxiety were positively associated with discrimination, while surgery status was a negatively associated variable. Parents with a college education or higher had significantly lower levels of perceived stigma compared to parents with a junior high school education (all P values <.05). These 4 factors explained 40.4% of the total model variance (F₈=9.726; P<.001; R²=0.450; adjusted R²=0.404).

Conclusions: Our findings highlight a concerning trend of diminished quality of life among parents of children with CL/P. Factors such as parents’ education level, surgery status, depression, and social anxiety are shown to influence the level of stigma experienced. Implementing comprehensive nursing care and providing presurgical support are effective strategies for alleviating parents’ social anxiety, reducing perceived stigma, and preventing depression.

(JMIR Form Res 2024;8:e53353) doi:10.2196/53353

KEYWORDS
stigma; social anxiety; depression; parents of children with cleft lip and palate; cleft lip; cleft palate; cross-sectional study
Introduction

Cleft lip with or without palate (CL/P) is a common congenital developmental malformation affecting the oral and maxillofacial region [1], with a global prevalence of about 1.62%–1.82% [2], and the highest prevalence in Asian countries [3]. There are 35,000 babies born with CL/P in China each year [4]. Palate repair surgery is usually done when babies are 10 to 24 months old, while lip surgery is generally performed at the age of 6 months [3]. The management of CL/P is long-term, beginning from birth and continuing into the late teenage years. Children born with CL/P not only face a visible facial disfigurement but also have other challenges related to the cleft, such as feeding difficulties, speech impediments, hearing issues, and psychological problems, which may compromise their ability to communicate effectively [5].

Children with CL/P are vulnerable to external discrimination because of the malformation. A child with CL/P is subjected to bullying, rejection, and social isolation, sometimes even from their own family [6]. The discrimination can have devastating health and well-being impacts, not only on children affected but also on people associated with the discriminated person. Having a baby with CL/P has variable psychological impacts on the parents [7]. Parents are often concerned about whether their children will be accepted as they grow up, and parents’ psychological state directly affects their children’s perception of the condition [6,8,9]. Some studies [10,11] have indicated that the perception of stigma experienced by the mother and children is consistent. Numerous studies [2,12,13] have shown that children with CL/P experience significant stigma, and limited engagement in social activities suggests that the parents of these children may experience high levels of stigma. Furthermore, Parents are troubled by their children’s conditions from birth to adulthood, while children often experience discrimination after their cognitive development has matured. Therefore, exploring ways to prevent these patients and their parents from feeling stigmatized is urgently necessary.

Currently, research on the perception of stigma among parents of children with CL/P is predominantly qualitative, emphasizing the exploration of factors contributing to parental stigma. However, there is limited reporting on assessments regarding the level and variation in the perception of stigma. In particular, there is no research on the perception of stigma in children with CL/P in China. Previous studies [11,14] have shown that superstition is the main reason for parents’ perception of stigma, and only 40% of parents believe that genetic problems cause CL/P. One study [11] found that 72% of mothers felt ashamed of having children with CL/P, and some even abandoned their babies. Therefore, it is crucial to examine the perception of stigma among parents of children with CL/P, pinpoint those experiencing heightened stigma levels, implement tailored nursing interventions based on sensitive factors, and mitigate the likelihood of public incidents. This study aims to investigate experiences of stigma in parents of children with CL/P and analyze the contributing factors.

Methods

Study Design and Participants

This cross-sectional study was conducted using a convenience sample of parents of children with CL/P, recruited from 2 hospitals in southern and western China (from the cities of Guizhou and Guangzhou) between June 2020 and January 2023. All parents of children who met the inclusion and exclusion criteria for surgical treatment in the Department of Stomatology were invited to complete a questionnaire. The inclusion criteria were as follows: parents of children (aged <15 years) with CL/P undergoing nasoalveolar molding therapy at the Department of Stomatology. These parents were invited to complete an anonymous questionnaire.

Exclusion criteria included parents of infants with additional birth defects or medical conditions, those with genetic diagnoses (eg, Down syndrome or Trisomy 21 syndrome), and individuals unable to complete the questionnaires due to critical illness or mental disorders.

Data Collection

After the doctors introduced the operational procedure or the nurses introduced the environment of the wards, the participants were invited to fill in the questionnaires. Uniform instructions were used to explain the study aims and their relevance to the participants. The participants were informed that the study was anonymous and voluntary and that they could quit anytime.

Variables and Instruments

The demographic data were collected via a questionnaire. Three questionnaires with good reliability and validity were used to collect the data on stigma, social anxiety, and depression.

Demographics Questionnaire

The demographic information questionnaire was designed by researchers after consulting the literature and mainly included parents’ age, role (father or mother), education level, and working status as well as children’s age, sex, type of CL/P, surgery status, and number of surgeries.

Disability Discrimination Perception Scale

Stigma was assessed using the Disability Discrimination Scale (DDPS), initially developed by Liu and Shen [15]. The questionnaire comprises 10 items (eg, “When participating in activities, I feel that people around me would not talk to me and avoid me”), designed to explore the perception of stigma among individuals with disabilities in their daily lives. It is also suitable for children aged 10 to 16 years. Respondents rated each item on a 5-point scale, ranging from 1 (very inconsistent) to 5 (completely consistent). Higher total scores indicate a greater perception of stigma, with scores ranging from 10 to 50. The scale demonstrates good internal consistency, as evidenced by a Cronbach’s reliability of 0.89. The criterion validity was 0.47, indicating good validity [15].

The Center for Epidemiological Studies Depression Scale

The level of depression was measured by the Center for Epidemiological Studies Depression Scale (CES-D) [16]. CES-D
consists of 20 items (eg, “I was bothered by things that usually don’t bother me”), and each item is scored from 0, indicating “rarely or none of the time (less than 1 day)” to 3, indicating “most or all of the time (5-7 days).” The total score of all items ranges from 12 to 60, with higher scores indicating a greater likelihood of depression; a total score of 15 points or less indicates no depression; 16 to 19 points indicate possible depression; and a score of 20 points or higher indicates definite depression [16]. The scale is suitable for screening people with depressive symptoms and can also be used to assess the severity of depressive symptoms. The Cronbach $\alpha$ of CES-D was 0.87, indicating good internal consistency. The split-half reliability was 0.85, the test-retest reliability was 0.70, and the criterion validity was 0.75. Therefore, CES-D had good reliability and validity.

**The Interaction Anxiousness Scale**

The Interaction Anxiousness Scale (IAS) was used to assess the subjective propensity to experience social anxiety independent of behavior [17]. The IAS contains 15 self-reported items (eg, “I want to be confident in social situations more”), which are answered on a 5-point scale, ranging from 1 (very inconsistent) to 5 (completely consistent). Total scores range from 15 to 75, with higher scores indicating higher levels of social concern. The Cronbach $\alpha$ reliability of the scale was 0.87, the test-retest reliability was 0.80, and the criterion validity was 0.48. In this study, according to the actual situation, 7 items related to sexual life were deleted, and 8 items were selected for evaluation. The reliability of the scale was retested, and Cronbach $\alpha$ reliability was 0.68, indicating acceptable reliability.

**Statistical Analysis**

IBM SPSS Statistics (version 25.0; IBM Corp) for Windows, was used for statistical analysis. Standard descriptive statistical values (means and SDs) were calculated. The independent samples 2-tailed $t$ test and ANOVA were used for the intergroup comparisons of parameter changes. Pearson analysis was used to examine the relationship between perception of stigma, social anxiety, and depression. Multiple linear regression was performed to assess the association between demographic factors, anxiety, depression, and perception of stigma. The significance level was set at 5%.

**Ethical Considerations**

The study was approved by the Ethics Committee for Clinical Studies at Guangzhou Women and Children's Medical Center, Guangzhou Medical University (NO.2022268A01). This study obtained informed consent from participants, who had the option to withdraw at any time. The research data were collected anonymously.

**Results**

In this study, a total of 110 questionnaires were distributed. Of the collected questionnaires, 104 were complete, resulting in a 94.6% response rate.

**Participant Characteristics**

Most participants were mothers aged 25-30 years and unemployed, with a high school diploma. Their children had CL/P, and surgery had not been performed. Unemployed parents constituted 71% (n=74) of the sample, with nearly half (n=14, 47%) of the employed parents being self-employed. Over half (n=61, 59%) of the respondents had completed high school. The children’s average age was 3.06 (SD 2.02) years. A total of 73 (72%) children had not undergone surgery, while most of those who had done the surgery (n=20, 65%) had undergone the procedure twice. A descriptive analysis of the general demographics, work-related characteristics, and disease-related information of the participants is shown in Table 1.
Table 1. Demographic characteristics of the participants (N=104).

<table>
<thead>
<tr>
<th>Variable and category</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>30 (29)</td>
</tr>
<tr>
<td>25-30</td>
<td>42 (40)</td>
</tr>
<tr>
<td>30-35</td>
<td>26 (25)</td>
</tr>
<tr>
<td>&gt;35</td>
<td>6 (6)</td>
</tr>
<tr>
<td><strong>Role</strong></td>
<td></td>
</tr>
<tr>
<td>Father</td>
<td>39 (38)</td>
</tr>
<tr>
<td>Mother</td>
<td>65 (63)</td>
</tr>
<tr>
<td><strong>Educational background</strong></td>
<td></td>
</tr>
<tr>
<td>Primary school or lower</td>
<td>6 (6)</td>
</tr>
<tr>
<td>Junior high school</td>
<td>23 (22)</td>
</tr>
<tr>
<td>Senior high school</td>
<td>61 (59)</td>
</tr>
<tr>
<td>College</td>
<td>7 (7)</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>7 (7)</td>
</tr>
<tr>
<td><strong>Working status</strong></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>74 (71)</td>
</tr>
<tr>
<td>Working</td>
<td>30 (29)</td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>74 (71)</td>
</tr>
<tr>
<td>Independent management</td>
<td>14 (14)</td>
</tr>
<tr>
<td>Enterprise unit</td>
<td>9 (9)</td>
</tr>
<tr>
<td>Public service unit</td>
<td>7 (7)</td>
</tr>
<tr>
<td><strong>Children's sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>60 (58)</td>
</tr>
<tr>
<td>Female</td>
<td>44 (43)</td>
</tr>
<tr>
<td><strong>Condition type</strong></td>
<td></td>
</tr>
<tr>
<td>Cleft lip</td>
<td>36 (35)</td>
</tr>
<tr>
<td>Cleft lip and palate</td>
<td>68 (65)</td>
</tr>
<tr>
<td><strong>Number of surgeries</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>73 (70)</td>
</tr>
<tr>
<td>1</td>
<td>7 (7)</td>
</tr>
<tr>
<td>2</td>
<td>20 (19)</td>
</tr>
<tr>
<td>≥3</td>
<td>4 (4)</td>
</tr>
<tr>
<td><strong>Status of surgery</strong></td>
<td></td>
</tr>
<tr>
<td>After surgery</td>
<td>73 (70)</td>
</tr>
<tr>
<td>Before surgery</td>
<td>31 (30)</td>
</tr>
</tbody>
</table>

Prevalence of Stigma, Social Anxiety, and Depression
The mean scores for the dimensions of perception of stigma, depression, and social anxiety were 22.97 (SD 9.21), 38.34 (SD 8.25), and 22.86 (SD 6.69), respectively. Additionally, 38% (n=40) of the participants exhibited a perception of stigma level beyond the moderate level, and every participant experienced evident depression, with 72% (n=75) displaying average levels of social anxiety.

Univariate Analyses of Factors Associated With Stigma
The demographic data, including parents’ age, role, education level, and working status as well as the children’s age, sex, CL/P type, surgery status, and the number of surgeries, were analyzed
using a one-way ANOVA and t tests. The results of the Kolmogorov-Smirnov test suggested that the continuous variables were reasonably and normally distributed. The t tests revealed that the parents whose children had undergone surgery experienced lower stigma levels compared to those whose children had not undergone surgery ($P=.01$). The perception of stigma scores differed significantly among parents having different educational backgrounds ($P<.001$) and between those whose children underwent different numbers of surgeries ($P=.04$). The specific results are detailed in Table 2.

Table 2. Univariate analyses of factors associated with stigma (N=104).

<table>
<thead>
<tr>
<th>Variable and category</th>
<th>Mean (SD)</th>
<th>$F$ test (df)</th>
<th>$t$ test (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td>1.58 (3)</td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>18-25</td>
<td>25.74 (9.93)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-30</td>
<td>21.14 (8.46)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-35</td>
<td>22.24 (8.30)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;35</td>
<td>23.50 (12.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Role</strong></td>
<td></td>
<td></td>
<td>0.32 (102)</td>
<td>.75</td>
</tr>
<tr>
<td>Father</td>
<td>23.28 (9.62)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>22.69 (8.96)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education background</strong></td>
<td></td>
<td>3.62 (4)</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Primary school or lower</td>
<td>22.50 (9.89)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior high school</td>
<td>24.52 (8.94)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior high school</td>
<td>24.21 (9.34)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>16.14 (2.73)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>13.43 (1.81)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Working status</strong></td>
<td></td>
<td></td>
<td>2.02 (102)</td>
<td>.06</td>
</tr>
<tr>
<td>Unemployed</td>
<td>24.00 (8.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working</td>
<td>20.23 (9.56)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
<td>1.36 (3)</td>
<td></td>
<td>.26</td>
</tr>
<tr>
<td>Unemployed</td>
<td>24.00 (8.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent management</td>
<td>19.43 (8.49)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterprise unit</td>
<td>21.89 (10.95)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public service unit</td>
<td>19.71 (10.95)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Children’s sex</strong></td>
<td></td>
<td></td>
<td>0.83 (102)</td>
<td>.41</td>
</tr>
<tr>
<td>Male</td>
<td>23.55 (9.47)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>22.05 (8.78)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Condition type</strong></td>
<td></td>
<td></td>
<td>0.34 (102)</td>
<td>.74</td>
</tr>
<tr>
<td>Cleft lip</td>
<td>23.33 (8.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleft lip and palate</td>
<td>22.69 (9.36)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of surgeries</strong></td>
<td></td>
<td>2.94 (3)</td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>0</td>
<td>24.34 (9.49)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17.57 (7.28)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18.95 (7.28)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>26.00 (7.07)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Status of surgery</strong></td>
<td></td>
<td></td>
<td>2.50 (102)</td>
<td>.01</td>
</tr>
<tr>
<td>Before surgery</td>
<td>24.34 (9.49)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After surgery</td>
<td>19.55 (7.47)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aNot applicable.
Correlation Analysis With Perception of Stigma, Social Anxiety, and Depression

The results showed that the perception of stigma was related to social anxiety ($r=0.54; P<.001$) and depression ($r=0.39; P<.001$); there was also a correlation between depression and social anxiety ($r=0.30; P=.002$). The specific results are shown in Table 3.

Table 3. Correlation analysis (Pearson r) with stigma, social anxiety, and depression. At $P=.01$ (2-tailed), the correlation was significant.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Stigma</th>
<th>Depression</th>
<th>Social anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>1</td>
<td>0.39</td>
<td>0.54</td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Depression</td>
<td>0.39</td>
<td>1</td>
<td>0.30</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>Social anxiety</td>
<td>0.54</td>
<td>0.30</td>
<td>1</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.002</td>
<td>—</td>
</tr>
</tbody>
</table>

*Not applicable.

Regression Analyses of Stigma

With the perception of stigma as the dependent variable, depression, social anxiety, and demographic data with differences between groups were included in the model as independent variables, and multiple linear regression analyses were performed. In the perception of stigma model, depression, social anxiety, educational background, and status of surgery were significant correlates explaining 40.4% of the total model variance ($F_8=9.726; P<.001$; $R^2=0.450$; adjusted $R^2=0.404$).

Table 4. Regression analyses of stigma.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Unstandardized coefficients (B)</th>
<th>SE</th>
<th>$t$ test (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant value</td>
<td>0.10</td>
<td>4.36</td>
<td>0.02 (95)</td>
<td>.98</td>
</tr>
<tr>
<td>Status of surgery</td>
<td>−5.85</td>
<td>2.22</td>
<td>−2.64 (95)</td>
<td>.01</td>
</tr>
<tr>
<td>Number of operations</td>
<td>2.08</td>
<td>1.14</td>
<td>1.82 (95)</td>
<td>.07</td>
</tr>
<tr>
<td>Depression</td>
<td>0.26</td>
<td>0.09</td>
<td>2.93 (95)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Social anxiety</td>
<td>0.52</td>
<td>0.12</td>
<td>4.48 (95)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Primary school or lower</td>
<td>0.23</td>
<td>3.30</td>
<td>0.07 (95)</td>
<td>.95</td>
</tr>
<tr>
<td>Senior high school</td>
<td>−0.29</td>
<td>1.75</td>
<td>−0.16 (95)</td>
<td>.87</td>
</tr>
<tr>
<td>College education</td>
<td>−6.53</td>
<td>3.09</td>
<td>−2.11 (95)</td>
<td>.04</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>−7.21</td>
<td>3.15</td>
<td>−2.29 (95)</td>
<td>.02</td>
</tr>
</tbody>
</table>

Discussion

This study found that parents of children with CL/P experienced a high level of perception of stigma and obvious social anxiety and depression. In addition, parents’ education level, children’s status of surgery, social anxiety, and depression affect parents’ levels of perceived stigma. This study complements the gap in the study of stigma perception in families of children with CL/P in China. Parents in our study exhibited a higher level of stigma perception. This could be attributed to the fact that most children underwent corrective surgery during infancy, making parents more attuned to discrimination, as they believe that visual impairments affect their children’s social lives. Thus, prevailing data suggest that Chinese clinical nurses should pay attention to the quality of life of parents of children with CL/P.

Our study showed that children’s sex and the type of CL/P could not affect parents’ discrimination perception. This is consistent with the findings of previous studies that reported no significant difference in the level of discrimination perception between patients of different sexes and CL/P types [10]. The results of our study showed no significant difference in the level of stigma perception between fathers and mothers. Fathers also had a high
level of stigma perception. However, previous studies [14,18] suggested that mothers of children with CL/P experienced severe stigma or discrimination while ignoring the possibility that fathers also experienced high levels of discrimination. Two other studies [19,20] pointed out that the main sources of stress among fathers were the treatment process, feeding, and social stigma. Therefore, in the care process of children with CL/P, we should not only focus on the mother’s psychological state but also consider the psychological state of the father. The importance of family nursing should also be emphasized.

Parents of children with CL/P experience obvious depression and anxiety. It is consistent with the findings of previous studies reporting that parents of newborns with CL/P had significantly higher levels of anxiety compared to control parents [20]. The study findings are consistent with the current literature, as depression or depressed mood in parents of children with CL/P is a common phenomenon [21,22]. Kumar [23] reported the highest incidence of depressive episodes, with 42% of parents of children with CL/P aged >10 years showing a strongly or very strongly elevated depressive disorders screening index. Sommer [24] has pointed out that mothers of children with CL/P believe their children have a negative impact on them. Their negative views and emotions are mainly derived from the long treatment times, the difficulty of treatment, the uncertainty of the infant’s future condition, and the fear of rejection. Family members of infants with CL/P experience stigma, anxiety, and worry, which may lead to individual negative emotions and coping styles, produce adverse intergroup relations, reduce subjective well-being and life satisfaction, and ultimately affect the quality of family life. Relieving the stigma perceived by parents of children with CL/P is the first step to relieving negative emotions like anxiety and depression in these parents. Therefore, we should pay attention to the physical rehabilitation of children as well as the mental health of children and their parents.

In this study, parents with junior high school education showed a high level of stigma perception, and the perception of stigma was lower among parents with a college education or higher. So far, no study has investigated the relationship between the education level of parents who have children with CL/P and their stigma perception. A previous study [6] showed that caregivers of children with attention deficit hyperactivity disorder, who have higher education levels, experience higher levels of stigma. However, the results of this study were inconsistent with a previous study, possibly because parents with a junior high school education have less knowledge of the condition and are more likely to focus on the appearance of their children. The difference could potentially be attributed to the fact that parents with higher education levels are more likely to access relevant information about the condition and feel more confident about the later recovery process. Hence, health care professionals should focus on parents with a high school education or lower, monitor variations in their perceived stigma levels, comprehend their psychological well-being across different stages of their children’s lives, and offer tailored psychological support.

Cleft surgery provides hope to those children and their parents, offering the possibility of re-enrollment in schools, employment opportunities, social acceptance, and improved prospects for marriage [6]. Undergoing surgery directly affects the level of the stigma perceived by parents. Parents of children without surgery perceive a higher level of stigma, which is consistent with previous studies [14] suggesting that parents of children with CL/P show obvious negative emotions in the first 3 months after the birth of their children without surgery. Therefore, the initial 3 months emerge as a pivotal phase for health care professionals to alleviate parental anxiety and deliver emotional and educational assistance to parents of infants with CL/P. Through comprehensive nursing and presurgical support, mothers can enhance their early feeding capabilities, expedite the bonding process with their infants, and facilitate the infants’ swift recovery. Familiarity with CL/P beforehand serves to diminish maternal stress during initial interactions, mitigating anxiety and perceptions of discrimination. Consequently, encouraging mothers to engage with CL/P-related charitable organizations or fostering interaction among children with CL/P in shared spaces can effectively reduce anxiety and perception of stigma levels.

This study has several limitations. All the participants in this study were family members of children who were willing to undergo surgery, and no family members of children who were unwilling to undergo treatment were included, which may have led to selective bias. Parents filled out the scale based on recall, which may introduce information bias. The sample size of this study was small, as it was limited to hospitals only from 2 provinces with a large gap in economic levels between the 2 places, and fewer fathers were included in the study. Future studies should increase the sample size and broaden the scope of the investigation.

Our findings highlight a concerning trend of diminished quality of life among parents of children with CL/P. Factors such as parents’ education level, the status of surgery in children, depression, and social anxiety are shown to influence the level of stigma experienced by these parents. Implementing comprehensive nursing care and providing presurgical support are effective strategies for alleviating parents’ social anxiety, reducing perceived stigma, and preventing depression.

Acknowledgments

The authors are sincerely grateful to all the participants who responded to the questionnaires and all the nurses and doctors for supporting this investigation.

This research was funded by the Guangzhou Municipal Science and Technology Bureau (grant 2023A03J0913 and 202103000088).
Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References


Abbreviations

CES-D: Center for Epidemiological Studies Depression Scale
CL/P: cleft lip with or without cleft palate
DDPS: Disability Discrimination Scale
IAS: Interaction Anxiousness Scale

©Yanan Zhang, Xinwen Zhang, Jinzhuo Jiang, Wanhua Xie, Daoman Xiang. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Health Behavior Change Intervention Preferences Expressed by American Indian Cancer Survivors From a Southwest Tribal Community: Semistructured Interview Study

Jennifer Erdrich1, MD, MPH, MFA; Felina M Cordova-Marks2, MPH, MSc, DrPH; William O Carson2, MPH; Jennifer W Bea2, PhD; William R Montfort3, PhD; Cynthia A Thomson2, RD, PhD

1Department of Surgery, College of Medicine, University of Arizona, Tucson, AZ, United States
2Department of Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States
3Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Jennifer Erdrich, MD, MPH, MFA
Department of Surgery
College of Medicine
University of Arizona
1501 North Campbell Avenue
Tucson, AZ, 85724-5018
United States
Phone: 1 520 626 2635
Email: jerdrich@surger.arizona.edu

Abstract

Background: While many factors, including social determinants of health, affect cancer mortality, one modifiable risk factor that may contribute to cancer disparities is obesity. The prevalence of obesity in the American Indian/Alaska Native population is 48.1% per the Centers for Disease Control and Prevention. The overall cancer mortality for the American Indian/Alaska Native population is 18% higher than the White population as reported by the American Cancer Society. Interventions tailored to American Indian/Alaska Native communities that promote healthy lifestyle behaviors after cancer diagnosis and prior to cancer surgery (prehab) might improve cancer outcomes for this population.

Objective: The aim of the study is to characterize the lifestyle behaviors of San Carlos Apache cancer survivors and identify preferences for the adaption of a prehab intervention.

Methods: Semistructured interviews and validated questionnaires were completed with San Carlos Apache cancer survivors (N=4), exploring their viewpoints on healthy lifestyle and cancer risk and preferences for program development. A thematic content analysis was conducted.

Results: Participants had an average BMI of 31 kg/m² and walked 53 minutes daily. The majority of participants reported a high willingness to change eating habits (n=3, 75%). All 4 reported willingness to participate in a diet and exercise program. Important themes and subthemes were identified: (1) cancer is perceived as a serious health condition in the community (N=4, 100%); (2) environmental exposures are perceived as cancer-causing threats (n=3, 75%); (3) healthy diet, exercise, and avoiding harmful substances are perceived as mitigating cancer risk (n=3, 75%); (4) barriers to healthy habits include distance to affordable groceries (n=3, 75%) and lack of transportation (n=2, 50%); (5) there is high interest in a prehab program geared toward patients with cancer (N=4, 100%); and (6) standard monitoring practiced in published prehab programs showed early acceptability with participants (N=4, 100%).

Conclusions: Collaboration with tribal partners provided important insight that can help inform the adaptation of a culturally appropriate prehab program for San Carlos Apache patients diagnosed with cancer.

(JMIR Form Res 2024;8:e51669) doi:10.2196/51669

KEYWORDS

Native American cancer disparities; diet; physical activity; prehabilitation; native; exercise; fitness; interviews; thematic analysis; lifestyle; Apache

https://formative.jmir.org/2024/1/e51669

JMIR Form Res 2024 | vol. 8 | e51669 | p.2161
(page number not for citation purposes)
Introduction

Cancer is the second leading cause of death for the American Indian/Alaska Native population with related disparities in cancer screening, incidence, stage at diagnosis, and timely treatment [1-8]. Consequences of colonialism such as geographic isolation, poverty, and social determinants of health have not only created steep disadvantages in the progression of cancer but also contributed to detrimental behaviors related to diet and exercise that place tribal members at higher risk for worse cancer outcomes [9]. There are 13 obesity-related cancer subtypes that account for 40% of all cancers diagnosed in the United States, which confers significant risk for the American Indian/Alaska Native population, given the 48.1% prevalence of obesity for this population [10,11]. Published data suggest that brief exercise interventions following diagnosis of solid tumor malignancies might lower cancer progression, specifically through the regulation of metabolism and the immune response [12-14]. Epidemiologic studies have suggested anti-inflammatory foods can improve cancer outcomes [15-18]. While anti-inflammatory diet and exercise trials are increasing for patients with cancer in the general population, few have involved patients from American Indian/Alaska Native populations [19].

The Partnership for Native American Cancer Prevention, established in 2002, is a program that strives to alleviate the unequal burden of cancer among Native Americans in the Southwest through research, training, and community outreach [20]. Through the Partnership for Native American Cancer Prevention, the investigators of this project partnered with the San Carlos Apache Healthcare Corporation (SCAHC) with approval from the Tribal Council. The San Carlos Apache reservation, established in 1871, spans 1.8 million acres of Sonoran Desert and alpine forests in Southeastern Arizona and is home to 10,815 people. The tribal government and SCAHC are committed to providing quality care that honors Apache traditions and promotes Apache values of well-being, goodness, balance, harmony, and beauty [21-23].

Lifestyle modifications in nutrition and exercise prior to oncologic treatment are key strategies of prehabilitation (prehab) and have been shown to optimize patients’ functional capacity and improve postoperative outcomes [24-26]. Results from this work are being used to propose the next evolution of this project, which is the adaptation of a culturally appropriate prehab program for San Carlos Apache patients diagnosed with cancer who are preparing for surgery. It was recognized that San Carlos Apache community members, specifically those who have experienced the cancer spectrum from diagnosis to survivorship would have the best insight for developing a program specifically tailored to the San Carlos Apache community. The purpose of this project was to conduct semistructured interviews and administer validated questionnaires to characterize lifestyle behaviors of San Carlos Apache cancer survivors and identify appropriate approaches for the adaption and delivery of a lifestyle intervention for San Carlos Apache patients diagnosed with cancer.

Methods

Study Design

The principles of community-based participatory research were followed throughout the study. SCAHC and Tribal Council members advised the project activities, which were conducted with their approval and collaboration. Qualitative semistructured interview questions were developed through an iterative process supervised by faculty experts in qualitative methods. The questions were then presented to tribal community members at SCAHC to ensure the content was appropriate. The questions reflected the content of published prehab programs and focused on understanding the most acceptable approach for program adaptation for San Carlos Apache patients. The questions were organized according to four areas: (1) knowledge of cancer and its causes and prevention, (2) perspectives on cancer and lifestyle, (3) diet and physical activity preferences, and (4) attitudes and beliefs regarding diet and physical activity interventions for survivors. In total, 32 semistructured interview questions were asked of participants, including potential subprompts. Interview data were complemented with facilitator administration of 2 validated questionnaires about nutrition and physical activity. These were the Rapid Eating Assessment for Participants-Shortened Version (REAPS) and the International Physical Activity Questionnaire (IPAQ), which were selected because of the validity, reliability, literacy, and low participant burden of these brief instruments [27-30].

Interviews were conducted by the principal investigator through password-protected Zoom (Zoom Video Communications). The participants accessed the platform through telephone dial-in. The camera feature was not used, names were avoided, and data were deidentified. The participants’ responses were audio-recorded with permission and transcribed. The interviewer took handwritten notes of the responses to cross-check the transcription and verify accurate audio recording. Minor errors were corrected on the final transcriptions.

An independent reviewer with expertise in qualitative methods and American Indian/Alaska Native health was consulted to perform a thematic content analysis of the deidentified transcripts using a matrix system grounded in a framework focused on determinants of health from an Indigenous perspective. This method has been used for analyzing qualitative data with other tribes and is described elsewhere [31]. Briefly, the matrix was used to organize the questions and responses into a coded summary chart, wherein similar responses were grouped together after an iterative process of grouping, regrouping, and refining. From this, the content was grouped into themes and subthemes. Direct quotes particularly illustrative of key themes were extracted. The reviewer summarized the principal findings for intervention planning.

Participants and Recruitment

Tribal members who are cancer survivors residing on the San Carlos Apache reservation, aged at least 18 years, and diagnosed with cancer within the previous 15 years were recruited for semistructured interviews. Recruitment efforts were centered on the recommendations of the social workers and physicians at SCAHC. The social workers at SCAHC engage all patients...
diagnosed with cancer, not only at diagnosis but also during survivorship, per their facility’s routine. This was recognized to position the social workers in contact with the highest volume of potential participants, who were approached on a rolling basis over the course of 3 months. Furthermore, social worker engagement was intended to minimize influence since the social workers are neither part of the university study team nor the clinical care team. The social workers introduced the study to the survivors, and for those who expressed interest in the study, the social workers verified that prospective participants gave permission to share their contact information. Each prospective participant underwent a brief screening by the study lead (JE) to verify eligibility following a telephone screening script. If individuals agreed to participate, they were consented and then scheduled for a semistructured interview on a separate day.

Ethical Considerations
This project was developed in collaboration with the San Carlos Apache Tribal Council and SCAHC. Human Subjects approval was obtained from the University of Arizona Institutional Review Board (study ID CR00001226), SCAHC Institutional Review Board, and San Carlos Apache Tribal Council Health Committee. All materials were approved by the Tribal Council and SCAHC. The tribal support letter was submitted to the University of Arizona. SCAHC Institutional Review Board and Tribal Council approved the content of this paper and its submission. Informed consent was obtained from all participants (verbal or written option, all participants chose verbal and were sent a hard copy of the written consent form). Upon completion, participants received a gift card as compensation for their time. Privacy and confidentiality were upheld in the data collection, and the data were deidentified.

Results

Demographics and Questionnaires on Physical Activity and Diet Behaviors
A total of 4 participants were interviewed from May to July 2022. SCAHC referred 12 potential participants; 8 were successfully contacted and 4 agreed to study participation (response rate of 4 of 12 potential participants or 33%). Interviews lasted 60-80 minutes. The average age of participants was 64 years, with 1 male patient and 3 female patients. The participants were diagnosed with cancer between 2015 and 2021. A family history of cancer was reported by 3 (75%) participants. All were overweight or obese with an average BMI of 31 kg/m². There were 2 (50%) participants who reported past smoking, and 1 (25%) who reported past moderate alcohol consumption. Based on the IPAQ, the participants reported an average of 158 minutes of vigorous activity daily, 90 minutes of moderate activity daily, and 225 minutes of sitting daily. All patients reported walking specifically as part of their physical activity routine and reported an average of 53 minutes of daily walking. Based on the REAPS, it was elicited that 2 (50%) participants ate 4 or more meals per week from restaurants or takeout, 3 (75%) ate 2 or more servings of fruit per day, 3 (75%) ate 2 or more servings of vegetables per day, 3 (75%) answered they usually or sometimes use processed meats, 2 (50%) regularly ate fried foods, 3 (75%) rarely ate sweets, and 2 (50%) reported drinking 16 ounces or more of nondiet soda or punch per day. On a scale of 1-5, for willingness to make changes in their eating habits where 1=very willing and 5=not at all willing, 3 (75%) participants answered positive willingness with a score of 1-2.

Semistructured Interviews: Qualitative Analysis
Overview
The following themes were revealed (Textbox 1): (1) cancer as a serious health condition in the community, (2) environmental exposure and contamination as a main cause of cancer, (3) personal choices as risk factors or prevention, (4) barriers, (5) acceptability of diet and physical activity programs, and (6) acceptability of monitoring and biospecimen collection.
Textbox 1. Qualitative semistructured interview themes and subthemes.

- Cancer as a serious health condition in the community
- Environmental contamination and exposure as a cause of cancer
  - Occupational exposure
  - Chemical in the community
  - Water contamination
- Personal choices as risk factors or prevention
  - The power of food, exercise, and lifestyle
- Barriers
  - Distance and transportation
  - The built environment
- Acceptability of diet and physical activity program
  - Considerations for program development
    - Support and encouragement
    - Central location
    - Education
- Acceptability of monitoring and biospecimen collection

Cancer as a Serious Health Condition in the Community
All participants affirmatively stated that cancer is a problem for the San Carlos Apache community. Participants defined cancer as a serious condition with the words “horrible,” “deadly,” and “life-threatening.” Based on this perception, all 4 participate in cancer screening and surveillance. Half of the participants endorsed the concept of cancer as potentially curable and the importance of health care in cancer prevention.

Environmental Exposure and Contamination as a Main Cause of Cancer
There were 3 (75%) participants who described an environmental issue as a main cause of cancer.

Occupational Exposure
Exposure to harmful chemicals through occupation was described by 2 (50%) participants as a main cause of cancer:

Some of it is asbestos for a lot of our workers who worked in these asbestos mines.

Another connected cancer to exposures in the workplace (workplace redacted for identity protection):

I used to be a [occupation redacted]. It was the environment we worked in. That’s why we’re doing an early retirement, [workplace redacted], because of our environment, what we ingested from the environment.

Chemicals in the Community
A history of chemicals in the community believed to be linked to cancer was vividly described by 1 (25%) participant:

I remember I was very young when they sprayed our area. I remember the planes hanging out in our area and how it killed a lot of our trees. I remember the one thing that really impacted me was the poppies. A lot of beautiful poppies that would grow all the way around our area where I lived, which was a mile at the time. We hardly have that any more ... There was a thing called agent orange ... I remember one time climbing there and I saw the planes. I remember them spraying something alongside the river.

Water Contamination
Water as a main cause of cancer was described by 2 (50%) participants with asbestos in the water and the pipelines themselves perceived as cancer-causing:

To me I think it’s drinking water from the faucet ... Way back in the years there were cancer causing pipelines that were installed. And this person told my Dad that you won’t be surprised when people from the reservation will be experiencing this cancer.

Asbestos was sprayed and caused contamination of our water. I’m sure, I mean those run right through the areas where they have mined and it seeps into our water system.

Personal Choices as Risk Factors or Prevention: The Power of Food, Exercise, and Lifestyle
The personal choices participants linked to cancer risk or prevention had a recurring connection to food, exercise, and lifestyle. Diet choices were believed to be a main cause of cancer by 2 (50%) participants, and 2 (50%) expressed that substance-free living is important to cancer risk reduction.
Nutrition is prevention.

Exercise regularly. Don’t smoke. Don’t drink. Keep down on your fatty foods.

I think it helps prevent it. Because a strong body kind of resists it, I believe.

Sugar and processed foods were specifically mentioned as impacting cancer:

I truly believe that the sugar we take in has a lot to do with it ... Processed food is a part of it. If we sit around and do nothing, that unhealthy food builds up in our body. However, if we exercise, it is burned off, then the food nourishes our body.

However, one participant did not believe food and physical activity impacted the risk of cancer based on her personal experience of self-perceived healthy diet and activity but a diagnosis of cancer:

I guess I have to say no from my own experience, because I did that, you know, I exercised like crazy. I ate well. I made sure everything—I grew up in a home where we had our own garden and made sure that we ate healthy food, and that’s how I grew up.

Barriers

Distance and Transportation

Distance and transportation were cited as barriers to healthy lifestyle behaviors. There were 2 (50%) participants who described how lack of transportation impacted their ability to eat healthier and potentially participate in programming. There were 3 (75%) participants who described the distance to a grocery store with affordable prices and fresh options as barriers to healthy eating.

Having the transportation probably because right now my ride is down and all of us have a hard time making our way there.

You know that this [name of grocery store redacted] could have more healthy food and not so much processed food. Prepackaged stuff. If they could make other things available out there. I hardly ever go to [redacted] except to get immediate things, you know ... Their stuff is very high compared to others but can’t make the distance to [other location] every day.

The Built Environment

Participants described how the hills on their reservation and the busy road add challenges to their ability to exercise outdoors.

It’s dangerous to even cross the street. When I do go out, I usually go up with someone early in the morning.

Having the time and a level place to walk ... it is hilly where I live.

Acceptability of Diet and Physical Activity Programs

Overview

The participants were asked about a theoretical diet and physical activity program in their community, and all stated they would participate. When asked if they would participate when first diagnosed with cancer, all said yes with half providing no stipulation and half expressing concern about it interfering with treatment or the side effects of treatment being prohibitive.

Oh, yes, I sure would. I would love that.

Yeah, as long as it did it not interfere with my treatment.

When asked if their participation would have been influenced if it were explained that participating during cancer treatment could potentially improve cancer outcomes, all responded with strong affirmation, including those who had initial apprehension about it interfering with treatment.

Considerations for Program Development

The participants were invited to brainstorm the type of diet and physical activity program (prehab) they would want for their community. There were recurring responses about the need for support and encouragement, a central location, and education. Participants emphasized the need for peer support and encouragement, particularly the importance of including a cancer survivor as part of the program team.

People who have experienced cancer might be someone to have there. They know their issues and what they went through and can share ... Like a group where people could learn what’s going on with them and know somebody is listening to them that knows what is going to happen ... I think I was lost when I was told about having cancer, and it was scary in ways, and nobody ever told me you’re not going to die from this one.

All participants referenced their community center as the preferred location for prehab. While there was consistency on the community center as the chosen central location, upon further probing, it became apparent that each district of the reservation has its own community center, and while the participants were uniform on the category of “community center,” they were referencing different candidate buildings. Reflective of the concern that lack of transportation is a barrier to a healthy lifestyle, one participant recommended that transportation to the central location be provided by the program. Participants also voiced interest in education, delivered either by peers or health professionals.

Somebody in oncology who can share what happens as nobody gets taught about these things. I was having to get on the computer to find out what this word means and the different services available. It’d be nice if someone had sat me down and kind of counseled me on it and what could happen.

First you have to educate them on the disease and what it does to their body. Then you have to cover the things that could be promoting it to happen. And lastly the things that could change that.

There was no clear consensus on how often the program should meet. One participant requested a daily program. The others recommended 1-3 sessions weekly. There was no consensus on how long each session should occur with a range of responses from 30 to 120 minutes, but 3 (75%) supported 1 hour as...
appropriate. Participants practiced a wide range of physical activities including treadmill, weights, biking, running, gardening, gathering acorns, and aerobics. All 4 mentioned walking as a favorable regular activity. There were 3 (75%) participants who described going to a fitness center as favorable. The participants were presented with a list of food items that have been included in other prehab programs and asked about their likelihood of consuming these if the food items were provided. On a scale of 1-5 with 1=absolutely would not eat and 5=absolutely would eat, the participants viewed fiber powder and flax seeds negatively, protein shakes neutrally, and walnuts as highly favorable, with all 4 giving a score of 4-5 likelihood of consuming walnuts as part of a prehab program.

Acceptability of Monitoring and Biospecimen Collection
All participants responded positively to the following forms of monitoring: measurement of heart rate, blood pressure, and weight; blood collection to measure glucose, cholesterol, and levels of inflammation; and biospecimen collection (ie, tumor tissue) to measure inflammation. Half brought up that this is a personal choice, and they were not sure if others would agree to monitoring.

I would. I don’t know about other people but I can. It’s a necessary part of what you do.

Yeah. It’s important to be monitored when you’ve got cancer.

The question about biospecimen collection generated more discussion. It was first explained that measuring inflammation in the cancer tissue before and after a prehab program is one way to see if the program is effective, and that it can be done without any extra procedures beyond those required for standard cancer treatment. Then the participants were asked, “Looking back, do you think you would be comfortable with your biopsy tissue and the tumor removed by the surgeon to be analyzed for inflammation to see if a diet and physical activity program can change cancer cells?” All said they would with 3 of the participants providing a yes without stipulation, and 1 person asking for further clarification. After repeating the explanation, the individual responded, “I definitely agree with that.”

Discussion
Principal Findings
To our knowledge, this is the first study to engage San Carlos Apache cancer survivors for their perspectives on lifestyle modifications in diet and physical activity and the cancer journey. It is the investigators’ long-term goal to work with the tribe and health facility to collaboratively develop a nutrition and physical activity prehab program that serves patients with cancer. Prior to programmatic development, the team recognized that it was critical to invite community members into the discussion early as a first phase of research to learn if such programming would be appealing and what background the community prioritizes. From these interviews, important themes and subthemes were highlighted: (1) cancer is perceived as a serious health condition in the community; (2) occupational exposures, chemical exposures, and water contamination on the reservation are perceived as cancer-causing threats; (3) healthy foods, regular exercise, and avoiding harmful substances are mostly perceived as measures an individual can use to try to prevent cancer, though there is recognition that it is not a guaranteed protection; (4) lack of transportation, distance to affordable groceries and recreational centers, and hilly and high-traffic roads are barriers to healthy habits for this community; (5) there is high interest in a prehab program geared toward patients with cancer that should include components of encouragement and education and take place at a central location; and (6) standard monitoring practiced in published prehab programs, which includes blood and tumor-tissue samples, is acceptable to the individual cancer survivors who participated in this study with the important point made that it is their personal view, and openly sharing a clear rationale may encourage wider acceptability.

Comparison to Prior Work
Lifestyle interventions for optimizing cancer outcomes are undergoing a paradigm shift in which earlier engagement (ie, prehab) holds promise for greater impact, including modulating inflammation. Obesity is considered a pro-inflammatory condition, and the synergy between obesity and inflammation is thought to influence cancer risk and progression. With an increased risk of obesity, American Indian/Alaska Native populations carry an increased risk of obesity-related cancers and worse outcomes [24-26]. Previous studies with participants from the general population have shown that prehab interventions have modulated inflammation in multiple cancer subtypes as measured by biomarkers in windows as brief as 3-5 weeks between diagnosis and operation [32-36]. The preoperative period is a unique window before treatment side effects have set in and might be more conducive to patient participation while additionally fostering healthy behaviors for treatment and survivorship ahead. Such studies have never been implemented for American Indian/Alaska Native populations. With a disproportionate burden of obesity and cancer mortality, prehab interventions during this diagnosis-to-surgery window of opportunity could provide gains in health equity for American Indian/Alaska Native populations, a group underrepresented in clinical research [37].

Leaders in the field have published their recommendation that well-designed interventions should include tumor-derived tissue from diagnostic biopsies and surgical resections to evaluate the effects of the intervention on the tumor microenvironment [12,38]. While this is an accepted measure with general populations, it is a delicate subject with American Indian/Alaska Native audiences, as some tribes have experienced breaches in the ethical conduct of research and misuse of their biospecimens [39]. Our team wanted to proactively address this component of prehab trials to understand whether this would be an acceptable method of measuring outcomes with San Carlos Apache participants. Prepared for apprehension, we were encouraged to learn that the participants saw the benefit of monitoring vitals, blood samples, and tissue samples. We hold this trust in the highest regard and comply with the University of Arizona’s rigorous process for tribal-related research (ABOR1-118 Tribal Consultation Policy), which recognizes the fundamental principles of tribal sovereignty and requires documented evidence of consultation and approval when...
conducting research with Native Nations [40]. These processes and special considerations are important to highlight for any team interested in research with a tribal partner, and prospective investigators should bear in mind that each individual tribe determines the process that must be followed at all steps of the investigation.

Dietary interventions rich in omega-fatty acids have been conducted outside the prehab context to alter inflammation in various disease conditions including cancer. These studies have featured omega-rich foods, like walnuts, which have a high content of polyunsaturated fatty acids and antioxidants that can improve blood lipid profiles, reduce oxidative stress, influence inflammatory biomarkers, and reduce tumor growth, offering a natural and tasty way of improving outcomes [41-54]. This single food intervention affords simplicity in delivery, has previously demonstrated high uptake and adherence, and was highly endorsed by the San Carlos Apache participants. One person commented that walnuts grow wild on the reservation and would be perceived as a more natural, traditional food option.

Strengths and Limitations
This study has limitations. With 4 participants, the findings may not be fully representative of all San Carlos Apache patients with cancer; however, the participants were consistent in most of their responses, which may have achieved saturation or sufficiency of themes. Qualitative researchers are recognizing that sufficiency depends on the rigor of the process and the richness of the data it generates, more than the absolute number of participants [55]. With the length of time these participants invested in the interviews and the candor of their responses, we believe this to be a strong and meaningful data set. Another limitation is that cancer survivors rather than patients who were newly diagnosed with cancer were interviewed. Survivors were selected because of their line of sight on the full spectrum of the cancer experience in their community, which had the advantage of providing reflections outside the acute exhaustion of a new diagnosis. This has the potential for recall bias and might not have rendered their real-time experience as accurately. Finally, a surprising result was the very high level of vigorous or moderate physical activity the participants reported, which is more active than average prehab patients and should be explored further.

Future Directions
Based on these qualitative data, San Carlos Apache members have demonstrated interest in a nutrition and exercise program delivered between diagnosis and oncologic surgery that includes encouragement, education, standard monitoring methods, and support in overcoming barriers to participation. The participants offered valuable insight on the specifics that would adapt published prehab models to their community. Such an intervention would attempt biochemical change using food as medicine and exercise as resilience, which is reflective of the wider momentum that American Indian/Alaska Native scholars are building in national discussions on the topics of Indigenous food sovereignty, decolonizing the diet, embracing traditional perspectives, and restoring balance [56-59].

Conclusions
American Indian/Alaska Native populations are known to be underrepresented in clinical trials and research, and even more importantly, they have never been included in the adaptation and participation of an intervention designed to modulate inflammation through diet and exercise prior to cancer surgery. San Carlos Apache would be the first tribal partner to conduct such a program and deliver the emerging science of prehab to a population that has yet to be included in this field of work. Future directions should include the findings from these interviews with continued tribal collaboration to develop a program specific to the San Carlos Apache community.

Acknowledgments
The authors acknowledge the cancer survivors of the San Carlos Apache community who participated in this study as well as the collaboration of the San Carlos Apache Healthcare Corporation and the Tribal Council. The authors would also like to thank Sherrie Harris and Marcy Langley for their support of the project through their roles at San Carlos Apache Healthcare Corporation. This research was supported by the National Cancer Institute of the National Institutes of Health (awards U54CA14392 and K08CA276137). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest
None declared.

References


20. NACP. The partnership for Native American Cancer Prevention. University of Arizona Cancer Center. 2024. URL: https://cancercenter.arizona.edu/researchers/collaborative-research/naacp [accessed 2024-03-26]


Abbreviations

IPAQ: International Physical Activity Questionnaire
REAPS: Rapid Eating Assessment for Participants-Shortened Version
SCAHC: San Carlos Apache Healthcare Corporation
Exploring the Implementation of Shared Decision-Making Involving Health Coaches for Diabetes and Hypertension Self-Management: Qualitative Study

Sungwon Yoon¹, PhD; Chao Min Tan², MScR; Jie Kie Phang², MPH; Venice Xi Liu¹, BSc; Wee Boon Tan², BSc; Yu Heng Kwan¹, PhD; Lian Leng Low², MBBS, MMed

¹Duke-NUS Medical School, Singapore, Singapore
²Centre for Population Health Research and Implementation, SingHealth Regional Health System, Singapore, Singapore

Corresponding Author:
Sungwon Yoon, PhD
Duke-NUS Medical School
8 College Rd
Singapore, 169857
Singapore
Phone: 65 65167666
Email: sungwon.yoon@duke-nus.edu.sg

Abstract

Background: An emerging focus on person-centered care has prompted the need to understand how shared decision-making (SDM) and health coaching could support self-management of diabetes and hypertension.

Objective: This study aims to explore preferences for the scope of involvement of health coaches and health care professionals (HCPs) in SDM and the factors that may influence optimal implementation of SDM from the perspectives of patients and HCPs.

Methods: We conducted focus group discussions with 39 patients with diabetes and hypertension and 45 HCPs involved in their care. The main topics discussed included the roles of health coaches and HCPs in self-management, views toward health coaching and SDM, and factors that should be considered for optimal implementation of SDM that involves health coaches. All focus group discussions were audio recorded, transcribed verbatim, and analyzed using thematic analysis.

Results: Participants agreed that the main responsibility of HCPs should be identifying the patient’s stage of change and medication education, while health coaches should focus on lifestyle education, monitoring, and motivational conversation. The health coach was seen to be more effective in engaging patients in lifestyle education and designing goal management plans as health coaches have more time available to spend with patients. The importance of a health coach’s personal attributes (eg, sufficient knowledge of both medical and psychosocial management of disease conditions) and credentials (eg, openness, patience, and empathy) was commonly emphasized. Participants viewed that addressing the following five elements would be necessary for the optimal implementation of SDM: (1) target population (newly diagnosed and less stable patients), (2) commitment of all stakeholders (discrepancy on targeted times and modality), (3) continuity of care (familiar faces), (4) philosophy of care (person-centered communication), and (5) faces of legitimacy (physician as the ultimate authority).

Conclusions: The findings shed light on the appropriate roles of health coaches vis-à-vis HCPs in SDM as perceived by patients and HCPs. Findings from this study also contribute to the understanding of SDM on self-management strategies for patients with diabetes and hypertension and highlight potential opportunities for integrating health coaches into the routine care process.

(JMIR Form Res 2024;8:e51848) doi:10.2196/51848

KEYWORDS
decision-making; diabetes; health coach; health coaching; healthcare professional; hypertension; patient; patient-centered care; person-centered care; qualitative research; self-management; shared decision-making
Introduction

Lifestyle factors, including engaging in adequate physical activity and consuming a healthy diet, are important in the management of diabetes and hypertension [1-3]. However, many patients with diabetes and hypertension do not achieve optimal lifestyle targets [4-6]. Effective interventions to help patients with diabetes and hypertension should be tailored for each patient, and this may be achieved through engaging in shared decision-making (SDM). In line with the growing emphasis on person-centered care, SDM is recognized as a way to empower patients with chronic diseases such as diabetes and hypertension. SDM involves patients and health care professionals (HCPs) collaboratively making a health care decision after discussing the treatment, management, and support packages and considering the patient’s preferences, priorities, and goals [7-9]. Although there is limited research on the effect of SDM on clinical outcomes for diabetes and hypertension [10], studies invariably suggest that SDM makes a positive difference to patients in their care. This includes better treatment adherence, increased patient coping, improved knowledge attainment, higher levels of patient satisfaction, and greater empowerment [10-12].

In addition to situations requiring treatment decisions, SDM may be useful in supporting healthy behavior change [13,14]. A feasibility study on decision tools in primary care to help initiate lifestyle change among patients with or at risk of coronary heart disease has shown the potential beneficial effects of paper-based tools for SDM in initiating behavior change [15]. Another feasibility study of an internet-based decision aid to encourage lifestyle change and adherence among people at moderate or high risk of coronary heart disease was found to increase participants’ ability to make clear decisions about making changes [16,17]. However, it was suggested that further impact may have been achieved if more comprehensive implementation strategies had been available for the interventions [16,17].

Despite the evidence supporting SDM, it is not widely practiced in clinical settings due to several reasons, such as low patient self-efficacy, a power imbalance between patients and physicians, and HCP’s limited time and knowledge [12,18-20]. To overcome these communication and resource barriers, several studies proposed the inclusion of health coaches to facilitate the SDM process in chronic care through continuous counseling dialogue with patients and exploration of patients’ situations and preferences in order to make informed decisions together on treatment and lifestyle [21,22]. Health coaches are individuals who aid patients in gaining the knowledge and confidence necessary to become engaged in their care and promote communication and collaborative decisions between patients and HCPs [23]. The practice of health coaching can differ in the type of coach, their training, and their level of involvement [24]. Nonetheless, randomized controlled trials have shown that health coaching can lead to enhanced self-management of diabetes and hypertension [25,26]. Furthermore, the experiences of patients and HCPs were found to be largely positive [27-29].

Although literature has documented the effectiveness and feasibility of SDM and health coaching, the evidence primarily comes from patients and HCPs who were willing to take part in or complete interventions [30-32]. There are fewer studies that have examined the viewpoints of potential end users’ perspectives regarding their preferences for and expectations of SDM [33,34], as well as the role and relationships of health coaches in patient care practice [35]. Obtaining the buy-in of patients and HCPs is crucial when developing a robust care model. To this end, we conducted a study to gather the viewpoints of patients and HCPs to gain insight on developing strategies for SDM programs that incorporate nurse-trained health coaches in primary care.

The aim of this formative study was to explore the perspectives and preferences of HCPs and patients with diabetes and hypertension concerning the respective professional roles of health coaches and HCPs in SDM, as well as the factors that should be considered for optimal implementation of a SDM model that involves patients, health coaches, and HCPs.

Methods

This was a qualitative study, reported following the Consolidated Criteria for Reporting Qualitative Research (COREQ) Guidelines [36].

Setting and Participants

This study was conducted in Singapore, a multiethnic city state where the majority of the population (80%) obtains health care from the public health care system [37]. Participants were mainly from SingHealth Cluster, which is the largest regional health care system in Singapore, offering a complete range of medical care for patients, including those with diabetes and hypertension. Eligible patient participants were those aged 40 years and older, diagnosed with diabetes and hypertension, and attending public primary care clinics. We identified eligible patients from a study cohort that investigates the clinical and cost-effectiveness of a behavioral intervention delivered through mobile health [38]. The participants were then approached through a phone call with the study aim and methods explained. On the other hand, eligible HCP participants were those responsible for managing patients with diabetes and hypertension in public primary care clinics, step-down care, and secondary or tertiary centers with at least 1 year of experience. We approached potential HCP participants through email and provided background information. Purposive sampling was adopted in terms of age (patients) and clinical experience (HCPs) to maximize diversity of perspectives.

Data Collection

We conducted focus group discussions (FGDs) with participants between February 2022 and May 2022. A semistructured topic guide was developed and subsequently pilot-tested to facilitate discussions on the roles of health coaches and HCPs in self-management, views toward health coaching and SDM, and factors that should be considered for optimal implementation of SDM that involves health coaches. All FGDs were carried out through web-based videoconferencing by facilitators (CMT and WBT) who were trained in social sciences and qualitative
research and did not have a personal relationship with the participants. Each FGD session lasted approximately 90 minutes for patient participants and 60 minutes for HCP participants. No repeat interviews were conducted, and transcripts were not returned to participants for further input. Data collection and analysis were an iterative process that continued until no new themes emerged. Field notes were taken to support the contextual interpretation of the data.

**Data Analysis**

All FGDs were audio-recorded and transcribed verbatim. Transcripts were checked for accuracy and thematically analyzed. A total of 2 coders (CMT and WBT) were assigned to code the patient FGDs, while 2 other coders (JKP and VXL) were assigned to code the HCP FGDs. The team adopted the 6 steps to thematic analysis suggested by Braun and Clarke [39], in which the coders first familiarize themselves with the data and generate initial codes independently before collating the codes into potential themes together. The themes were constantly reviewed, refined, and reclassified to ensure the best fit of themes to the data. Discrepancies between coders were resolved through consensus meetings involving all study team members. The FGDs were conducted until thematic saturation occurred at the 17th and 18th FGDs with patients and HCPs, respectively. Storing and managing data during data analysis was done using NVivo (version 12; QSR International).

**Ethical Considerations**

This study was approved by the SingHealth Centralized Institutional Review Board (2019/2468). Participants provided verbal informed consent before the study began. The study team maintained data confidentiality by redacting personally identifiable information from interview transcripts and generating unique study identifiers, which were linked to participant identifiable information only through a password-protected file. Participants were reimbursed SG $60 (US $44.70) to defray the cost of their participation in this research.

**Results**

**Participant Characteristics**

Out of 89 patients approached, 39 were recruited for the study, with the most frequent reasons for the decline being difficulties in participating in web-based interviews or schedule unavailability. The recruited patients participated in 17 FGDs. Their age ranged from 43 to 68 years, with 74% (29/39) being male candidates and 64% (25/39) being Chinese. Concurrently, we approached 52 HCPs and recruited 45 HCPs who participated in 18 FGDs, with schedule unavailability as the main reason for the decline. Approximately 65% (29/45) were clinicians, followed by 16% (7/45) being nurses. The range of clinical experience in managing chronic diseases of the HCPs was from 1 year to 28 years (Table 1).
Table 1. Characteristics of focus group participants (N=84).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patients (n=39)</th>
<th>HCPs(^a) (n=45)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGDs(^c), n</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>29 (74)</td>
<td>11 (24)</td>
</tr>
<tr>
<td>Female</td>
<td>10 (26)</td>
<td>34 (76)</td>
</tr>
<tr>
<td><strong>Age (years), range</strong></td>
<td>43-68</td>
<td>25-56</td>
</tr>
<tr>
<td><strong>Years of clinical experience, range</strong></td>
<td>N/A(^c)</td>
<td>1-28</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>25 (64)</td>
<td>36 (80)</td>
</tr>
<tr>
<td>Indian</td>
<td>5 (13)</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Malay</td>
<td>9 (23)</td>
<td>5 (11)</td>
</tr>
<tr>
<td><strong>Profession, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clinicians</td>
<td>N/A</td>
<td>29 (65)</td>
</tr>
<tr>
<td>Pharmacists</td>
<td>N/A</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Dieticians</td>
<td>N/A</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Nurses</td>
<td>N/A</td>
<td>7 (16)</td>
</tr>
<tr>
<td>Administrator</td>
<td>N/A</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Diagnosis, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes only</td>
<td>16 (41)</td>
<td>N/A</td>
</tr>
<tr>
<td>Diabetes and hypertension</td>
<td>23 (59)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\)HCP: health care professional.
\(^b\)FGD: focus group discussion.
\(^c\)N/A: not applicable.

Through our analysis, we identified three themes that represent (1) the participants’ perspectives concerning the professional roles of health coaches and HCPs in SDM, (2) the perceived importance of health coaches’ credentials and attributes, and (3) a total of 5 essential elements to be considered for optimal implementation of SDM. Figure 1 presents a visual summary that suggests how SDM involving health coaches could be applied in clinical settings to facilitate diabetes and hypertension self-management, based on the findings.
Perceived Preference for Professional Roles of HCP and Health Coach in SDM

While there were some commonalities in the roles of HCPs and health coaches in the SDM model of care, a distinction in the extent of their responsibilities was evident.

Patient Education

Many of the patient participants and HCP participants recognized the shared responsibilities of HCPs and health coaches in improving patients’ understanding of their conditions in order to decide on the self-management strategies that fit best according to their individual situations and capacities. However, their expectations of the specific roles that HCPs and health coaches would play in patient education differed. The health coach was seen to be more effective in engaging patients in lifestyle education and informing patients on healthy lifestyle choices, while the HCPs are expected to educate patients on medication and alternative treatment options for their conditions. The participants explained that the difference in expectations was based on the perceived amount of time availability the professionals have with the patients and the background of the professionals.

[The health coach] is a single point of contact that I can refer to, who is an expert in this area, and I can leverage on that to achieve the goals that I want. Knowing that there is somebody associated with you, and you can engage with it helps a lot. Whether it’s about physical activity, food intake, [or] the discipline you need to get in order to achieve the goals, if I know I can reach out to someone to talk about it, it will definitely make a difference. [Patient 37, Patient FGD #13]

I will tell her what the best option as a physician is, based on our guidelines. However, I will tell her other possible options if let’s say she doesn’t want the recommended option. As a physician, it’s our duty to tell patients the options they have and the pros and cons of each option. [HCP 36, HCP FGD #12]

Goal Setting

Both patient participants and HCP participants agreed that setting actionable goals would be crucial to improving clinical outcomes. However, the brief consultation sessions in primary care settings were inadequate for patients to develop personalized care plans with their HCPs. Thus, patient participants saw the value of involving the health coach to work with them to set actionable goals while taking into consideration their personal circumstances in order to devise an appropriate self-management plan that aligns with the expectations of the HCPs. At the same time, HCPs suggested that the health coach plays an important role in identifying any lapses and bridging the gaps between the treatment offered by HCPs and the health care preferences and goals of the patients.

A doctor’s goals may be different from a patient’s goals. Sometimes it’s hard for us to assess the ideas, values, and preferences during our short 10 to 15 minutes [consultations]. If the health coach informs us, that will be good so that the patient, doctor, and health coach can be on the same page to help the patient achieve his or her goals. [HCP 36, HCP FGD #12]

[A health coach should] connect with the patient to discover what the problems are and also be aware of the patient’s environment. Then align these with the expectations that HCPs might have of the patient. So, the health coach’s duty is basically to identify all of these, in order to make things easier. [Patient 25, Patient FGD #10]

Patient Empowerment

According to patients, “feeling empowered” entailed remaining motivated to engage in behaviors that promote their health. To
this end, they preferred to work with the health coaches and mutually establish achievable goals related to behaviors such as diet and exercise. The partnership that patients form with their health coach allows patients to feel supported throughout their self-management journey, which motivates them to be more engaged and adherent to the chosen care plan. Likewise, HCP participants recognized the significance of such a partnership, in which the patients are free to express their expectations and wishes for care while deciding on a care plan that would be tailored to their individual needs and preferences.

> It would be more of a motivating aspect because the health coach can help set specific Agendas, target them, and communicate with patients...In this way, we are more motivated to try hard to meet the targets. Because when seeing the doctor, you know, he/she will just tell you to lose weight but [the] health coach can motivate you for sure. [Patient 25, Patient FGD #10]

> In an ideal setting, [health coaches] need to understand what their targets are, and what the patient thinks of their health conditions. From there, see what the patient is willing and able to do and what their plans are going forward. [HCP 09, HCP FGD #1]

### Importance of Health Coach’s Credentials and Attributes

Both patient participants and HCP participants emphasized the importance of health coaches’ credentials and attributes that would influence their acceptance of health coaching as part of routine patient care. Most participants mentioned that a desirable health coach would need to possess sufficient knowledge of both medical and psychosocial management of diabetes and hypertension so that health coaches could offer appropriate guidance to patients.

> Health coach should be medically trained to give correct advice. I mean, apart from medication and disease management [which is the ambit of HCPs], the health coach should be able to provide psychosocial counseling. [HCP 38, HCP FGD #8]

A health coach’s personal attributes were equally stressed; most participants noted that a health coach should demonstrate positive personality traits such as openness, patience, and empathy to effectively improve a patient’s willingness to consider the recommended health practices as suggested by the health coach.

> How do I become open to the health coach? Firstly, I think the health coaches themselves must be very caring and full of empathy, to disarm all the unhappiness of the patient, maybe then patients will be willing to tell the coach about their story. [Patient 51, Patient FGD #17]

### 5 Essential Elements for Optimal Implementation of SDM

A total of 5 elements have been identified that participants believe should be considered for optimal implementation of the SDM model of care involving health coaches.

#### Target Population

Most participants (patients and HCPs) shared that SDM should target a certain segment of patients to maximize its benefits. They believed that individuals with newly diagnosed chronic conditions and those with poorly controlled diabetes and hypertension would benefit more from the proposed SDM model of care involving health coaches. This is because these patient groups may have inadequate knowledge or support to effectively manage their conditions. Thus, the education, guidance, and discussion provided by the health coach could prepare them with the necessary knowledge to begin their self-management journey effectively.

> Maybe this will help for those who just are diagnosed with diabetes, and you know, those at a loss or don’t know what to do. But if you’re working with a “seasoned” patient, they know what to do, what to expect and all. [Patient 17, Patient FGD #3]

#### Commitment of all Stakeholders

Participants asserted that in order to effectively facilitate SDM and achieve the mutually established goals of improving self-management, all stakeholders involved in this communication process should be equally committed. They expressed commitment to engaging in open dialogues and establishing partnerships among all 3 parties, including patients, HCPs, and health coaches, at targeted times and through preferred modalities. While patient participants expressed a preference for frequent check-ins with the health coach (eg, monthly), HCP participants felt that meetings between health coaches and HCPs should be arranged on a case-by-case basis, depending on the urgency and complexity of the patient’s conditions due to their high workload. Most participants were open to the varied modes of communication, including in-person and web-based means to facilitate the SDM to cater to different situational needs (eg, web-based call for a brief check-in and an in-person call for an in-depth dialogue).

> I think doctors need to focus on the very complex cases for communication with health coaches. We can’t put too much effort into handling every single chronic patient because the workload will be too high. [HCP 02, HCP FGD #2]

#### Continuity of Care

Care continuity was identified by participants as a critical factor in facilitating SDM. Both HCP participants and patient participants expressed a preference for having the same health coach for follow-up appointments to foster a sense of rapport, continuity of care, and motivate patients to carry out the plan of care. Additionally, patient participants felt that following up with the same health coach would help them build trust with their health coach and disclose their concerns, strengths, and limitations in their self-management journey.

> In a polyclinic setting, the doctor will change every appointment. If the same health coach can provide consistent support and counseling to the patient, and if the patient has someone who is checking on him,
he will want to take better care of his chronic conditions. [HCP 02, HCP FGD #2]

I think the health coach should be someone that at least I know and have some level of good relationship...so I can treat the coach like a friend and open up. [Patient 32, Patient FGD #9]

**Philosophy of Care**

Participants highlighted the importance of open communication and person-centered inquiry to facilitate mutual understanding among all parties involved in SDM. They valued strategies that could allow patients to set personal goals, negotiate, and discuss challenges. Thus, the philosophy of care was focused on supporting patients to make informed choices and engaging patients in discussion to develop a care plan that is tailored to individuals, as opposed to simply offering generic health advice that may not be as effective in motivating patients.

Ask the patient what they want first, because if it’s not something that they want, it’s not likely that they will cooperate with us [HCPs or health coaches] even though it is what we want from them. [HCP 49, HCP FGD #18]

**Faces of Legitimacy**

Notably, our interviews revealed that patients generally prioritized the advice given by their physicians because they perceived their physician’s advice to be more important and reliable than that of health coaches and other HCPs (eg, dieticians, nurses, and physiotherapists). This could pose a challenge to SDM when consensus cannot be built about the preferred self-management plan among all parties involved and patients are less receptive to exploring other treatment options and recommendations unless they are endorsed by a physician.

I’ll still take the final instructions from the doctor. Frankly speaking, the health coach might be knowledgeable in terms of some medical information, but they are still not reliable. [Patient 17, Patient FGD #3]

Sometimes it also depends on which healthcare provider is approaching the patient. A lot of times, our patients defer to what we intended so if the doctors don’t say like “Oh you need to do this” then they won’t really cooperate because doctor’s recommendations take precedence over whatever other professionals are seeking to help. [HCP 49, HCP FGD #18]

To mitigate this, some of our HCP participants, who are physicians, suggested that reinforcement and endorsing the advice from the health coaches and other HCPs during follow-up appointments would be important in increasing patients’ trust in other providers and promoting effective communication in SDM.

One thing that physicians like me could do is to reinforce what the health coach has taught the patients. Then the patient would realize that, Oh yes, that [advice given by health coach] is very important. [HCP 49, HCP FGD #18]

**Discussion**

**Principle Findings**

This study explored the preferences and perspectives of both patients and HCPs on how SDM involving health coaches could help patients make informed decisions about their health and improve self-management of their diabetes and hypertension. While some perspectives varied across patients and HCPs, we identified three unified themes, including (1) the perceived preference for and expectation of the roles of HCPs and health coaches in SDM, (2) the importance of health coaches’ credentials and attributes, and (3) the elements necessary for effective implementation of SDM. The findings gained from this study offered key insights to support efforts to optimally implement SDM involving health coaches for patients with chronic conditions.

The lack of patient education [40,41] and psychosocial support [42-45] can hinder patients’ ability to self-manage their diabetes and hypertension, which eventually results in suboptimal control and negative health outcomes. In this study, patient participants and HCP participants agreed that the primary responsibility of a health coach is to educate patients on healthy lifestyle choices and provide several self-management options before setting actionable goals that align with the patient’s needs and preferences, while an HCP is expected to provide medication education and offer alternative treatment options for their conditions. In this regard, SDM provides a platform for patients, health coaches, and HCPs to engage in conversations that enable information to be shared and address each party’s expectations for care [13]. In addition, the involvement of health coaches in SDM has been shown to improve self-management by fostering greater patient involvement in their care and designing care plans that take into account their unique treatment goals and preferences [46]. When patients have a better understanding of their options and have the autonomy to express their preferences and wishes for care, they are more likely to be satisfied with the eventual plan of care and adhere to it [47]. Moreover, our findings showed that the involvement of a health coach would offer patients a sense of support through their self-management journey and motivate them to take charge of their diabetes and hypertension self-care. This finding reflects previous studies that found integrative health coaching improved patients’ psychosocial outcomes, resulting in reduced perceived barriers to self-management, enhanced perceptions of social support, and ultimately improved clinical outcomes in patients with diabetes and hypertension [48-50]. The results of this study offer valuable insights into the distinct responsibilities of health coaches and HCPs in SDM and chronic disease management and highlight potential areas of emphasis in patient coaching, workflow, and collaborative efforts.

A common theme running through the FGDs was participants’ keen interest in the credentials and characteristics of health coaches. Participants stressed the importance of a health coach’s positive attitude and knowledge in both the medical and psychosocial aspects of disease management in order to engage in a partner relationship in SDM. Indeed, a health coach’s professional expertise and personal traits, such as openness and
empathy, can improve the therapeutic relationship and ultimately enhance self-management skills [30,51]. Therefore, it is essential for health coaches to receive appropriate clinical training and education as well as possess a strong capacity for empathy [52]. Beyond the credentials and characteristics of health coaches, studies on SDM also stressed the importance of a supportive and caring environment with adequate interaction time as key aspects of the patient-provider partnership in chronic care [53,54]. Many of the physicians interviewed in this study often mentioned that they are unable to cover all aspects of the patient’s self-care due to the brief consultation sessions they have with the patients. The time required for information sharing and clarifying patients’ values, needs, and preferences could impact the already-pressured clinical setting [13]. Therefore, it was suggested that the involvement of health coaches in SDM would help to prioritize discussions about specific aspects of diabetes and hypertension self-management (ie, medical and lifestyle) and allow patients to benefit from the enhanced support from their health coaches, who have more time to work with them on modifying their lifestyle and achieving better control of their conditions. Our findings underscore the significance of health coaches’ competencies to ensure that health coaches can fulfill their core responsibilities and the potential benefit of involving health coaches in SDM to support patients further in their self-management journey.

Lastly, our participants believed that newly diagnosed or less stable patients could benefit the most from SDM involving health coaches and emphasized the importance of continuity of care through the same coach. They also recognized that open communication and person-centered inquiry would be crucial for improving the quality of SDM. Indeed, previous studies demonstrated that open communication and consistent coaching improved decision quality, knowledge, and risk perception among patients with diabetes and hypertension [10,55]. Despite these findings supporting the use of SDM to support healthy behavior change among patients with chronic disease, patients in this study still held a strong belief in the traditional approach of “doctor knows best” (faces of legitimacy), with many patients relying disproportionately on physicians for decisions [56]. Patients’ reliance on physicians for decisions may pose a challenge to the SDM process involving health coaches since patients may prioritize the advice of their physicians over that of health coaches. For health care institutions that wish to implement SDM for chronic disease management, we suggest distinguishing the roles of HCPs and health coaches in chronic disease management to ensure successful implementation of SDM. Institutions can also consider educating patients about the unique and valuable contributions that health coaches can make in their care [57] to reinforce their trust in health coaches. As observed in this study, the health coaches’ involvement in SDM was important in offering personalized support to patients to modify their lifestyle and self-management in order to achieve behavioral change and gain better control of their diabetes and hypertension. Future research should aim to identify factors that affect patients’ engagement and trust with health coaches to enable successful implementation of this SDM model for chronic disease management.

Limitations
This study has a few limitations. The perspectives of health coaches were not included, which may limit the comprehensiveness of the results. Furthermore, participants’ preferences and expectations were not examined by subgroups such as HCP’s professional roles or patients’ confidence levels in self-management and cultural backgrounds. Further research focusing on these aspects may prove useful for a richer understanding of the SDM implementation. Despite these limitations, this study provided valuable insights into the SDM model of care, highlighting how patients, HCPs, and health coaches can collaborate and the factors needed to be considered for robust implementation of the SDM for patients with diabetes and hypertension.

Conclusions
Our findings examined the viewpoints of potential end users’ perspectives regarding their preferences for and expectations of SDM from patients and HCPs. Our analysis identified the appropriate roles of health coaches vis-à-vis HCPs in SDM and underscored the importance of a health coach’s credentials and personal attributes. At the same time, the five elements for optimal implementation of SDM can be used to guide future efforts to contextualize SDM and integrate health coaches into routine primary care to support diabetes and hypertension treatment.

Acknowledgments
The authors thank the participants who generously provided their time and shared their perspectives. We would also like to thank Cassandra and Valen from the Centre for Population Health Research and Implementation for their support for the study. This study is supported by the National Research Foundation, Singapore, under its AI Singapore Programme (AISG-GC-2019-001-2A). This study is also supported by the Singapore Ministry of Health’s National Medical Research Council under its HPHSR Clinician Scientist Award (HCSAINV21jun-0004) and the Ministry of Education’s Academic Research Fund Tier 1 Funding (2022-MOET1-0005).

Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Consolidated criteria for reporting qualitative research (COREQ): 32-item checklist.
[DOCX File, 23 KB - formative_v8i1e51848_app1.docx]

References


45. McCoy MA. Examining relationships among person factors, psychosocial factors, and coping patterns in people living with type 2 diabetes mellitus in North Central Appalachia. West Virginia University. 2021. URL: https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=11189&context=etd [accessed 2024-03-08]


Abbreviations

- COREQ: Consolidated Criteria for Reporting Qualitative Research
- FGD: focus group discussion
- HCP: health care professional
- SDM: shared decision-making
The Effect of a Combined Intermittent Fasting Healthy Plate Intervention on Anthropometric Outcomes and Body Composition Among Adults With Overweight and Obesity: Nonrandomized Controlled Trial

Shazana Rifham Abdullah¹, MPH, DrPH; Ruziana Mona Wan Mohd Zin¹, BSc; Nur Hayati Azizul¹, MBBS; Nur Suffia Sulaiman¹, MPH, DrPH; Norhayati Mustafa Khalid¹, MSc; Roshan Jahn Mohd Salim Mullahi Jahn¹, MSc; Muhamad Khairul Nazrin Khalil¹, MB Ch BAO; Norhashimah Abu Seman¹, PhD; Nur Azlin Zainal Abidin¹, MD; Azizan Ali¹, MSc; You Zhuan Tan¹, MD; Azahadi Omar², MCommHealth; Zamtira Seman², MSc; Abqariyah Yahya³, PhD; Mohd Fairulnizal Md Noh¹, MSc, PhD

¹Nutrition, Metabolism and Cardiovascular Research Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia
²Sector for Biostatistic and Data Repository, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia
³Department of Social and Preventive Medicine, Faculty of Medicine, Universiti Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Shazana Rifham Abdullah, MPH, DrPH
Nutrition, Metabolism and Cardiovascular Research Centre, Institute for Medical Research
National Institutes of Health, Ministry of Health
Persiaran Setia Murni, Setia Alam
Shah Alam, 40170
Malaysia
Phone: 60 333627487
Email: shazana.a@moh.gov.my

Abstract

Background: Adult obesity and overweight pose a substantial risk to global public health and are associated with various noncommunicable diseases. Although intermittent fasting (IF) is increasingly used as a relatively new dietary strategy for weight loss, the effectiveness of 2 days per week of dry fasting remains unknown.

Objective: This study aims to evaluate the effectiveness of a combined dry IF and healthy plate (IFHP) and healthy plate (HP) intervention in improving anthropometric outcomes and body composition.

Methods: This nonrandomized controlled trial involved 177 adults who were overweight and obese. Among them, 91 (51.4%) were allocated to the IFHP group and 86 (48.6%) were allocated to the HP group. The overall study duration was 6 months (October 2020 to March 2021). The intervention was divided into 2 phases: supervised (3 months) and unsupervised (3 months). The data were collected at baseline, after the supervised phase (month 3), and after the unsupervised phase (month 6). Anthropometric (weight, height, waist circumference, and hip circumference) and body composition (body fat percentage, body fat mass, skeletal muscle mass, and visceral fat area) data were measured at all 3 data collection points. Sociodemographic data were obtained using a questionnaire at baseline.

Results: Most participants were female (147/177, 83.1%) and Malay (141/177, 79.7%). After 3 months, there were significant reductions in weight (difference −1.68; P<.001), BMI (difference −0.62; P<.001), body fat percentage (difference −0.921; P<.001), body fat mass (difference −1.28; P<.001), and visceral fat area (difference −4.227; P=.008) in the IFHP group, whereas no significant changes were observed in the HP group. Compared to baseline, participants in the IFHP group showed a significant decrease in weight (difference −1.428; P=.003), BMI (difference −0.522; P=.005), body fat percentage (difference −1.591; P<.001), body fat mass (difference −1.501; P<.001), visceral fat area (difference −7.130; P<.001), waist circumference (difference −2.304; P=.001), and hip circumference (difference −1.908; P=.002) at month 6. During the unsupervised phase, waist (IFHP difference −3.206; P<.001, HP difference −2.675; P=.004) and hip (IFHP difference −2.443; P<.001, HP difference −2.896; P<.001) circumferences were significantly reduced in both groups (P<.01), whereas skeletal muscle mass (difference 0.208;
Introduction

Background

Obesity, defined as a disease by the American Medical Association in June 2013 [1], poses greater and long-lasting harm to the collective health of adults and children worldwide. Despite being prioritized in health policies for decades, the prevalence of obesity continues to rise across countries and socioeconomic statuses. Globally, the obesity rate has almost tripled since 1975 with 39% and 13% of adults being overweight and obese, respectively [2]. A similar pattern has been observed in Malaysia where the prevalence of overweight and obesity in adults had increased by 12.6% from 2011 to 2019 [3]. In 2019, the National Health and Morbidity Survey reported that half of the adult population in Malaysia was either overweight (30.4%) or obese (19.7%) compared with 29.4% and 15.1% in 2011, respectively [3].

Obesity is a leading health problem that increases the risk of multiple health conditions such as cardiovascular diseases, type 2 diabetes mellitus, dyslipidemia, asthma, infertility, and certain types of cancers. In addition, it has adverse effects on the economy and national productivity. In 2019, the economic impact of overweight and obesity across 161 countries was estimated to be an average of 2.19% of the gross domestic product. If the current trends continue, it is projected to rise to 3.29% of the gross domestic product by 2060 [4]. Obesity has other associated indirect costs and losses, such as lost workdays, economic burdens stemming from premature mortality, and reduced work productivity [4].

Dietary modification is a keystone of weight management. As the foundation of weight loss, a state of negative balance must be achieved. Over the decades, there has been much debate on how to attain and maintain this negative energy balance by researching the most practical and feasible methods for weight loss. Studies have demonstrated that different dietary regimes are not only effective in lowering weight but also in improving the risks of chronic diseases such as type 2 diabetes mellitus, hypertension, and nonalcoholic fatty liver disease [5-8]. Intermittent fasting (IF) is a form of calorie restriction that consists of various eating plans that cycle between fasting and nonfasting states over a defined period. The effectiveness of wet IF in reducing weight and improving cardiometabolic effects has been proven in recent studies [9,10]. However, the benefits of dry fasting (except for Ramadan fasting) are not well documented. Wet IF is a form of fasting that restricts the consumption of all types of food and drink except water [11], whereas dry IF is defined as a complete fast in which no food or fluid intake is allowed [12].

A practical and easy method for restricting calorie intake is by controlling the portion size of a meal. This method has been widely practiced worldwide and differs based on the culture and eating habits [13,14]. In Malaysia, the Malaysian Healthy Plate is a portion control method that was created to translate the messages in the Malaysian Dietary Guidelines 2010 and Malaysian Food Pyramid 2010 [15]. It is a single-meal guide that suggests food intake following the quarter-quarter-half concept, which divides the plate into a quarter plate of grains or grain products; a quarter plate of fish, poultry, meat, or egg; and a half plate of fruits and vegetables [16].

Objectives

Although numerous studies have proven the effectiveness of conventional IF in reducing body weight [17], the efficacy of 2 nonconservative days per week of dry IF remains unclear. Furthermore, the reported benefits of the Malaysian Healthy Plate on improving health are still lacking, despite being widely practiced in Malaysia since 2016. To date, no other study has reported the effects of combined IF and a healthy plate (HP) dietary protocol on health. Hence, this nonrandomized controlled trial aims to determine the effectiveness of combined IF and HP (IFHP) and HP interventions in improving anthropometric outcomes and body composition. We hypothesized that there would be a significant improvement in the outcomes in both groups, with more prominent changes observed in the IFHP group because of the added IF intervention.

Methods

Study Population and Design

This nonrandomized controlled study involved a total of 177 participants with overweight or obesity. The participants were divided into 2 intervention groups, the combined IFHP group and the HP group, based on their designated workplace (the National Institutes of Health in Setia Alam; the Institute for Medical Research, Jalan Pahang; and the Institute Kementerian Kesihatan Malaysia [Teknologi Makmal Perubatan], Jalan Pahang). The distance between Jalan Pahang and Setia Alam is approximately 40 km. The allocation was conducted in a manner that aimed to prevent contamination bias and was chosen based on the practicality of monitoring the participants. The study population exhibited a rather homogeneous set of sociodemographic characteristics, environmental conditions, facility settings, and job types.

P=.04) and visceral fat area (difference −2.903; P=.003) were significantly improved in the IFHP group only. No significant difference in the between-group comparison was detected throughout the intervention (all P>.05).

Conclusions: A combined IFHP intervention was effective in improving anthropometric outcomes and body composition in adults with overweight and obesity.

International Registered Report Identifier (IRRID): RR2-10.2196/33801.

(KEYWORDS intermittent fasting; dry fasting; healthy plate; obesity; overweight

Introduction

Background

Obesity, defined as a disease by the American Medical Association in June 2013 [1], poses greater and long-lasting harm to the collective health of adults and children worldwide. Despite being prioritized in health policies for decades, the prevalence of obesity continues to rise across countries and socioeconomic statuses. Globally, the obesity rate has almost tripled since 1975 with 39% and 13% of adults being overweight and obese, respectively [2]. A similar pattern has been observed in Malaysia where the prevalence of overweight and obesity in adults had increased by 12.6% from 2011 to 2019 [3]. In 2019, the National Health and Morbidity Survey reported that half of the adult population in Malaysia was either overweight (30.4%) or obese (19.7%) compared with 29.4% and 15.1% in 2011, respectively [3].

Obesity is a leading health problem that increases the risk of multiple health conditions such as cardiovascular diseases, type 2 diabetes mellitus, dyslipidemia, asthma, infertility, and certain types of cancers. In addition, it has adverse effects on the economy and national productivity. In 2019, the economic impact of overweight and obesity across 161 countries was estimated to be an average of 2.19% of the gross domestic product. If the current trends continue, it is projected to rise to 3.29% of the gross domestic product by 2060 [4]. Obesity has other associated indirect costs and losses, such as lost workdays, economic burdens stemming from premature mortality, and reduced work productivity [4].

Dietary modification is a keystone of weight management. As the foundation of weight loss, a state of negative balance must be achieved. Over the decades, there has been much debate on how to attain and maintain this negative energy balance by researching the most practical and feasible methods for weight loss. Studies have demonstrated that different dietary regimes are not only effective in lowering weight but also in improving the risks of chronic diseases such as type 2 diabetes mellitus, hypertension, and nonalcoholic fatty liver disease [5-8]. Intermittent fasting (IF) is a form of calorie restriction that consists of various eating plans that cycle between fasting and nonfasting states over a defined period. The effectiveness of wet IF in reducing weight and improving cardiometabolic effects has been proven in recent studies [9,10]. However, the benefits of dry fasting (except for Ramadan fasting) are not well documented. Wet IF is a form of fasting that restricts the consumption of all types of food and drink except water [11], whereas dry IF is defined as a complete fast in which no food or fluid intake is allowed [12].

A practical and easy method for restricting calorie intake is by controlling the portion size of a meal. This method has been widely practiced worldwide and differs based on the culture and eating habits [13,14]. In Malaysia, the Malaysian Healthy Plate is a portion control method that was created to translate the messages in the Malaysian Dietary Guidelines 2010 and Malaysian Food Pyramid 2010 [15]. It is a single-meal guide that suggests food intake following the quarter-quarter-half concept, which divides the plate into a quarter plate of grains or grain products; a quarter plate of fish, poultry, meat, or egg; and a half plate of fruits and vegetables [16].

Objectives

Although numerous studies have proven the effectiveness of conventional IF in reducing body weight [17], the efficacy of 2 nonconservative days per week of dry IF remains unclear. Furthermore, the reported benefits of the Malaysian Healthy Plate on improving health are still lacking, despite being widely practiced in Malaysia since 2016. To date, no other study has reported the effects of combined IF and a healthy plate (HP) dietary protocol on health. Hence, this nonrandomized controlled trial aims to determine the effectiveness of combined IF and HP (IFHP) and HP interventions in improving anthropometric outcomes and body composition. We hypothesized that there would be a significant improvement in the outcomes in both groups, with more prominent changes observed in the IFHP group because of the added IF intervention.

Methods

Study Population and Design

This nonrandomized controlled study involved a total of 177 participants with overweight or obesity. The participants were divided into 2 intervention groups, the combined IFHP group and the HP group, based on their designated workplace (the National Institutes of Health in Setia Alam; the Institute for Medical Research, Jalan Pahang; and the Institute Kementerian Kesihatan Malaysia [Teknologi Makmal Perubatan], Jalan Pahang). The distance between Jalan Pahang and Setia Alam is approximately 40 km. The allocation was conducted in a manner that aimed to prevent contamination bias and was chosen based on the practicality of monitoring the participants. The study population exhibited a rather homogeneous set of sociodemographic characteristics, environmental conditions, facility settings, and job types.
This study enrolled participants aged 19 to 59 years with a BMI ≥23 kg/m² (overweight or obese) who were ready to participate in the intervention (assessed through readiness to participate in screening) and provided informed consent. The exclusion criteria included the following: (1) a recent involvement in weight loss programs (eg, IF, diet changes, physical activity changes, or activities that were performed constantly to reduce weight); (2) having any eating disorder; (3) diagnosed with diabetes and hypertension (on medication) or other metabolic health disturbance conditions; (4) taking any medication or supplements that can affect the study outcomes; (5) pregnancy; and (6) a lack of capacity or language skills to independently follow the protocol. A detailed description of the study design and methodology has been published elsewhere [18].

Sample Size Calculation
Sample size calculation was conducted using the PS: Power and Sample Size Calculation program (version 3.1.6; William D Dupont and Walton D Plummer, Jr). The sample size was calculated based on a between-group mean difference (δ) of 5% (SD 10%) in weight loss, 95% CI (±α= .05), and a power of study of 80% (1−β=.80). According to a review by Ryan and Yockey [19], a minimum weight loss of 5% is needed to improve cardiometabolic risk such as hypertension, diabetes mellitus, and hyperlipidemia. The minimum sample size required for this study was 64 participants for each arm. After considering 40% attrition, the calculated sample size for each arm was 90 participants, and a total of 180 participants were required for this study.

Dietary Protocol
Participants in the combined IFHP group were required to observe dry fasting from dawn to dusk for 2 days a week (Mondays and Thursdays) and practice HP for the rest of the week. During the fasting days, they were encouraged to have a meal before dawn, and no food or drink was allowed following that meal (approximately 13 hours) until sunset. Smoking and sexual activity were also forbidden during the fasting day, following the Islamic fasting obligation. For the remaining 5 days of the week, they were asked to apply the HP concept while consuming meals.

In the HP group, the participants were required to practice the HP concept daily, which entailed the division of plate portions into a quarter for protein, a quarter for complex carbohydrates, and a half for fruits and vegetables. Although they were advised to practice the HP concept for all 3 main meals per day, compliance with the dietary protocol was considered if they applied the concept to at least 1 main meal per day. The adherence to dietary protocols was monitored by trained research assistants through a daily record of food intake picture (1 meal per day) and a weekly fasting record.

Study Procedures
There were two stages of the recruitment phase: (1) health screening and (2) readiness to participate in screening. Those who were interested in joining the study were screened for the inclusion and exclusion criteria by the study team. Eligible volunteers were screened for their readiness to participate in the study, namely their motivation, enthusiasm, and willingness to commit to study protocols. Only participants who were ready to commit were included in the study.

The overall duration of the intervention was 6 months, which included 3 months of a supervised phase and another 3 months of an unsupervised phase. During the supervised phase, a reminder to fast was sent to their mobile phone on the eve of fasting days and compliance was recorded twice weekly. For HP, all participants were required to send at least 1 picture of their meal every day to the research assistants. On the contrary, no fasting reminder or meal pictures were exchanged during the unsupervised phase.

The data collection was conducted at 3 time points: baseline (before the intervention started), month 3 (at the end of the supervised phase), and month 6 (at the end of the unsupervised phase). During baseline data collection, the participants were asked to answer questions on sociodemographic factors. Anthropometric measurements were taken, including body composition analysis, at all 3 points of data collection.

Measurements of body weight and height were conducted using a Seca electronic column scale (Seca GmbH and Co KG) in kg and cm to the nearest 0.1 kg and 0.1 cm, respectively. Body weight was measured in light clothing and participants were required to remove their outer garments and shoes. BMI was calculated by dividing weight by height squared (kg/m²) and categorized into overweight (23.0-27.4 kg/m²), obese I (27.5-32.4 kg/m²), obese II (32.5-37.4 kg/m²), and obese III (≥37.5 kg/m²) based on the cutoff points for public health action in Malaysia [20]. Waist and hip circumferences were measured using a Seca measuring tape (Seca GmbH and Co KG) to the nearest 0.1 cm, with the participants standing. Waist circumference was measured at the midpoint between the top of the iliac crest and the lower margin of the last palpable rib, whereas hip circumference was measured at the widest diameter around the buttocks. To minimize measurement error, 2 measurements were taken for each parameter and the average was calculated. The waist-to-hip ratio was calculated by dividing the waist measurement by the hip measurement.

For body composition analysis, parameters such as body fat percentage, body fat mass, muscle mass, and visceral fat area were measured. These parameters were measured using a tetrapolar bioimpedance multifrequency InBody 770 analyzer (Biospace). Personal profiles (age, height, weight, and sex) were entered after reading the measurement. Before the measurement, the participants were asked to remove all metal items and stand barefoot on the device in a supine position. They then grasped the handles of the units with their thumbs and palms while maintaining direct contact with the electrode. In total, 8 polar tactile electrodes were used in the bioelectrical impedance analysis (BIA): 2 for each thumb, 2 for the palms, and 2 for the front and 2 for the back of each foot. The body composition results were computed using proprietary prediction algorithms that were built into the firmware of the device and applied only to the device being studied. The total time required for the measurement was approximately 2 minutes [21].

The International Physical Activity Questionnaire (IPAQ)–Short Form was used to assess the physical activity of the participants.
over the past week. The questionnaire has been validated for use among adults in 12 countries [22]. In this study, we used the Malay version of the IPAQ-Short Form, which was validated using the data from the 2011 National Health and Morbidity Survey [23]. The participants were asked to record the number of days in the previous week that they engaged in specific activities (vigorous and moderate activities and walking) for at least 10 minutes as well as the amount of time (in min) that they engaged in a particular activity on an average day. Energy expenditure or metabolic equivalent of task (MET) minutes per week was used to determine the physical activity based on the IPAQ scoring protocol. To calculate the MET scores for each activity, the total minutes spent on vigorous activity, moderate activity, and walking over the past week were multiplied by 8, 4, and 3.3, respectively. The total physical activity score was calculated as the sum of the MET scores for each of the 3 activity categories. All participants were asked to maintain their usual physical activity throughout the study period. The physical activity score was included as one of the controlled variables in the analysis.

The Food Frequency Questionnaire (FFQ) was used to measure the participants’ dietary intake over the previous month. The questionnaire consisted of questions covering the frequency of consuming cereals and cereal products, fast food, meat and meat products, fish and seafood, eggs, legumes and legume products, milk and milk products, vegetables, fruits, drinks, alcoholic drinks, confectionaries, bread spreads, and flavor intake. We used the validated Malay version of the FFQ, which consists of 165 items [24], and the participants required approximately 30 minutes to answer the questions at each point of data collection. We calculated the calorie intake using nutrient data that were extracted from the Nutritionist Pro nutrition analysis software (version 7.8.0; Axxya Systems). The calorie intake changes observed in this study confirmed the dietary protocol adherence.

**Statistical Analyses**

Statistical analysis was performed using the SPSS software (version 25; IBM Corp). For continuous variables, the normality of distribution was tested using the Kolmogorov-Smirnov test. Normally distributed data were presented as mean (SD), whereas the median (IQR) summarized the skewed data. For categorical variables, frequencies were calculated and presented as percentages. All variables were compared using the independent 2-tailed t test or Mann-Whitney U test for continuous variables and the chi-square or Fisher exact test (n ≤ 5 in any cell) for categorical variables. To determine the changes in outcome following the intervention, a generalized estimating equation (GEE) analysis was performed, which was adjusted for possible confounders such as age, ethnicity, sex, and physical activity. The GEE statistical analysis has been widely used owing to its robustness and ability to analyze correlated, nonnormally distributed data [25]. Compared with the repeated measures ANOVA, GEE provides more flexibility in handling missing data and accommodates unbalanced designs. The time effect was analyzed for all 3 comparisons: baseline versus month 3, baseline versus month 6, and month 3 versus month 6. To control for the overall type I error rate, we adjusted for multiple comparisons using the Bonferroni correction method in the GEE analysis. For the calorie intake analysis, we excluded participants with extreme calorie intake (<500 and >3500 kcal/day for women and <800 and >4000 kcal/day for men) [26]. All statistical tests were 2-sided, and the significance level was set at .05.

**Ethical Considerations**

This study was approved by the Medical Research and Ethics Committee, Ministry of Health Malaysia (NMRR-19-3261-51726) and was conducted in full conformity with the current revision of the Declaration of Helsinki and the International Council for Harmonisation Guidelines for Good Clinical Practice. Before recruitment, all participants were thoroughly briefed about the potential risks involved with participating in this study, and written informed consent was obtained from them. The participants had the autonomy to discontinue their involvement in the study at any point. The data were treated as strictly confidential, and each participant was assigned a unique anonymous identifier. Each participant received a reimbursement of RM30 (US $6) after each data collection, resulting in a total compensation of RM60 (US $13) per participant across the study period. This study was registered with ClinicalTrials.gov (NCT05034653).

**Results**

**Flow of Study Participants**

Overall, 302 volunteers were interested in joining the study, of which 99 (32.8%) were excluded because of the failure to meet the eligibility criteria during the first screening stage. Another 27 volunteers were excluded during the screening process in stage 2 for multiple reasons, such as pursuing further education, commitment issues, or relocating. Hence, a total of 177 participants were recruited in this study: 91 (51.4%) in the IFHP group and 86 (48.6%) in the HP group. During the supervised phase of the intervention, a total of 28 participants withdrew from the study (IFHP: 16/28, 57%; HP: 12/28, 43%), whereas 27 participants withdrew during the unsupervised phase (IFHP: 12/27, 44%; HP: 15/27, 56%). The reasons for dropout included the inability to commit to the study intervention (24/55, 44%); pregnancy (13/55, 24%); transfer to a different workplace (8/55, 15%); started medication for hypertension, diabetes, or hypercholesterolemia (5/55, 9%); and other reasons (5/55, 9%). Finally, 63 and 59 participants completed the study in the IFHP and HP groups, respectively (Figure 1).
Sociodemographic Characteristics

Table 1 summarizes the characteristics of the participants based on the intervention groups at baseline. A majority of the participants were female (147/177, 83.1%) and Malay (141/177, 79.6%) with a mean age of 34.47 (SD 7.40) years. Most participants (121/177, 68.4%) had a diploma or degree as the highest educational status. In terms of background illness history, 5.1% (9/177) and 1.1% (2/177) of the participants reported being diagnosed with hyperlipidemia and hypertension, respectively. However, none of the participants were taking any medication for the illnesses. Participants in the IFHP group had a higher mean weight (72.50, SD 18.55 kg) than those in the HP group (70.60, SD 16.28 kg); however, this difference was not statistically significant ($P=.13$). In addition, no significant differences in physical activity, daily calorie intake, other body composition, and anthropometric parameters were found between the participants from both intervention groups (Table 1). Detailed $t$ test results are presented in Multimedia Appendix 1.
Table 1. Baseline characteristics of the participants based on the intervention groups.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=177)</th>
<th>Intervention groups</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IFHP&lt;sup&gt;a&lt;/sup&gt; (n=91)</td>
<td>HP&lt;sup&gt;b&lt;/sup&gt; (n=86)</td>
</tr>
<tr>
<td><strong>Sociodemographics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>30 (16.9)</td>
<td>18 (19.8)</td>
<td>12 (14)</td>
</tr>
<tr>
<td>Female</td>
<td>147 (83.1)</td>
<td>73 (80.2)</td>
<td>74 (86)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>4 (2.3)</td>
<td>0 (0)</td>
<td>4 (4.7)</td>
</tr>
<tr>
<td>Indian</td>
<td>12 (6.8)</td>
<td>4 (4.4)</td>
<td>8 (9.3)</td>
</tr>
<tr>
<td>Malay</td>
<td>141 (79.6)</td>
<td>80 (87.9)</td>
<td>61 (70.9)</td>
</tr>
<tr>
<td>Others</td>
<td>20 (11.3)</td>
<td>7 (7.7)</td>
<td>13 (15.1)</td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>34.47 (7.40)</td>
<td>33.82 (7.50)</td>
<td>35.15 (7.27)</td>
</tr>
<tr>
<td><strong>Highest education status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary school</td>
<td>24 (13.6)</td>
<td>13 (14.3)</td>
<td>11 (12.8)</td>
</tr>
<tr>
<td>Diploma or degree</td>
<td>121 (68.4)</td>
<td>62 (68.1)</td>
<td>59 (68.6)</td>
</tr>
<tr>
<td>Master or PhD</td>
<td>32 (18.1)</td>
<td>16 (17.6)</td>
<td>16 (18.6)</td>
</tr>
<tr>
<td><strong>Background illness, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>2 (1.1)</td>
<td>1 (1.1)</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>9 (5.1)</td>
<td>6 (6.6)</td>
<td>3 (3.5)</td>
</tr>
<tr>
<td>Heart disease</td>
<td>1 (0.6)</td>
<td>0 (0)</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>Others</td>
<td>1 (0.6)</td>
<td>0 (0)</td>
<td>1 (1.2)</td>
</tr>
<tr>
<td>Physical activity (MET&lt;sup&gt;e&lt;/sup&gt;-min/wk), median (IQR)</td>
<td>829.50 (769.48)</td>
<td>706 (1055.70)</td>
<td>627 (629.50)</td>
</tr>
<tr>
<td>Calorie intake (kcal/d), mean (SD)</td>
<td>2093.68 (645.08)</td>
<td>2082.03 (667.01)</td>
<td>2106.14 (625.21)</td>
</tr>
<tr>
<td><strong>Anthropometrics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body weight (kg), median (IQR)</td>
<td>71.70 (18.70)</td>
<td>72.50 (18.55)</td>
<td>70.60 (16.28)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>158.36 (7.09)</td>
<td>158.67 (7.13)</td>
<td>158.04 (7.08)</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), median (IQR)</td>
<td>28.43 (6.50)</td>
<td>28.48 (7.25)</td>
<td>28.39 (5.70)</td>
</tr>
<tr>
<td><strong>BMI category, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>63 (35.6)</td>
<td>32 (35.2)</td>
<td>31 (36)</td>
</tr>
<tr>
<td>Obese class I</td>
<td>69 (39.0)</td>
<td>32 (35.2)</td>
<td>37 (43)</td>
</tr>
<tr>
<td>Obese class II</td>
<td>32 (18.1)</td>
<td>20 (22.0)</td>
<td>12 (14)</td>
</tr>
<tr>
<td>Obese class III</td>
<td>13 (7.3)</td>
<td>7 (7.6)</td>
<td>6 (7)</td>
</tr>
<tr>
<td>Waist circumference (cm), mean (SD)</td>
<td>92.46 (10.88)</td>
<td>93.49 (10.70)</td>
<td>91.36 (11.03)</td>
</tr>
<tr>
<td>Hip circumference (cm), mean (SD)</td>
<td>108.86 (8.66)</td>
<td>109.39 (9.02)</td>
<td>108.31 (8.29)</td>
</tr>
<tr>
<td><strong>Body composition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body fat percentage (%), mean (SD)</td>
<td>44.09 (6.17)</td>
<td>44.21 (6.36)</td>
<td>43.96 (6)</td>
</tr>
<tr>
<td>Body fat mass (kg), mean (SD)</td>
<td>33.26 (8.96)</td>
<td>34.08 (9.37)</td>
<td>32.37 (8.45)</td>
</tr>
<tr>
<td>Skeletal muscle mass (kg), mean (SD)</td>
<td>22.71 (4.90)</td>
<td>23.36 (4.80)</td>
<td>22.13 (4.96)</td>
</tr>
<tr>
<td>Visceral fat area (cm&lt;sup&gt;2&lt;/sup&gt;), mean (SD)</td>
<td>166.52 (40.44)</td>
<td>168.13 (41.32)</td>
<td>164.77 (39.62)</td>
</tr>
</tbody>
</table>
Changes in Anthropometric and Body Composition

Baseline Versus Month 3

Within the IFHP group, there was a significant reduction in weight (estimated margin [EM] mean difference −1.680; \(P<.001\)), BMI (EM mean difference −0.620; \(P<.001\)), body fat percentage (EM mean difference −0.921; \(P<.001\)), body fat mass (EM mean difference −1.280; \(P<.001\)), and visceral fat area (EM mean difference −4.227; \(P=.008\)) at month 3 compared to baseline. Meanwhile, no significant changes were observed in skeletal muscle mass (\(P=.43\)), waist circumference (\(P=.45\)), and hip circumference (\(P=.80\); Table 2).

Table 2. Comparison of the anthropometric and body composition changes between the intermittent fasting healthy plate (IFHP) and healthy plate (HP) groups at baseline, month 3, and month 6.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Time</th>
<th>Time effect</th>
<th>Group effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline, mean (SE)</td>
<td>Baseline vs month 3, EM mean difference ((P) value)</td>
<td>Baseline vs month 6, EM mean difference ((P) value)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>78.83 (1.96)</td>
<td>−1.68 (&lt;.001)</td>
<td>−1.428 (.003)</td>
</tr>
<tr>
<td>HP</td>
<td>76.96 (2.55)</td>
<td>−0.317 (.49)</td>
<td>−0.133 (&gt;.99)</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>29.45 (0.64)</td>
<td>−0.62 (&lt;.001)</td>
<td>−0.522 (.005)</td>
</tr>
<tr>
<td>HP</td>
<td>28.69 (0.7)</td>
<td>−0.101 (.78)</td>
<td>−0.023 (&gt;.99)</td>
</tr>
<tr>
<td>Body fat percentage (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>39.7 (0.72)</td>
<td>−0.921 (&lt;.001)</td>
<td>−1.591 (&lt;.001)</td>
</tr>
<tr>
<td>HP</td>
<td>39.57 (0.97)</td>
<td>−0.168 (.98)</td>
<td>−0.175 (&gt;.99)</td>
</tr>
<tr>
<td>Body fat mass (kg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>31.38 (1.36)</td>
<td>−1.28 (&lt;.001)</td>
<td>−1.501 (&lt;.001)</td>
</tr>
<tr>
<td>HP</td>
<td>30.58 (1.5)</td>
<td>−0.206 (.78)</td>
<td>−0.072 (&gt;.99)</td>
</tr>
<tr>
<td>Skeletal muscle mass (kg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>26.34 (0.4)</td>
<td>−0.144 (.43)</td>
<td>0.064 (&gt;.99)</td>
</tr>
<tr>
<td>HP</td>
<td>25.62 (0.87)</td>
<td>0.151 (.98)</td>
<td>0.171 (&gt;.99)</td>
</tr>
<tr>
<td>Visceral fat area (cm(^2))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>148.88 (6.29)</td>
<td>−4.227 (.008)</td>
<td>−7.13 (&lt;.001)</td>
</tr>
<tr>
<td>HP</td>
<td>148.51 (7.41)</td>
<td>−0.118 (&gt;.99)</td>
<td>−0.597 (&gt;.99)</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>98.14 (1.56)</td>
<td>0.902 (.45)</td>
<td>−2.304 (.001)</td>
</tr>
<tr>
<td>HP</td>
<td>93.9 (1.85)</td>
<td>0.823 (.45)</td>
<td>−1.852 (.10)</td>
</tr>
<tr>
<td>Hip circumference (cm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFHP</td>
<td>109.13 (1.3)</td>
<td>0.535 (.80)</td>
<td>−1.908 (.002)</td>
</tr>
<tr>
<td>HP</td>
<td>106.91 (1.48)</td>
<td>1.861 (&lt;.001)</td>
<td>−1.034 (.03)</td>
</tr>
</tbody>
</table>

P values <.05 are considered statistically significant. P values were derived from the GEE test adjusted for sex, age, ethnicity, and physical activity.
body fat mass (EM mean difference $-0.206; P=.78$), and visceral fat area (EM mean difference $-0.118; P>.99$). However, these changes were not statistically significant. The hip circumference significantly increased at month 3 compared to baseline (EM mean difference 1.861; $P<.001$).

There were no significant changes in other parameters (all $P>.05$; Table 2).

**Baseline Versus Month 6**

After 6 months, participants in the IFHP group showed significant reduction in weight (EM mean difference $-1.428; P=.003$), BMI (EM mean difference $-0.522; P=.005$), body fat percentage (EM mean difference $-1.591; P<.001$), body fat mass (EM mean difference $-1.501; P<.001$), body fat area (EM mean difference $-7.130; P<.001$), waist circumference (EM mean difference $-2.304; P<.001$), and hip circumference (EM mean difference $-1.908; P=.002$). However, there were no significant changes observed in skeletal muscle mass ($P>.99$; Table 2).

Moreover, hip circumference (EM mean difference $-1.034; P=.03$) showed a significant reduction among participants in the HP group after 6 months. Although there was a reduction in weight, BMI, body fat percentage, body fat mass, and visceral fat area, it was not statistically significant (all $P>.99$). The increase in skeletal muscle mass was also not statistically significant ($P>.99$; Table 2).

**Month 3 Versus Month 6**

Comparing month 6 to month 3, there were significant improvements in skeletal muscle mass (EM mean difference $0.208; P=.04$), visceral fat area (EM mean difference $-2.903; P=.003$), waist circumference (EM mean difference $-3.206; P<.001$), and hip circumference (EM mean difference $-2.443; P<.001$) in the IFHP group. Otherwise, there were no significant changes in other parameters (all $P>.05$; Table 2).

Furthermore, participants in the HP group showed a significant reduction in waist circumference (EM mean difference $-2.675; P=.004$) and hip circumference (EM mean difference $-2.896; P<.001$) at month 6 compared to month 3. However, no other significant changes were observed (all $P>.05$; Table 2).

In view of the between-group comparison, there were no significant differences between the IFHP and HP groups observed at all time points (all $P>.05$; Table 2).

**Calorie Intake Changes**

There was a significant reduction in calorie intake at month 3 and month 6 compared to baseline in both intervention groups (Table 3). Compared to baseline, participants in the IFHP group consumed an average of 325 kcal and 266 kcal fewer calories per day after 3 months (EM mean difference $-325.09; P=.002$) and 6 months (EM mean difference $-266.70; P=.006$), respectively. Similarly, there was a significant reduction in calorie intake at month 3 (EM mean difference $-320.11; P=.004$) and month 6 (EM mean difference $-437.79; P<.001$) compared to baseline in the HP group. However, there were no significant changes in calorie intake between month 3 and month 6 in both the IFHP ($P>.99$) and HP ($P=.73$) groups. There was also no significant difference in the between-group comparison throughout the intervention ($P=.94$; Table 3).

**Discussion**

**Principal Findings**

To our knowledge, this is the only study to investigate the effect of IFHP regimens on anthropometric outcomes and body composition worldwide. This study found that a combined IFHP diet had a notable effect on weight, BMI, body fat percentage, body fat mass, and visceral fat area after 3 and 6 months of intervention. Although there were no changes in weight and BMI in the IFHP group at month 6 compared to those at month 3, visceral fat area was greatly reduced, whereas skeletal muscle mass showed a substantial improvement. After 6 months, a notable reduction in hip and waist circumference was observed among participants in both groups. The improvement was seen during the unsupervised period as there were no notable changes...
in these parameters at month 3. This study showed mixed findings when comparing supervised and unsupervised phases. After completing the supervised phase at month 3, the IFHP group showed substantial reductions in weight, BMI, body fat percentage, body fat mass, and visceral fat area, but not in skeletal muscle mass, waist circumference, or hip circumference. Meanwhile, only skeletal muscle mass, waist circumference, hip circumference, and visceral fat area improved significantly during the unsupervised phase. Overall, there was no notable difference between groups throughout the study period.

Comparisons With Previous Works, Interpretations, and Implications

The short- and long-term effectiveness of the 2-day per week fasting in reducing weight has been reported in various studies [27-30]. Generally, the concept of a 5:2 IF diet is defined in most studies as restricting calorie intake to approximately 25% of the baseline energy intake twice a week, which makes it slightly different from our IF protocol. Furthermore, any zero-calorie beverages, such as plain water, were allowed during the fasting period. In 2013, a randomized controlled study was conducted to determine the efficacy of combined fasting (Islamic Sunnah Fasting for 2 days per week) and calorie restriction (reduction of 300-500 kcal/day from habitual energy intake) among older men for 3 months. The study reported a significant interaction effect in body weight, BMI, fat percentage, fat mass, and fat-free mass after 3 months [31].

Studies reported in a recent systematic review and meta-analysis showed a positive effect of a portion control diet on weight and body composition [32]. The concept of the plate model, which was first introduced by the Community Nutrition Group of the British Dietetic Association and the Swedish Diabetic Association, provides a simple and practical visual tool to demonstrate the healthy portion of a meal intake [33]. Although the division and type of food in a meal differ according to culture and eating habits, the foundation concepts of a “plate model” or HP are calorie restriction and healthy redistribution of macronutrients, such as carbohydrates, proteins, and fibers. Our study found that among those on the HP diet alone, there were reductions in weight, BMI, body fat percentage, body fat mass, and visceral fat area after 3 and 6 months, but the changes were not statistically significant. This suggests that the addition of the IF diet 2 days per week to the HP diet had a greater impact on weight loss and improvement in body composition.

Notably, there was >4 cm² and 7 cm² loss of visceral fat area among participants in the IFHP group after 3 and 6 months, respectively. Our findings are in agreement with those of previous studies [34,35]. In a study conducted among overweight older men, there was a significant reduction in the visceral fat area following a 6-week time-restricted feeding intervention [35]. The role of fasting in facilitating lipolysis has been studied extensively. Fasting stimulates intracellular lipolysis, which is triggered by several hormonal changes such as decreased levels of insulin and increased cortisol, catecholamines, and growth hormone levels. Lipolysis is also stimulated by the sympathetic innervation of the adipose tissue [36]. A classical study supported the role of these hormones in triggering the lipolysis of adipose tissue following fasting by reporting that hypophysectomy or adrenalectomy in rats caused a reduction in plasma nonesterified fatty acid levels [37].

Central adiposity, represented by abnormal waist circumference, waist-to-hip ratio, and waist-to-height ratio, is highly associated with a higher risk of noncommunicable diseases such as cardiovascular disease and cancer [38-40]. Although both intervention groups showed significant reductions in waist and hip circumference after 6 months, greater mean differences were observed among participants in the IFHP group. Previous studies reported a similar significant reduction in waist and hip circumference after IF [34,41,42]. The London Ramadan Study is an observational study conducted during Ramadan to explore the health consequences of Ramadan IF. Besides other findings, this study showed a significant decrease in waist (mean difference ~1.93 cm; *P*<.01) and hip circumference measurements (mean difference ~2.86 cm; *P*<.01) [41]. In the IFHP group, we observed a significant decrease in both parameters during the unsupervised phase of the intervention (from month 3 to month 6). Although there was an insignificant increase in weight and BMI during this period, body fat percentage and visceral fat area were significantly reduced, which may explain the improvement in waist and hip measurements.

Apart from the weekly fasting record and daily meal picture, the reduction in calorie intake served as one of the adherence monitoring methods for the dietary protocols used in our study. The significant reduction in calorie intake at months 3 and 6 explained the improvement observed in the anthropometric outcomes and body composition parameters, especially among participants in the IFHP group. The findings also confirmed the principle of calorie restriction in IF and HP without the need to properly count the calorie intake allowable to be consumed per day to cause weight loss. A recent study reported no significant difference between time-restricted feeding and calorie restriction interventions in calorie reduction and weight loss after 3 months [43]. This finding provides an advantage to both interventions as they are more practical and easier to comply with.

Improvements in clinical end points and cardiometabolic biomarkers following changes in body composition have been reported in various studies [44,45]. However, for weight changes, it has been suggested that different degrees of weight loss contribute to different biomarker responses [46]. On the basis of the findings of a randomized controlled trial conducted among obese Americans, a 5% weight loss significantly improved the plasma concentrations of some cardiometabolic risk factor parameters, such as insulin, triglyceride, glucose, leptin, and alanine transaminase, but did not have a significant effect on other parameters (low-density lipoprotein, high-density lipoprotein, free fatty acids, and adiponectin). Only after 16% weight loss, the plasma concentration of free fatty acids, C-reactive protein, and adiponectin improved significantly. Moreover, this study demonstrated that liver and adipose tissue sensitivity improved with a 5% weight reduction and remained stable, whereas muscle insulin sensitivity continued to improve with weight losses of 11% and 16%, respectively [46].

The main advantages of the BIA method in the assessment of body composition are its noninvasiveness, lack of the necessity...

https://formative.jmir.org/2024/1/e51542

JMIR Form Res 2024 | vol. 8 | e51542 | p.2191
for highly qualified personnel, short duration, and active involvement of the participants. Furthermore, the repeatability of the BIA test is crucial for observing changes in the body composition of an individual, either in a clinical setting or for research purposes [47]. However, this test also has its limitations. Several factors have been shown to influence the measurement parameters, including body structure, obesity status, body temperature, hydration status, water and electrolyte imbalance, physical activity, and errors while performing the procedure [47]. This study tried to reduce some of these limitations by properly preparing for the examination, such as instructing the participants to fast at least 6 hours before the procedure and ensuring compliance with the procedure during measurements. Apart from the abovementioned limitations, the literature has reported inconsistent results among ethnicities. On the basis of current data, BIA is still not recommended for African American individuals, but studies have shown that it provided valid results in other populations, including Asian populations, by using race-specific equations [48].

Strengths and Limitations

This study has several strengths. To our knowledge, this study is 1 of 2 studies that were conducted to investigate the health impact of dry IF twice per week. However, although the other study [31] was conducted among overweight older men, our study allowed for higher generalization of the findings as a broader range of age group (19 to 59 years) of participants and both sexes were involved. Two days of voluntary dry fasting per week (on Mondays and Thursdays) is commonly practiced in countries with a predominantly Muslim population, including Malaysia. The positive implications on health demonstrated in our study will encourage people to practice this IF protocol as a weight management method beyond religious obligation. Those who are interested in practicing but are not accustomed to the 13-hour dry fast, such as non-Muslims, must take special precautions to avoid dehydration, lethargy, and electrolyte imbalance. In addition, although the Malaysian Healthy Plate has been introduced since 2016, there is no study that investigates its effectiveness in improving health. Our findings have added new evidence to the body of knowledge and evaluated the outcome of the Malaysian Healthy Plate at the same time.

The main limitation of this study was the impact of the movement control order implemented during the COVID-19 pandemic on the findings of the study and the participants’ compliance with the intervention. As civil servants, many of our participants worked from home during most of the intervention period. Staying at home may limit their physical activity, influence their food intake control, reduce motivation toward weight loss, and expose them to unhealthy eating. The negative consequences of social lockdowns on weight management and weight-related behavior have been reported in several studies [49,50]. Furthermore, the use of FFQ to measure dietary consumption has several limitations. Despite the requirement for good memory and literacy among participants, the food-based FFQ used in our study may also have led to response errors if the participants did not prepare their food and were unaware of the ingredients. Moreover, the dietary intake tends to be underestimated because various seasonings and culinary oils that significantly contribute to energy and nutrient intakes are not considered when calculating dietary intakes [51].

Conclusions

To the best of our knowledge, this study is the first to introduce a protocol for healthy eating that combines 2-day per week dry IF with the Malaysian Healthy Plate. Our study found that a combined dry IF and HP diet was effective in reducing weight and improving body composition. Although practicing the HP protocol alone has been demonstrated to reduce weight and improve body composition to some extent, the inclusion of a dry IF component is necessary to further amplify these effects. The significant reduction in calorie consumption in both groups validates the efficacy of both dietary protocols as a feasible and practical method for calorie restriction.

Acknowledgments

The authors would like to thank the Director General Health of Malaysia for granting permission to publish this paper. The authors acknowledge the participants and are grateful to the study teams for their support and contribution to this study. This study was funded by the Ministry of Health Malaysia Research grant NMRR-19-3261-51726.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions

SRA, NHA, RMWMZ, NSS, NMK, RJMSMJ, MKNK, NAS, NAZA, AA, YZT, AO, ZS, and MFMN conceptualized the study and contributed to the design and implementation of the research. SRA drafted the first version of the manuscript. SRA, RMWMZ, ZS, and AY conducted data analyses. All authors reviewed and edited the subsequent drafts of the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.
References


11. Capritto A. What is water fasting? VeryWell Fit. URL: https://www.verywellfit.com/is-water-fasting-safe-4588873 [accessed 2021-08-09]


31. Sundfør TM, Svendsen M, Tonstad S. Effect of intermittent versus continuous energy restriction on weight loss, maintenance.


35. Goodman HM, Knobil E. Growth hormone and fatty acid mobilization: the role of the pituitary, adrenal and thyroid.


40. Teng NI, Shahar S, Rajab NF, Manaf ZA, Johari MH, Ngah WZ. Improvement of metabolic parameters in healthy older adults: a prospective study of 0.5 million people.


Abbreviations
BIA: bioelectrical impedance analysis
EM: estimated margin
FFQ: Food Frequency Questionnaire
GEE: generalized estimating equation
HP: healthy plate
IF: intermittent fasting
IFHP: intermittent fasting healthy plate
IPAQ: International Physical Activity Questionnaire
MET: metabolic equivalent of task

©Shazana Rifham Abdullah, Ruziana Mona Wan Mohd Zin, Nur Hayati Azizul, Nur Suffia Sulaiman, Norhayati Mustafa Khalid, Roshan Jahn Mohd Salim Mullahi Jahn, Muhamad Khalirul Nazrin Khalil, Norhashimah Abu Senam, Nur Azlin Zainal Abidin, Azizan Ali, You Zhuan Tan, Azahadi Omar, Zamtira Seman, Abqariyah Yahya, Mohd Fairulnizal Md Noh. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Effect of the COVID-19 Pandemic on Gambling Behavior in Mainland Chinese Gamblers in Macau: Cross-Sectional Survey Study

Jinquan Zhou\(^1\), PhD; Hong-Wai Ho\(^2\); ChiBiu Chan\(^3\)

\(^1\)Macao Polytechnic University, Macao, China
\(^2\)Centre for Gaming and Tourism Studies, Macao Polytechnic University, Macao, China
\(^3\)Faculty of Business, Macao Polytechnic University, Macao, China

Corresponding Author:
Jinquan Zhou, PhD
Macao Polytechnic University
R de Luís Gonzaga Gomes
Macao
China
Phone: 853 88593326
Email: jqzhou@mpu.edu.mo

Abstract

Background: This study examined the effects of the COVID-19 pandemic on the gambling behavior of individuals who were already actively engaged in such pursuits. We aimed to uncover the intricate consequences of the pandemic on this specific demographic, emphasizing the importance of understanding the complex connection between public health concerns such as the COVID-19 pandemic and gambling behavior from a public health perspective. In addition to identifying immediate impacts, this study holds significance in assessing potential long-term public health implications for the broader gambling industry.

Objective: This study investigated how the COVID-19 pandemic has affected the gambling behavior of Mainland Chinese tourists in Macau from a public health perspective. We aimed to understand the changing patterns of gambling habits within this specific demographic by comparing their behavior before and during the pandemic, with a particular emphasis on the evolving dynamics of gambling and their public health consequences. This study provides a detailed exploration of the impact and implications of global health emergencies on this particular demographic’s gambling behaviors and preferences.

Methods: This study used a robust cross-sectional analysis involving a sample of 334 Mainland Chinese gamblers with prior experiences in casinos in Macau. The sample deliberately encompassed individuals involved in gambling before and during the COVID-19 pandemic. Data were collected through carefully designed questionnaires to gather information on gambling habits, preferences, and observed behavioral changes in the sample.

Results: This study unveiled a notable shift in Mainland Chinese gamblers’ behavior during the COVID-19 pandemic. A considerable number of participants opted for web-based platforms over traditional land-based casinos, resulting in reduced budgets, less time spent on gambling, and decreased participation in social gambling. Remarkably, there was a notable surge in online gambling, indicating a noteworthy adaptability of gamblers to changing circumstances. These findings emphasize the dynamic nature of gambling habits during global public health emergencies, revealing the resilient and evolving preferences of Mainland Chinese gamblers in response to the challenges posed by the pandemic.

Conclusions: This study highlights the negative impact of the COVID-19 pandemic on casino gambling, notably evident in a significant decline in Mainland Chinese tourists visiting Macau for gambling. There is a noticeable shift from traditional gambling to web-based alternatives, with individuals seeking options within the pandemic constraints. Furthermore, the findings point out an increase in gambling among the younger generation and behavioral changes in individuals with mood disorders. The findings of this study emphasize the critical need for proactive measures to address evolving gambling preferences and associated risks during public health crises; furthermore, these findings underscore the importance of adaptive strategies within the gambling industry, as well as the necessity for effective public health interventions and regulatory frameworks to respond to unprecedented challenges with efficacy and precision.

(JMIR Form Res 2024;8:e45700) doi:10.2196/45700
KEYWORDS
Chinese gamblers; gambling behavior; online gambling; COVID-19; Macau

Introduction

Background
Gambling addiction is widely acknowledged as a significant public health concern [1,2]. Some common types of gambling-related harm include financial difficulties, mental health problems, relationship disruption, emotional distress, cultural harm, and also involvement in criminal activities [3]. In Macau, gambling plays a vital role in driving economic growth, tourism, and job creation. Since 2006, Macau has risen to prominence as the world’s largest gambling market, surpassing Las Vegas [4,5], with Mainland Chinese tourists constituting a substantial 93% of the city’s gambling clientele [6]. While this success has undoubtedly contributed significantly to tourism and economic development, it has also brought forth a concerning issue—the rise of gambling addiction among the influx of Chinese tourists. Despite their pivotal role in Macau’s prosperity, the substantial number of Chinese tourists engaging in gambling activities raises pressing concerns regarding the prevalence of problem gambling within this demographic.

The COVID-19 pandemic, caused by SARS-CoV-2, has been a global health crisis since late 2019 and continues to affect the world today [7]. Similar to many other regions, Macau has not been immune to the rapid spread of the virus. In response, the region implemented stringent measures to control the transmission, including a mandatory 14-day quarantine and a required nucleic acid test for entrants. This pandemic has evolved into a worldwide public health crisis, with profound economic repercussions on a global scale [8]. Macau’s gambling industry, in particular, faced significant challenges, experiencing a substantial decrease in gambling revenue [9]. This decline underscores the impact of the pandemic on gamblers’ behavior and the challenges faced by the sector due to public health measures. The highly contagious nature of the coronavirus prompted a crisis in environmental health for tourists and gamblers [10]. In response, public health authorities prioritized anti–COVID-19 measures, ranging from recommendations to enforceable mandates. In Macau, the local government adopted a stringent approach, implementing a zero-tolerance policy with mandatory quarantine and nucleic acid testing at entry and exit points. Furthermore, casino operators adopted measures to limit the concentration of gamblers and reduce the risk of virus transmission. Aligned with the Chinese government, the authorities in Macau implemented a health code system for public health and safety. The COVID-19 pandemic significantly impacted the casinos in Macau, leading to lasting negative effects on the gambling and tourism industries. Some casinos were forced to close due to the challenges posed by the pandemic [11].

The growing global prevalence of gambling and its potential impact on public health prompted us to analyze the evolving patterns of gambling. Chinese gamblers constitute the primary course of patronage for casinos across Asia [12]. While gambling is prohibited in Mainland China, Macau stands out as a legal gaming destination due to its Special Administrative Region (SAR) status under the “one country, two systems” principle. This study holds particular relevance in understanding the behavior of Chinese gamblers within regional and global public health environments. The COVID-19 pandemic has significantly influenced gambler behavior, including factors such as travel restrictions [13], risk assessment [14], alternative attractions, the impact of destination marketing strategies [15], and the rise of online gambling [16]. Macau’s distinctive position as a major gambling hub, heavily dependent on tourism, presents an intriguing case to investigate the public health implications of COVID-19 and gambling addiction. However, there is a notable research gap in understanding the changes in gambling behavior among Mainland Chinese gamblers in Macau during the pandemic. Examining these shifts is crucial to addressing the existing knowledge gap in this context and gaining a comprehensive understanding of the effects of the pandemic on gambling patterns and preferences. This knowledge will enable the implementation of evidence-based public health policies and practices that prioritize the well-being and safety of both local residents and tourists in the postpandemic era.

We aimed to investigate the gambling behaviors of Chinese gamblers during the COVID-19 pandemic, specifically by comparing their past and current behaviors through a cross-sectional analysis. Furthermore, this study explores the factors that influence these behaviors within the context of public health measures. The findings offer valuable insights and recommendations for industry executives and policy makers, facilitating a comprehensive understanding of the shifts in gambling patterns during public health crises. Considering the importance of public health, these insights could also serve as a crucial resource for guiding strategic adjustments in the gambling industry.

The Literature
COVID-19 Impact on Tourism: Historical Context and Consequences in Macau
Gambling is a recognized public health issue with wide-ranging implications. The recent COVID-19 pandemic has brought forth significant public health implications for the gambling industry worldwide [16]. In response to the pandemic, the Macau SAR government implemented various proactive measures to safeguard public health, including specific guidelines and regulations for casinos [17]. Following the confirmation of the first COVID-19 case on January 22, 2020, the local government temporarily suspended casino and entertainment businesses for half a month starting February 4, 2020. Furthermore, nonessential services were postponed, with entry bans for nonresidents implemented from March 18, 2020, except for individuals from Mainland China, Hong Kong, Taiwan, and nonresident workers. Effective from May 11, 2020, entry requirements included a negative COVID-19 test within 7 days or a Macau Guangdong Health Code [18]. Macau gradually resumed normal casino operations in August 2020, but foreign tourists were required to undergo nucleic acid testing within 7
days before arrival. Preventive measures within casinos included mask-wearing mandates, exclusion of individuals with fever or acute cough, health declarations, and health codes. Entry was denied to those with influenza-like symptoms, and visitors’ health conditions were closely monitored. Inside the casinos, efforts were made to prevent close contact, avoid overcrowding, suspend promotional activities, and limit the number of individuals to ≤50% of the venue’s capacity. A minimum distance of 1 m was maintained between gamblers.

The ongoing pandemic and the subsequent prevention strategies have contributed to the resurgence of travelers visiting Macau for gambling and leisure purposes. The rapid transmission of COVID-19 has led to a decline in consumer economic activity, directly impacting the gambling and tourism industry. The COVID-19 pandemic has had a devastating effect on the global tourism sector, as strict measures have been taken to contain the spread of SARS-CoV-2. Consequently, there has been a substantial decline in global tourist arrivals [19]. For instance, in Macau, total visitor spending (excluding gambling) in the first quarter of 2020 experienced a substantial year-on-year decline of 70.4%, reaching MOP 5.01 billion (MOP $1=US $0.12). Per capita expenditure also decreased by 4.8%, with each visitor spending approximately MOP 1555 during that period [20].

On the basis of the aforementioned information, it is evident that COVID-19 has had a significant impact on consumer behavior and public health, particularly in relation to the gambling industry and the broader tourism sector. The pandemic has resulted in negative effects on gambling spend and public health, such as a decrease in gambling revenue, changes in the consumption patterns of tourists, and a decline in the overall tourism industry [21]. Moreover, the public health implications of COVID-19 have not only affected the emotional and psychological well-being of gamblers but also led to changes in their behavioral patterns due to safety concerns [22]. The altered behavior and psychological needs resulting from the COVID-19 pandemic are likely to influence the perception of safety among tourists and subsequently affect their willingness to gamble and their gambling behavior from a public health perspective.

**Gambling Behavior**

Gambling is part of a gambler’s leisure activities and social relations [23]. Given the unique characteristics of gambling compared to other consumption behaviors, the gambling process carries the possibility of influencing income. Rational individuals who engage in gambling derive enjoyment from the overall journey of the activity, regardless of whether they win or lose money during the process. Gambling as a consumption item exhibits a distinct attribute wherein the potential for income fluctuation during consumption is observed [24]. Gambling is recognized as an activity through which individuals have the opportunity to pursue potential financial gains or explore avenues for earning income [25]. Furthermore, it is essential to consider gambling behavior as a continuous variable. The involvement of individuals in gambling displays inherent flexibility, indicating the potential for modifications influenced by various factors. Within this continuum, it is noteworthy that gamblers often demonstrate a propensity for frequent gambling [1] and engage in a wider variety of games [26,27].

Gambling behavior is associated with gambling experience, frequency, and morbidity [28-30]. More frequent gambling, participation in more types of gambling, and the cost of gambling are associated with pathological gambling behavior [26,27]. The South Oaks Gambling Screen, developed by Lesieur and Blume [31], was designed to measure and assess pathological gambling behavior, particularly focusing on gambling frequency. Raylu and Oei [32] developed the Gambling Urge Scale to measure pathological gambling behavior with gambling adventures. Hasking and Oei [33] tested the pressure measurement of pathological gambling behavior using COPE questionnaire.

Moreover, gambling behavior is influenced by several factors, including psychological activity and the psychological characteristics of participation in gambling [34]. In particular, the theory of adaptation is used as a supplement to the exposure hypothesis to study the impact of the environment on gambling behavior. Social adaptation will lead gamblers to gradually realize that their chances of winning money are negative in the long run. In response, gamblers adjust the amount of money they bet and even their gambling behavior (such as abandoning gambling). People’s prolonged lack of exposure to gambling facilities can lead to a gradual reduction in interest in gambling games or a modification of their behavior [35].

In addition, gambling is a limited rational economic behavior influenced by the environment and culture. Binde [36] noted that due to cultural and historical influences, gambling participation has not improved despite various restrictions. After analyzing the psychological characteristics of gambling participants, they found that gambling intention leads to attributes such as the concept of luck, the neglect of probability, the “illusion of success,” and “the illusion of control.” Participation will be influenced by personal psychological characteristics such as risk aversion, ambiguity aversion, risk tolerance, loss aversion, and optimism. However, the National Research Council found that approximately 98% of Americans had no clinical problems [37]. The association between gambling intention and the frequency of gambling, as well as problem gambling, has been consistently observed in research studies [38,39]. Despite the implementation of diverse restrictions, it is important to acknowledge that participation in gambling behavior has not witnessed significant improvement, which can be attributed to the influence of cultural and historical factors [40]. In the realm of gambling research, it is widely acknowledged that gambling behavior is intricately linked to both the experience and frequency of gambling [28,30]. These prior studies highlight the significance of individuals’ prior exposure to gambling activities as a key determinant influencing their subsequent gambling behavior. Consequently, researchers have introduced the variable of experience as an additional determinant to elucidate the intricate complexities underlying gambling behavior [41].

Chinese gamblers have unique gambling characteristics [42]. The Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition), defines gambling disorder as a behavioral
addiction characterized by “persistent and recurrent problematic gambling behavior leading to clinically significant distress or impairment” [43]. Despite most forms of gambling being banned in Mainland China (except for state-run lotteries), the prevalence of gambling addiction in Mainland China is estimated to be higher than that in many Western countries, affecting approximately 2.5% to 4% of the adult population [44]. The participation rate in various forms of gambling in China is relatively high, which may be influenced by the cultural acceptance of gambling [24]. This cultural acceptance of gambling as a form of entertainment and social activity has contributed to its widespread practice among Chinese individuals [45-47]. However, it is important to acknowledge that the Chinese government has been making considerable efforts to combat illegal cross-border and online gambling activities, aiming to mitigate potential negative consequences such as financial strain and addiction [48,49].

Familiarity and recognition of gambling are the factors that attract Chinese individuals to casinos [30]. In contrast to tourists from various regions [6], the Chinese community stands out for its strong emphasis on gambling, as it is widely regarded as a preferred form of entertainment [45]. Furthermore, the Gambling Among Members of Ethnic Communities in Sydney Project found that Chinese participants thought gambling was a routine social activity rather than a way to escape the problems of everyday life [51]. A multisite study on older Chinese-Canadian gambling patterns and associated predictors showed that higher education levels and higher life satisfaction reduced the likelihood of gambling [52]. Some studies have found that Chinese (especially men) individuals have higher rates of gambling and problem gambling than the general population [53]. Chinese people who participate in gambling spend more money per week than the general community, and the participation rate of casino game consoles in the Chinese community is lower than that in other communities [54]. Galletti [55] reported that Chinese gamblers often gamble differently than the average North American gambler. Chinese gamblers often adjust their bets drastically based on their perception of luck, essentially outperforming the average North American gambler. Chinese gamblers often adjust their bets drastically based on their perception of luck, essentially outperforming the average North American gambler regarding the frequency of participation in various forms of gambling and gambling bets [32].

The COVID-19 pandemic has had several notable effects on gambling behavior from a public health standpoint. First, previous studies have reported significant reductions in the number of individuals engaging in gambling activities due to the pandemic and public health interventions [22,56]. In addition, the implementation of travel restrictions aligned with public health recommendations has influenced gambling behavior by imposing limitations on the gamblers’ ability to travel to their preferred destinations [13]. Furthermore, individuals have been more cautious in assessing the risks associated with traveling abroad [14], leading to considerations of alternative attractions or tourism products that notably increased during the pandemic [16]. In addition to the aforementioned impacts, there has been a notable increase in online gambling, particularly among the younger generation [22]. It is also important to note that during the COVID-19 pandemic, online gambling was associated with an elevated risk of problem gambling [57].

**Methods**

**Study Design**

This study used a quantitative design to investigate the gambling behavior of Mainland Chinese gamblers in Macau during the COVID-19 pandemic. Given the challenges of distributing questionnaires directly to individuals amidst the pandemic, a web-based data collection method was used. The survey team consisted of members from the marketing department of Macau casinos who actively engaged with customers through their social instant SMS text messaging platform (WeChat groups, Tencent Holdings Limited). This approach allowed the study team to effectively gather data on the gambling activities of these customers in Macau both before and after the pandemic.

**Data Collection**

Owing to the travel restrictions imposed during the COVID-19 pandemic, conducting on-site questionnaire distribution became challenging. Therefore, this study used web-based surveys through the social instant SMS text messaging platform “WeChat” to investigate gamblers who had visited Macau to gamble before the pandemic and whether they had visited Macau during the pandemic. The study used a carefully designed questionnaire using a closed-question format to gather data. To ensure a representative sample, a targeted sampling approach was adopted to reach out to individuals with relevant experiences and characteristics regarding the research topic. The survey was conducted from February 24 to March 2, 2021. A total of 350 questionnaires were distributed among the active groups on WeChat, a popular instant SMS text messaging and social media platform in Mainland China. Following a careful review of the responses received, 334 questionnaires were considered valid, representing a response rate of 95.4%.

**Measures**

We used the following measurement tools to assess gambling behavior in this study. The South Oaks Gambling Screen, developed by Lesieur and Blume [31] in 1987, was used to measure pathological gambling behavior based on gambling frequency. The Gambling Urge Scale [32] was used to measure pathological gambling behavior characterized by excessive risk-taking [58]. The COPE (Pressure Measurement of Pathological Gambling Behavior Scale) was used to assess the pressure associated with pathological gambling behavior [33]. In addition, the gambling experience scale is used to gauge various aspects of gambling behavior [28,29,59]. The gambling behavior scale encompassed the following 3 dimensions: gambling amount (3 items: “I often choose to bet large amounts when I gamble”; “I can control the amount of money I bet”; and “I can bet on budget when I gamble”); gambling mood (3 items: “when I am depressed”; “I will increase the number of times I engage in gambling when I feel stressed and nervous”; and “I will prolong my gambling time, I feel self-blame for my gambling behavior and consequences”); and social gambling (3 items: “I can give up or delay critical social activities because I turn to gambling”; “If someone invites me to gamble together,
I will not hesitate to get involved”; and “I will most likely participate in gambling-related activities if it is necessary for socializing”.

The basic sociodemographic variables included in this study were sex (female or male); marital status (married or unmarried); age (grouped into 3 categories: ≤25 years, 26-35 years, ≥36 years); monthly income (divided into 4 groups: monthly income of <RMB 10,000, RMB 10,001 to RMB 20,000, RMB 20,001 to RMB 30,000, and ≥RMB 30,001); education level (categorized as high school or below, junior college degree, and bachelor’s degree or above); and gambling before the age of 21 years (indicated as either yes or no).

In addition to the sociodemographic variables, the respondents were asked to provide information about their behavior during the COVID-19 pandemic. The questionnaire included the following items: participation in gambling activities during the pandemic (with options of yes or no); online gambling during the pandemic (with options of yes or no); pre-pandemic gambling frequency (once a week, once a month, once a day, or irregular); gambling budget for each prepandemic participation (<RMB 10,000, RMB 10,001 to RMB 20,000, RMB 20,001 to RMB 30,000, ≥RMB 30,001, or no budget); gambling budget for each gambling participation after the pandemic (<RMB 10,000, RMB 10,001 to RMB 20,000, RMB 20,001 to RMB 30,000, ≥RMB 30,001, or no budget); duration of stay at the casino (<6 hours, 6-12 hours, 1 day-2 days, 2-3 days, or >3 days); and habits (with options of smoking, drinking, not smoking, or not drinking).

Analysis

The methodological approach of this study involved the use of an web-based survey design to collect data on gambling behavior from Mainland Chinese gamblers. Specifically, this study focused on examining both online gambling and on-site gambling activities. The survey was developed and distributed through WeChat groups and consisted of 3 sections. The first section encompassed statements related to on-site and online gambling and other gambling activities. This section included questions about the frequency of gambling in the past year and the gambling budget both before and during the pandemic. The second section used 9 scales to measure various aspects of gamblers’ behavior. Respondents were asked to rate their responses on a 5-point scale. The last section of the survey was dedicated to collecting demographic data from the respondents and providing information about their characteristics.

Ethical Considerations

The methodology and survey questionnaire used in this study were approved by the institutional review board of the Center for Gaming and Tourism Studies, Macao Polytechnic University (ethics approval number: RP/CJT-02/2023/E01). All procedures complied with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The key aspects of the research ethics of human subjects are confidentiality, privacy, and consent. We obtained informed consent from all participants after fully informing them about the study. Privacy and confidentiality of participants’ personal information were strictly maintained throughout the research process, with data anonymization techniques used during reporting to protect their identities. Throughout the study, attention was given to addressing any potential conflicts of interest and upholding the highest ethical standards to protect the rights of the participants. No compensation was provided for participation in any aspect of the study.

Results

Sample Characteristics

The descriptive data for the sample is presented in Table 1. Among the 334 respondents, 49.1% (n=164) reported participating in gambling activities before the age of 21 years, while 50.9% (n=170) indicated that they did not engage in such activities. Of the 334 respondents, 48.8% (n=163) indicated their involvement in casino gambling activities during the pandemic, while 51.2% (n=171) reported abstaining from such activities. Furthermore, 52.4% (175/334) of the respondents reported participating in online gambling, whereas 47.6% (159/334) individuals reported nonparticipation in online gambling. Of the respondents, 64.4% (215/334) of individuals identified as male. Regarding educational attainment, 24.9% (83/334) of the respondents indicated that their education level was high school or below, 41.6% (139/334) respondents reported that they were enrolled in college, and 33.5% (112/334) respondents held bachelor’s degrees or higher qualifications. A substantial majority of participants, 77.5% (259/334) stated that they were married.

In terms of age distribution, the largest proportion of respondents (192/334, 57.5%) were aged 26 to 35 years, followed by those aged 36 to 50 years (96/334, 28.7%). Regarding monthly income, most respondents (177/334, 53%) earned <RMB 10,000; in addition, 95 (28.4%) of the 334 respondents reported a monthly income between RMB 10,001 and RMB 20,000, while 44 (13.2%) fell within the income range of RMB 20,001 to RMB 30,000.

Before the pandemic, the gambling frequency of the 334 respondents varied as follows: 126 (37.7%) individuals engaged in gambling once a month, 81 (24.3%) individuals gambled once a week, 23 (6.9%) individuals gambled once a day, and 104 (31.3%) individuals gambled irregularly. As for the allocation of the gambling budget, 202 (60.5%) respondents allocated <RMB 10,000, 76 (22.8%) respondents allocated between RMB 10,001 and RMB 20,000, 45 (13.5%) respondents allocated between RMB 20,001 and RMB 30,000, and 11 (3.3%) respondents allocated ≥RMB 30,001.

During the pandemic, the gambling budget distribution among the 334 respondents was as follows: 156 (46.7%) respondents allocated <RMB 10,000, 79 (23.7%) respondents allocated between RMB 10,001 and RMB 20,000, 37 (11.1%) respondents allocated between RMB 20,001 and RMB 30,000, 18 (5.4%) respondents allocated ≥RMB 30,001, and 44 (13.2%) respondents did not have any specific budget.
Table 1. Profile of the sample (N=334).

<table>
<thead>
<tr>
<th>Item and response</th>
<th>Frequency, n (%)</th>
<th>Cumulative (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gambling before the age of 21 years</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>164 (49.1)</td>
<td>49.1</td>
</tr>
<tr>
<td>No</td>
<td>170 (50.9)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Gambling in casino</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>163 (48.8)</td>
<td>48.8</td>
</tr>
<tr>
<td>No</td>
<td>171 (51.2)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Gambling online</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>175 (52.4)</td>
<td>52.4</td>
</tr>
<tr>
<td>No</td>
<td>159 (47.6)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male individuals</td>
<td>215 (64.4)</td>
<td>64.4</td>
</tr>
<tr>
<td>Female individuals</td>
<td>119 (35.6)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school and below</td>
<td>83 (24.9)</td>
<td>24.9</td>
</tr>
<tr>
<td>Junior college</td>
<td>139 (41.6)</td>
<td>66.5</td>
</tr>
<tr>
<td>Bachelor’s degree and above</td>
<td>112 (33.5)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Habit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>91 (27.2)</td>
<td>27.2</td>
</tr>
<tr>
<td>Drinking</td>
<td>102 (30.5)</td>
<td>57.8</td>
</tr>
<tr>
<td>None of the above</td>
<td>141 (42.2)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>259 (77.5)</td>
<td>77.5</td>
</tr>
<tr>
<td>Unmarried</td>
<td>75 (22.5)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Gambling frequency (before pandemic)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a week</td>
<td>81 (24.3)</td>
<td>24.3</td>
</tr>
<tr>
<td>Once a month</td>
<td>126 (37.7)</td>
<td>62</td>
</tr>
<tr>
<td>Once a day</td>
<td>23 (6.9)</td>
<td>68.9</td>
</tr>
<tr>
<td>Irregular</td>
<td>104 (31.3)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤25</td>
<td>46 (13.8)</td>
<td>13.8</td>
</tr>
<tr>
<td>26 to 35</td>
<td>192 (57.5)</td>
<td>71.3</td>
</tr>
<tr>
<td>≥36</td>
<td>96 (28.7)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Monthly income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤RMB 10,000</td>
<td>177 (53)</td>
<td>53</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000</td>
<td>95 (28.4)</td>
<td>81.4</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000</td>
<td>44 (13.2)</td>
<td>94.6</td>
</tr>
<tr>
<td>≥RMB 30,001</td>
<td>18 (5.4)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Gambling budget (before the pandemic)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤RMB 10,000</td>
<td>202 (60.5)</td>
<td>60.5</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000</td>
<td>76 (22.8)</td>
<td>83.2</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000</td>
<td>45 (13.5)</td>
<td>96.7</td>
</tr>
<tr>
<td>≥RMB 30,001</td>
<td>11 (3.3)</td>
<td>100</td>
</tr>
</tbody>
</table>
Regarding the duration spent in casinos, the following patterns were observed among the 334 respondents: 172 (51.5%) individuals spent <6 hours in casinos, 96 (28.7%) individuals dedicated 6 to 12 hours, 48 (14.4%) individuals engaged in gambling for 1 to 2 days, and 18 (5.4%) individuals extended their visits for >2 to 3 days in casinos. These findings elucidate the diverse levels of engagement and time commitment exhibited by the surveyed individuals during their casino experiences.

Measurement Model

First, we tested the questionnaire for its reliability. The measurement of gambling behavior had a Cronbach α coefficient of 0.824 for gambling amount, 0.824 for gambling mood, 0.804 for social gambling, and 0.858 for gambling involvement, indicating a high level of reliability. Next, the scale was further validated through an exploratory factor analysis. The Kaiser-Meyer-Olkin measure of sampling adequacy was found to be 0.940, indicating that the scale was suitable for factor analysis. The Bartlett Test of Sphericity yielded a chi-square value of 2105.821 ($P$ < .001), supporting the suitability of the scale for factor analysis. The dimensionality reduction analysis was performed using SPSS software (version 26.0; IBM Corp). The validity test demonstrated that each question factor had loadings ranging from 0.659 to 0.813, indicating a high level of conceptual validity for the scale. The total number of explanatory variables accounted for 72.65% of the variance, indicating that the measured indicators provided substantial explanatory power. Overall, the analysis revealed a high level of reliability of the measured indicators.

Variance Analysis

We conducted binary logistic regression analysis based on unweighted data to examine group comparisons of gambling behaviors. Chi-square tests were used to explore potential independent variables associated with the changes in gambling behavior, which served as the dependent variable. The results of the logistic regression analysis are presented as odds ratios with 95% CIs. All statistical analyses were performed using SPSS software.

First, we analyzed the samples based on whether respondents of different genders participated in Macau casinos during the pandemic and engaged in online gambling. The descriptive statistics of the sample are presented in Table 2. A total of 64.4% (215/334) male individuals and 35.6% (119/334) female individuals responded to the survey. Among the 334 participants, a total of 175 (52.4%) respondents were involved in online gambling, of which 136 (77.7%) were male individuals and 39 (22.3%) were female individuals. During the pandemic, 78.4% (116/334) of male individuals and 21.6% (32/334) of female individuals were involved in online gambling. In addition, 27 (8.1%) of the 334 respondents, comprising 20 (74.1%) male individuals and 7 (25.9%) female individuals, did not visit Macau; 15 (4.5%) respondents, comprising 10 (66.7%) male individuals and 5 (33.3%) female individuals, traveled to Macau for gambling. During the pandemic, 171 (51.2%) of the 334 respondents, with 89 (52%) male individuals and 82 (48%) female individuals, did not travel to Macau. This shows that the impact of the pandemic on gambling is enormous.
The basic statistics related to traveling to Macau and participating in online gambling are presented in Table 3. Gender differences were evident, with notable variations in gambling habits among individuals aged 21 years, frequency of gambling before the COVID-19 pandemic, monthly income, gambling budget (before the pandemic), gambling budget (during the pandemic), and the length of stay at casinos ($P < .01$). Furthermore, marital status exhibited significant differences in traveling to Macau for gambling ($P < .01$), whereas no notable differences were observed in online gambling. However, there were no significant differences in terms of education or age between those who traveled to Macau and those who engaged in online gambling.

An independent sample test was conducted to compare the gender of individuals who traveled to Macau for gambling (Table 4). The results revealed that there was a significant difference in the gambling amount for married individuals who traveled to Macau and participated in gambling during the pandemic ($P < .05$). However, no significant difference was observed in traveling to Macau for gambling during the pandemic. These findings suggest that the pandemic has had a significant impact on the gambling behavior of married gamblers. In addition, regardless of marital status, there were no significant differences in gambling amounts during the pandemic.

### Table 2. Survey respondents and the final study populations by gambling experienced group (N=334).

<table>
<thead>
<tr>
<th>GONLINE</th>
<th>Male individuals</th>
<th>Female individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
<td>Total</td>
</tr>
<tr>
<td>GINVOIVE</td>
<td>(n=148), n (%)</td>
<td>(n=175), n (%)</td>
</tr>
<tr>
<td>Yes</td>
<td>116 (78.4)</td>
<td>136 (77.7)</td>
</tr>
<tr>
<td>No</td>
<td>20 (74.1)</td>
<td>10 (66.7)</td>
</tr>
<tr>
<td>Total</td>
<td>136 (77.7)</td>
<td>79 (49.7)</td>
</tr>
</tbody>
</table>

### Notes

- **GONLINE**: gambling online.
- **GINVOIVE**: gambling in casinos.

The basic statistics related to traveling to Macau and participating in online gambling are presented in Table 3. Gender differences were evident, with notable variations in gambling habits among individuals aged 21 years, frequency of gambling before the COVID-19 pandemic, monthly income, gambling budget (before the pandemic), gambling budget (during the pandemic), and the length of stay at casinos ($P < .01$). Furthermore, marital status exhibited significant differences in traveling to Macau for gambling ($P < .01$), whereas no notable differences were observed in online gambling. However, there were no significant differences in terms of education or age between those who traveled to Macau and those who engaged in online gambling.

An independent sample test was conducted to compare the gender of individuals who traveled to Macau for gambling (Table 4). The results revealed that there was a significant difference in the gambling amount for married individuals who traveled to Macau and participated in gambling during the pandemic ($P < .05$). However, no significant difference was observed in traveling to Macau for gambling during the pandemic. These findings suggest that the pandemic has had a significant impact on the gambling behavior of married gamblers. In addition, regardless of marital status, there were no significant differences in gambling amounts during the pandemic.
Table 3. Gambling behavior by demographic characteristics among all participants (N=334).

<table>
<thead>
<tr>
<th>Item and comparison group</th>
<th>GINVOLVE&lt;sup&gt;a&lt;/sup&gt;</th>
<th>&lt;sup&gt;P&lt;/sup&gt; value</th>
<th>GONLINE&lt;sup&gt;b&lt;/sup&gt;</th>
<th>&lt;sup&gt;P&lt;/sup&gt; value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male individuals (n=215)</td>
<td>126 (58.6)</td>
<td>&lt;.01</td>
<td>136 (63.3)</td>
<td>79 (36.7)</td>
</tr>
<tr>
<td>Female individuals (n=119)</td>
<td>37 (31.1)</td>
<td></td>
<td>39 (32.8)</td>
<td>80 (67.2)</td>
</tr>
<tr>
<td>Gambling before the age of 21 years</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Yes (n=164)</td>
<td>135 (82.3)</td>
<td></td>
<td>139 (84.8)</td>
<td>25 (15.2)</td>
</tr>
<tr>
<td>No (n=170)</td>
<td>28 (16.5)</td>
<td></td>
<td>36 (21.2)</td>
<td>134 (78.8)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>.90</td>
<td></td>
<td>.58</td>
</tr>
<tr>
<td>High school or below (n=83)</td>
<td>42 (50.6)</td>
<td></td>
<td>47 (56.6)</td>
<td>36 (43.4)</td>
</tr>
<tr>
<td>Junior college (n=139)</td>
<td>66 (47.5)</td>
<td></td>
<td>73 (52.5)</td>
<td>66 (47.5)</td>
</tr>
<tr>
<td>Bachelor or above (n=112)</td>
<td>55 (49.1)</td>
<td></td>
<td>55 (49.1)</td>
<td>57 (50.9)</td>
</tr>
<tr>
<td>Habit</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Smoking (n=91)</td>
<td>52 (57.1)</td>
<td></td>
<td>56 (61.5)</td>
<td>35 (38.5)</td>
</tr>
<tr>
<td>Drinking (n=102)</td>
<td>69 (67.6)</td>
<td></td>
<td>74 (72.5)</td>
<td>28 (27.5)</td>
</tr>
<tr>
<td>None of the above (n=141)</td>
<td>42 (29.8)</td>
<td></td>
<td>45 (31.9)</td>
<td>96 (68.1)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td>.01</td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Married (n=259)</td>
<td>136 (52.5)</td>
<td></td>
<td>142 (54.8)</td>
<td>117 (45.2)</td>
</tr>
<tr>
<td>Unmarried (n=75)</td>
<td>27 (36)</td>
<td></td>
<td>33 (44)</td>
<td>42 (56)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td>.25</td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>≤25 (n=46)</td>
<td>27 (58.7)</td>
<td></td>
<td>29 (63)</td>
<td>17 (37)</td>
</tr>
<tr>
<td>26 to 35 (n=192)</td>
<td>94 (49)</td>
<td></td>
<td>100 (52.1)</td>
<td>92 (47.9)</td>
</tr>
<tr>
<td>≥36 (n=96)</td>
<td>42 (43.8)</td>
<td></td>
<td>46 (47.9)</td>
<td>50 (52.1)</td>
</tr>
<tr>
<td>Gambling frequency (before pandemic)</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Once a week (n=81)</td>
<td>55 (67.9)</td>
<td></td>
<td>61 (75.3)</td>
<td>20 (24.7)</td>
</tr>
<tr>
<td>Once a month (n=126)</td>
<td>80 (63.5)</td>
<td></td>
<td>86 (68.3)</td>
<td>40 (31.7)</td>
</tr>
<tr>
<td>Once a day (n=23)</td>
<td>16 (69.6)</td>
<td></td>
<td>13 (56.5)</td>
<td>10 (43.5)</td>
</tr>
<tr>
<td>Irregular (n=104)</td>
<td>12 (11.5)</td>
<td></td>
<td>15 (14.4)</td>
<td>89 (85.6)</td>
</tr>
<tr>
<td>Monthly income&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>≤RMB 10,000 (n=177)</td>
<td>69 (39)</td>
<td></td>
<td>79 (44.6)</td>
<td>98 (55.4)</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000 (n=95)</td>
<td>57 (60)</td>
<td></td>
<td>58 (61.1)</td>
<td>37 (38.9)</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000 (n=44)</td>
<td>28 (63.6)</td>
<td></td>
<td>29 (65.9)</td>
<td>15 (34.1)</td>
</tr>
<tr>
<td>≥RMB 30,001 (n=18)</td>
<td>9 (50)</td>
<td></td>
<td>9 (50)</td>
<td>9 (50)</td>
</tr>
<tr>
<td>Gambling budget (before the pandemic)</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td>&lt;.01</td>
</tr>
<tr>
<td>≤RMB 10,000 (n=202)</td>
<td>76 (37.6)</td>
<td></td>
<td>90 (44.6)</td>
<td>112 (55.4)</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000 (n=76)</td>
<td>47 (61.8)</td>
<td></td>
<td>45 (59.2)</td>
<td>31 (40.8)</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000 (n=45)</td>
<td>34 (75.6)</td>
<td></td>
<td>33 (73.3)</td>
<td>12 (26.7)</td>
</tr>
<tr>
<td>≥RMB 30,001 (n=11)</td>
<td>6 (54.5)</td>
<td></td>
<td>7 (63.6)</td>
<td>4 (36.4)</td>
</tr>
<tr>
<td>Gambling budget (during the pandemic)</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td>&lt;.01</td>
</tr>
<tr>
<td>≤RMB 10,000 (n=156)</td>
<td>76 (48.7)</td>
<td></td>
<td>86 (55.1)</td>
<td>70 (44.9)</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000 (n=79)</td>
<td>49 (62)</td>
<td></td>
<td>51 (64.6)</td>
<td>28 (35.4)</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000 (n=37)</td>
<td>20 (54.1)</td>
<td></td>
<td>20 (54.1)</td>
<td>17 (45.9)</td>
</tr>
</tbody>
</table>
We conducted an ANOVA on the variables related to gambling in Macau (Table 5). The results indicated that individuals who traveled to Macau for gambling during the pandemic reported spending >2 days at casinos and gambling significantly higher amounts compared to those who stayed for 1 to 2 days. In addition, the social gambling behavior of those who stayed between 6 and 12 hours and those who stayed for >2 days was significantly higher than that of individuals who stayed for 1 day to 2 days.

The budget for each gambling session (before the pandemic) indicated notable differences among the different income groups. The group with a budget of ≥RMB 30,001 reported gambling significantly higher amounts compared to the group with a budget of RMB 20,001 to RMB 30,000 and the group with a budget of <10,000 or the group without a budget. There were significantly more gamblers aged <25 years and 26 to 35 years in Macau during the pandemic than those aged >36 years. Among gamblers who gambled more often than before the pandemic, the group that gambled once a day reported significantly higher rates of gambling than the other frequency groups.

Furthermore, an ANOVA was conducted on variables related to online gambling (Table 6). Regarding academic qualifications, the influence of gambling sentiment on gamblers with a college degree was considerably higher than that on gamblers with a high school education or below. In terms of the frequency of gambling before the pandemic, the group that gambled daily exhibited significantly higher rates compared to the other frequency groups. Among online gamblers, those aged 26 to 35 years accounted for a larger proportion of gambling compared to those aged >35 years; those aged <25 years and 26 to 35 years significantly engaged in more social gambling compared to those aged >35 years. Among non–online gamblers, those aged <25 years and those aged 26 to 35 years had a significantly larger presence than those in the 35-year age group.

Online gambling budgets vary during the COVID-19 pandemic. Nevertheless, the group with a gambling budget of >RMB 30,001 still maintained significantly higher gambling amounts compared to the group without a budget for online gambling. In terms of the duration spent at the casino participating in gambling, the group spending >2 days at the casino stayed more than twice as long as the group spending 1 day to 2 days.

---

**Table 4. Gambling behavior independence tests**

<table>
<thead>
<tr>
<th>Gambling in casinos</th>
<th>Levin test</th>
<th>Mean equivalence t test (2-tailed)</th>
<th>Mean squared error difference (SE)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F test (df)</td>
<td>P value</td>
<td>t test (df)</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variance</td>
<td>4.923 (332)</td>
<td>.03</td>
<td>-2.717 (169)</td>
<td>.007</td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variance</td>
<td>c</td>
<td></td>
<td>-3.089 (114.972)</td>
<td>.003</td>
</tr>
</tbody>
</table>

**Table 5. Gambling behavior independence tests**

<table>
<thead>
<tr>
<th>Item and comparison group</th>
<th>GINVOICE&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value</th>
<th>GONLINE&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes, n (%)</td>
<td>No, n (%)</td>
<td>Yes, n (%)</td>
<td>No, n (%)</td>
</tr>
<tr>
<td>≥RMB 30,001 (n=18)</td>
<td>16 (88.9)</td>
<td>2 (11.1)</td>
<td>15 (83.3)</td>
<td>3 (16.7)</td>
</tr>
<tr>
<td>No budget (n=44)</td>
<td>2 (4.5)</td>
<td>42 (95.5)</td>
<td>3 (6.8)</td>
<td>41 (93.2)</td>
</tr>
<tr>
<td>Time of stay at the casino</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 hours (n=172)</td>
<td>70 (40.7)</td>
<td>102 (59.3)</td>
<td>74 (43)</td>
<td>98 (57)</td>
</tr>
<tr>
<td>6 to 12 hours (n=96)</td>
<td>58 (60.4)</td>
<td>38 (39.6)</td>
<td>63 (65.6)</td>
<td>33 (34.4)</td>
</tr>
<tr>
<td>1 day to 2 days (n=48)</td>
<td>25 (52.1)</td>
<td>23 (47.9)</td>
<td>28 (58.3)</td>
<td>20 (41.7)</td>
</tr>
<tr>
<td>&gt;2 days (n=18)</td>
<td>10 (55.6)</td>
<td>8 (44.4)</td>
<td>10 (55.6)</td>
<td>8 (44.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>GINVOICE: gambling in casinos.

<sup>b</sup>GONLINE: gambling online.

<sup>c</sup>RMB $1=US $1 0.14.
Table 5. ANOVA multiple comparisons of the involved gambling behavior.

<table>
<thead>
<tr>
<th>Time to stay at the casino</th>
<th>Mean deviation</th>
<th>σ</th>
<th>P value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AGB1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 day to 2 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 hours</td>
<td>0.274</td>
<td>0.213</td>
<td>.20</td>
<td>−0.147 to 0.696</td>
</tr>
<tr>
<td>6 to 12 hours</td>
<td>0.308</td>
<td>0.219</td>
<td>.16</td>
<td>−0.125 to 0.741</td>
</tr>
<tr>
<td>&gt;2 days</td>
<td>0.760&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.343</td>
<td>.03</td>
<td>0.083 to 1.437</td>
</tr>
<tr>
<td><strong>AGB3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 day to 2 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 hours</td>
<td>0.390</td>
<td>0.199</td>
<td>.053</td>
<td>−0.004 to 0.783</td>
</tr>
<tr>
<td>6 to 12 hours</td>
<td>0.497&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.205</td>
<td>.02</td>
<td>0.093 to 0.902</td>
</tr>
<tr>
<td>&gt;2 days</td>
<td>0.880&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.320</td>
<td>.007</td>
<td>0.248 to 1.513</td>
</tr>
</tbody>
</table>

Gambling budget (before the pandemic)

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AGB1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤RMB 10,000&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000</td>
<td>0.240</td>
<td>0.185</td>
<td>.20</td>
<td>−0.126 to 0.606</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000</td>
<td>−0.128</td>
<td>0.231</td>
<td>.58</td>
<td>−0.584 to 0.329</td>
</tr>
<tr>
<td>≥RMB 30,001</td>
<td>0.362&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.620</td>
<td>.03</td>
<td>0.138 to 2.587</td>
</tr>
<tr>
<td>No budget</td>
<td>0.021</td>
<td>0.165</td>
<td>.90</td>
<td>−0.305 to 0.347</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000</td>
<td>≤RMB 10,000</td>
<td>0.128</td>
<td>0.231</td>
<td>.58</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000</td>
<td>0.368</td>
<td>0.263</td>
<td>.16</td>
<td>−0.151 to 0.887</td>
</tr>
<tr>
<td>≥RMB 30,001</td>
<td>0.490&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.647</td>
<td>.02</td>
<td>0.212 to 2.768</td>
</tr>
<tr>
<td>No budget</td>
<td>0.149</td>
<td>0.249</td>
<td>.55</td>
<td>−0.343 to 0.641</td>
</tr>
<tr>
<td>≥RMB 30,001</td>
<td>≤RMB 10,000</td>
<td>−0.362&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.620</td>
<td>.03</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000</td>
<td>−0.122</td>
<td>0.632</td>
<td>.08</td>
<td>−2.371 to 0.127</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000</td>
<td>−0.490&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.647</td>
<td>.02</td>
<td>−2.768 to −0.212</td>
</tr>
<tr>
<td>No budget</td>
<td>−0.341&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.627</td>
<td>.03</td>
<td>−2.579 to −0.104</td>
</tr>
</tbody>
</table>

Age (years)

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AGB2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 to 35</td>
<td>0.351&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.170</td>
<td>.04</td>
<td>0.017 to 0.687</td>
</tr>
<tr>
<td>≤25</td>
<td>0.477&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.212</td>
<td>.03</td>
<td>0.059 to 0.895</td>
</tr>
</tbody>
</table>

Gambling frequency (before the pandemic)

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AGB1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambling in casino, variable, and comparison group</td>
<td>Mean deviation</td>
<td>σ</td>
<td>P value</td>
<td>95% CI</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----------------</td>
<td>----</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td><strong>Once a week</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a month</td>
<td>0.011</td>
<td>0.213</td>
<td>.96</td>
<td>−0.410 to 0.432</td>
</tr>
<tr>
<td>Once a day</td>
<td>0.739&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.370</td>
<td>.047</td>
<td>0.009 to 1.471</td>
</tr>
<tr>
<td>Irregular</td>
<td>0.026</td>
<td>0.193</td>
<td>.90</td>
<td>−0.356 to 0.407</td>
</tr>
<tr>
<td><strong>Once a month</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a week</td>
<td>−0.011</td>
<td>0.213</td>
<td>.96</td>
<td>−0.432 to 0.410</td>
</tr>
<tr>
<td>Once a day</td>
<td>0.728&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.353</td>
<td>.04</td>
<td>0.032 to 1.425</td>
</tr>
<tr>
<td>Irregular</td>
<td>0.014</td>
<td>0.157</td>
<td>.93</td>
<td>−0.296 to 0.325</td>
</tr>
<tr>
<td><strong>Once a day</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a week</td>
<td>−0.739&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.370</td>
<td>.047</td>
<td>−1.471 to −0.009</td>
</tr>
<tr>
<td>Once a month</td>
<td>−0.728&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.353</td>
<td>.04</td>
<td>−1.425 to −0.032</td>
</tr>
<tr>
<td>Irregular</td>
<td>−0.714&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.341</td>
<td>.04</td>
<td>−1.387 to −0.041</td>
</tr>
</tbody>
</table>

<sup>a</sup>AGB1: gambling amount.
<sup>b</sup>P<.05.
<sup>c</sup>AGB3: social gambling.
<sup>d</sup>RMB $1=US $1 0.14.
<sup>e</sup>AGB2: gambling mood.
Table 6. ANOVA multiple comparisons of the online gambling behavior least significant difference ANOVA analysis.

<table>
<thead>
<tr>
<th>Gambling online, variable, and comparison group</th>
<th>Mean deviation</th>
<th>σ</th>
<th>P value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGB2&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or below</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior college</td>
<td>−0.340&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.163</td>
<td>.04</td>
<td>−0.663 to −0.018</td>
</tr>
<tr>
<td>Bachelor’s degree or above</td>
<td>−0.268</td>
<td>0.173</td>
<td>.12</td>
<td>−0.610 to 0.074</td>
</tr>
<tr>
<td><strong>Gambling frequency (before the pandemic)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGB1&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a week</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a month</td>
<td>0.167</td>
<td>0.155</td>
<td>.28</td>
<td>−0.139 to 0.473</td>
</tr>
<tr>
<td>Once a day</td>
<td>0.663&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.283</td>
<td>.02</td>
<td>0.105 to 1.222</td>
</tr>
<tr>
<td>Irregular</td>
<td>−0.042</td>
<td>0.267</td>
<td>.87</td>
<td>−0.569 to 0.485</td>
</tr>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤25</td>
<td>0.356</td>
<td>0.220</td>
<td>.11</td>
<td>−0.078 to 0.790</td>
</tr>
<tr>
<td>26 to 35</td>
<td>0.350&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.165</td>
<td>.04</td>
<td>0.024 to 0.676</td>
</tr>
<tr>
<td>AGB2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤25</td>
<td>0.530&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.205</td>
<td>.01</td>
<td>0.125 to 0.936</td>
</tr>
<tr>
<td>26 to 35</td>
<td>0.332&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.154</td>
<td>.03</td>
<td>0.029 to 0.637</td>
</tr>
<tr>
<td>AGB3&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤25</td>
<td>0.411&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.203</td>
<td>.045</td>
<td>0.010 to 0.813</td>
</tr>
<tr>
<td>26 to 35</td>
<td>0.393&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.153</td>
<td>.01</td>
<td>0.092 to 0.695</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 to 35</td>
<td>0.539&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.223</td>
<td>.02</td>
<td>0.099 to 0.980</td>
</tr>
<tr>
<td>≥36</td>
<td>0.557&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.237</td>
<td>.02</td>
<td>0.089 to 1.026</td>
</tr>
<tr>
<td><strong>Gambling budget (before the pandemic)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No budget</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambling online, variable, and comparison group</td>
<td>Mean deviation</td>
<td>σ</td>
<td>P value</td>
<td>95% CI</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------------</td>
<td>---</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>≤RMB 10,000&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.973</td>
<td>0.546</td>
<td>.08</td>
<td>−0.106 to 2.052</td>
</tr>
<tr>
<td>RMB 10,001 to RMB 20,000</td>
<td>0.922</td>
<td>0.553</td>
<td>.10</td>
<td>−0.170 to 2.013</td>
</tr>
<tr>
<td>RMB 20,001 to RMB 30,000</td>
<td>0.683</td>
<td>0.576</td>
<td>.24</td>
<td>−0.454 to 1.820</td>
</tr>
<tr>
<td>≥RMB 30,001</td>
<td>0.778&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.588</td>
<td>.047</td>
<td>0.016 to 2.339</td>
</tr>
</tbody>
</table>

**Time of stay at the casino**

Yes

AGB3

<table>
<thead>
<tr>
<th></th>
<th>Mean deviation</th>
<th>σ</th>
<th>P value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;2 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 hours</td>
<td>−0.508</td>
<td>0.290</td>
<td>.08</td>
<td>−1.080 to 0.064</td>
</tr>
<tr>
<td>6 to 12 hours</td>
<td>−0.400</td>
<td>0.293</td>
<td>.17</td>
<td>−0.978 to 0.178</td>
</tr>
<tr>
<td>1 day to 2 days</td>
<td>−0.781&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.317</td>
<td>.02</td>
<td>−1.407 to −0.155</td>
</tr>
</tbody>
</table>

<sup>a</sup>AGB2: gambling mood.
<sup>b</sup>P<.05.
<sup>c</sup>AGB1: gambling amount.
<sup>d</sup>AGB3: social gambling.
<sup>e</sup>RMB $1=US $1 0.14.

**Discussion**

**Principal Findings**

Recognition is growing that gambling poses an increasing public health concern. Despite being beneficial for the economy and profitable for businesses, gambling causes personal and social harm to both individuals and communities. In particular, gambling presents a significant global public health challenge with the rise of online gambling [60]. The effects of gambling on public health are evident across various levels, including personal, interpersonal, and societal dimensions [61]. In this cross-sectional study, we investigated the changes in the gambling behavior of Mainland Chinese gamblers visiting Macau and participating in online gambling during the COVID-19 pandemic. In addition, we analyzed other factors that influenced gambling behavior during the pandemic, such as certain sociodemographic characteristics. We used data from a sample of 334 respondents to conduct a comparative analysis of gamblers’ gambling behaviors. The main findings of this study are presented subsequently.

First, the COVID-19 pandemic has had a significant and wide-ranging impact on gambling behavior. Among the 334 respondents in this study, a substantial proportion of 171 (51.2%) individuals reported not participating in gambling during the pandemic. This finding underscores the profound influence of the pandemic on individuals’ gambling behavior. Specifically, the willingness of gamblers to visit Macau for gambling tourism has been affected due to the various consequences of public health measures [56]. The effects of gambling on public health are evident across various levels, including personal, interpersonal, and societal dimensions [61]. In this cross-sectional study, we investigated the changes in the gambling behavior of Mainland Chinese gamblers visiting Macau and participating in online gambling during the COVID-19 pandemic. In addition, we analyzed other factors that influenced gambling behavior during the pandemic, such as certain sociodemographic characteristics. We used data from a sample of 334 respondents to conduct a comparative analysis of gamblers’ gambling behaviors. The main findings of this study are presented subsequently.

First, the COVID-19 pandemic has had a significant and wide-ranging impact on gambling behavior. Among the 334 respondents in this study, a substantial proportion of 171 (51.2%) individuals reported not participating in gambling during the pandemic. This finding underscores the profound influence of the pandemic on individuals’ gambling behavior. Specifically, the willingness of gamblers to visit Macau for gambling tourism has been affected due to the various consequences of public health measures [56]. Notably, 20 (74.2%) male individuals and 7 (26%) female individuals of the respondents refrained from gambling in Macau and instead opted for an alternative approach, which included online gambling. This finding aligns with the conclusions drawn by Sachdeva et al [16], indicating that individuals consider various factors that influence their ultimate gambling behavior in the face of changing circumstances, especially during a pandemic.

Second, it is noteworthy that, among the 334 respondents, 175 (52.4%) individuals engaged in online gambling, with 136 (77.7%) being male individuals and 39 (22.3%) being female individuals. This indicates the proportion of participants who turned to online gambling, particularly during the pandemic. Notably, 20 (74.2%) male individuals and 7 (26%) female individuals of the respondents refrained from gambling in Macau and instead opted for an alternative approach, which included online gambling. This finding aligns with the conclusions drawn by Sachdeva et al [16], indicating that individuals consider various factors that influence their ultimate gambling behavior in the face of changing circumstances, especially during a pandemic.

Third, the results revealed significant differences among Mainland Chinese gamblers who engaged in both travel to Macau and online gambling, with regard to gender, gambling habits, prepandemic frequency of play, monthly income, gambling budget (before and during the pandemic), and length of stay at the casino (P<.01). Notably, marital status demonstrated significant differences in terms of traveling to Macau for gambling (P<.05), whereas no significant differences in terms of marital status were observed in online gambling. However, we did not find any significant differences regarding the education and age of gamblers in this study. Despite the changes in participation methods and behaviors during the pandemic, this finding aligns with the conclusions drawn by...
Chan et al [50], suggesting that problem gambling may be prevalent among Chinese gamblers, regardless of whether they participate in gambling at land-based casinos or web-based platforms.

To examine the difference in gambling behavior among individuals who visited casinos, we found from this study that there was a significant difference in gambling amount among married individuals who traveled to Macau for gambling ($P<0.05$). However, there was no significant difference in the gambling budget of married gamblers before and during the pandemic. Moreover, regardless of marital status, there was no significant difference in gambling expenditure during the pandemic. Among the gamblers who visited Macau for gambling during the pandemic, those who stayed longer at the casino and participated in gambling exhibited higher levels of gambling expenditure compared to those who stayed for 1 day to 2 days. This suggests that high-value gamblers tended to extend their stay at the casinos before the pandemic and had a tendency toward problem gambling [57]. Regarding casino gambling, respondents who stayed at the casinos for 6 to 12 hours or >2 days demonstrated a significantly greater level of involvement than those who stayed for 1 day to 2 days. Before the pandemic, individuals with gambling budgets of >RMB 30,001 were considerably higher than those with other budgets, indicating that high-value gamblers had higher budgets allocated to gambling before the pandemic. In terms of age demographics, gamblers aged <25 years and those aged 26 to 35 years were considerably more prevalent than those aged >36 years during the pandemic, indicating that young gamblers were more prevalent than gamblers in other stages before the pandemic. Among the gamblers who gambled regularly before the pandemic, those who gambled once a day exhibited a significantly higher frequency compared to those who gambled once a week, once a month, or irregularly. Furthermore, in the context of social gambling, it was observed that those aged <25 years and those aged 26 to 35 years had significantly higher frequency of gambling compared to those aged 35 years. This indicates a greater inclination of gamblers in these age groups to participate in gambling for social purposes [22]. Consequently, the pandemic has shaped their behaviors within the casino environment during this period of public health crisis.

When examining the difference in gambling behavior of online gamblers, the results reveal a significant difference in the number of gamblers based on educational attainment, particularly between college-educated individuals and those with a high school education or below. This discrepancy suggests that the gambling sentiment and mindset of gamblers in this study may have been influenced by the pandemic, leading to changes in their behavior. Regarding the frequency of gambling participation before the COVID-19 pandemic, the number of gamblers who gambled once a day was considerably higher than that in the other groups, suggesting that gamblers who gambled frequently also placed more bets. Regarding age, among the gamblers participating in online gambling, those aged 26 to 35 years exhibited significantly higher gambling expenditure compared to those aged 35 to 45 years; those aged <25 years and 26 to 35 years demonstrated significantly higher gambling mood compared to those aged 35 to 35 years; furthermore, those aged <25 years or 26 to 35 years displayed considerably high involvement in social gambling, indicating the increased participation of the younger generation in online gambling and social gambling activities compared to older age groups [22]. The budget allocation for each gambling session during the pandemic revealed that the group with a budget >RMB 30,001 was significantly larger than the group with other budget allocations, reflecting the importance of online gambling as a prevalent form of gambling in the aftermath of the pandemic. Regarding the duration of gambling participation, the group that engaged in gambling for >2 days was significantly larger than the group that participated for 1 day to 2 days. This finding suggests that online gambling serves as a platform that fulfills the socialization needs of gamblers over a prolonged period.

Gambling has gained widespread popularity among adolescents worldwide [23,62]. Excessive or problematic gambling can give rise to various public health implications, including financial consequences, addiction, and mental health issues. A public health approach is essential to address and mitigate the harm associated with gambling [63,64]. The harm caused by gambling can manifest in various ways, affecting both individuals and communities. Considering the economic significance of the gaming industry, it is advisable to adopt a comprehensive and coordinated approach that engages policy makers, gaming operators, and the broader community to effectively prevent problem gambling among both local citizens and tourists. Public health initiatives are recommended to involve the implementation of player protection measures, public awareness and education, provision of support services, self-regulation within the industry, and promotion of responsible gambling practices to create safer environments. Addressing gambling addiction issues at both personal and societal levels can lead to the minimization of negative impacts associated with gambling, the safeguarding of vulnerable populations, and the fostering of overall well-being in individuals and communities.

**Limitations**

This study has several limitations that should be acknowledged. First, the data collected for this study were obtained from a survey conducted with gamblers on a social instant SMS text messaging platform (WeChat), which may introduce the possibility of selection bias and limit the generalizability of our findings. In addition, the relatively small sample size used in this study may restrict the generalizability of the results.

Second, the scale designed to measure gambling behavior in this study was simplified based on previous research. This simplification may have overlooked important dimensions of gambling behavior, potentially impacting the comprehensiveness of our findings. Furthermore, the study did not categorize the different types of gambling games, making it difficult to ascertain the specific preferences of gamblers, such as table games, slot machines, online gambling, and sports betting. Future research should consider using more comprehensive scales to capture the complexity of gambling behavior and incorporate the categorization of different types of gambling activities.
Third, while the cross-sectional design used in this study is useful for examining associations at a specific point in time, a time-series approach would provide a more accurate reflection of the dynamic changes in individuals’ gambling behavior. Future research should consider longitudinal designs to gain a better understanding of how gambling behaviors evolve and to capture the impact of various factors on these changes. Furthermore, it is important to recognize that the COVID-19 pandemic has significantly altered the macroenvironment, including work and income situations, which may have influenced changes in gambler behavior. Future research should include relevant control variables to account for these potential confounding factors and to obtain a more comprehensive understanding of the impact of the pandemic on gambling behavior.

Conclusions
The primary conclusion of this paper is that a high proportion of the original Mainland Chinese gamblers opted not to visit Macau for gambling purposes after the outbreak of the COVID-19 pandemic. While the reasons for this shift may vary, the public health crisis has resulted in a decreased reliance on land-based gambling and a greater inclination toward online gambling. Consequently, there has been a reduction in the amount of money spent at gambling establishments and a decrease in the duration of casino visits. Casinos have traditionally served as highly popular social venues for gamblers. However, the antipandemic measures implemented by governments and casino operators have, to some extent, affected the social gambling behavior of individuals, leading to a shift away from traditional casinos and a greater emphasis on online gambling platforms.

Additionally, the COVID-19 pandemic has led to a notable increase in online gambling activities among Mainland Chinese gamblers. Particularly, the younger generations have been significantly affected by this observed phenomenon. Online gambling has gained popularity during the pandemic, with a substantial increase in both the amount and frequency of gambling observed. The restrictions imposed by the pandemic have had an impact on gamblers’ behavior, affecting their mood while participating in gambling activities. Notably, individuals with a middle level of education have experienced a significant impact on their gambling behavior. The findings of this study indicate that the pandemic has led to a decrease in land-based gambling, creating an environment that is conducive to controlling problem gambling; however, there appears to be a shift in gambling behavior from traditional casinos toward online gambling platforms as a direct consequence of the pandemic. The impact of the COVID-19 pandemic on gambling behavior underscores the significance of implementing targeted public health strategies and interventions to address the changing dynamics of gambling during public health emergencies.

Acknowledgments
This work was supported by the Macao Polytechnic University (Research Project RP/CJT-02/2023)

Data Availability
The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

References


32. Raylu N, Oei TP. The gambling urge scale: development, confirmatory factor validation, and psychometric properties. Psychol Addict Behav 2004 Jun;18(2):100-105. [doi: 10.1037/0893-164x.18.2.100]


Abbreviations

SAR: Special Administrative Region

©Jinquan Zhou, Hong-Wai Ho, ChiBiu Chan. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: Since 2016, the government of Bangladesh has been piloting a health protection scheme known as Shasthyo Surokhsha Karmasuchi (SSK), which specifically targets households living below the poverty line. This noncontributory scheme provides enrolled households access to inpatient health care services for 78 disease groups. Understanding patients’ experiences with health care utilization from the pilot SSK scheme is important for enhancing the quality of health care service delivery during the national-level scale-up of the scheme.

Objective: We aimed to evaluate patient satisfaction with the health care services provided under the pilot health protection scheme in Bangladesh.

Methods: A cross-sectional survey was conducted with the users of the SSK scheme from August to November 2019. Patients who had spent a minimum of 2 nights at health care facilities were selected for face-to-face exit interviews. During these interviews, we collected information on patients’ socioeconomic characteristics, care-seeking experiences, and level of satisfaction with various aspects of health care service delivery. To measure satisfaction, we employed a 5-point Likert scale (very satisfied, 5; satisfied, 4; neither satisfied nor dissatisfied, 3; dissatisfied, 2; very dissatisfied, 1). Descriptive statistics, statistical inferential tests (t-test and 1-way ANOVA), and linear regression analyses were performed.

Results: We found that 55.1% (241/438) of users were either very satisfied or satisfied with the health care services of the SSK scheme. The most satisfactory indicators were related to privacy maintained during diagnostic tests (mean 3.91, SD 0.64), physicians’ behaviors (mean 3.86, SD 0.77), services provided at the registration booth (mean 3.86, SD 0.62), confidentiality maintained regarding diseases (mean 3.78, SD 0.72), and nurses’ behaviors (mean 3.60, SD 0.83). Poor satisfaction was identified in the interaction of patients with providers about illness-related information (mean 2.14, SD 1.40), availability of drinking water (mean 1.46, SD 0.76), cleanliness of toilets (mean 2.85, SD 1.04), and cleanliness of the waiting room (mean 2.92, SD 1.09). Patient satisfaction significantly decreased by 0.20 points for registration times of 16-30 minutes and by 0.32 points for registration times of >30 minutes compared with registration times of ≤15 minutes. Similarly, patient satisfaction significantly decreased with an increase in the waiting time to obtain services. However, the satisfaction of users significantly increased if they received a complete course of medicines and all prescribed diagnostic services.

Conclusions: More than half of the users were satisfied with the services provided under the SSK scheme. However, there is scope for improving user satisfaction. To improve the satisfaction level, the SSK scheme implementation authorities should pay attention to reducing the registration time and waiting time to obtain services and improving the availability of drugs and prescribed diagnostic services. The authorities should also ensure the supply of drinking water and enhance the cleanliness of the facility.
KEYWORDS
Shasthyo Surokhsa Karmasuchi; health care services; health care utilization; satisfaction; below poverty line; Bangladesh; patient satisfaction; physician behavior

Introduction

Globally, more than half of the population encounters difficulties in accessing essential health care services, with the majority residing in low- and middle-income countries (LMICs) [1]. These nations experience substantial challenges in financing health care [2-5]. Consequently, health care financing in these countries heavily relies on out-of-pocket spending by households, leading to increased financial distress on families during their illness [2,3,6]. In many instances, the most affected are those in poverty, and they lack access to health care services when they are unwell [7]. Similar to other LMICs, out-of-pocket spending for health care in Bangladesh is notably high. Recent evidence indicates that 68.5% of the total health care expenditure is shouldered by households through out-of-pocket payments [8]. Another recent study reported that such high out-of-pocket payments resulted in 24.6% of households experiencing catastrophic health expenditure when estimated using the 10% threshold of the budget share method. Furthermore, in 2016, over 8.5 million people were pushed into poverty due to health care expenses [9]. Moreover, the incidence of catastrophic health expenditure is more concentrated among the poorest households (16.5%) compared to the richest (9.2%) [10]. To reduce the burden of health care among the population and progress toward universal health coverage, the Government of Bangladesh has developed the Health Care Financing Strategy 2012–2032, intending to provide financial protection for health care to all citizens by 2032 [11]. As a component of this strategy, the Health Economics Unit of the Ministry of Health and Family Welfare of the Government of Bangladesh has been implementing a social health protection scheme known as “Shasthyo Surokhsa Karmasuchi (SSK)” since 2016. Although there is a comprehensive plan to cover the entire population of the country within the financing scheme, the current implementation is limited to a noncontributory scheme focusing on the below-poverty-line population. The scheme is being piloted in 3 subdistricts: Kalihati, Madhupur, and Ghatail under Tangail District. The scheme has enrolled almost 1,00,000 households that have access to inpatient health care services from Upazila Health Complexes (UzHCs) of the respective Upazilas (subdistricts) and district hospitals. Participation in the scheme is mandatory for households identified as being below the poverty line, and their enrollment is noncontributory, meaning that these enrolled households are not required to pay any fees for services. Notably, the scheme does not offer purchasing services to the above-poverty-line population. The government established a pool fund, allocating BDT 1000 (US$12) per household per year as a premium (BDT 84.5 = US$ 1, August 2019, Bangladesh Bank). This measure ensures access to inpatient health care services for the enrolled below-poverty-line households, covering 78 different disease groups. The annual coverage limit for each household is BDT 50,000 (US $592). Under the scheme, inpatient health care is delivered through UzHCs, serving as the first access point for the insured beneficiaries to receive health care services. Through a structured referral system, the beneficiaries can also access services at the Tangail District Hospital. The scheme ensures that insured patients receive free diagnostic services and medicines through hospitals, contracted diagnostic centers, and pharmacies. The SSK management authority, scheme operator, hospitals, contracted diagnostic centers, and pharmacies play crucial roles in the implementation of the scheme [12].

Although the scheme provides free inpatient care services to the member households, the health care utilization under the SSK scheme is notably low. A study revealed that less than half of the beneficiary households used health care services under the SSK scheme [13]. Several factors may contribute to this low utilization rate. For instance, quality of care might be a significant factor among the various important determinants of health service utilization. Quality of service is recognized as one of the key components in achieving universal health coverage by its definition [14]. Traditionally, the quality of health care services was primarily assessed based on professional practice standards. However, in the recent decades, patients’ perceptions of health care have emerged as an important indicator for evaluating the quality of health care services. Various studies have demonstrated that health service utilization is closely linked with users’ perceptions of the quality of health care provided [15-17]. Consequently, patient satisfaction is considered as an important aspect of performance improvement of the delivered health care services, alongside clinical effectiveness. It is a multidimensional aspect where patients’ perceptions and attitudes shape their overall health care–seeking experience [18,19]. Several factors, including registration time and process, waiting time to obtain health care services, interpersonal communication, and availability of basic amenities within health care facilities, can influence patient satisfaction with health care services [20-23]. Increased utilization and satisfaction of any insurance scheme are associated with improved quality of health care services. However, the literature provides mixed evidence. For example, a study in India found no significant difference in satisfaction levels between insured and uninsured hospitalized patients [24]. Conversely, regarding the overall quality of care provided under the National Health Insurance Scheme of Ghana, a significant portion of insured patients reported higher satisfaction compared with uninsured patients [25]. Evidence from Nigeria indicated that most patients were satisfied with the service delivery of their national health insurance scheme [26-28]. In Ethiopia, a study revealed that approximately 55% of enrollees were satisfied with the community-based health insurance scheme [29], whereas another study from the same country indicated that over 90% of households were satisfied with the community-based health insurance scheme [30]. Different Vietnamese studies have reported poor satisfaction among beneficiaries regarding service coverage and quality of care.
under national health insurance [31,32]. A recent study conducted on a self-financed health insurance scheme in Bangladesh showed that, overall, members of the scheme were satisfied with the health care services; however, their satisfaction levels could be improved in several aspects of health care service delivery [33].

Despite the pilot implementation of the SSK scheme since 2016 and its low utilization, no research has been conducted on service users’ experiences and levels of satisfaction with the scheme. Gaining a better understanding of beneficiaries’ experiences and levels of satisfaction with the health care service provided under the pilot SSK scheme is crucial. This insight can help identify the gaps in the quality of health care services provided. Such evidence will be useful for the key stakeholders of the health protection scheme, allowing them to make necessary changes in the service delivery process and related aspects to enhance the quality of health care services provided under the scheme. As a result, this study was conducted to address 2 central research questions: (1) What is the level of satisfaction among the beneficiaries of the SSK scheme? and (2) What are the factors influencing their satisfaction level? In addressing these research questions, this study aimed to assess patients’ levels of satisfaction with the services offered by the SSK scheme in Bangladesh.

Methods

Study Design

A cross-sectional exit patient survey was designed to gain insights into the experiences of insured patients with various aspects of the service delivery process and the quality of services provided under the pilot SSK scheme. Every second patient who had been admitted for at least 2 nights at a scheme-designated facility was selected and interviewed at the time of discharge.

Study Setting and Sample

The study was conducted in the UzHCs of Kalihati, Madhupur, and Ghatail Upazilas (subdistricts), and Tangail District Hospital of Tangail District. Insured inpatients were interviewed after discharge from the health care facilities. The survey of the respondents took place between August 18 and November 16, 2019, on working days, from Saturday to Thursday. Every second discharged inpatient from the male and female wards was interviewed. To ensure the quality of the data, a maximum of 4 patients were interviewed each day at an SSK hospital. A total of 438 discharged inpatients aged 18 years or older were interviewed from 3 UzHCs (Kalihati, n=128; Madhupur, n=176; and Ghatail, n=134) and Tangail District Hospital (n=88).

Data Collection Process

A semistructured questionnaire was designed and pretested before data collection. Face-to-face interviews were conducted with the insured patients and, in certain cases, with attendants of patients at the time of discharge. An attendant was considered as a respondent when the patient was not involved with the various dimensions of the service delivery process during the inpatient episode owing to the physical condition.

The questionnaire covered demographic and socioeconomic details of the respondents and households, health care utilization, and various dimensions of satisfaction related to the SSK scheme. These dimensions included the registration process at the SSK booth, the dignity of patients during treatment, clear communication with health care providers, privacy during treatment, the quality of basic amenities, the availability of drugs and supplies, and the availability of prescribed diagnostic services.

Four experienced research assistants were employed for patient recruitment and conducting the interviews. Prior to the interviews, written informed consent was obtained from all participants, and their participation was entirely voluntary. Completed interviews were cross-checked among the interviewers and further reviewed by the supervisor to ensure data quality and to address any associated issues, if needed, during the data collection.

Study Variables

We collected information on various background characteristics of the patients, including age, sex, education level, current employment status, current marital status, and family size. For measuring satisfaction levels, we considered several dimensions of health care delivery under the SSK scheme.

The first dimension was hospitalization-related factors. It included self-reported illnesses and length of stay. Self-reported illnesses were categorized into 3 groups: communicable, noncommunicable, and others (ie, obstetrics and injury). Communicable diseases encompass illnesses caused by viruses or bacteria that spread through contact, bodily fluids, blood products, insect bites, or the air. Noncommunicable diseases, on the other hand, are those that do not transmit between individuals and often necessitate long-term treatment.

The second dimension was service utilization–related aspects. It included waiting time for registration, waiting time to obtain health care services, behavior of health care providers (including physicians, nurses, and other staff, such as ward boys and cleaners), interaction of health care providers with patients, privacy during diagnostic services, and confidentiality of the health care provided.

The third dimension was facility environment and basic amenity–related factors. It included cleanliness of health facilities, waiting rooms, and toilets, and availability of drinking water.

The satisfaction measurement items demonstrated a satisfactory level of internal consistency, as indicated by an overall Cronbach α coefficient of 0.77 out of 1.0 [34].

Satisfaction Measurements

Patient satisfaction was measured with a collective outcome of 14 different items. The selection of items for measurements was devised based on a literature review of patient satisfaction with the insurance scheme as well as previous systematic reviews [26,27,33,35-40]. The existing literature has examined various aspects of health service delivery from the patients’ viewpoints, encompassing domains such as patient-provider interactions, the physical environment, and internal management processes.
We selected items that revolved around these domains as they encompassed the most influential satisfaction constructs. The 14 items are presented in Textbox 1.

Each considered item was rated on a 5-point Likert scale (very satisfied, 5; satisfied, 4; neither satisfied nor dissatisfied, 3; dissatisfied, 2; very dissatisfied, 1). The total satisfaction score of respondents for all items ranged from a minimum of 14 to a maximum of 70. Furthermore, we included an item in the questionnaire to assess the overall satisfaction (on a scale of 5) with the services at the SSK facility.

Textbox 1. Items for patient satisfaction.

<table>
<thead>
<tr>
<th>Satisfaction items</th>
</tr>
</thead>
<tbody>
<tr>
<td>How will you rate the behavior of the authority of Shasthyo Surokhsha Karmasuchi (SSK) at the registration booth?</td>
</tr>
<tr>
<td>What is your opinion about the time taken for completing registration?</td>
</tr>
<tr>
<td>What is your opinion about the waiting time before consultation with the service provider?</td>
</tr>
<tr>
<td>How will you rate the behavior of the service provider during your treatment at this hospital?</td>
</tr>
<tr>
<td>How will you rate the behavior of nurses during your treatment at this hospital?</td>
</tr>
<tr>
<td>How will you rate the behavior of the aya/ward boy during your treatment at this hospital?</td>
</tr>
<tr>
<td>How will you rate the interaction with the service provider about your illness and treatment?</td>
</tr>
<tr>
<td>How will you rate the doctor’s attitude toward listening to your problems?</td>
</tr>
<tr>
<td>How will you rate the privacy maintained during diagnostic tests?</td>
</tr>
<tr>
<td>What is your opinion about the privacy maintained during consultation?</td>
</tr>
<tr>
<td>What is your opinion about the cleanliness of this hospital?</td>
</tr>
<tr>
<td>How will you rate the cleanliness of the waiting room of this hospital?</td>
</tr>
<tr>
<td>How will you rate the cleanliness of the toilets of this hospital?</td>
</tr>
<tr>
<td>What is your opinion regarding the availability of drinking water in the hospital?</td>
</tr>
</tbody>
</table>

Statistical Analysis

We analyzed the data using Stata version 16 (StataCorp) [41]. We performed both descriptive analysis and statistical inferential tests to measure the association between dependent and independent variables. In the descriptive analysis, background characteristics of the study participants and health care facility utilization–related characteristics were presented in terms of frequency (n) and percentage (%) with 95% CIs. Moreover, we performed a t-test for variables with 2 categories and 1-way ANOVA for variables with more than 2 categories to test the significant differences in average satisfaction levels across the demographic and socioeconomic characteristics related to the SSK scheme.

To identify factors associated with patients’ average scores for satisfaction with the services under the SSK scheme, a linear regression analysis was performed. We estimated the satisfaction level for each patient by taking the average of the reported satisfaction levels in the 14 items. In the univariate unadjusted regression model, the dependent variable was the mean satisfaction score and the independent variables were age, gender, education, employment status, marital status, family size, self-reported illness, length of hospitalization, registration time, waiting time to obtain services, status of receiving drugs and supplies, and status of receiving diagnostic services. However, in the multivariate regression model, we included independent variables that had a significant association with the satisfaction score (ie, P values ≤.05) in the univariate regression models. We considered P values of <.05 as statistically significant in our analysis.

Ethics Approval

This study was approved by the Research Review Committee and Ethical Review Committee of the icddr,b (protocol#: PR-17047). Participants in the study were recruited and interviewed after obtaining written informed consent, and their participation was voluntary.

Results

Descriptive Statistics

A total of 438 patients aged 18 years and above were interviewed at the studied facilities (Table 1), and 60.1% (263/438) of the patients were female. According to education level, 60.9% (267/438) of patients had no education, whereas 24.2% (106/438) and 14.8% (65/438) had primary and secondary levels of education, respectively. Moreover, 67.8% (297/438) of patients were not involved with income generation. In terms of marital status, 83.3% (365/438) of patients were married. Moreover, 54.1% (237/438) were from a household consisting of more than 4 members.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Value (N=438), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant variable</strong></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>158 (36.1)</td>
</tr>
<tr>
<td>18-44</td>
<td>202 (46.1)</td>
</tr>
<tr>
<td>45-64</td>
<td>78 (17.8)</td>
</tr>
<tr>
<td>&gt;64</td>
<td>175 (39.9)</td>
</tr>
<tr>
<td>Male</td>
<td>263 (60.1)</td>
</tr>
<tr>
<td>Female</td>
<td>267 (61.0)</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
</tr>
<tr>
<td>No education</td>
<td>267 (61.0)</td>
</tr>
<tr>
<td>Primary</td>
<td>106 (24.2)</td>
</tr>
<tr>
<td>Secondary or higher</td>
<td>65 (14.8)</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>141 (32.2)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>197 (44.0)</td>
</tr>
<tr>
<td>Retired or student</td>
<td>100 (22.8)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>11 (2.5)</td>
</tr>
<tr>
<td>Married</td>
<td>365 (83.3)</td>
</tr>
<tr>
<td>Widowed, divorced, or separated</td>
<td>62 (14.2)</td>
</tr>
<tr>
<td>Family size</td>
<td></td>
</tr>
<tr>
<td>≤4 members</td>
<td>201 (45.9)</td>
</tr>
<tr>
<td>&gt;4 members</td>
<td>237 (54.1)</td>
</tr>
<tr>
<td>Self-reported illness</td>
<td></td>
</tr>
<tr>
<td>Communicable</td>
<td>135 (30.8)</td>
</tr>
<tr>
<td>Noncommunicable</td>
<td>274 (62.6)</td>
</tr>
<tr>
<td>Others (ie, obstetrics and injury)</td>
<td>29 (6.6)</td>
</tr>
<tr>
<td><strong>Hospital service utilization variable</strong></td>
<td></td>
</tr>
<tr>
<td>Length of hospitalization (days)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>151 (34.5)</td>
</tr>
<tr>
<td>3-4</td>
<td>208 (47.5)</td>
</tr>
<tr>
<td>&gt;4</td>
<td>79 (18.0)</td>
</tr>
<tr>
<td>Registration time (min)</td>
<td></td>
</tr>
<tr>
<td>≤15</td>
<td>290 (66.2)</td>
</tr>
<tr>
<td>16-30</td>
<td>91 (20.8)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>57 (13.0)</td>
</tr>
<tr>
<td>Waiting time to get services (min)</td>
<td></td>
</tr>
<tr>
<td>≤15</td>
<td>258 (58.9)</td>
</tr>
<tr>
<td>16-30</td>
<td>84 (19.2)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>96 (21.9)</td>
</tr>
<tr>
<td>Status of getting drugs and supplies</td>
<td></td>
</tr>
<tr>
<td>Partially received</td>
<td>91 (20.8)</td>
</tr>
</tbody>
</table>
According to self-reported diseases, 62.6% (274/438) of patients reported the reason for hospitalization as noncommunicable disease, 30.8% (135/438) reported the reason as communicable disease, and 7.0% (29/438) reported the reason as other health problems (ie, obstetrics and injury). Regarding the length of hospitalization, 47.5% (208/438) of patients were admitted for 3-4 days, 34.5% (151/438) were admitted for 2 days, and 18.0% (79/438) were admitted for more than 4 days. Among the respondents, 66.2% (290/438) mentioned that they had completed their registration process within 15 minutes, and 58.9% (258/438) waited for 15 minutes or less to get services. The majority of patients (347/438, 79.2%) received all prescribed medicines and supplies free from the SSK pharmacy. Regarding laboratory services, 74.4% (326/438) of patients reported that they received diagnostic services as prescribed. More details of the descriptive statistics are shown in Table 1.

### Level of Satisfaction by Different Items

Patient satisfaction with the items considered while using the SSK scheme is shown in Table 2. A total of 14 satisfaction items were used to examine patient satisfaction. The highest average score on satisfaction was related to “privacy maintained during diagnostic tests” (mean 3.91, SD 0.64), followed by “physicians’ behaviors” (mean 3.86, SD 0.77), “services at the SSK registration booth” (mean 3.86, SD 0.62), “confidentiality maintained about diseases” (mean 3.78, SD 0.72), and “services from nurses” (mean 3.6, SD 0.83). Among service-related items, a lower level of satisfaction was reported for the interaction of service providers with patients (mean 2.14, SD 1.4). Among the items in the environment and basic amenities domain, comparatively higher satisfaction was found for the cleanliness of the health facility (mean 3.43, SD 0.76), followed by the cleanliness of the waiting room (mean 2.92, SD 1.09) and toilets (mean 2.85, SD 1.04). The lowest level of satisfaction was reported for the availability of drinking water (mean 1.46, SD 0.76).

### Table 2. Patient satisfaction with health care services at Shasthya Suroksha Karmasuchi facilities by different items (N=438).

<table>
<thead>
<tr>
<th>Item</th>
<th>Very satisfied, n (%)</th>
<th>Satisfied, n (%)</th>
<th>Neutral, n (%)</th>
<th>Dissatisfied, n (%)</th>
<th>Very dissatisfied, n (%)</th>
<th>Overall score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Services at the SSK® registration booth (reception)</td>
<td>30 (6.9)</td>
<td>338 (77.2)</td>
<td>51 (11.6)</td>
<td>15 (3.4)</td>
<td>4 (0.9)</td>
<td>3.86 (0.62)</td>
</tr>
<tr>
<td>2. Registration time</td>
<td>53 (12.1)</td>
<td>191 (43.6)</td>
<td>93 (21.2)</td>
<td>48 (11.0)</td>
<td>53 (12.1)</td>
<td>3.33 (1.19)</td>
</tr>
<tr>
<td>3. Waiting time to get health care services</td>
<td>74 (16.9)</td>
<td>146 (33.3)</td>
<td>78 (17.8)</td>
<td>61 (13.9)</td>
<td>79 (18.0)</td>
<td>3.17 (1.36)</td>
</tr>
<tr>
<td>4. Physicians’ behaviors</td>
<td>52 (11.9)</td>
<td>314 (71.7)</td>
<td>42 (9.6)</td>
<td>20 (4.6)</td>
<td>10 (2.3)</td>
<td>3.86 (0.77)</td>
</tr>
<tr>
<td>5. Nurses’ behaviors</td>
<td>29 (6.6)</td>
<td>265 (60.5)</td>
<td>94 (21.5)</td>
<td>40 (9.1)</td>
<td>10 (2.3)</td>
<td>3.60 (0.83)</td>
</tr>
<tr>
<td>6. Other staff behaviors</td>
<td>12 (2.7)</td>
<td>249 (56.9)</td>
<td>117 (26.7)</td>
<td>44 (10.1)</td>
<td>16 (3.7)</td>
<td>3.45 (0.85)</td>
</tr>
<tr>
<td>7. Interaction of health care providers with patients regarding illness</td>
<td>30 (6.9)</td>
<td>83 (19.0)</td>
<td>32 (7.3)</td>
<td>65 (14.8)</td>
<td>228 (52.1)</td>
<td>2.14 (1.40)</td>
</tr>
<tr>
<td>8. Empathy of health care providers</td>
<td>38 (8.7)</td>
<td>173 (40.0)</td>
<td>126 (28.8)</td>
<td>69 (15.8)</td>
<td>32 (7.3)</td>
<td>3.27 (1.06)</td>
</tr>
<tr>
<td>9. Privacy during diagnostics among patients who got diagnostic tests</td>
<td>47 (12.9)</td>
<td>250 (68.7)</td>
<td>54 (14.8)</td>
<td>13 (3.6)</td>
<td>0 (0.0)</td>
<td>3.91 (0.64)</td>
</tr>
<tr>
<td>10. Confidentiality of diseases</td>
<td>40 (9.1)</td>
<td>292 (66.7)</td>
<td>83 (19.0)</td>
<td>17 (3.9)</td>
<td>6 (1.4)</td>
<td>3.78 (0.72)</td>
</tr>
<tr>
<td>11. Cleanliness of the health facility</td>
<td>10 (2.3)</td>
<td>226 (51.6)</td>
<td>151 (34.5)</td>
<td>45 (10.3)</td>
<td>6 (1.4)</td>
<td>3.43 (0.76)</td>
</tr>
<tr>
<td>12. Cleanliness of the waiting room</td>
<td>4 (0.9)</td>
<td>163 (37.2)</td>
<td>133 (30.4)</td>
<td>71 (16.2)</td>
<td>67 (15.3)</td>
<td>2.92 (1.09)</td>
</tr>
<tr>
<td>13. Cleanliness of toilets</td>
<td>8 (1.8)</td>
<td>132 (30.1)</td>
<td>134 (30.6)</td>
<td>113 (25.8)</td>
<td>51 (11.6)</td>
<td>2.85 (1.04)</td>
</tr>
<tr>
<td>14. Availability of drinking water</td>
<td>2 (0.5)</td>
<td>16 (3.7)</td>
<td>11 (2.5)</td>
<td>125 (28.5)</td>
<td>284 (64.8)</td>
<td>1.46 (0.76)</td>
</tr>
</tbody>
</table>

^aSSK: Shasthya Suroksha Karmasuchi.
Overall Patient Satisfaction With Health Care Services at SSK Facilities

Considering the response to the overall patient satisfaction with the services at SSK facilities, 8.5% (37/438) reported being very satisfied and 46.6% (204/438) reported being satisfied with the services received under the SSK scheme. On the other hand, 31.3% (137/438) of respondents reported feeling neither satisfied nor dissatisfied. Moreover, 8.9% (39/438) were dissatisfied and 4.8% (21/438) were very dissatisfied (Multimedia Appendix 1).

Patient Satisfaction by Socioeconomic and Hospital Service Utilization Characteristics

Patient satisfaction levels significantly varied across different groups of age, sex, marital status, illness type, registration time, waiting time, status of receiving drugs, and status of getting diagnostic tests (Table 3). Patients aged between 45 and 64 years were comparatively more satisfied (mean 3.28, 95% CI 3.21-3.34) with services under the SSK scheme, and the difference in the satisfaction level across the age groups was statistically significant ($P<.001$). Male patients were significantly ($P=.01$) more satisfied (mean 3.24, 95% CI 3.17-3.31) than female patients. Married and widowed, divorced, or separated individuals were more satisfied than unmarried individuals, and the difference was statistically significant ($P<.001$). However, there was no significant difference in satisfaction by education level, employment status, or household size.

Patients with noncommunicable diseases had a higher satisfaction level (mean 3.22, 95% CI 3.17-3.28) than patients with other illnesses, and the difference in the satisfaction level was statistically significant ($P=.008$). Satisfaction scores decreased with increases in the length of hospitalization, registration time, and waiting time. The satisfaction level was significantly ($P=.006$) higher among patients who received all prescribed drugs from the scheme (mean 3.20, 95% CI 3.15-3.26). Similarly, the satisfaction level was higher among patients who received all prescribed diagnostic or laboratory services compared with other groups (mean 3.22, 95% CI 3.17-3.27), and the difference in the satisfaction level across the groups was statistically significant ($P<.001$).
Table 3. Satisfaction scores by patient and service characteristics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Score, mean (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-44</td>
<td>3.03 (2.95-3.10)</td>
<td>&lt;.001(^a)</td>
</tr>
<tr>
<td>45-64</td>
<td>3.28 (3.21-3.34)</td>
<td></td>
</tr>
<tr>
<td>&gt;64</td>
<td>3.18 (3.08-3.28)</td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td>.01(^b)</td>
</tr>
<tr>
<td>Male</td>
<td>3.24 (3.17-3.31)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3.12 (3.06-3.18)</td>
<td></td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td>.21(^a)</td>
</tr>
<tr>
<td>No education</td>
<td>3.19 (3.13-3.24)</td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>3.17 (3.07-3.28)</td>
<td></td>
</tr>
<tr>
<td>Secondary or higher</td>
<td>3.08 (2.96-3.20)</td>
<td></td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
<td>.47(^a)</td>
</tr>
<tr>
<td>Employed</td>
<td>3.22 (3.14-3.30)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>3.16 (3.09-3.23)</td>
<td></td>
</tr>
<tr>
<td>Retired and student</td>
<td>3.12 (3.03-3.21)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td>&lt;.001(^a)</td>
</tr>
<tr>
<td>Unmarried</td>
<td>2.65 (2.42-2.88)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>3.19 (3.14-3.24)</td>
<td></td>
</tr>
<tr>
<td>Widowed, divorced, or separated</td>
<td>3.12 (2.99-3.24)</td>
<td></td>
</tr>
<tr>
<td><strong>Family size</strong></td>
<td></td>
<td>.21(^b)</td>
</tr>
<tr>
<td>≤4</td>
<td>3.20 (3.14-3.26)</td>
<td></td>
</tr>
<tr>
<td>&gt;4</td>
<td>3.14 (3.08-3.21)</td>
<td></td>
</tr>
<tr>
<td><strong>Self-reported illness</strong></td>
<td></td>
<td>.008(^a)</td>
</tr>
<tr>
<td>Communicable</td>
<td>3.07 (2.98-3.16)</td>
<td></td>
</tr>
<tr>
<td>Noncommunicable</td>
<td>3.22 (3.17-3.28)</td>
<td></td>
</tr>
<tr>
<td>Others (ie, obstetrics and injury)</td>
<td>3.10 (2.93-3.27)</td>
<td></td>
</tr>
<tr>
<td><strong>Length of hospitalization (days)</strong></td>
<td></td>
<td>.13(^a)</td>
</tr>
<tr>
<td>2</td>
<td>3.12 (3.03-3.21)</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>3.17 (3.11-3.23)</td>
<td></td>
</tr>
<tr>
<td>&gt;4</td>
<td>3.26 (3.16-3.35)</td>
<td></td>
</tr>
<tr>
<td><strong>Registration time (min)</strong></td>
<td></td>
<td>&lt;.001(^a)</td>
</tr>
<tr>
<td>≤15</td>
<td>3.25 (3.20-3.30)</td>
<td></td>
</tr>
<tr>
<td>16-30</td>
<td>3.05 (2.95-3.16)</td>
<td></td>
</tr>
<tr>
<td>&gt;30</td>
<td>2.93 (2.80-3.07)</td>
<td></td>
</tr>
<tr>
<td><strong>Waiting time to get services (min)</strong></td>
<td></td>
<td>&lt;.001(^a)</td>
</tr>
<tr>
<td>≤15</td>
<td>3.31 (3.25-3.37)</td>
<td></td>
</tr>
<tr>
<td>16-30</td>
<td>3.01 (2.91-3.12)</td>
<td></td>
</tr>
<tr>
<td>&gt;30</td>
<td>2.92 (2.83-3.01)</td>
<td></td>
</tr>
<tr>
<td><strong>Status of getting drugs and supplies</strong></td>
<td></td>
<td>.006(^b)</td>
</tr>
<tr>
<td>Partially received</td>
<td>3.04 (2.96-3.13)</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Score, mean (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------</td>
<td>---------</td>
</tr>
<tr>
<td>All received</td>
<td>3.20 (3.15-3.26)</td>
<td>&lt;.001a</td>
</tr>
<tr>
<td>Status of getting laboratory services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not prescribed</td>
<td>2.94 (2.81-3.06)</td>
<td></td>
</tr>
<tr>
<td>Partially received</td>
<td>3.16 (3.02-3.30)</td>
<td></td>
</tr>
<tr>
<td>All received</td>
<td>3.22 (3.17-3.27)</td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>3.17 (3.12-3.21)</td>
<td></td>
</tr>
</tbody>
</table>

aOne-way ANOVA.

Determinants of Patient Satisfaction With Services Provided Under the SSK Scheme

Our analysis demonstrated noteworthy associations between satisfaction scores and various factors (Table 4). The satisfaction score was significantly higher by 0.13 points in patients aged between 45 and 64 years than in patients aged between 18 and 44 years. Additionally, the satisfaction score was significantly higher by 0.34 points in married patients than in unmarried patients. Moreover, the satisfaction score was significantly higher by 0.15 points in patients seeking care for noncommunicable diseases than in patients seeking care for communicable diseases. We found a significant negative association of the satisfaction score with extended registration and waiting time for obtaining services. Conversely, a positive association was observed with the status of receiving all drugs, supplies, and diagnostic services. The satisfaction score was significantly lower by 0.18 points in patients with a registration time of 16-30 minutes and by 0.33 points in patients with a registration time of >30 minutes than in patients with a registration time of ≤15 minutes. Similarly, the satisfaction score was significantly lower by 0.30 points in patients who waited for 16-30 minutes to obtain services and by 0.36 points in patients who waited for >30 minutes to obtain services than in patients who waited for ≤15 minutes to obtain services. Moreover, the satisfaction score was significantly higher by 0.13 points in patients who received the complete course of prescribed medicines from the SSK pharmacy than in patients who received partial medicines and supplies. Likewise, the satisfaction score was significantly higher by 0.26 points in patients who received partial diagnostic services and by 0.28 points in patients who received full diagnostic services than in patients who were not prescribed diagnostic services.
Table 4. Determinants of patient satisfaction with services under the Shasthyo Surokhsha Karmasuchi scheme.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unadjusted coefficient(^b), value (95% CI)</th>
<th>P value</th>
<th>Adjusted coefficient(^b), value (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-44 (reference)</td>
<td>N/A(^c)</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>45-64</td>
<td>0.25 (0.15 to 0.35)</td>
<td>&lt;.001</td>
<td>0.13 (0.03 to 0.22)</td>
<td>.009</td>
</tr>
<tr>
<td>&gt;64</td>
<td>0.15 (0.02 to 0.28)</td>
<td>.02</td>
<td>0.03 (−0.10 to 0.16)</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.12 (0.02 to 0.21)</td>
<td>.01</td>
<td>0.09 (0.00 to 0.18)</td>
<td>.043</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No education (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>−0.02 (−0.13 to 0.09)</td>
<td>.74</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Secondary or higher</td>
<td>−0.10 (−0.24 to 0.02)</td>
<td>.11</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>−0.06 (−0.17 to 0.05)</td>
<td>.26</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Retired or student</td>
<td>−0.10 (−0.22 to 0.03)</td>
<td>.13</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.54 (0.25 to 0.83)</td>
<td>&lt;.001</td>
<td>0.34 (0.08 to 0.61)</td>
<td>.01</td>
</tr>
<tr>
<td>Widowed, divorced, or separated</td>
<td>0.47 (0.16 to 0.78)</td>
<td>.003</td>
<td>0.26 (−0.03 to 0.55)</td>
<td>.08</td>
</tr>
<tr>
<td><strong>Family size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4 (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>&gt;4</td>
<td>0.06 (−0.03 to 0.15)</td>
<td>.21</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Self-reported illness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicable (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Noncommunicable</td>
<td>0.15 (0.05 to 0.25)</td>
<td>.003</td>
<td>0.10 (0.01 to 0.19)</td>
<td>.03</td>
</tr>
<tr>
<td>Others (ie, obstetrics and injury)</td>
<td>0.03 (−0.16 to 0.23)</td>
<td>.74</td>
<td>0.01 (−0.16 to 0.18)</td>
<td>.91</td>
</tr>
<tr>
<td><strong>Length of hospitalization (days)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>0.05 (−0.05 to 0.15)</td>
<td>.33</td>
<td>0.02 (−0.07 to 0.11)</td>
<td>.63</td>
</tr>
<tr>
<td>&gt;4</td>
<td>0.14 (0.00 to 0.27)</td>
<td>.045</td>
<td>0.06 (−0.06 to 0.18)</td>
<td>.32</td>
</tr>
<tr>
<td><strong>Registration time (min)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤15 (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>16-30</td>
<td>−0.20 (−0.31 to −0.08)</td>
<td>&lt;.001</td>
<td>−0.18 (−0.28 to −0.09)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&gt;30</td>
<td>−0.32 (−0.46 to −0.19)</td>
<td>&lt;.001</td>
<td>−0.33 (−0.45 to −0.21)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Waiting time to get services (min)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤15 (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>16-30</td>
<td>−0.30 (−0.41 to −0.18)</td>
<td>&lt;.001</td>
<td>−0.30 (−0.40 to −0.20)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&gt;30</td>
<td>−0.39 (−0.49 to −0.28)</td>
<td>&lt;.001</td>
<td>−0.36 (−0.46 to −0.26)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Status of getting drugs and supplies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partially received (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>All received</td>
<td>0.16 (0.05 to 0.27)</td>
<td>.01</td>
<td>0.13 (0.04 to 0.23)</td>
<td>.008</td>
</tr>
</tbody>
</table>
The study showed that patient satisfaction was the highest regarding the privacy and confidentiality maintained by providers during diagnostic tests and the patients’ diseases. The finding is similar to that in a study conducted in Bangladesh [33] among the beneficiaries of a community-based health insurance scheme. Another study conducted among adult patients at a general hospital in Ethiopia also reported that patient privacy and confidentiality maintained by health care providers were significantly associated with higher satisfaction levels [47]. Our study found that patients were satisfied with providers’ behaviors, particularly physicians’ and nurses’ behaviors, which influenced the overall level of patient satisfaction. Although not directly comparable, the proportion of patients satisfied with the behavior of providers was higher than the proportion reported in a study conducted in rural Bangladesh (84% vs 69%) [45]. Previous studies have also reported that the behavior of health care providers toward patients is directly connected with patient satisfaction [33,43,48].

Regarding interactions with health care providers, our study found that two-thirds of patients were not satisfied. This might be the result of patients not knowing about their illnesses from physicians during their treatment episodes. It is evident from the literature that patients’ satisfaction levels are influenced by healthy interpersonal communication with health care providers as this maintains a better physician-patient relationship [43]. A previous study conducted in Bangladesh showed that more than half of the surveyed patients could not ask questions to their providers about their illness [49]. However, as all patients in our study were inpatients and stayed at the facility for at least 2 days, it is unlikely that patients could not ask their providers about their illness.

Patient experiences with the cleanliness of health facilities and toilets and the availability of drinking water were not positive. Previous studies revealed that the health facility environment and cleanliness were crucial aspects of patient satisfaction [33,50-52]. Moreover, evidence indicates that since environmental contamination is directly connected with nosocomial infection, the physical environment can lead to the dissatisfaction of patients at health facilities instead of increasing satisfaction [33,50-52].

We found that patient age was significantly associated with the level of satisfaction. Another study conducted in Bangladesh [44] reported significant variation in the average satisfaction score across patient age, which is similar to our findings. Two other studies conducted among beneficiaries of health insurance schemes also reported similar findings that age was significantly associated with the level of satisfaction [30,53]. Lower waiting times for registration and health care were significantly associated with patient satisfaction. The findings are consistent with the literature that patients’ satisfaction levels are influenced by healthy interpersonal communication with health care providers as this maintains a better physician-patient relationship [43].

### Discussion

#### Principal Results and Comparison With Prior Work

We found that 55.1% (241/438) of patients were either very satisfied or satisfied with the services provided by the SSK health protection scheme. The mean satisfaction score was 3.17 out of 5, which means that, on average, the satisfaction level among the patients was slightly above the level of neither satisfied nor dissatisfied. Regarding the 14 considered items for measuring satisfaction, most of the patients were either very satisfied or satisfied with services at the SSK center (368/438, 84.0%), physicians’ behaviors (366/438, 83.6%), and privacy maintained during diagnostic services (297/364, 81.6%). On the other hand, majority of the patients were either very dissatisfied or dissatisfied with the availability of drinking water (409/438, 93.4%) and interaction with health care providers (293/438, 66.9%) regarding the illness. In multiple regression analysis, we found that receiving prescribed drugs and diagnostic services, the waiting time for registration, and the waiting time for getting treatment were the strongest predictors of patient satisfaction.

Health financing schemes are becoming popular to maintain and improve the health of the population in LMICs [2,6,42]. The SSK health protection scheme has been introduced to increase the access of the poor population to inpatient health care services and ensure financial protection against expenditure to alleviate poverty or extreme poverty induced by out-of-pocket payments for health care in Bangladesh. Although several studies have been conducted on patient satisfaction with health care utilization in different settings in Bangladesh [33,43-46], patient satisfaction with services under the SSK health protection scheme has not been studied thus far. The mean satisfaction score in our study was higher than that in a study conducted to assess satisfaction with the service quality of UzHCs among the uninsured population (3.17 vs 2.75) [44]. The SSK scheme provides health care to members through selected UzHCs; however, compared with nonmembers, insured patients are supposed to receive all prescribed medicines and diagnostic services from private providers contracted by the scheme [13]. The situation is different for other UzHCs where the SSK scheme is not being implemented. The availability of medicines and diagnostic services under the SSK scheme might have increased the satisfaction level among the insured patients.

Our study showed that patient satisfaction was the highest regarding the privacy and confidentiality maintained by providers during diagnostic tests and the patients’ diseases. The number of observations was 438, R-square value was 0.319, and adjusted R-square value was 0.293. The dependent variable is the average satisfaction score of 14 items.

### Table: Status of getting laboratory services

<table>
<thead>
<tr>
<th>Status of getting laboratory services</th>
<th>Unadjusted coefficient, value (95% CI)</th>
<th>P value</th>
<th>Adjusted coefficient, value (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not prescribed (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Partially received</td>
<td>0.22 (0.04 to 0.41)</td>
<td>.02</td>
<td>0.26 (0.09 to 0.43)</td>
<td>.002</td>
</tr>
<tr>
<td>All received</td>
<td>0.29 (0.17 to 0.41)</td>
<td>&lt;.001</td>
<td>0.28 (0.17 to 0.39)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unadjusted coefficient, value (95% CI)</th>
<th>P value</th>
<th>Adjusted coefficient, value (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status of getting laboratory services</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not prescribed (reference)</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Partially received</td>
<td>0.22 (0.04 to 0.41)</td>
<td>.02</td>
<td>0.26 (0.09 to 0.43)</td>
<td>.002</td>
</tr>
<tr>
<td>All received</td>
<td>0.29 (0.17 to 0.41)</td>
<td>&lt;.001</td>
<td>0.28 (0.17 to 0.39)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

The number of observations was 438, R-square value was 0.319, and adjusted R-square value was 0.293.

N/A: not applicable.
with the findings that prolonged waiting times for registration and services are associated with lower client satisfaction [54,55]. Patients who received care for noncommunicable diseases were significantly more satisfied than patients having communicable diseases. This might be because people having noncommunicable diseases require regular medications, which are common and available through the contracted pharmacy. Such availability of medicines might have increased patient satisfaction. Similarly, SSK beneficiaries who received all prescribed medicines and diagnostic services were significantly more satisfied. According to the benefits package of the SSK scheme, patients should receive all prescribed medicines and diagnostic services for 78 disease groups. However, 20.8% (91/438) of patients reported that they received partial medicines and 8.7% (38/438) reported that they received partial diagnostic services. It might have happened that some of the prescribed medicines or diagnostic tests were not correlated with the 78 disease categories and therefore were not provided under the scheme. However, evidence indicates that medicines and diagnostic tests are associated with higher out-of-pocket expenditure and lead to falling into poverty [9,56,57]. Scheme beneficiaries are provided free essential medicines and free diagnostic services, and they have a low chance of incurring treatment costs and experience low risks of catastrophic health expenditure, impoverishment, and further impoverishment [9], thus increasing their satisfaction with the services under the scheme. However, other variables, such as education level, employment status, family size, and length of hospitalization, were not significantly associated with satisfaction levels. This might be because the SSK scheme targets the below-poverty-line population having relatively similar socioeconomic characteristics; thus, their perceptions of satisfaction do not vary across these factors. These findings are consistent with the findings of other studies conducted in India [24] and Turkey [37].

This is the first study to explore patient satisfaction with the pilot SSK scheme in Bangladesh. Furthermore, we included patients from all 4 facilities under the SSK scheme rather than selecting them purposively. The findings of this study will help SSK implementation authorities to understand the patient experience of the service delivery process and the quality of health care provided under the SSK scheme.

Limitations
The design of this study was observational in nature, which did not allow us to establish any causal inference with satisfaction and other characteristics under the SSK scheme without a control group. The study only focused on the point of view of the beneficiaries, and we did not explore the providers’ views in this context. The survey collected self-reported satisfaction information from patients, which is highly susceptible to social desirability bias as patients might give responses that please health care providers instead of truly reflecting their satisfaction. However, we interviewed patients at hospital premises in the absence of any providers to minimize such bias.

Conclusions
Our findings demonstrate that more than half of the patients were overall satisfied with the services provided under the SSK scheme. However, there is room for improvement in several dimensions, such as the cleanliness of the waiting room and toilets and the availability of drinking water. Furthermore, attention should be paid to minimizing the waiting time for registration and accessing health care services, and improving providers’ skills on interaction with patients. The results of this study could help stakeholders make necessary changes in the identified determinants of satisfaction related to health service delivery of the SSK scheme. Such changes will enhance the quality of services as well as increase utilization of the scheme in the target population, ultimately advancing progress toward achieving universal health coverage.

Acknowledgments
This study received support from the United States Agency for International Development (USAID) under the terms of USAID’s Research for Decision Makers (RDM) activity cooperative agreement number AID-388-A-17-00006. The views expressed herein do not necessarily reflect the views of the US Government or USAID. The icddr,b acknowledges with gratitude the commitment of the USAID to its research efforts and funding for this study, icddr,b is also thankful to the Governments of Bangladesh, Canada, Sweden, and the United Kingdom for providing core and unrestricted support. Furthermore, the authors are grateful to Health Economics Unit, Ministry of Health and Family Welfare of the Government of Bangladesh, for its cooperation during this study. All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability
The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
MZH, MGR, and MEC contributed to conceptualizing, analyzing, writing, revising, and finalizing the manuscript with the support of OA, SA, GGM, and MWA. All authors have read, revised, and approved the final version of the manuscript.
None declared.

Multimedia Appendix 1
Overall satisfaction with the inpatient care services under the Shasthyo Suroksha Karmasuchi (SSK) scheme.

References


57. Ghosh S. Catastrophic Payments and Impoverishment Due to Out-of-Pocket Health Spending. Economic and Political Weekly 2011:63-70 [FREE Full text]

Abbreviations

LMIC: low- and middle-income country
SSK: Shasthyo Surokhsha Karmasuchi
UzHC: Upazila Health Complex

Edited by A Mavragani; submitted 10.06.23; peer-reviewed by MA Islam, S Bidmon; comments to author 13.09.23; revised version received 07.10.23; accepted 22.11.23; published 24.04.24.

Please cite as:
Hasan MZ, Rabhani MG, Akter O, Melhi GG, Ahmed MW, Ahmed S, Chowdhury ME
Patient Satisfaction With the Health Care Services of a Government-Financed Health Protection Scheme in Bangladesh: Cross-Sectional Study
JMIR Form Res 2024;8:e49815
URL: https://formative.jmir.org/2024/1/e49815
PMID:38655783
Abstract

Background: Health organizations face the critical task of executing and overseeing comprehensive health care. To address the challenges associated with this task, evidence-based dashboards have emerged as valuable tools. Since 2016, the regional health organizations of Quebec, Canada, have been responsible for ensuring implementation of the Quebec Alzheimer Plan (QAP), a provincial plan that aims to reinforce the capacity of primary care services to detect, diagnose, and treat persons with dementia. Despite the provincial scope of the QAP, the diverse material and human resources across regions introduce variability in the interest, utility, and specific needs associated with these dashboards.

Objective: The aim of this study was to assess the interest and utility of dashboards to support the QAP implementation, as well as to determine the needs for improving these aspects according to the perspectives of various types of professionals involved across regions.

Methods: An evaluative study using qualitative methods was conducted within a collaborative research approach involving different stakeholders, including the ministerial advisor and the four project managers responsible for supporting the implementation of the QAP, as well as researchers/scientific advisors. To support these organizations, we developed tailored, 2-page paper dashboards, detailing quantitative data on the prevalence of dementia, the use of health services by persons with dementia, and achievements and challenges of the QAP implementation in each organization’s jurisdiction. We then conducted 23 focus groups with the managers and leading clinicians involved in the implementation of the QAP of each regional health organization. Real-time notes were taken using a structured observation grid. Content analysis was conducted according to different regions (organizations...
Introduction

Health organizations worldwide are encouraged to continually improve their performance by identifying areas in need of enhancement and implementing interventions to address them [1-3]. The learning health system approach posits that a system “learns” through a reflective or cyclical process that engages a community in an empirical analysis of data related to a problem, leading to the discovery of new knowledge and practices [1-4]. Dashboards are an important tool to support a learning health system as they allow tracking the performance of an organization or a specific process and identifying areas in need of improvement [5]. These visual and interactive tools are used to monitor, measure, control, and analyze the performance and outcomes of an organization or a specific process [6,7] and can assist decision-makers in making changes based on the obtained results [8,9]. Dashboards may include alerts, customization options, and contextual information [7,10].

Dashboards can serve decision-makers as well as health managers and professionals in focusing on the most critical activities, monitoring trends, analyzing and identifying areas in need of improvement, and making data-driven decisions [7]. They can foster reflection on the causes explaining results and adjusting actions to better achieve the intended goals [11]. In this sense, their use can be beneficial for improving the quality of care, reducing costs, and increasing efficiency in health care [12]. Additionally, dashboards reduce cognitive load, task completion time, and errors, and enhance situational awareness and adherence to evidence-based practice guidelines [12]. However, the diverse needs of different users (decision-makers, managers, or clinicians) can make it challenging to choose indicators in a dashboard that must be concise to fulfill its function [13,14]. These needs vary and depend on the clinical or managerial context, population characteristics, and the role of professionals [10], which can create information overload [6]. Indeed, health care managers in Canada often report the presence of too many indicators in dashboards, leading to confusion [13,15]. It is necessary to adapt the dashboards to professional contexts and consider the multiplicity of uses in their development and evaluation [16].

The use of a dashboard in national programs such as an Alzheimer Plan further increases the need for considering the multiplicity of uses, as these dashboards need to take into consideration regional differences, especially with respect to resources and populations. Dashboards have been used in the implementation of national programs for health improvement in many countries, including the United States, Canada, and Australia [17-19], for evaluating major neurocognitive disorders [20,21] or for improving care for these individuals [22]. To our knowledge, no study has reported the use of dashboards in the context of implementing an Alzheimer Plan.

Quebec is one of the first Canadian provinces to have developed and implemented an Alzheimer Plan [23]. The Quebec Alzheimer Plan (QAP; Plan ministériel sur les troubles neurocognitifs majeurs) gives interdisciplinary primary care clinic clinicians the responsibility to identify, assess, diagnose, treat, and follow individuals with dementia and their care partners [24,25]. The person-centered approach of the QAP is anchored in primary care, with specialized services supporting more complex clinical situations [24,25]. The QAP includes an implementation strategy supported by the development of professional and organizational capacities, the deployment of partnership governance, ambitious change management, and an independent evaluation by our research team [24,25]. For the generalization of the QAP to the entire province in 2016, the Ministry of Health and Social Services assigned the responsibility for implementing the QAP to the regional health organizations [23,25,26]. This involves training and mentoring doctors and health care professionals in each interdisciplinary primary care clinic in their territory and supporting implementation of the QAP. Four QAP project managers, one for each integrated university health and social services network overseeing regional health organizations [27], served as an interface between the Ministry of Health and Social Services and the health organizations in their territory. These four project managers support organizations to promote change and the sharing of experiences among organizations in their territory.
Health care needs and resources vary significantly from one organization to another based on regional factors such as geographic distribution and demographic characteristics of the population [1,28-30], as well as the availability of local health and social services [1,29,31,32]. For example, the higher population density [1] in urban regions can increase both health and social services needs and resources [1,32]. Although these regions benefit from more resources, access to emergency services, psychosocial services, services for older adults, and mental health services remains suboptimal in almost all regions [29]. Health organizations in more remote areas generally cover very large territories [33]. These organizations face challenges such as geographic accessibility issues [31,34], a lack of local resources across the care continuum [35-37], and an aging population with complex health needs [28,38]. It is important to consider these regional differences in the development of dashboards supporting the implementation of a large-scale Alzheimer Plan [39,40]. Thus, in 2020, the ministry entrusted our research team with the development of dashboards to support regional health organizations in their implementation activities. The objective of this study was to evaluate the interest and utility of these dashboards and to assess the needs for improving them according to various types of health professionals in different regions.

Methods

Design

We conducted an evaluative study based on descriptive qualitative methods [41].

Context Surrounding the Development of Dashboards

For dashboard development, we used a collaborative approach [42] at the core of learning health system approaches. Such an approach can lead to more relevant, efficient, and sustainable results than a more traditional approach in terms of development and problem-solving in the relevant communities and organizations [43].

In collaboration with the ministerial advisor and the four project managers of the QAP, we selected a format and relevant indicators for this exercise. These choices of format and indicators were then validated with members of the QAP advisory committee, composed of managers, leading clinicians, and anyone involved in the QAP implementation from regional organizations in the province.

A format of a maximum 2-sided sheet was selected (see Multimedia Appendix 1). The choice of a paper rather than digital format was made owing to its simplicity, accessibility (easily accessible to all without depending on technology or connectivity), and instantaneous visibility (without requiring complex navigation). These advantages were prioritized despite the knowledge of disadvantages such as the need for manual updates and space limitations. This support is particularly useful experimentally, even though we acknowledge that it will need to evolve into a digital modality during potential routinization in regular managerial processes.

On the first page, nine indicators were selected from provincial clinical-administrative data from the Quebec Integrated Chronic Disease Surveillance System developed by the Institut national de santé publique du Québec [44]. Indicator selection was based on a conceptual framework on the quality of care for dementia, covering a continuum of care from primary care to emergency use and hospitalizations [45]. Eight of these indicators (type of physician most regularly visited, percentage of people with at least one visit to a family doctor, average number of visits to a family doctor per person, percentage of people with at least one emergency room visit, percentage of people with at least one hospitalization, average number of days hospitalized per person, percentage of people with at least one hospitalization with an alternative care level—representing patients who are hospitalized but no longer requiring acute care and waiting for long-term placement—and average number of days in care level) were measured in 2019-2020 in the population of individuals with dementia 65 years and older, and one indicator (prevalence) was measured between 2000-2001 and 2019-2020 in the population of individuals with dementia 40 years and older. While these surveillance data are highly useful for tracking diseases, they are sometimes underused by local actors due to their excessive quantity and the lack of means for rapid analysis [44]. Data visualization, facilitated by dashboards, makes these data more accessible, facilitating decision-making.

On the second page, we presented findings of a thematic analysis of two sources: annual reports to the Ministry of Health and Social Services produced by the managers involved in the QAP between 2017 and 2019, as well as notes of meetings between the ministerial advisor and the four QAP project managers with the managers involved in the QAP in 2020. While quantitative data highlight high-level trends, qualitative data served to enrich the dashboards by providing context to the figures, offering a more comprehensive and engaging perspective [46].

Finally, we chose to present specific results for the population of each health organization and for the entire province. This choice was made to contextualize the results of each organization to those of the entire province.

Data Collection

In spring 2022 (April to June), the ministerial advisor and the four QAP project managers (EM, CF, JD, and CM) organized 90-minute virtual meetings with the managers and leading clinicians involved in the QAP implementation as well as other persons involved in the QAP (eg, research personnel) in each of the 24 regional health organizations. These meetings aimed to share the new directions of the QAP and explore the progress and challenges of the pandemic with each organization. The last 30 minutes of these meetings were devoted to presenting regional dashboards and facilitating group interviews by the research team. The research team had approximately 10 minutes to present the dashboards (5 minutes for quantitative aspects and 5 minutes for qualitative aspects) and approximately 20 minutes to conduct a group interview with the participants in the meetings to assess the interest and utility of the dashboards as well as to assess the needs for improvement of these dashboards. The interview guide included four open-ended questions and prompts: “Are these results surprising to you?” “How will this dashboard be useful to you?” “Are there any
results missing that you would have wanted in this dashboard?” and “Do you have any questions?”

Four members of the research team alternatively facilitated the group interviews (GAL, MG, YC, and IV). One member (GAL) was present in 20 group interviews, ensuring consistency, reinforcing study fidelity and credibility [47]. Two members were always present: one to present the quantitative part of the dashboards (GAL or IV) and the other to present the qualitative part (MG or YC). In addition, one member was designated the main note-taker, while the second member supported the main note-taker and had to read and complete the notes or suggest modifications on a document shared live and online by the main note-taker.

The note-takers were instructed to note dynamics between the participants and to indicate the participant who put forth each observation noted. Thus, a code was assigned to the interlocutor: manager (ie, director, deputy director, coordinator, unit or service head), leading clinician (ie, territorial nurse, physician, or social worker), other participant (ie, research agent or research coordinator), or an unidentified person (when note-takers had not identified the interlocutor).

Consent was obtained verbally at the onset of meetings, and no audio or video recordings were conducted to streamline the analyses to protect confidentiality, especially given the involvement of project managers in organizing the meetings.

Ethical Considerations

This study received ethical approval from the ethics committee of the Centre intégré universitaire de santé et de services sociaux de l’Estrie - Centre hospitalier universitaire de Sherbrooke (MP-31-2021-3701) to ensure that all research activities involving human participants adhered to ethical standards and guidelines. The participants provided informed consent before participating in the study, emphasizing their voluntary participation, the purpose of the research, and the confidentiality of their information. All procedures involving human subjects were conducted in accordance with the ethical standards of the ethics committee and the principles outlined in the Declaration of Helsinki. The research team prioritized participant well-being, privacy, and the responsible handling of data throughout the study.

Analysis

Overview

Content analysis was performed on the observation notes [48]. First, a research team member (GAL) read all the notes several times to become familiar with the content. Subsequently, a preliminary version of the coding manual was generated by analyzing the observation notes. Each segment of the notes taken in the observation grid, deemed relevant and related to the research question, was coded and then grouped under conceptual categories. Once the preliminary coding manual was developed, all conceptual categories were defined. A second research team member (MG), who participated in the majority of group interviews, reviewed the manual to ensure its accuracy and consistency. The coding manual was then iteratively revised during the analysis [49] until consensus was reached among all research team members (GAL, MG, YC, and IV).

Categorization by Region

Each regional health organization involved in the QAP was categorized into one of three groups using a method used by the Ministry of Health and Social Services [50].

Thus, 13 organizations were categorized in a “university/peripheral regions” group, referring to regions located in a university city where there is a medical school or on the outskirts of such cities. Four organizations were categorized in a “mixed regions” group, where a lack of resources applies to only one part of the territory or where only a part of the territory is considered rural or remote. Seven organizations were categorized in a “remote or isolated regions” group, including organizations not falling into any of the above categories and for which the entire territory is considered remote from urban centers or even isolated. A list of organizations grouped by region is presented in Multimedia Appendix 2.

Similarities and differences were then noted between different types of regions (university/peripheral, mixed, and remote/isolated).

Categorization by Type of Participants

As interest, perceived utility, and needs in terms of dashboards could vary among different types of participants (managers, leading clinicians, or other participants), and their degree of participation also varied greatly between group interviews, a specific analysis for each type of participant was conducted.

Similarities and differences were then noted between different types of professionals (managers, leading clinicians, and other participants).

Results

Participants

Overall, 23 group interviews were conducted. One organization from a university/peripheral region could not participate due to a significant change in the local governance of the QAP, and two organizations from remote/isolated regions participated in the same group interview. Eighty-two individuals participated in the group interviews: 44 were managers, 32 were leading clinicians, and 7 were other participants (mainly research coordinators).

Similarities Across Regions

Interest

Dashboards received positive reactions from participants in all regions. Some indicators sparked more discussion than others. These discussions focused on whether the results were surprising or not and on participants providing justifications for these results. Participants from all regions were particularly interested in the following indicators: alternative care level, prevalence, emergency room visits, and the type of physician most regularly visited. Participants from all regions were also highly interested in qualitative results, but the varied nature of these results does
not allow for a more specific analysis of these discussions (Table 1).

### Table 1. Codes for the main discussion points related to interest categorized by regions and by type of professionals.

<table>
<thead>
<tr>
<th>Discussion points</th>
<th>University/peripheral organizations (n=13)</th>
<th>Mixed organizations (n=4)</th>
<th>Remote/isolated organizations (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive reactions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants found the results not surprising</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Participants found the results interesting</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Participants did not ask for a copy of the dashboard</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Participant appreciated the format of the dashboard</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

| Results that were discussed |                       |                           |                                     |
|------------------------------|-----------------------|---------------------------|                                     |
| Only qualitative findings were available | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| Prevalence was not available | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Alternate level of care (delayed discharge) indicator | 0 | 4 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| Prevalence indicator | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 0 | 0 |
| Emergency indicators | 0 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Most regularly seen physicians indicator | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Qualitative findings | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| References (emergency or hospital) | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Hospitalization indicators | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

\[a\] The numbers in each cell represent the number of times the codes were observed during the discussions.

\[b\] UI: unidentified.

\[c\] M: managers.

\[d\] CL: leading clinicians.

\[e\] O: other.

\[f\] Despite the fact that no indicators of references were presented, the discussions revolved around references to emergency or hospitalizations.

### Utility

Participants from all regions offered several reflections on the utility of the dashboards. Primarily, participants believed that their tailored dashboard would be useful for identifying successes and challenges specific to their territory. Moreover, the participants thought the dashboards could mobilize different stakeholders, including the top management of regional health organizations or clinicians from interdisciplinary primary care clinics in their region. In all cases, the reception of dashboards suggested a desire for broader processes of critical thinking, self-examination, and learning on the part of participants (Table 2).
Table 2. Codes for the main discussion points related to utility categorized by regions and by type of professionals.\(^a\)

<table>
<thead>
<tr>
<th>Discussion points</th>
<th>University/peripheral organizations (n=13)</th>
<th>Mixed organizations (n=4)</th>
<th>Remote/isolated organizations (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UI(^b)  M(^c)  LC(^d)  O(^e)</td>
<td>UI  M  LC  O</td>
<td>UI  M  LC  O</td>
</tr>
<tr>
<td>Planning or implementation support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dashboards could be useful for the local Quebec Alzheimer Plan committees</td>
<td>1  2  2  0</td>
<td>0  2  0  0</td>
<td>0  0  0  0</td>
</tr>
<tr>
<td>Dashboards could be useful for clinician training</td>
<td>1  1  2  0</td>
<td>0  2  0  0</td>
<td>0  0  0  0</td>
</tr>
<tr>
<td>Dashboards are useful to identify challenges</td>
<td>1  1  0  0</td>
<td>0  1  0  0</td>
<td>0  1  0  0</td>
</tr>
<tr>
<td>Dashboards could be useful to follow patients</td>
<td>0  0  0  0</td>
<td>0  1  0  0</td>
<td>0  1  0  0</td>
</tr>
<tr>
<td>Participants did not react when prompted regarding the dashboard usefulness for the local Quebec Alzheimer Plan committees or training</td>
<td>1  0  0  0</td>
<td>0  0  0  0</td>
<td>0  0  0  0</td>
</tr>
<tr>
<td>Mobilization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobilize directions</td>
<td>1  3  1  0</td>
<td>0  2  0  0</td>
<td>1  2  0  0</td>
</tr>
<tr>
<td>Mobilize family physicians</td>
<td>0  0  1  0</td>
<td>0  1  0  0</td>
<td>0  2  0  0</td>
</tr>
<tr>
<td>Mobilize primary care clinicians (practicing in interdisciplinary groups)</td>
<td>0  1  1  0</td>
<td>0  0  0  0</td>
<td>1  0  0  0</td>
</tr>
<tr>
<td>Mobilize family physicians (practicing in solo practices or who are uncooperative)</td>
<td>0  0  0  0</td>
<td>0  0  0  0</td>
<td>1  0  0  0</td>
</tr>
<tr>
<td>Mobilize champions (physicians)</td>
<td>0  1  0  0</td>
<td>0  0  0  0</td>
<td>0  0  0  0</td>
</tr>
<tr>
<td>Mobilize key actors (no precision on who these key actors are)</td>
<td>0  1  1  0</td>
<td>0  1  0  0</td>
<td>0  0  0  0</td>
</tr>
<tr>
<td>Other uses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To modify the perceived role of nurses</td>
<td>0  0  1  0</td>
<td>0  0  0  0</td>
<td>0  0  0  0</td>
</tr>
<tr>
<td>To know what is done elsewhere</td>
<td>0  1  0  0</td>
<td>0  0  0  0</td>
<td>1  0  0  0</td>
</tr>
<tr>
<td>Research purposes</td>
<td>0  0  0  1</td>
<td>0  0  0  0</td>
<td>0  0  0  0</td>
</tr>
</tbody>
</table>

\(^a\)The numbers in each cell represent the number of times the codes were observed during the discussions.
\(^b\)UI: unidentified.
\(^c\)M: managers.
\(^d\)CL: leading clinicians.
\(^e\)O: other.

**Specific Needs**

**Comparisons**

Participants from all regions made proposals or asked clarifying questions suggesting specific needs. The participants appreciated the element of comparisons presented, but many suggested for the dashboard to also present the temporal trend of indicators. Other types of comparisons were also proposed depending on the organizations, without unanimity among regions. For example, participants from a mixed organization would have liked to have a comparison to the metropolis of the province (Montreal), while participants from another mixed organization would prefer to have a comparison to a territory that is geographically or demographically similar or, conversely, to be compared to a territory that is completely different (Table 3).
### Table 3. Codes for the main discussion points related to needs and questions, categorized by regions and by type of professionals.\(^a\)

<table>
<thead>
<tr>
<th>Discussion points</th>
<th>University/peripheral organizations (n=13)</th>
<th>Mixed organizations (n=4)</th>
<th>Remote/isolated organizations (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UI(^b) M(^c) LC(^d) O(^e)</td>
<td>UI M LC O</td>
<td>UI M LC O</td>
</tr>
<tr>
<td><strong>Comparisons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparisons over time (evolution)</td>
<td>0 1 0 0 1 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Comparisons to the province are useful</td>
<td>0 0 1 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Comparisons with another territory would be better</td>
<td>0 2 1 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Comparisons with a specific territory are not important; what matters is to go beyond territorial characteristics</td>
<td>0 1 0 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Comparison with the province is over too large a scale (no alternative offered)</td>
<td>0 1 0 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Comparisons with a similar territory in terms of aging population would be better</td>
<td>0 1 0 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Comparisons to a completely different territory would be interesting</td>
<td>0 0 0 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Stratifications: stratify by smaller regions within the territory</td>
<td>0 1 0 0 0 0 1</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td><strong>Identify specific patients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifying patients registered to an interdisciplinary primary care clinic versus those in other practices would be important</td>
<td>1 1 2 0 0 2 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Identifying orphan patients would be important</td>
<td>0 3 3 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Identifying patients who are known to home care services would be important</td>
<td>0 3 1 0 0 1 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Identifying patients with undiagnosed mental health issues would be important</td>
<td>0 1 0 0 0 0 0</td>
<td>0 1 0 0</td>
<td>0 1 0 0</td>
</tr>
</tbody>
</table>

\(^a\)The numbers in each cell represent the number of times the codes were observed during the discussions.
\(^b\)UI: unidentified.
\(^c\)M: managers.
\(^d\)CL: leading clinicians.
\(^e\)O: other.

**Targeted Populations and Other Indicators**

Participants from all regions expressed the need to produce results concerning patients not registered in interdisciplinary primary care clinics or without a family doctor (orphan patient), and to produce results on services received by individuals with dementia from health care professionals other than fee-for-service physicians (Table 3). Proposals particularly concerned visits with nurses, social workers, pharmacists, and occupational therapists, without a real consensus on the type of professional (Table 4).
<table>
<thead>
<tr>
<th>Discussion points</th>
<th>University/peripheral organizations (n=13)</th>
<th>Mixed organizations (n=4)</th>
<th>Remote/isolated organizations (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UIb</td>
<td>Mc</td>
<td>LCd</td>
</tr>
<tr>
<td>Other health professional indicators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicators from other professionals would be important; no precision on the type of professionals</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Indicators from social workers would be important</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Indicators from occupational therapists would be important</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Indicators from nurses would be important</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Indicators from pharmacists would be important</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Other indicators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care follow-up indicators (eg, evaluation, support to the care partner)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Reasons for the consultation or the hospitalization</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Home care service use</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Trajectory indicators</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Referrals for behavioral and psychological symptoms of dementia</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Number of beds</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Number of physicians trained for dementia care</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Number of human resources (before/after the pandemic)</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Antidementia medications</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Satisfaction of patients</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Timeliness of the diagnosis</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other propositions or questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on how to diagnose, what is a diagnosis allowing</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Diagnostic tools and training specific for their population (Indigenous)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10-year projections of prevalence</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dashboards presented monthly</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>More information on how patients were identified</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Divergences Between Regions

Interest
Participants from university/peripheral regions were particularly interested in results on hospitalizations and continuity of care (visits to family physicians). These same results were less emphasized by participants from other regions (mixed or remote/isolated) (Table 1).

Utility
Participants from university/peripheral regions were the first to propose that these dashboards could serve as a planning tool for their local QAP steering committees as well as for training clinicians. Following this observation, participants in subsequent focus groups were all invited to directly address this point. Not all responded enthusiastically to this opportunity, and responses varied among participants within the same region, whether peripheral/university, mixed, or remote/isolated. This indicates that perceived utility varies among participants, even among those from university/peripheral regions. Participants from university/peripheral regions also mentioned that these dashboards could be useful for research purposes. This difference likely reflects the level of engagement of regional teams with the QAP (Table 2).

Specific Needs
Stratifications
Participants from remote/isolated or mixed organizations proposed stratifying data based on smaller territories. They suggested that presenting data for their vast territory did not allow for detecting variations they perceived between different subregions (Table 3).

Targeted Populations and Other Indicators
Participants from university/peripheral and mixed regions made several suggestions targeting vulnerable patients or other diverse indicators. For example, they proposed stratifying dashboards or having specific data for patients known in home support or support for elderly autonomy, patients of solo-practicing physicians, and patients with mental health disorders not identified before being diagnosed with dementia (Table 3). Several specific indicators were proposed, but these proposals varied widely between different organizations, whether from university/peripheral or mixed regions. These propositions included: reasons or complexity of hospitalization or consultation, follow-up in interdisciplinary primary care, trajectory indicators, customer satisfaction, home care follow-up, number of clinicians trained under the QAP, timely diagnosis, and projections for the next 10 years (Table 4).

Participants from remote/isolated organizations expressed a need for more information and culturally adapted training for their region. In particular, they expressed the need for diagnostic tools and training adapted to their population, which includes many people from Indigenous communities.

Similarities and Differences According to Type of Professionals
Leading clinicians who participated in the focus groups mainly came from university/peripheral or mixed organizations. Additionally, university/peripheral organizations were the only ones in which other types of participants (research agents or coordinators) took part in the focus group.

In general, managers and leading clinicians shared the same interest in the prevalence indicator. However, leading clinicians expressed needs that managers and other participants did not express, particularly on follow-up indicators in Family Medicine Groups (interdisciplinary primary care teams), reasons for consultations/visits in primary care or emergencies, and patient satisfaction.

Some managers and research agents or coordinators would have appreciated more information on how individuals with dementia were identified for quantitative indicators. Finally, only one leading clinician and the research agents or coordinators suggested using these dashboards for research purposes.

Discussion

Principal Findings
This study determined that dashboards are of interest to those responsible for implementing an Alzheimer Plan in regional health organizations, and their format and content were appreciated regardless of the region or profession of the participants [51]. The study also highlighted specific needs regarding these dashboards that transcend all regions. Participants from all regional health organizations expressed a need for data on orphan patients or those not registered in interdisciplinary primary care clinics, as well as data on indicators related to services offered by other health care and social professionals (ie, nurses, social workers, pharmacists, and occupational therapists). These needs seem to express a desire to better delineate the challenges posed by orphan patients.
in the organization of health care across the province and to better monitor how interprofessional collaboration takes shape in interdisciplinary primary care clinics (Family Medicine Groups in Quebec).

We also identified interests and specific needs among participants from different regions and different types of professionals. The interest in hospitalization and continuity of care expressed by participants from university/peripheral regions (more urban) can be explained by the higher population density and a diversity of health resources in these regions [31]. Additionally, urban regions often have a greater diversity of health care and social professionals, specialists, and medical technologies available to meet the population’s needs [52]. In contrast, in remote or isolated regions, resources and the number of professionals may be more limited [36,38]. This higher resource availability in urban settings could explain higher interest in continuity of care in urban settings.

Another significant specific need that varied between regions is the need for diagnostic tools and adapted training expressed by participants from remote and isolated regions. The geographical, demographic, and socioeconomic characteristics of these regions make them unique and require different approaches to meet their health needs [1]. Health care professionals in these regions often face different challenges in providing health care for people with dementia [35,36]. This specific need may arise from the observation that both in Quebec and Canada, there is a higher proportion of Indigenous communities in rural areas [53,54], requiring cultural adaptation of care delivery and supporting tools. This is especially important considering the data availability in these regions [55].

Two regions only received qualitative dashboards. Quantitative surveillance data were not available for these organizations either due to their small populations posing risks to data dissemination or due to the fee-for-service mode of physician payment, on which surveillance data rely, but which is less frequent in these regions [56]. The inclusion of qualitative data has proven to be a significant asset in addressing these limitations.

Finally, organizations that span a larger territory (often from remote/isolated or mixed regions) expressed the need for a more granular analysis [42]. Several participants wanted more precise data for smaller territories. However, an ethical constraint prevented us from producing dashboards for smaller territories. Producing a single dashboard for all 24 regional health organizations remains a challenge to explore in future work.

The collaborative approach among researchers, scientific advisors from the National Institute of Public Health of Quebec, QAP project managers, and the QAP ministerial advisor is a major strength of the study [42]. This approach ensured a good understanding of the QAP implementation and formed an alliance between different stakeholders. The research and scientific advisor team could identify the most relevant data, while project managers—as the true points of contact with leading clinicians and managers of different regional health organizations—ensured that the messages were meaningful and well understood by all. The ministerial advisor and the four project managers responsible of the QAP also ensured that the collected and presented data allowed all stakeholders to align with the QAP orientations.

**Limitations**

The study also has limitations, particularly in terms of participant acquiescence biases during focus groups. Although the meetings were organized for guidance purposes, managers and leading clinicians from regional health organizations may have felt the need to demonstrate their progress to the ministerial advisor and project managers, as well as to researchers, due to their apparent proximity to the decision-makers and in front of their colleagues. To counter this acquiescence bias, we reiterated the independence of the research team from decision-makers at the beginning of each meeting. Another limitation is the categorization of different regions. Such categorizations are often arbitrary and inconsistent and do not consider the diversity of each region, especially for larger territories, which are often more remote [57]. However, this categorization aligns with that of the Ministry of Health and Social Services and reflects the organization of resources across the different regions of Quebec [50]. Finally, the relatively short time for conducting focus groups (30 minutes) could be considered a limitation. However, in addition to achieving data saturation during focus groups, we conducted these focus groups across all health organizations involved in the QAP and involved several stakeholders (at least one manager was always present, leading clinicians participated in several focus groups, and even research professionals participated in a few), ensuring good representativity of the results.

**Future Work and Recommendations**

Considering that the use of dashboards for an Alzheimer Plan is not documented in the scientific literature, future research should focus on the adoption of this tool in different regional health organizations in Quebec and at the governmental level. Studies on the use of dashboards exist for the evaluation of major neurocognitive disorders and the improvement of care offered to these individuals [20-22], but no study specifically mentions the use of a dashboard for an Alzheimer Plan. This will facilitate the transition from an experimental dashboard to a regular tool for managing or monitoring the implementation of an Alzheimer Plan. The composition of focus group participants was different for each region, which may have influenced the results and conclusions. The presence of more leading clinicians in university/peripheral regions may have led to discussions more focused on clinical aspects, while the predominance of managers in mixed or remote/isolated regions may have led to discussions more focused on logistical and organizational challenges related to service delivery in these regions. Other demographic characteristics of participants, such as gender and sex, could have influenced discussions and analyses. It would be interesting for future studies to analyze these differences. Furthermore, no patient or caregiver was part of the focus group, a major element of a learning health system. With increasing incentives to include citizens in health innovations, it will be a great opportunity to study the impact of their perspectives on developing and using dashboards.
Conclusion
In conclusion, dashboards are part of a learning health system and are a very useful tool for reporting on the challenges and issues related to the implementation of an Alzheimer Plan. However, it is important to consider the differences in the utility and information needs of various regions and types of professionals when developing dashboards to enable an adapted, efficient, and equitable implementation of an Alzheimer Plan that extends to a diverse set of organizations with varied resources. Taking these differences into account in the development of dashboards supporting the implementation of an Alzheimer Plan allows for better meeting the needs of all individuals with major neurocognitive disorders and providing optimal and equitable care, regardless of their region of residence.

Acknowledgments
We would like to thank Laura Rojas-Rozo, Mary Henein, Juliette Champoux-Pellegrin, and Sylvie Muller for formatting the dashboards; Louis Rochette from the Institut national de santé publique du Québec for extracting quantitative data and enthusiastically responding to related questions; and the managers, clinicians, and other participants who provided many insights to better support them in the implementation of the Quebec Alzheimer Plan. This work was supported by the Canadian Consortium on Neurodegeneration in Aging (CNA-163902). The Consortium is funded by a grant from the Canadian Institutes of Health Research (CIHR) and by funding from several partners.

Data Availability
All data generated or analyzed during this study are included in this published article and its appendix files. The comprehensive data set, including anonymized participant information and research materials, is available upon reasonable request. Researchers seeking access to the data for academic, noncommercial purposes can contact the corresponding author to facilitate the sharing of relevant materials. The aim is to promote transparency, reproducibility, and collaboration within the scientific community while respecting the privacy and confidentiality of the participants.

Authors’ Contributions
GAL and IV designed the study. All authors participated in the development of regional dashboards. GAL, MG, YC, IV, EM, CF, JD, and CM contributed to data collection and interpretation of results. GAL and ALC drafted the initial manuscript. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Example of an anonymized dashboard translated in English from French.
[PDF File (Adobe PDF File), 257 KB - formative_v8i1e55064_app1.pdf ]

Multimedia Appendix 2
Categorization of the regional health organizations.
[DOCX File , 21 KB - formative_v8i1e55064_app2.docx ]

References
5. Le système de santé apprenant ACTE. Alliance pour des communautés en santé. URL: [https://www.alliancecan.org/fr/Le-syst%C3%A8me-de-sante%CC%A8me-de-sant%C3%A9-apprenant-ACTE](https://www.alliancecan.org/fr/Le-syst%C3%A8me-de-sante%CC%A8me-de-sant%C3%A9-apprenant-ACTE) [accessed 2024-04-26]


44. Concannon D, Herbst K, Manley E. Developing a data dashboard framework for population health surveillance: widening access to clinical trial findings. JMIR Form Res 2019 Apr 04;3(2):e11342 [FREE Full text] [doi: 10.2196/11342] [Medline: 36100380]
50. Les régions SARROS. SARROS. URL: https://www.equipesarros.ca/regions/ [accessed 2020-02-25]

https://formative.jmir.org/2024/1/e55064


Abbreviations

QAP: Quebec Alzheimer Plan (Plan ministériel sur les troubles neurocognitifs majeurs)

©Genevieve Arsenault-Lapierre, Alexandra Lemay-Compagnat, Maxime Guillette, Yves Couturier, Victoria Massamba, Isabelle Dufour, Eric Maubert, Christine Fournier, Julie Denis, Caroline Morin, Isabelle Vedel. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction: Retrospective Cohort Study

Pin Zhang1,2,*, MM; Lei Wu1,*, MM; Ting-Ting Zou1, MM; ZiXuan Zou1, MM; JiaXin Tu1, MM; Ren Gong3, PhD; Jie Kuang1,*, PhD

1Jiangxi Provincial Key Laboratory of Preventive Medicine, School of Public Health, Nanchang University, Nanchang, China
2School of Public Health and Management, Nanchang Medical College, Nanchang, China
3Department of Cardiology, The Second Affiliated Hospital of Nanchang University, Nanchang, China
*these authors contributed equally

Corresponding Author:
Jie Kuang, PhD
Jiangxi Provincial Key Laboratory of Preventive Medicine
School of Public Health
Nanchang University
461 BaYi Road
Nanchang, 330006
China
Phone: 86 0791 8636228
Email: kuangjie@ncu.edu.cn

Abstract

Background: The incidence of major adverse cardiovascular events (MACEs) remains high in patients with acute myocardial infarction (AMI) who undergo percutaneous coronary intervention (PCI), and early prediction models to guide their clinical management are lacking.

Objective: This study aimed to develop machine learning–based early prediction models for MACEs in patients with newly diagnosed AMI who underwent PCI.

Methods: A total of 1531 patients with AMI who underwent PCI from January 2018 to December 2019 were enrolled in this consecutive cohort. The data comprised demographic characteristics, clinical investigations, laboratory tests, and disease-related events. Four machine learning models—artificial neural network (ANN), k-nearest neighbors, support vector machine, and random forest—were developed and compared with the logistic regression model. Our primary outcome was the model performance that predicted the MACEs, which was determined by accuracy, area under the receiver operating characteristic curve, and F1-score.

Results: In total, 1362 patients were successfully followed up. With a median follow-up of 25.9 months, the incidence of MACEs was 18.5% (252/1362). The area under the receiver operating characteristic curve of the ANN, random forest, k-nearest neighbors, support vector machine, and logistic regression models were 80.49%, 72.67%, 79.80%, 77.20%, and 71.77%, respectively. The top 5 predictors in the ANN model were left ventricular ejection fraction, the number of implanted stents, age, diabetes, and the number of vessels with coronary artery disease.

Conclusions: The ANN model showed good MACE prediction after PCI for patients with AMI. The use of machine learning–based prediction models may improve patient management and outcomes in clinical practice.

(JMIR Form Res 2024;8:e48487) doi:10.2196/48487

KEYWORDS
acute myocardial infarction; percutaneous coronary intervention; machine learning; early prediction; cardiovascular event
**Introduction**

Acute myocardial infarction (AMI) is a common clinical acute and severe disease with rapid onset, rapid progression, and high mortality [1-3]. In 2017, there were approximately 695,000 new cases of AMI in the United States, and it is estimated that 325,000 people will have recurrent events [4]. There are approximately 500,000 new cases of AMI in China every year, and 2.5 million patients have a history of myocardial infarction [5]. As technology has advanced, percutaneous coronary intervention (PCI) has become the primary approach for treating AMI. Although PCI can significantly reduce the fatality rate of AMI, the rate of major adverse cardiovascular events (MACEs) among patients after PCI is still very high, which seriously affects the clinical outcomes of patients [6-10]. A study by Copeland-Halperin et al [11] showed that the incidence of MACEs in patients with AMI one year after PCI was 17.8% [11].

Identifying patients with AMI undergoing PCI who are at high risk of MACEs may help clinical decision-making incorporate timely measures to improve clinical outcomes. Some studies, such as Global Registry of Acute Coronary Event [12], Thrombolysis in Myocardial Infarction Risk [13,14], and Acute Catheterization and Urgent Intervention Triage StrategY-PCI [15], as well as studies that generated the Mayo Clinic PCI Risk and the China Acute Myocardial Infarction scoring systems, have explored the risks after PCI [16]. Despite these advances, individualized prediction of MACEs remains challenging with low specificity and positive predictive accuracy, and most of the methods rely on traditional parameter models, such as logistic regression, to screen for variables and build a series of risk-scoring models.

In recent years, machine learning methods that rely on a strong self-learning capability, such as random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), and artificial neural network (ANN) have become increasingly prevalent in prognostic prediction [1,13,17,18]. By calling various functions, these models can extract and integrate information from all kinds of complex data to make better predictions. A study of a consecutive cohort of patients with hypertrophic cardiomyopathy (HCM) presented a machine learning–based model to identify individual patients with HCM at high risk of developing advanced heart failure symptoms. The results showed that the 5-year risk prediction of progressive heart failure in patients with HCM can be estimated [19].

We found that machine learning models, such as RF, ANN, SVM, and KNN, perform well in clinical prognosis prediction research. Thus, this study sought to develop a machine learning–based model, integrating clinical, anatomical, and laboratory features, to predict MACEs in patients who have recently been diagnosed with AMI after their first PCI and improve overall patient outcomes by implementing earlier management.

**Methods**

**Study Design, Setting and Participant Selection**

This retrospective cohort study was conducted at the Department of Cardiovascular Medicine, the Second Affiliated Hospital of Nanchang University (a teaching tertiary hospital), in Jiangxi Province, China. We collected electronic medical records of patients with AMI who underwent PCI for the first time from January 2018 to December 2019. These patients were followed up through December 2021.

The inclusion criteria of the participants were as follows:

- The patient was ≥18 years of age.
- This was the patient’s first clinically diagnosed AMI (clinical evidence of AMI as evident from the detection of a rise or fall of cardiac troponin values and at least one of the following symptoms of myocardial ischemia: symptoms of acute myocardial ischemia, new ischemic electrocardiogram (ECG) changes, and development of pathological Q waves).
- PCI was performed for the first time at this hospital.
- Among the left main artery, left circumflex branch, left anterior descending branch, and right coronary artery, at least one had stenosis ≥50%.
- Complete medical records and follow-up data were available.

The following exclusion criteria were applied:

- History of PCI and coronary artery bypass grafting treatment
- Complications from other heart diseases requiring surgical procedures, such as heart bypass
- Recent active bleeding
- An intracerebral mass or an aneurysm

We adopted the “Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research” to guide the reporting of our study [20].

**Data Collection, Definition of Outcomes, and Predictor Variables**

Data were collected from electronic health records, including demographic characteristics, clinical investigations, the first laboratory tests, and disease-related events. MACEs were defined as cardiomyopathies (excluding infectious, familial, alcohol, and drug-related cardiomyopathies), hypertensive heart disease, recurrent myocardial infarction, heart failure, sudden cardiac death, revascularization, malignant arrhythmia, and stent thrombosis [21]. Abnormal Q waves were identified by the clinician based on ECG results. Left ventricular ejection fraction (LVEF) was defined as normal (more than 50%), mildly abnormal (40% to 50%), moderately abnormal (30% to 40%), and severely abnormal (less than 30%) [22]. According to the number of diseased coronary vessels and implanted stents, they were classified as I, II, III, and IV.
Ethics Approval
This study was reviewed and approved by the Second Affiliated Hospital of Nanchang University Medical Ethics Committee (No. Review 2017 No. (098)).

Data Preprocessing for Machine Learning Model Development
All analyses were performed with R software (version 4.0.1; R Core Team). The patients were randomly assigned to training (n=953, 70%) and testing (n=409, 30%) data sets by calling the createDataPartition function using the random number method, and chi-square tests showed that there was no statistical difference between them ($\chi^2_1=2.169; P=.14$). We developed machine learning models using the training data set. We analyzed the missing and out-of-range values with imputation methods. We used multiple imputation with chained equations to assign any missing predictor values [23]. The imputation processes were performed separately in the training and testing sets after the data were split. To improve the accuracy of the machine learning models and increase the speed of finding the optimal solution by gradient descent, we standardized and normalized all input variables before the model was built. To alleviate the problem of imbalanced classification samples, we adopted the random oversampling method. We used the ROSE package in R to generate new balanced training data. After random oversampling, the number of patients with MACE in the training data sets changed from 186 to 471.

Predictor Selection for Model Development
The model was built using demographic information (age and sex), personal comorbidities (diabetes and peripheral arterial disease), preoperative PCI (LVEF, the number of diseased vessels, and abnormal Q waves), serological examination (beta 2 microglobulin, B-type brain natriuretic peptide, glucose, serum creatinine clearance, and estimated glomerular filtration rate), and the characteristics of PCI (the number of implanted stents; n=65; Table S1 in Multimedia Appendix 1). A total of 12 variables with significant differences in the univariate analysis were included in the model development (Table 1).
### Table 1. Baseline characteristics of the study patients (N=1362).

<table>
<thead>
<tr>
<th>Variables</th>
<th>MACE(^a) (n=252)</th>
<th>Non-MACE (n=1110)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65</td>
<td>101 (40.08)</td>
<td>543 (48.92)</td>
<td>.04</td>
</tr>
<tr>
<td>65</td>
<td>94 (37.30)</td>
<td>332 (29.91)</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>57 (22.62)</td>
<td>235 (21.17)</td>
<td></td>
</tr>
<tr>
<td><strong>Diabetes, n (%)</strong></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>75 (29.76)</td>
<td>261 (23.51)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>177 (70.24)</td>
<td>849 (76.49)</td>
<td></td>
</tr>
<tr>
<td><strong>Vascular disease, n (%)</strong></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>111 (44.05)</td>
<td>569 (51.26)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>141 (55.95)</td>
<td>541 (48.74)</td>
<td></td>
</tr>
<tr>
<td><strong>Abnormal Q wave, n (%)</strong></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>125 (49.60)</td>
<td>480 (43.24)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>127 (50.40)</td>
<td>630 (56.76)</td>
<td></td>
</tr>
<tr>
<td><strong>LVEF(^b), n (%)</strong></td>
<td></td>
<td></td>
<td>.005</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>167 (66.27)</td>
<td>832 (74.95)</td>
<td></td>
</tr>
<tr>
<td>40%-50%</td>
<td>57 (22.62)</td>
<td>188 (16.94)</td>
<td></td>
</tr>
<tr>
<td>30%-40%</td>
<td>19 (7.54)</td>
<td>65 (5.86)</td>
<td></td>
</tr>
<tr>
<td>&lt;30%</td>
<td>9 (3.57)</td>
<td>25 (2.25)</td>
<td></td>
</tr>
<tr>
<td><strong>Vessels with coronary artery disease, n (%)</strong></td>
<td>45 (17.86)</td>
<td>288 (25.95)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>I</td>
<td>75 (29.76)</td>
<td>370 (33.33)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>123 (48.81)</td>
<td>418 (37.66)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>9 (3.57)</td>
<td>34 (3.06)</td>
<td></td>
</tr>
<tr>
<td><strong>Implanted stent number, n (%)</strong></td>
<td>10 (3.97)</td>
<td>40 (3.60)</td>
<td>.004</td>
</tr>
<tr>
<td>No stent</td>
<td>106 (42.06)</td>
<td>594 (53.51)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>84 (33.33)</td>
<td>301 (27.12)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>37 (14.68)</td>
<td>114 (10.27)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>15 (5.95)</td>
<td>61 (5.50)</td>
<td></td>
</tr>
<tr>
<td><strong>Brain natriuretic peptide (pg/μL), mean (SD)</strong></td>
<td>684.36 (997.90)</td>
<td>518.27 (773.65)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Serum creatinine clearance (mL/min), mean (SD)</strong></td>
<td>65.19 (30.18)</td>
<td>71.87 (44.35)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>EGFR(^c) (ml/min), mean (SD)</strong></td>
<td>75.68 (28.92)</td>
<td>80.55 (31.82)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Beta 2 microglobulin (mg/L), mean (SD)</strong></td>
<td>3.23 (3.61)</td>
<td>2.72 (5.51)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Glucose (mmol/L), mean (SD)</strong></td>
<td>7.22 (3.32)</td>
<td>6.68 (3.00)</td>
<td>.02</td>
</tr>
</tbody>
</table>

\(^a\)MACE: major adverse cardiovascular events.

\(^b\)LVEF: left ventricular ejection fraction.

\(^c\)EGFR: estimated glomerular filtration rate.

---

**Model Testing and Performance Evaluation**

Based on a previous application of the model [24], the parameter range of the model was preset, and the GridSearchCV function was used to select the optimal parameters of each machine learning model.

To minimize potential overfitting in the above machine learning models, we called the trainControl function in the caret package of R language for 7-fold cross-validation during the development process. The model performance was assessed for accuracy, recall, precision, area under the receiver operating characteristic curve (AUC), and \(F_1\)-score in the testing data set. We identified
the important predictors through importance analysis of the variables. Logistic regression analysis was used to compare the absolute value of the coefficients of variables; RF was used to measure the importance of features by calculating information gain through entropy; and the ANN method was used to calculate the relative importance of variables based on the generalized weight method.

**Statistical Analysis**

The following R packages for machine learning approaches were used: caret, randomForest, and neuralnet. Baseline characteristics were compared with the Wilcoxon rank sum test for continuous variables and the chi-square test for categorical variables. We considered $P<.05$ (2-sided) to be statistically significant.

**Results**

A total of 1531 patients were screened; 140 patients who did not undergo PCI for the first time were excluded; 19 patients were lost to follow-up; and 1362 patients who were successfully followed up were included in this analysis (Figure 1). The mean follow-up time was 28.0 (SD 11.0) months (median 29.9 months). A total of 252 MACEs were observed, including 128 cases of recurrent myocardial ischemia and 117 cases of myocardial infarction and reinfarction. The positive rates of MACEs were 4.63%, 11.38%, 14.54%, and 18.50% at 30 days, 6 months, 1 year, and 3 years after PCI, respectively. MACEs occurred in 203 (18.7%) male patients and 49 (17.8%) female patients. As shown in Figure 2, the survival rate of the sample population decreased rapidly in the first 3 months after PCI, especially 30 days after PCI, and there was no difference in the log-rank test of the survival curve between male and female patients.

Table 1 shows the baseline characteristics of the MACE group and the non-MACE group. Age, diabetes, peripheral and cerebrovascular history, LVEF, abnormal Q wave, the number of vessels with coronary artery disease, the number of implanted stents, brain natriuretic peptide, serum creatinine, estimated glomerular filtration rate, beta 2 microglobulin, and glucose were significantly different between the 2 groups ($P<.001$). The nonsignificant differences in variables between the 2 groups are shown in Table S1-S6 in Multimedia Appendix 1.

Table 2 shows the performance of the 3 models with 7-fold cross-validation. ANN, KNN, SVM, RF, and logistic regression exhibited the best to worst performance in terms of their AUC, accuracy, recall, and $F_1$-score. However, KNN performed best in terms of precision. The average accuracy, recall, precision, AUC, and $F_1$-score of the ANN model were 80.52%, 81.33%, 69.94%, 83.68%, and 79.47%, respectively.

In the testing data set, the ANN model showed a higher AUC than RF and logistic regression. Figure 3 shows that the AUCs of the ANN, RF, KNN, SVM, and logistic regression models were 0.805, 0.798, 0.772, 0.727, and 0.718, respectively; the average accuracy for the above 3 models was 0.821, 0.741, and 0.729, respectively, and the average $F_1$-scores were 0.804, 0.722, and 0.709, respectively.

The 10 most important predictors in the ANN model are shown in Table 3. These were LVEF (0.27), the number of implanted stents (0.14), age (0.13), diabetes (0.10), the number of vessels with coronary artery disease (0.09), vascular disease (0.08), brain natriuretic peptide (0.05), glucose (0.05), beta 2 microglobulin (0.04), and abnormal Q wave (0.02).

**Figure 1.** Flowchart for patient enrollment. AMI: acute myocardial infarction; MACE: major adverse cardiovascular event.
Figure 2. Prognostic survival curve of patients with acute myocardial infarction undergoing percutaneous coronary intervention.

Table 2. Comparison of models for predicting major adverse cardiovascular events based on 7-fold cross-validation.

<table>
<thead>
<tr>
<th>Models</th>
<th>Accuracy, mean (SD)</th>
<th>Recall, mean (SD)</th>
<th>Precision, mean (SD)</th>
<th>AUC(^a), mean (SD)</th>
<th>(F_1)-score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>72.37 (2.05)</td>
<td>67.33 (8.42)</td>
<td>59.62 (8.34)</td>
<td>73.52 (2.37)</td>
<td>71.11 (6.01)</td>
</tr>
<tr>
<td>K-nearest neighbors</td>
<td>81.44 (2.22)</td>
<td>80.23 (1.56)</td>
<td>70.22 (7.23)</td>
<td>81.87 (3.32)</td>
<td>77.95 (5.70)</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>74.91 (3.03)</td>
<td>80.03 (1.76)</td>
<td>65.94 (7.02)</td>
<td>78.68 (1.82)</td>
<td>76.41 (5.92)</td>
</tr>
<tr>
<td>Random forest</td>
<td>73.44 (1.58)</td>
<td>71.23 (1.56)</td>
<td>61.22 (7.23)</td>
<td>74.87 (2.12)</td>
<td>71.92 (6.30)</td>
</tr>
<tr>
<td>Artificial neural network</td>
<td>80.52 (1.13)</td>
<td>81.33 (0.56)</td>
<td>69.94 (7.02)</td>
<td>83.68 (1.82)</td>
<td>79.47 (4.57)</td>
</tr>
</tbody>
</table>

\(^a\)AUC: area under the receiver operating characteristic curve.

Figure 3. The area under the receiver operating characteristic (ROC) curve of artificial neural network (ANN), random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression models.
Discussion

Principal Findings

In this study, we developed a machine learning–based model integrating clinical, anatomical, and laboratory test features to predict MACEs in patients with newly diagnosed AMI after their first PCI. The major findings suggest that the ANN model had higher predictive accuracy (accuracy of 87.99%, AUC of 0.81, and F1-score of 0.71), compared to RF, KNN, SVM, and logistic regression.

Among the patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%).

Our results indicated that the 3-year prognostic risk among patients with AMI undergoing their first PCI was mainly related to age, ECG characteristics, ventricular ejection ability, coronary artery lesions, stent implantation after PCI, and some serological variables. Yang et al [32] found that the risk ratio of hospital deaths after PCI was 3.723 (95% CI 2.86-4.84) for South Korean patients aged >65 years relative to those aged ≤65 years. A Korean multicenter AMI National Institutes of Health–registered project found that the MACE rate, 3 years after PCI, among patients with AMI with an LVEF <40% was 3.34 times that of the control group [33]. Fam et al [34] conducted a retrospective study on patients with clinical AMI in Asian multiethnic groups and found that the risk of MACEs among patients with diabetes, 2 years after PCI, was 1.84 times higher than that among patients without diabetes [34]. Diabetes is a chronic metabolic disease, and long-term diabetes is often accompanied by bleeding disorders, vascular endothelial dysfunction, small artery lesions, high blood sugar [35], hemostatic disorders [36], endothelial dysfunction, and a series of other changes [37]. These characteristics will accelerate the process of atherosclerotic disease deterioration. The number of coronary artery lesions

Table 3. Importance of each variable in the artificial neural network model.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left ventricular ejection fraction</td>
<td>0.27</td>
</tr>
<tr>
<td>The number of implanted stents</td>
<td>0.14</td>
</tr>
<tr>
<td>Age</td>
<td>0.13</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.10</td>
</tr>
<tr>
<td>The number of vessels with coronary artery disease</td>
<td>0.09</td>
</tr>
<tr>
<td>Vascular disease</td>
<td>0.08</td>
</tr>
<tr>
<td>Brain natriuretic peptide</td>
<td>0.05</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.05</td>
</tr>
<tr>
<td>Beta 2 microglobulin</td>
<td>0.04</td>
</tr>
<tr>
<td>Abnormal Q wave</td>
<td>0.02</td>
</tr>
</tbody>
</table>
and the number of stents implanted in a patient are also positively correlated with the risk of postoperative MACEs to a certain extent. This may be because a higher number of vessels with coronary artery disease and the number of implanted stents tend to indicate a more serious condition, leading to a worse prognosis for the patients. Hongbo et al [38] found that the probability of a poor prognosis in patients with multiple coronary artery lesions was 20.0%, compared with 6.98% in patients with single coronary artery lesions [38].

The results of the machine learning model showed that predictors like LVEF, number of implanted stents, and age were more important to the model. LVEF is a common variable that reflects left ventricular function, and patients with a low LVEF have a significantly higher MACE rate [39]. An increase in age can lead to the aggravation of atherosclerosis [40]. The number of implanted stents may be related to the severity of the disease and the extent of the infarction [41]. This reminds us that we should pay special attention to the prognosis of patients with AMI who have a low LVEF value, older age, and more implanted stents in clinical practice.

Study Limitations
This study has some limitations. First, there may have been an issue of survival bias in the study, as patients with missing follow-up data were excluded. Second, the data have missing values. We have filled missing values with multiple imputation; however, imputation with these techniques could synthetically reduce the variance in these variables and may have affected the accuracy of the constructed model. Finally, although the models were internally validated with data from the same hospital, further work should include validation with external data from other hospitals or centers.

Conclusions
This study revealed that the ANN model showed good MACE prediction performance for patients with AMI after PCI, and it identified the most important predictors, which may aid in clinical decision-making and improve outcomes. This model needs to be externally validated in larger populations and multicenter settings.

Acknowledgments
The authors are grateful to the following individuals for contributing data: Li Li, Nan Wang, LiFang Deng, ShanLan Yang, and Bin Liu. This study was supported by the National Natural Science Foundation of China (grants 81960611, 82160645, 82360667, and 81260441), Natural Science Foundation of Jiangxi Province (grants 20192BAA208005 and 20212BAB206091), and the National Undergraduate Training Program for Innovation and Entrepreneurship (grant 202010403019).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Additional statistics.
[DOCX File, 53 KB - formative_v81e48487_app1.docx]

References


**Abbreviations**

AMI: acute myocardial infarction

ANN: artificial neural network

AUC: area under the receiver operating characteristic curve

ECG: electrocardiogram

HCM: hypertrophic cardiomyopathy

HORIZONS-AMI: Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction

KNN: k-nearest neighbors

LVEF: left ventricular ejection fraction

MACE: major adverse cardiovascular event

PCI: percutaneous coronary intervention

RF: random forest

SVM: support vector machine
Please cite as:
Zhang P, Wu L, Zou TT, Zou Z, Tu J, Gong R, Kuang J
Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction: Retrospective Cohort Study
JMIR Form Res 2024;8:e48487
URL: https://formative.jmir.org/2024/1/e48487
doi:10.2196/48487
PMID:38170581

©Pin Zhang, Lei Wu, Ting-Ting Zou, ZiXuan Zou, JiaXin Tu, Ren Gong, Jie Kuang. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Differences in Psychological Inflexibility Among Men With Erectile Dysfunction Younger and Older Than 40 Years: Web-Based Cross-Sectional Study

Junichi Saito¹, PhD; Hiroaki Kumano², MD, PhD; Mohammad Ghazizadeh³, MD, PhD; Chigusa Shimokawa³, MA; Hideki Tanemura³, MBA

¹Comprehensive Research Organization, Waseda University, Saitama, Japan
²Faculty of Human Sciences, Waseda University, Saitama, Japan
³Logos Science Corp, Tokyo, Japan

Corresponding Author:
Junichi Saito, PhD
Comprehensive Research Organization
Waseda University
2-579-15, Mikajima, Tokorozawa
Saitama, 359-1192
Japan
Phone: 81 429498113
Email: tekuteke@aoni.waseda.jp

Abstract

Background: Psychological inflexibility is a core concept of acceptance and commitment therapy (ACT), which is a comprehensive, transdiagnostic interpretation of mental health symptoms. Erectile dysfunction (ED) is a condition that affects male sexual performance, involving the inability to achieve and maintain a penile erection sufficient for satisfactory sexual activity. Psychosocial factors primarily influence ED in men younger than 40 years, whereas biological factors are more likely to be the underlying cause in older men.

Objective: This web-based cross-sectional study examined differences in depression, anxiety, and psychological inflexibility among men with ED younger and older than 40 years in a Japanese population.

Methods: We used a web-based survey to gather data from various community samples. ED was assessed by the International Index of Erectile Function - 5 (IIEF-5) questionnaire, while depression, anxiety, and psychological inflexibility were evaluated by the Patient Health Questionnaire-9 (PHQ-9), General Anxiety Disorder-7 (GAD-7), Acceptance and Action Questionnaire-II (AAQ-II), Cognitive Fusion Questionnaire (CFQ), and Valuing Questionnaire–Obstacle Subscale (VQ-OB) questionnaires. The chi-square test estimated the scores of PHQ-9 and GAD-7 among men with ED, comparing those younger than 40 years and those older than 40 years. Additionally, a two-way ANOVA was conducted with ED severity and age group as independent variables, assessing psychological inflexibility.

Results: Valid responses from 643 individuals (mean age 36.19, SD 7.54 years) were obtained. Of these, 422 were younger than 40 years (mean age 31.76, SD 5.00 years), and 221 were older than 40 years (mean age 44.67, SD 2.88 years). There was a statistical difference in the prevalence of depression as judged by PHQ≥10 between men with ED younger and older than 40 years (P<.001). On the other hand, there was no difference in the prevalence of anxiety as judged by GAD≥10 (P=.12). The two-way ANOVA revealed that the interactions for CFQ (P=.04) and VQ-OB (P=.01) were significant. The simple main effect was that men with ED younger than 40 years had significantly higher CFQ (P=.01; d=0.62) and VQ-OB (P<.001; d=0.87) scores compared to those older than 40 years in moderate ED and severe ED. Additionally, it was found that men younger than 40 years with moderate to severe ED had significantly higher CFQ (P=.01; d=0.42) and VQ-OB (P=.02; d=0.38) scores compared to men younger than 40 years without ED. On the other hand, no interaction was found for AAQ-II (P=.16) scores.

Conclusions: To the best of our knowledge, this web-based cross-sectional study is the first to examine the relationship between psychological inflexibility and ED. We conclude that men with moderate and severe ED younger than 40 years have higher psychological inflexibility and might be eligible for ACT.

(JMIR Form Res 2024;8:e45998) doi:10.2196/45998

https://formative.jmir.org/2024/1/e45998
erectile dysfunction; acceptance and commitment therapy; psychological inflexibility; depression; anxiety; men; cross-sectional study; psychological; utility; psychosocial; men; therapy; impotence; erection

**Introduction**

The efficacy of acceptance and commitment therapy (ACT) has been evaluated in numerous randomized controlled studies exploring various targeted conditions [1]. There is supporting evidence for ACT across various physical illnesses (eg, chronic pain [2], diabetes [3], epilepsy [4], cancer [5], and irritable bowel syndrome [6]). Many of these studies use a transdiagnostic method to analyze psychological issues within individual health conditions. Psychological inflexibility is a core concept of ACT, which is a comprehensive, transdiagnostic interpretation of mental health symptoms [7]. Psychological inflexibility highlights two interrelated processes: cognitive fusion and experiential avoidance. Cognitive fusion represents the phenomenon by which individuals are influenced by the literal meaning of their thoughts instead of viewing them as transient internal states [8]. Experiential avoidance represents an attempt or desire to suppress unwanted internal experiences, such as emotions, thoughts, memories, and bodily sensations [9]. These processes are obstacles to one’s valued living activities, decreasing well-being [10].

Erectile dysfunction (ED) is a condition that affects male sexual performance, involving the inability to achieve and maintain a penile erection sufficient for satisfactory sexual activity [11]. Several reviews and clinical guidelines are available for ED. However, many of these approaches to assessing and treating ED purely from a medical perspective seldom address the psychosocial components of ED [12]. Pharmacological treatment alone does not respond to all the concurrent factors of ED, including anxiety, loss of self-confidence, depressed mood, difficulties in a couple’s communication, relationship disputes, or a partner’s sexual problems [13]. Recent systematic reviews have shown that combining phosphodiesterase-5 inhibitors with psychological treatment exhibits significant potential for treating ED [14].

There is a widespread assumption that psychosocial factors primarily influence ED in men younger than 40 years, whereas biological factors are more likely to be the underlying cause of ED in older men. Moore et al [15] showed different symptom patterns among patients with ED according to age groups. They reported that younger men had comparatively more significant depressive symptoms, with lower relationship satisfaction, more negative reactions from partners, and lower job satisfaction. Given these findings, it is possible that people younger than 40 years are more psychologically inflexible than those older than 40 years and that ACT is more effective for them. However, studies on ACT for ED remain limited, with only a few identified. Therefore, this cross-sectional study used assessments to evaluate depression, anxiety, and psychological inflexibility in men younger and older than 40 years. However, it is well known that ED prevalence varies across geographical groups [16,17]; therefore, it is essential to research ED etiology according to different racial, cultural, religious, and socioeconomic backgrounds. There might be many potential patients with ED in Japan, so we conducted a web-based survey for this study.

**Methods**

**Participants**

To gather data from a wide range of community samples, we used a web-based survey, conducted with the assistance of a marketing research service provider (Rakuten Insight, Inc) in Japan. Based on the International Index of Erectile Function – 5 (IIEF-5) cutoff point [18], participants of all severities were recruited to include a certain percentage of patients of all ages. All enrolled participants followed the following criteria: (1) male; (2) aged 20 to 50 years; and (3) married or living with a fixed sexual partner for more than 6 months. The exclusion criteria were as follows: (1) sexual dysfunction caused by Peyronie disease or other organic lesions of the external genitalia; (2) prostate cancer, hypertensive disease, cardiac disease, cerebrovascular disease, chronic kidney disease, and diabetes; and (3) a history of sertraline or other medicines that may influence erection and psychological symptoms.

Participants were first instructed that this survey would be administered anonymously, and their responses were not compulsory. Then, those participants who agreed to participate in this research responded to the surveys. Participants were given points to exchange for items within the survey company’s system as a reward.

**Ethical Considerations**

This study was approved by the Waseda University Academic Research Ethical Review Committee (2019-363). The study protocol followed the guidelines for epidemiological studies in accordance with the Declaration of Helsinki.

**Measurements**

**International Index of Erectile Function-5 (IIEF-5)**

The Japanese version of IIEF-5 is a 5-item self-report questionnaire designed to measure erectile function [18]. Items are rated on a 5-point Likert-type scale, ranging from 1 to 5. The total score can range from 5 to 25, with high scores meaning high erectile function. Based on the original validation studies, the total score can then be interpreted as suggesting “no ED” (22-25), “mild ED” (17-21), “mild-to-moderate ED” (12-16), “moderate ED” (8-11), and “severe ED” (5-7).

**Patient Health Questionnaire-9 (PHQ-9)**

The Japanese version of the Patient Health Questionnaire-9 (PHQ-9) is a 9-item self-report questionnaire designed to measure depression [19]. Items are rated on a 4-point Likert-type scale, ranging from 0 to 3. The total score can range from 0 to 27, with high scores meaning high depression. Based on the original validation studies, the total score can then be interpreted as suggesting no depression (0-4), mild depression (5-9), moderate depression (10-14), moderately severe depression (15-19), severe depression (20-27), or major depression (28-30).
or severe depression (20-27). A cutoff score of 10 is suggested as indicating a possible diagnosis of depressive disorder.

**Generalized Anxiety Disorder-7 (GAD-7)**

The Japanese version of the Generalized Anxiety Disorder-7 (GAD-7) questionnaire is a 7-item self-report questionnaire designed to measure generalized anxiety disorder [20]. Items are rated on a 4-point Likert-type scale, ranging from 0 to 3. The total score can range from 0 to 21, with high scores meaning high anxiety. Based on the original validation studies, the total score can then be interpreted as suggesting no anxiety (0-4), mild anxiety (5-9), moderate anxiety (10-14), or severe anxiety (14-21). A cutoff score of 10 is suggested as indicating a possible diagnosis of generalized anxiety disorder.

**Acceptance and Action Questionnaire-II (AAQ-II)**

The Japanese version of the Acceptance and Action Questionnaire-II (AAQ-II) is a 7-item self-report questionnaire designed to measure experiential avoidance [21]. Items are rated on a 7-point Likert-type scale, ranging from 1 to 7. The total score can range from 7 to 49, with high scores meaning high experiential avoidance.

**Cognitive Fusion Questionnaire (CFQ)**

The Japanese version of the Cognitive Fusion Questionnaire (CFQ) is a 7-item self-report questionnaire designed to measure cognitive fusion [22]. Items are rated on a 7-point Likert-type scale, ranging from 1 to 7. The total score can range from 7 to 49, with high scores meaning high cognitive fusion.

**Valuing Questionnaire–Obstacle Subscale (VQ-OB)**

The Japanese version of the Valuing Questionnaire–Obstacle Subscale (VQ-OB) is a 5-item self-report questionnaire designed to measure obstruction of valued living [23]. Items are rated on a 7-point Likert-type scale, ranging from 0 to 6. The total score can range from 0 to 30, with high scores meaning high obstruction of valued living.

**Statistical Analysis**

We used mean (SD) values to describe numerical data and counts and percentages to describe categorical data. The chi-square tests estimated categorical data, and numerical data were estimated by t tests. A two-way ANOVA test was used to assess the differences in men with ED aged younger and older than 40 years regarding psychological inflexibility and the interaction between them. Post hoc tests were conducted using the Holm method to control for type I errors. Cohen d index was calculated as effect sizes, serving as standardized indicators unaffected by sample sizes. All tests were 2-tailed, and a statistical difference was assumed when the P value was <.05. All statistical analyses were conducted through IBM SPSS Statistics (version 25.0; IBM Corp).

**Results**

We obtained valid responses from 643 individuals (mean age 36.19, SD 7.54 years). Of these, 422 were younger than 40 years (mean age 31.76, SD 5.00 years), and 221 were older than 40 years (mean age 44.67, SD 2.88 years). Table 1 shows the demographic characteristics of participants by age difference. No statistical difference was found in ED severity, phosphodiesterase-5 inhibitors use, and marriage status between men with ED younger and older than 40 years.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>&lt;40 years of age (n=422)</th>
<th>≥40 years of age (n=221)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>31.76 (5.00)</td>
<td>44.67 (2.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>Single</td>
<td>77 (18.25)</td>
<td>35 (15.84)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>345 (81.75)</td>
<td>186 (84.16)</td>
<td></td>
</tr>
<tr>
<td>Duration of marriage (years), mean (SD)</td>
<td>4.87 (3.90)</td>
<td>10.43 (7.53)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>IIEF-5 severity, n (%)</td>
<td></td>
<td></td>
<td>.84</td>
</tr>
<tr>
<td>No ED</td>
<td>84 (19.91)</td>
<td>49 (22.17)</td>
<td></td>
</tr>
<tr>
<td>Mild to mild-to-moderate ED</td>
<td>226 (53.55)</td>
<td>112 (50.68)</td>
<td></td>
</tr>
<tr>
<td>Moderate to severe ED</td>
<td>112 (26.54)</td>
<td>60 (27.15)</td>
<td></td>
</tr>
<tr>
<td>PDE-5 use, n (%)</td>
<td></td>
<td></td>
<td>.63</td>
</tr>
<tr>
<td>Not using</td>
<td>337 (79.86)</td>
<td>180 (81.45)</td>
<td></td>
</tr>
<tr>
<td>Using</td>
<td>85 (20.14)</td>
<td>41 (18.55)</td>
<td></td>
</tr>
</tbody>
</table>

aIIEF-5: International Index of Erectile Function-5.
bPDE-5: phosphodiesterase-5 inhibitor.

The prevalence of depression as judged by PHQ≥10 among men younger than 40 years was 39.81% (168/422), and it was 24.89% (55/221) among those older than 40 years. There was a statistical difference in the prevalence of depression between the two groups (P<.001). In addition, the prevalence of anxiety, as judged by GAD≥10, was 27.25% (115/422) among men.
younger than 40 years, and it was 21.72% (48/221) among those older than 40 years. There was no difference in the prevalence of anxiety between men with ED in the two age groups (\(P=.12\)). Table 2 illustrates these results.

<table>
<thead>
<tr>
<th>Questionnaires and characteristics</th>
<th>&lt;40 years of age (n=422)</th>
<th>≥40 years of age (n=221)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHQ-9</strong>a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No depression or mild depression (PHQ-9&lt;10), n (%)</td>
<td>254 (60.19)</td>
<td>166 (75.11)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prevalence of depression (PHQ-9≥10), n (%)</td>
<td>168 (39.81)</td>
<td>55 (24.89)</td>
<td></td>
</tr>
<tr>
<td><strong>GAD-7</strong>b</td>
<td></td>
<td></td>
<td>.13</td>
</tr>
<tr>
<td>No anxiety or mild anxiety (GAD-7&lt;10), n (%)</td>
<td>307 (72.75)</td>
<td>173 (78.52)</td>
<td></td>
</tr>
<tr>
<td>Prevalence of anxiety (GAD-7≥10), n (%)</td>
<td>115 (27.25)</td>
<td>48 (21.72)</td>
<td></td>
</tr>
</tbody>
</table>

aPHQ-9: Patient Health Questionnaire-9.
bGAD-7: Generalized Anxiety Disorder-7.

The two-way ANOVA was performed with ED severity and age (<40 or >40 years) as independent variables and the scores of AAQ, CFQ, and VQ-OB as dependent variables. The results showed no significant differences in AAQ-II (\(P=.14\)), CFQ (\(P=.08\)), and VQ-OB (\(P=.30\)) scores attributed to ED severity. Moreover, no difference in ED severity or psychological inflexibility depending on the duration of the marriage was found. On the other hand, there were significant differences in the scores of CFQ (\(P=.04\)) and VQ-OB (\(P=.004\)) attributed to age. As the interactions were significant for CFQ (\(P=.04\)) and VQ-OB (\(P=.01\)) scores, the simple main effect was examined. It was found that men with ED younger than 40 years had significantly higher CFQ (\(P=.01; d=0.62\)) and VQ-OB (\(P<.001; d=0.87\)) scores compared to those older than 40 years, in cases of moderate and severe ED. Additionally, it was found that men with moderate to severe ED younger than 40 years had significantly higher CFQ (\(P=.01; d=0.42\)) and VQ-OB (\(P=.02; d=0.38\)) scores compared to men with no ED younger than 40 years. These results are illustrated in Table 3 and Figures 1 and 2.

<table>
<thead>
<tr>
<th>Parameters and factors</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>(F) test (df)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AAQ-II</strong>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>281.60</td>
<td>140.80</td>
<td>1.91 (2,637)</td>
<td>.14</td>
</tr>
<tr>
<td>Age</td>
<td>204.74</td>
<td>204.74</td>
<td>2.78 (1,637)</td>
<td>.10</td>
</tr>
<tr>
<td>ED severity x age</td>
<td>267.27</td>
<td>133.64</td>
<td>1.81 (2,637)</td>
<td>.16</td>
</tr>
<tr>
<td><strong>CFQ</strong>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>433.63</td>
<td>216.82</td>
<td>2.49 (2,637)</td>
<td>.08</td>
</tr>
<tr>
<td>Age</td>
<td>348.12</td>
<td>348.12</td>
<td>3.99 (1,637)</td>
<td>.04</td>
</tr>
<tr>
<td>ED severity x age</td>
<td>534.92</td>
<td>267.46</td>
<td>3.07 (2,637)</td>
<td>.04</td>
</tr>
<tr>
<td><strong>VQ-OB</strong>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>74.56</td>
<td>37.28</td>
<td>1.19 (2,637)</td>
<td>.30</td>
</tr>
<tr>
<td>Age</td>
<td>263.36</td>
<td>263.36</td>
<td>8.38 (1,637)</td>
<td>.004</td>
</tr>
<tr>
<td>ED severity x age</td>
<td>250.82</td>
<td>125.41</td>
<td>3.99 (2,637)</td>
<td>.01</td>
</tr>
</tbody>
</table>

aAAQ-II: Acceptance and Action Questionnaire-II.
bCFQ: Cognitive Fusion Questionnaire.
cVQ-OB: Valuing Questionnaire–Obstacle Subscale.
Figure 1. Results of the two-way ANOVA and the simple main effect of the Cognitive Fusion Questionnaire (CFQ). ED: erectile dysfunction. $P=0.01$ for “a,” “b,” and “c”.

Figure 2. Results of the two-way ANOVA and the simple main effect of the Valuing Questionnaire–Obstacle Subscale (VQ-OB). ED: erectile dysfunction. $P<0.001$ for “a” and $P=0.02$ for “b”.

Discussion

Principal Findings

This cross-sectional study evaluated depression, anxiety, and psychological inflexibility in men younger and older than 40 years with ED. There was no statistical difference in demographic characteristics between the two groups. The average age of the participants was 31.76 (SD 5.00) years in men younger than 40 years; the population was assumed to have mainly mild-to-moderate psychogenic ED. On the other hand, the average age of the participants was 44.67 (SD 2.88) years in men older than 40 years; the population was assumed to have mostly mild-to-moderate organic ED.

Depression was found in both groups. The results of our study were consistent with a previous study [15], which found that younger men had comparatively greater depressive symptoms. In contrast, the prevalence of anxiety was not different between the two age groups. One possible reason is that the anxiety in men with ED is not general anxiety but specific anxiety about sexual situations. Masters and Johnson [24] highlighted the
central role of sexual performance anxiety in couples presenting with sexual dysfunction [24]. In treating sexual dysfunctions, Kaplan [25] emphasizes the importance of addressing specific sources of sexual anxiety, such as fear of failure and not pleasing one’s partner [25]. Although a Japanese version does not exist now, it may be necessary to use a questionnaire like the Erectile Performance Anxiety Index [26].

Men with ED younger than 40 years had significantly higher CFQ and VQ-OB scores than those older than 40 years in cases of moderate and severe ED. Furthermore, men with moderate-to-severe ED younger than 40 years had significantly higher CFQ and VQ-OB scores compared to men without ED. These results partly support our hypothesis that men younger than 40 years are more psychologically inflexible than those older than 40 years. Cognitive fusion might be the critical component of ACT for ED. For example, the fusion with sexual performance anxiety, such as “I might fail again,” makes it impossible to pay attention to the sexual partner, which results in erectile failure. It is also consistent with Barlow’s theory [27]. Barlow [27] proposed a model for the interaction of anxiety and cognitive interference. This model examines how anxiety and cognitive interference interact, particularly in a sexual context, where a lack of control over one’s arousal diverts attention from erotic arousal to physical arousal and the negative consequences associated with failure to attain an erection.

On the other hand, there were no significant differences in the scores of the AAQ-II, which might be related to psychometric issues with AAQ-II. To date, the most used self-report measure of psychological inflexibility, especially experience avoidance, has been the AAQ-II. There was no significant difference in ED severity and psychological inflexibility depending on the duration of the marriage. However, various issues regarding the AAQ-II have emerged from the existing literature [28]. The authors found that the AAQ-II faced challenges in distinguishing distress (like negative affect and neuroticism) from experiential avoidance. For clinical application, researchers have expanded the range of measures for psychological inflexibility. They have developed specific versions of the AAQ-II tailored to different populations or disorders, with currently over 20 available versions (examples include those for the workplace, tinnitus, irritable bowel syndrome, exercise, and epilepsy). The disorder-specific AAQ-II variants indicate greater incremental validity in their targeted areas than the general AAQ-II [29]. Thus, developing a questionnaire on ED-related psychological inflexibility might be necessary.

There are some limitations to this study. First, this study used a cross-sectional approach, indicating merely “associations” rather than “causality” between psychological inflexibility and ED. Further controlled experimental and longitudinal studies are essential to delve deeper into the impact of psychological inflexibility on ED. Second, in this study, no responses were obtained from the partners of men with ED. Including the partners in the assessment and treatment of ED is recommended. It is desirable to obtain responses from partners in future studies. Finally, the specific racial or ethnic and socioeconomic profiles of the participants may restrict the broader applicability of the findings. The study was also conducted during the COVID-19 epidemic, which may have influenced the results.

Conclusions
To the best of our knowledge, this web-based cross-sectional study was the first to examine the relationship between psychological inflexibility and ED. We conclude that men with moderate and severe ED younger than 40 years have higher psychological inflexibility and might be eligible for the ACT. In addition, developing a Japanese version of the questionnaire is necessary to measure ED-related anxiety and psychological inflexibility.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to the Waseda University Academic Research Ethical Review Committee’s data-sharing policy but are available from the corresponding author upon reasonable request.

Conflicts of Interest
This research was funded by Logos Science Corp, Ltd, Tokyo, Japan. MG, CS, and HT are members of the Logos Science Corp, Ltd.

References


Abbreviations

AAQ-II: Acceptance and Action Questionnaire - II
ACT: Acceptance and Commitment Therapy
CFQ: Cognitive Fusion Questionnaire
ED: Erectile Dysfunction
GAD-7: Generalized Anxiety Disorder - 7
IIEF-5: International Index of Erectile Function - 5
PHQ-9: Patient Health Questionnaire - 9
VQ-OB: Valuing Questionnaire–Obstacle Subscale
Determining Distinct Suicide Attempts From Recurrent Electronic Health Record Codes: Classification Study

Kate H Bentley, PhD; Emily M Madsen, BSc; Eugene Song, BA; Yu Zhou, MPP; Victor Castro, MS; Hyunjoon Lee, MSc; Younga H Lee, PhD; Jordan W Smoller, MD, SCD

Corresponding Author:
Kate H Bentley, PhD
Center for Precision Psychiatry
Department of Psychiatry
Massachusetts General Hospital
185 Cambridge Street
2nd Floor
Boston, MA, 02114
United States
Phone: 1 6177247741
Email: kbentley@mgh.harvard.edu

Abstract

Background: Prior suicide attempts are a relatively strong risk factor for future suicide attempts. There is growing interest in using longitudinal electronic health record (EHR) data to derive statistical risk prediction models for future suicide attempts and other suicidal behavior outcomes. However, model performance may be inflated by a largely unrecognized form of “data leakage” during model training: diagnostic codes for suicide attempt outcomes may refer to prior attempts that are also included in the model as predictors.

Objective: We aimed to develop an automated rule for determining when documented suicide attempt diagnostic codes identify distinct suicide attempt events.

Methods: From a large health care system’s EHR, we randomly sampled suicide attempt codes for 300 patients with at least one pair of suicide attempt codes documented at least one but no more than 90 days apart. Supervised chart reviewers assigned the clinical settings (ie, emergency department [ED] versus non-ED), methods of suicide attempt, and intercode interval (number of days). The probability (or positive predictive value) that the second suicide attempt code in a given pair of codes referred to a distinct suicide attempt event from its preceding suicide attempt code was calculated by clinical setting, method, and intercode interval.

Results: Of 1015 code pairs reviewed, 835 (82.3%) were nonindependent (ie, the 2 codes referred to the same suicide attempt event). When the second code in a pair was documented in a clinical setting other than the ED, it represented a distinct suicide attempt 3.3% of the time. The more time elapsed between codes, the more likely the second code in a pair referred to a distinct suicide attempt event from its preceding code. Code pairs in which the second suicide attempt code was assigned in an ED setting at least 5 days after its preceding suicide attempt code had a positive predictive value of 0.90.

Conclusions: EHR-based suicide risk prediction models that include International Classification of Diseases codes for prior suicide attempts as a predictor may be highly susceptible to bias due to data leakage in model training. We derived a simple rule to distinguish codes that reflect new, independent suicide attempts: suicide attempt codes documented in an ED setting at least 5 days after a preceding suicide attempt code can be confidently treated as new events in EHR-based suicide risk prediction models. This rule has the potential to minimize upward bias in model performance when prior suicide attempts are included as predictors in EHR-based suicide risk prediction models.
Introduction

Suicide is the tenth leading cause of death in the United States, with more than 48,000 suicide deaths annually [1]. Over the past 20 years, the suicide rate has increased by over 35% [2]. Most people who die by suicide have recently interacted with the health care system, with over half having a health care visit in the month prior to death [3,4]. Health care systems thus offer a key opportunity to identify people at high risk for suicide. Unfortunately, clinicians are poor at predicting who will make a suicide attempt [5] and traditionally studied risk factors perform no better than chance at predicting future suicidal behavior [6].

Recent work has focused on developing and validating machine learning models that use routinely collected electronic health record (EHR) data to predict future suicidal behavior [7]. Such models have demonstrated high levels of accuracy, exceeding that seen with clinician prediction and usual clinical risk factors [6,8-10]. EHR-based suicide risk prediction models, however, face one significant challenge that to date has not been adequately addressed. Suicide attempt is generally the outcome of interest in these models and is typically defined by International Classification of Diseases (ICD) diagnostic codes [11,12]. Within a given patient’s EHR, a suicide attempt code may be given multiple times across distinct health care encounters, often over very short periods of time (eg, days and weeks). Such “recurrent” codes may represent either distinct, new events (ie, multiple suicide attempts) or refer to the same event (ie, a single suicide attempt). The latter may occur when, for example, after making a suicide attempt, a patient has an emergency department (ED) visit followed by an inpatient hospitalization or outpatient follow-up encounters, with one (or multiple) suicide attempt codes assigned at each. In the absence of manual reviews of the narrative notes within patients’ EHRs, which cannot be performed at scale, it can be challenging to determine whether such recurrent suicide attempt codes, especially when documented over short time periods, refer to independent, distinct suicide attempts. Failure to make this important distinction can result in a form of “data leakage” in which the outcome to be predicted is included among features used for the prediction. This can result in substantial inflation of model performance [13].

To address this issue, some researchers have restricted model development to predict only the first occurrence of a suicide attempt code in a patient’s EHR [14-16]. This approach has a major limitation, however, in that a past suicide attempt is among the strongest known predictors of future suicidal behavior [17]. Thus, models that predict only the first documented suicide attempt ignore the subset of patients who may be at highest risk and thus of greatest clinical concern: those with a prior suicide attempt. Another approach is to include any previous suicide attempt codes as predictors of a subsequent suicide attempt code [18-21] thus including potential “repeat attempters” in these models. This approach, however, poses a significant risk of artificially inflating model performance if subsequent codes do not in fact refer to new suicide attempts. In other words, if a suicide attempt code instance used as an outcome actually indexes an attempt that was included a predictor, model performance will be inflated.

To minimize the risk of data leakage while retaining the option of including prior attempts as predictors, we aimed to develop an automated rule for determining whether recurrent suicide attempt codes in the EHR refer to distinct events. Such a rule might be based on relevant variables including clinical setting (eg, a suicide attempt code documented in the ED may be more likely to refer to a new suicide attempt event than one given in a non-ED setting), method (eg, suicide attempt codes that specify different methods may be more likely to refer to distinct events than codes specifying the same method), and time (eg, the more time elapsed between 2 suicide attempt codes, the less likely it may be that the codes refer to the same event). Here, we conducted a comprehensive manual EHR chart review to derive an automated rule that could identify criteria for selecting distinct suicide attempts with high confidence.

Methods

Data Source

The data source for this study was the Mass General Brigham (MGB) Research Patient Data Registry [22]. This registry covers 6.7 million patients treated in MGB-affiliated hospitals including the Massachusetts General Hospital and Brigham and Women’s Hospital in Boston.

Ethics Approval

This research was approved by the MGB institutional review board, which granted a waiver of informed consent (protocol #2018P001508).

Case Definition and Inclusion Criteria

Details of the development of our EHR-based case definitions for suicide attempt in the MGB health care system are reported elsewhere [14,15]. In brief, we first identified candidate ICD, Ninth Revision (ICD-9) and ICD, Tenth Revision (ICD-10) codes that are likely to capture suicide attempts. Next, expert clinicians conducted manual chart reviews of 670 patients (over 3000 narrative notes) to determine a final set of codes that capture suicide attempts with a positive predictive value (PPV) of >0.70: for ICD-9, E95*, 965*, 967*, 969*, and 881*, and for ICD-10, X71*-X83*, T14.91*, T36*-T50*, T56.9, T57.9, T58.0, T58.1, T58.9, T59.9, T60.9, T61.0, T61.1,
T61.9, T62.9, T63.9, T64.0, T64.8, and T65.9, where the fifth is 2).

For this study, we randomly selected a sample of 300 patients with 2 suicide attempt codes documented at least one but no more than 90 days apart (the “narrow sample”). This interval was chosen to capture codes that were given within a narrow time frame and thus potentially enriched for being “leaked” codes. In a sensitivity analysis, we randomly selected a second, smaller sample of 100 patients with 2 suicide attempt codes documented at least 1 day apart but with no other restrictions on intercode interval (the “broad sample”). A total of 31 patients appeared in both narrow and broad samples. Patients for whom we were unable to confidently locate the narrative notes corresponding to documented suicide attempt codes (eg, no narrative notes available within 30 days of the suicide attempt code date, narrative notes recorded on paper and never migrated to the EHR) were excluded after the sampling process.

Results

Descriptive Statistics

The mean number of suicide attempt codes per patient in the narrow sample was 3.38 (SD 4.62; range 1–47). A total of 225 (75%) patients had <4 codes and 281 (93.7%) had <10 codes. A total of 210 (20.7%) code pairs had a second code reflecting a subsequent encounter for a condition for which the patient had received active treatment (indicated by a seventh “D” character).

Regarding how often the codes in a pair referred to distinct suicide attempts, of the 300 patients in the narrow sample, only 81 (27%) had more than one confirmed (by manual chart review) suicide attempt captured by the reviewed code pairs. Of the 1015 code pairs, only 180 (17.7%) referred to 2 distinct suicide attempt events. Table S1 in Multimedia Appendix 1 presents an example of sampled codes (and the variables assigned to each code and code pair) for a deidentified patient.

For clinical setting, the most common code pair types were non-ED/non-ED (n=542, 53.4%) followed by ED/ED (n=274, 27%). Regarding the 749 total non-ED codes, the most commonly represented clinical setting was inpatient (n=411, 54.9% of all non-ED codes), followed by other or unclear setting (n=149, 19.9% codes), intensive or critical care units (n=134, 17.9% codes), and outpatient (n=55, 7.3% codes). For suicide attempt method, the majority of code pairs (n=766, 75.5%) comprised 2 codes that referred to the same method. The median interval between codes in each code pair, across all codes, was 1 day. Among code pairs that referred to distinct suicide attempt events, the median interval was 35 days.

PPVs

Clinical Setting

Non-ED/ED code pairs (23 total code pairs) had the highest PPVs (0.96, 95% CI 0.87-1.04) for distinct suicide attempt events (Table 1). ED/ED pairs (274 total code pairs) had the second-highest PPVs (0.49, 95% CI 0.43-0.55). When the second code in a pair was assigned in a non-ED setting, PPVs were low (below 0.10).

In a sensitivity analysis, we excluded codes or encounters documented in inpatient settings with a prior code on the previous day from an inpatient or critical or intensive care setting in which each code in the pair was documented to 2 distinct suicide attempts (dichotomous variable indicating distinct or not distinct suicide attempts), (2) clinical setting in which each code in the pair was documented (dichotomous variable indicating ED or non-ED [eg, outpatient and inpatient setting], (3) suicide attempt method of each code in the pair (categorical variable with 6 categories derived from previous literature: poisoning, cutting or piercing, hanging or strangulation or suffocation, jumping, firearm, and other [which included codes with no specified method]), and (4) time elapsed (in days) between codes in each pair [23]. When there were multiple encounters with suicide attempt codes on the same day, these variables were assigned to codes at the day level; see Table S1 in Multimedia Appendix 1 for an example of how we combined multiple same-day encounters.

Data Analysis

We defined PPV as the probability that the second code in a pair referred to a new suicide attempt independent of the first code in the pair. To mimic the approach that would likely be taken in building predictive models, each code pair was treated independently (ie, we did not account for the nested nature of code pairs within patients). First, for the narrow sample, we calculated (in Excel [Microsoft]) PPVs and 95% CIs by clinical setting, suicide attempt method, and intercode interval, respectively. For clinical setting, we calculated the PPVs for 4 possible code pair types: (1) both codes documented in the ED (ED/ED), (2) first code ED and second code non-ED (ED/non-ED), (3) first code non-ED and second code ED (non-ED/ED), and (4) neither code ED (non-ED/non-ED). For suicide attempt method, we calculated the PPVs of 2 possible code pair types: (1) same suicide attempt method for codes in a pair and (2) different suicide attempt methods for codes in a pair. For intercode interval, we first calculated PPVs for all 7-day intervals from 1 to 91 days, followed by collapsing across intervals from 92 days on. We then calculated the PPVs for time intervals within each of the 6 (4 clinical settings and 2 suicide attempt methods) code pair types. To derive our proposed rule, we set our benchmark PPV to 0.90. For each of the 6 code pair types, we determined the minimum time elapsed between codes (ie, interval floor) at which the PPV was at least 0.90. For a sensitivity analysis, we computed the same series of PPVs for the broad sample.

Procedure

Under the supervision of JWS (a senior clinician with expertise in the treatment of suicidal behavior), 2 study team members (EMM and ES) manually reviewed the EHR clinical encounter data (including narrative notes) relevant to each pair of suicide attempt codes (“code pair”) per sampled patient (1015 in the narrow sample and 300 in the broad sample; 1253 unique codes across the 2 samples). Each code pair comprised a given suicide attempt code and the immediately (temporally) preceding code in a patient’s EHR. All applicable code pairs per patient were examined (including other code pairs with >90-day intervals for patients in the narrow sample). Chart reviewers assigned the following variables to each code pair: (1) whether the code pair referred to 2 distinct suicide attempts (dichotomous variable indicating distinct or not distinct suicide attempts), (2) clinical setting in which each code in the pair was documented (dichotomous variable indicating ED or non-ED [eg, outpatient and inpatient setting], (3) suicide attempt method of each code in the pair (categorical variable with 6 categories derived from previous literature: poisoning, cutting or piercing, hanging or strangulation or suffocation, jumping, firearm, and other [which included codes with no specified method]), and (4) time elapsed (in days) between codes in each pair [23]. When there were multiple encounters with suicide attempt codes on the same day, these variables were assigned to codes at the day level; see Table S1 in Multimedia Appendix 1 for an example of how we combined multiple same-day encounters.

For this study, we randomly selected a sample of 300 patients with 2 suicide attempt codes documented at least one but no more than 90 days apart (the “narrow sample”). This interval was chosen to capture codes that were given within a narrow time frame and thus potentially enriched for being “leaked” codes. In a sensitivity analysis, we randomly selected a second, smaller sample of 100 patients with 2 suicide attempt codes documented at least 1 day apart but with no other restrictions on intercode interval (the “broad sample”). A total of 31 patients appeared in both narrow and broad samples. Patients for whom we were unable to confidently locate the narrative notes corresponding to documented suicide attempt codes (eg, no narrative notes available within 30 days of the suicide attempt code date, narrative notes recorded on paper and never migrated to the EHR) were excluded after the sampling process.
setting. For example, if a patient was given suicide attempt codes on three consecutive days in an inpatient setting, we only used the day 1 code. This resulted in 792 (versus 1015) analyzed code pairs. The results were overall very similar to when we did not exclude contiguous inpatient codes (Multimedia Appendix 2).

Table 1. Code pairs in the narrow sample defined by the clinical setting (ED\textsuperscript{a} or non-ED) of the first and second codes in each pair.

<table>
<thead>
<tr>
<th>First code clinical setting</th>
<th>Second code clinical setting</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Interval between codes (days), median (Q1\textsuperscript{b}, Q3\textsuperscript{c})</th>
<th>Interval between codes (days), mean (SD)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV\textsuperscript{d} (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-ED</td>
<td>Non-ED</td>
<td>542 (53.4)</td>
<td>1 (1, 1)</td>
<td>3.58 (8.64)</td>
<td>14</td>
<td>0.03 (0.01-0.04)</td>
</tr>
<tr>
<td>ED</td>
<td>Non-ED</td>
<td>176 (17.3)</td>
<td>1 (1, 3)</td>
<td>6.47 (34.23)</td>
<td>10</td>
<td>0.06 (0.02-0.09)</td>
</tr>
<tr>
<td>Non-ED</td>
<td>ED</td>
<td>23 (2.3)</td>
<td>52 (13, 125)</td>
<td>154.09 (286.63)</td>
<td>22</td>
<td>0.96 (0.87-1.04)</td>
</tr>
<tr>
<td>ED</td>
<td>ED</td>
<td>274 (27)</td>
<td>5 (1, 36)</td>
<td>53.79 (211.77)</td>
<td>134</td>
<td>0.49 (0.43-0.55)</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>1015</td>
<td>1</td>
<td>21.05 (122.43)</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}ED: emergency department.
\textsuperscript{b}Q1: first quartile.
\textsuperscript{c}Q3: third quartile.
\textsuperscript{d}PPV: positive predictive value.

**Suicide Attempt Method**

For suicide attempt method (same versus different method for 2 codes in a code pair), the PPVs were below 0.25 (Table 2).

Table 2. Code pairs defined by whether the first and second codes referred to the same or a different suicide attempt method.

<table>
<thead>
<tr>
<th>First and second code</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV\textsuperscript{a} (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same method</td>
<td>766 (75.5)</td>
<td>128</td>
<td>0.17 (0.14-0.19)</td>
</tr>
<tr>
<td>Different method</td>
<td>249 (24.5)</td>
<td>52</td>
<td>0.21 (0.16-0.26)</td>
</tr>
<tr>
<td>Overall</td>
<td>1015</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PPV: positive predictive value.

**Intercode Interval**

Table 3 presents PPVs for code pairs broken down by 7-day (week-long) intervals; the majority (n=797, 78.5%) of code pairs had an intercode interval of 7 days or less. The more days elapsed between 2 codes, the larger the PPV (and, fewer code pairs per strata). Table S7 in Multimedia Appendix 4 presents PPVs for code pairs broken down by interval and clinical setting (non-ED/non-ED, ED/non-ED, non-ED/ED, ED/ED), and Table S8 in Multimedia Appendix 5 presents PPVs for code pairs broken down by interval and suicide attempt method (same versus different). In another sensitivity analysis, given that ICD-9 is no longer used, we also computed all PPVs reported in Tables 1-3 when excluding code pairs with at least one ICD-9 coded event. The same pattern of findings held, with 95% CIs for all PPVs overlapping with those in Tables 1-3.
Table 3. Code pairs defined by intercode interval.

<table>
<thead>
<tr>
<th>Intercode interval</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV&lt;sup&gt;a&lt;/sup&gt; (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-7 days</td>
<td>797 (78.5)</td>
<td>31</td>
<td>0.04 (0.03-0.05)</td>
</tr>
<tr>
<td>8-14 days</td>
<td>48 (4.7)</td>
<td>19</td>
<td>0.40 (0.26-0.53)</td>
</tr>
<tr>
<td>15-21 days</td>
<td>31 (3)</td>
<td>17</td>
<td>0.55 (0.37-0.72)</td>
</tr>
<tr>
<td>22-28 days</td>
<td>20 (2)</td>
<td>14</td>
<td>0.70 (0.50-0.90)</td>
</tr>
<tr>
<td>29-35 days</td>
<td>17 (1.7)</td>
<td>10</td>
<td>0.59 (0.35-0.82)</td>
</tr>
<tr>
<td>36-42 days</td>
<td>17 (1.7)</td>
<td>15</td>
<td>0.88 (0.73-1.04)</td>
</tr>
<tr>
<td>43-49 days</td>
<td>10 (1)</td>
<td>8</td>
<td>0.80 (0.55-1.05)</td>
</tr>
<tr>
<td>50-56 days</td>
<td>18 (1.8)</td>
<td>14</td>
<td>0.78 (0.59-0.97)</td>
</tr>
<tr>
<td>57-63 days</td>
<td>5 (4.9)</td>
<td>3</td>
<td>0.60 (0.17-1.03)</td>
</tr>
<tr>
<td>64-70 days</td>
<td>7 (0.7)</td>
<td>6</td>
<td>0.86 (0.60-1.12)</td>
</tr>
<tr>
<td>71-77 days</td>
<td>2 (0.2)</td>
<td>2</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>78-84 days</td>
<td>9 (0.9)</td>
<td>9</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>85-91 days</td>
<td>6 (0.6)</td>
<td>6</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>92+ days</td>
<td>28 (2.7)</td>
<td>26</td>
<td>0.93 (0.83-1.02)</td>
</tr>
<tr>
<td>Overall</td>
<td>1015</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

<sup>a</sup>PPV: positive predictive value.

As shown in Figure 1, across all code pairs, pairs with an interval of at least 53 days had a PPV of 0.90 (range 0.88-0.93). The interval floors meeting our benchmark PPV (at least 0.90) within each of the 6 code pair types are also labeled in Figure 1 (clinical setting) and Figure 2 (suicide attempt method). For non-ED/ED code pairs (23 code pairs), an interval floor of 1 day had a PPV of 0.96. When both codes were assigned in the ED (271 code pairs), PPV reached 0.90 when the intercode interval was at least 5 days. When the second code in a pair was documented in an ED (regardless of the setting in which the first code was documented), PPV was 0.91 when the intercode interval was 5 days (the PPV was 0.89 for 4 days). Thus, whenever the second code in a pair was documented in an ED at least 5 days after the previous code, the probability that the second code referred to an independent suicide attempt was at least 90%.

Figure 1. PPVs for interval floors by code pair types defined by clinical setting. The labeled data points indicate the interval floor at which the PPV was at least 0.90 (or the maximum PPV). Gray lines reflect PPVs for interval floors across all code pair types. Red lines refer to code pairs documented in ED (first code) and ED (second code) settings (ED/ED). Blue lines are non-ED/ED code pairs; purple lines ED/non-ED; and green lines non-ED/non-ED. ED: emergency department; PPV: positive predictive value.
Sensitivity Analysis: Broad Sample

Results from the same series of analyses in the broad sample are presented in Multimedia Appendix 6. Of the 100 patients, 45 (45%) had more than 1 confirmed suicide attempt. Of the 300 code pairs, 86 (28.7%) referred to 2 distinct suicide attempts. The median interval between codes in each pair was also 1 day. Among code pairs that referred to distinct suicide attempts, the median interval was 133 days. Overall, we found a similar pattern of PPVs (in almost all cases overlapping 95% CIs) to those from the narrow sample. Across all code pairs in the broad sample, those with an interval of at least 37 days had a PPV of 0.90 (range 0.87-0.93). When both codes were given in the ED (86 pairs), PPV reached 0.90 when the interval was at least 2 days.

Discussion

Primary Findings

Machine learning suicide risk prediction models that leverage routinely collected EHR data can outperform clinician assessment [8] and have the potential to improve how patients at risk for suicide are identified and treated. These models are typically trained using ICD codes to label suicide attempts. An under-appreciated challenge when building these models, however, is that ICD codes indexing a single suicide attempt are often used repeatedly across multiple encounters. This could create a substantial problem for models that incorporate prior suicide attempts, an established risk factor, in predicting subsequent attempts or suicidal behavior.

Some investigators side-step this issue by restricting model predictions to only the first occurrence of a suicide attempt code. This approach, however, limits the utility of prediction models by ignoring prior attempts, the best-known risk factor for suicidal behavior, and limiting their application to a subset of those at risk; prior studies indicate that nearly one-quarter of those who engage in deliberate self-harm have recurrent episodes within 3 years [24]. Here we aimed to develop a portable, automated rule for determining when recurrent suicide attempt codes refer to distinct suicide attempt events in a patient’s history. Based on chart review of clinical encounters corresponding to 1015 unique ICD code pairs, we found that, for patients with more than 1 documented suicide attempt code, repeat codes most often (>80% of the time) reflected nonindependent events, underscoring the high frequency of “leaked” suicide attempt codes. When collapsing across all clinical settings, repeat codes needed to be documented at least 53 days after the preceding code in order to refer (with probability >90%) to a new, distinct suicide attempt. However, repeat codes documented in an ED at least 5 days after the preceding suicide attempt code were likely (probability >90%) to refer to a new, distinct suicide attempt.

The most informative variables for determining whether recurrent suicide attempt codes referred to distinct suicide attempts were the clinical setting in which the codes were documented and the time elapsed between codes. First, regarding clinical setting, when a suicide attempt code was documented in an ED after the preceding code, it referred to a new suicide attempt more than half the time. Suicide attempt codes documented in non-ED settings, accounting for most of the second codes among all code pairs, however, were highly unlikely to refer to a new suicide attempt (probability <5%). This may be due to the fact that the vast majority (nearly three-quarters) of non-ED codes occurred in inpatient or intensive or critical care units, where patients may be treated over the course of several days or longer, potentially accumulating multiple suicide attempt codes that all refer to the same index event that may have prompted inpatient or intensive treatment. This pattern of findings, for one, highlights the considerable risk of treating all recurrent suicide attempt codes (especially those from non-ED settings) as distinct events, and the potential importance of using a simple rule, such as that proposed here, to identify probable distinct suicide attempt events.
Along these lines, the more time elapsed between 2 suicide attempt codes, the more likely it was the codes referred to distinct events. Combining these 2 variables—clinical setting and time elapsed—provided a simple rule for determining whether recurrent suicide attempt codes refer to distinct events with at least 90% probability. Although the accuracy of our proposed rule (at least 5 days elapsed between a code given in the ED and the preceding code) may differ in other health care systems, we recommend that others consider taking into account these 2 variables when incorporating recurrent suicide attempt codes in EHR-based suicide risk prediction models.

Perhaps surprisingly, whether the coded suicide attempt method for 2 codes in a pair was the same or different did not provide value in identifying distinct suicide attempt events. However, in the relatively small proportion of code pairs (24.5%) that referred to different methods, the most common “profile” was 1 code with a specific method (eg, poisoning and cutting or piercing) and the other code with method categorized as “other” (not a different specific method); notably, the “other” category included codes lacking any specified method. Thus, the fact that method did not help identify distinct events may largely reflect inconsistencies in how or whether the suicide attempt method is coded by providers. In contrast, neither of the other 2 variables examined (clinical setting nor intercode interval) should be impacted by irregular coding practices, and thus may also be more scalable and reliable for other health care systems planning to use this or a similar rule.

Our derived rule (at least 5 days elapsed between a code from the ED and the preceding code) may have more impact on certain suicide-related prediction tasks than others. For example, it may be especially relevant when estimating patients’ risk of repeat suicidal behavior, for example after an ED visit for suicidal behavior, which could influence clinical decision-making at the point of care (eg, about discharge home or to outpatient care versus hospitalization). This rule may have less impact for other related prediction tasks, such as estimating patients’ risk of suicidal behavior after nonsuicide-related outpatient visits or broader population-based prediction efforts [25]. These results may also be less relevant for models that solely predict fatal self-harm or suicide deaths [26,27]. Future work should systematically evaluate the performance and clinical utility of models that do and do not incorporate the proposed rule for incorporating recurrent suicide attempt codes across a range of prediction goals and clinical contexts.

Our results must be considered in the context of a few key limitations. First, some of the sampled patients may have presented to hospitals outside of the MGB system for suicide attempts. In these cases, the corresponding diagnostic codes and contextual information were either unavailable or only sporadically recorded in narrative notes at subsequent clinical encounters within MGB. We also excluded sampled patients for whom chart reviewers could not confidently match data pulled from the MGB Research Patient Data Registry to the narrative notes.

Conclusions
This analysis indicates that EHR-based suicide attempt prediction models that include ICD codes for prior attempts as a predictor may be highly susceptible to bias due to data leakage in model training. Our proposed rule for circumventing this issue should minimize this bias and its inflationary effect on model performance metrics. The key variables included in our rule (clinical setting and time elapsed between codes) are widely available in health system data warehouses and should be easily integrated into EHR-based models. It is also possible that the approach taken in this study may be relevant for developing and refining machine learning models aimed to predict other episodic events of interest that can be repeatedly documented in the health record, such as unintentional overdose, domestic abuse, or episodes of violence. If effectively implemented into existing and future suicide risk prediction models, this rule could increase the robustness and validity of machine-learning based approaches to identifying the individuals at highest risk for suicide, and ultimately advance suicide prevention efforts in health care contexts on a large scale.

Acknowledgments
The authors thank the Enterprise Research Infrastructure and Services at MGB for their in-depth support and for the provision of the research patient data registry and the ERISOne Linux cluster. This work was supported by grants (NIMH R01 MH117599; JWS) and (K23MH120436; KHB) from the National Institute of Mental Health, and a gift from the Tommy Fuss Fund (JWS). JWS is a member of the Leon Levy Foundation Neuroscience Advisory Board, the Scientific Advisory Board of Sensorium Therapeutics (with equity), and has received an honorarium for an internal seminar Tempus Labs. He is principal investigator of a collaborative study of the genetics of depression and bipolar disorder sponsored by 23andMe for which 23andMe provides analysis time as in-kind support but no payments.

Data Availability
The data used in this study cannot be made publicly available due to restrictions relating to the use of EHR data.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Sampled codes or code pairs and designations per manual chart review for an example (deidentified) patient.
Multimedia Appendix 2
Sensitivity analysis (excluding contiguous codes from inpatient settings).

Multimedia Appendix 3
Code pairs in the narrow sample defined by specific category of suicide attempt method (poisoning, cutting or piercing, hanging or strangulation or suffocation, jumping, firearm, or other) of the first and second code in each code pair.

Multimedia Appendix 4
Code pairs in the narrow sample defined by both the clinical setting (ED or non-ED) of and the interval (in days) between the first and second codes in each pair.

Multimedia Appendix 5
Code pairs in the narrow sample defined by both suicide attempt method (same or different) and the interval (in days) between first and second codes in each pair.

Multimedia Appendix 6
Sensitivity analysis (results for broad sample).

References


22. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a research patient data repository. AMIA Annu Symp Proc 2006;2006:1044 [FREE Full text] [Medline: 17238663]


Abbreviations
- **ED**: emergency department
- **EHR**: electronic health record
- **ICD**: International Classification of Diseases
- **MGB**: Mass General Brigham
- **PPV**: positive predictive value
©Kate H Bentley, Emily M Madsen, Eugene Song, Yu Zhou, Victor Castro, Hyunjoon Lee, Younga H Lee, Jordan W Smoller. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Synergizing the Behavior Change Wheel and a Cocreative Approach to Design a Physical Activity Intervention for Adolescents and Young Adults With Intellectual Disabilities: Development Study

Laura Maenhout, MSc; Julie Latomme, PhD; Greet Cardon, PhD; Geert Crombez, PhD; Geert Van Hove, PhD; Sofie Compernolle, PhD

1Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium
2Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium
3Department of Special Needs Education, Ghent University, Ghent, Belgium

Corresponding Author:
Laura Maenhout, MSc
Department of Movement and Sports Sciences
Ghent University
Watersportlaan 2
Ghent, 9000
Belgium
Phone: 32 92646312
Email: laura.maenhout@ugent.be

Abstract

Background: There is a need for physical activity promotion interventions in adolescents and young adults with intellectual disabilities. Current interventions have shown limited effectiveness, which may be attributed to the absence of theory and a population-specific development. Combining a planning model (including theory) and cocreation with the target audience during intervention development could potentially address this gap.

Objective: This study aimed to report the systematic development of the Move it, Move ID! intervention by describing how the 8 different steps of the Behavior Change Wheel (BCW) were applied and present the results that emerged from those steps. In doing so, the (theoretical) content of the intervention is described in detail.

Methods: A total of 23 adolescents and young adults (aged 14-22 years) with mild to moderate intellectual disabilities were designated as cocreators of the intervention. Across 2 groups, 6 similar cocreation sessions were organized in each. The content and sequence of the sessions were structured to align with the 8 steps of the BCW. All sessions were recorded and transcribed verbatim. Both a deductive (ie, steps of the BCW) and inductive (ie, resonating the voice of the participants) analysis approach were applied specifically focusing on identifying and describing the findings within each of the BCW steps.

Results: After behavioral analysis (steps 1-4), 10 intervention goals were chosen and linked to Capability, Opportunity, and Motivation–Behavior components (theory within the BCW) that needed to be addressed. Psychological capability, social opportunity, and reflective motivation were emphasized as the first targets to focus on. A key finding was the urge for real-life social connectedness and social integration, which makes the social component as part of physical activity a central theme to focus on within intervention development. Judgments on the most suitable intervention functions (step 5) and behavior change techniques (step 7) were explained. When discussing the mode of delivery of the intervention (step 8), it was underscored that solely relying on a mobile health app would not fulfill participants’ social needs. Hence, the chosen intervention adopts a dyadic approach in which young individuals with intellectual disabilities are matched with peers without intellectual disabilities to engage in physical activities together, with a mobile app playing a supportive role in this partnership.

Conclusions: The transparent description of the development process highlights why certain intervention components and behavior change techniques were chosen and how they are intertwined by means of the selected intervention design. This paper provides a detailed blueprint for practitioners wanting to integrate the BCW and its associated behavior change techniques, in combination with actively involving the target group, into their intervention development for people with intellectual disabilities.

(JMIR Form Res 2024;8:e51693) doi:10.2196/51693
Both reviews concluded that only a small number of individuals meet the PA guidelines, indicating that they are less active than the general population [18,19].

**Current PA Interventions and Their Effectiveness**

Although PA research in people with IDs has been growing over the last decade, this field has been underresearched. A PA Series in *The Lancet* (2021) stated that, between 1999 and 2019, <5% of all articles in the 5 highest-impact medical journals focused on people with disabilities (not limited to IDs) and <7% of these addressed PA or health [2]. A systematic review from 2019 on the effectiveness of PA interventions among participants with IDs of all ages identified only three effective randomized controlled trials out of 9 [20]: (1) a 10-week progressive resistance training program in adolescents and young adults (aged 14-22 years) with Down syndrome in Australia [21], (2) a 12- to 16-month multicomponent diet and PA program in adults (aged 20-66 years) with mild to moderate IDs in Sweden [22], and (3) an 8-month PA and fitness program in “fast-walking” older adults with mild to moderate IDs (aged >40 years) in the Netherlands [23]. The success of these randomized controlled trials was mainly attributed to the following factors: (1) practical support from others (eg, a mentor) in guiding and helping participants with IDs through the intervention and for increasing motivation, (2) establishing a routine that involves regular PA as well as the adaptability of an intervention to the specific routines of both carers and participants, (3) the simplicity of an intervention by adapting interventions to the specific needs of the participants, and (4) familiarity with the intervention [20]. None of the 9 interventions in the systematic review by Hassan et al [20] included a technological component (ie, eHealth or mobile health [mHealth]). However, there seems to be no reason why digital interventions would not be feasible in this target group [24]. In the study by Ptomey et al [25], 95% of the participants, aged 14.9 (SD 2.2) years on average, reported that using a tablet computer was easy and enjoyable. It is then no surprise that, in recent years, there has been a growing interest in the development of digital interventions for individuals with IDs [26-30].

The number of effective PA interventions for people with IDs remains limited. A potential reason for the limited effectiveness is currently attributed to the lack of a theoretical framework for intervention development and the difficulty in concretizing behavior change techniques (BCTs) in an understandable way for this population [10,20,31,32]. A 2017 systematic review on the use of BCTs in lifestyle change interventions for people with IDs, for example, concluded that 73% of the studies aiming to improve PA in the target group did not use any theoretical framework [31]. Nevertheless, the use of a theoretical framework is an important prerequisite for intervention effectiveness [2,20,33-36]. Furthermore, when examining theory-based interventions for people with IDs, concerns have been raised regarding the suitability of the theories used (eg, social cognitive theory, theory of planned behavior, and self-determination theory) as a starting point for designing interventions for this specific target group. These theories may...
not sufficiently address the specific challenges faced by people with IDs. More specifically, these theories tend to be specific and detailed, yet they may not encompass the complete spectrum of potential influences on behavior within this particular target group and often concentrate on individual-level factors [20,31,32].

Applying the Behavior Change Wheel and a Cocreative Approach to Build Theory-Based PA Interventions

The Behavior Change Wheel (BCW) is a planning model aimed at guiding a scientific and systematic intervention development process [33,37]. The BCW contains a behavioral theory at its heart, the Capability, Opportunity, and Motivation–Behavior (COM-B) model, which encompasses the full range of influences contributing to the behavior of interest [33,38]. A total of 3 behavioral components are summarized in the COM-B model, which states that, for each behavior to occur, individuals need capability (physical and psychological), opportunity (physical and social), and motivation (reflective and automatic) [33,37,39]. The COM-B model is in turn linked to the Theoretical Domains Framework (TDF) [33,40], which subdivides the COM-B model into 14 domains. The BCW further formulates 9 intervention functions linked to 93 BCTs [41] and 7 policy types with the aim of modifying each of the 3 COM-B components and, thus, changing behavior. The COM-B model describes the minimal factors that behavioral scientists agree on to achieve behavior change and has been developed with interdisciplinary research in mind [37]. It is an open model and relatively easy to communicate, especially with vulnerable groups. In recent years, this model has demonstrated applicability in the context of PA among people with IDs and their carers [32,42]. This study chose the BCW as a planning model for intervention development because of its practical use and feasibility in combination with a cocreative approach. Current lifestyle modification approaches for this target group lack a robust foundation addressing their unique needs [10,32,43]. Therefore, deeply engaging with this group and customizing approaches to promote their PA is vital. Unfortunately, individuals with IDs are seldom heard in research, and interviews with caregivers often take precedence, potentially overshadowing their authentic experiences [43]. Neglecting the perspectives of individuals with IDs can undermine intervention acceptability, comprehensibility, and feasibility [10,44-48]. To clarify, previous intervention studies have reported that some BCTs may be too complex for the target group (eg, self-monitoring through the use of pedometers) [31,49]. The cocreative approach (in combination with the BCW planning model) in this study will aid in determining which BCTs might be most appropriate for people with IDs or adapting them if necessary through collaboration.

Aims

This paper aimed to (1) report the systematic development of the Move it, Move ID! intervention by describing how the different steps of the BCW were applied and (2) present the results that emerged from those steps. In doing so, we described the (theoretical) content of the Move it, Move ID! intervention in detail.

Methods

Participants and Recruitment

It was prioritized to focus on young people with IDs as cocreators rather than their parents or teachers because of the historical pattern of marginalization in previous research on intervention development [43]. Through purpose sampling, 2 class groups of adolescents or young adults with mild to moderate IDs aged between 13 and 22 years (ie, age of special needs secondary education in Flanders, Belgium) were recruited to participate in the cocreation sessions. In February 2021 and March 2021, a total of 2 physical education (PE) teachers from different special needs schools in Flanders were contacted to explain the purpose and design of the project via email and phone. They were asked whether they were interested in involving one of their classes in cocreating a PA promotion intervention. Each PE teacher subsequently suggested 1 class group to take part. All adolescents from the selected classes (classes A and B) were invited to participate during the first visit, in which written informed consent from all participants and passive consent from their parents were obtained (Table 1). Class A comprised 14 adolescents aged between 17 and 22 years with a mild to moderate level of ID (mean age 20.33, SD 1.94 years; 3/14, 21% female). Class B comprised 9 adolescents aged between 14 and 15 years with mild IDs (mean age 14.22, SD 0.44 years; 6/9, 67% girls). This aligned with cocreation guidelines, which recommend groups of 10 to 12 cocreators [45,50]. A detailed description of (the recruitment of) participants, as well as the ethical process (next subsection), can be found in the study by Maenhout et al [51].
<table>
<thead>
<tr>
<th>BCW</th>
<th>Researchers’ tasks</th>
<th>Cocreation part with participants with IDs&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1: understand the behavior</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1: define the health problem in behavioral terms</td>
<td>Determine the health problem in behavioral terms using the literature: insufficient PA&lt;sup&gt;b&lt;/sup&gt; in people with IDs</td>
<td>No input was gathered from the participants with IDs in the first 2 steps as we relied on the literature to define the health problem and select the target behavior. Furthermore, the PI&lt;sup&gt;c&lt;/sup&gt; is currently affiliated with the Department of Movement and Sports Sciences (Ghent University), which is why we focused on PA.</td>
</tr>
<tr>
<td>Step 2: select the target behavior</td>
<td>Select the target behavior: increasing PA levels in adolescents and young adults with IDs</td>
<td>No input was gathered from the participants with IDs in the first 2 steps as we relied on the literature to define the health problem and select the target behavior. Furthermore, the PI&lt;sup&gt;c&lt;/sup&gt; is currently affiliated with the Department of Movement and Sports Sciences (Ghent University), which is why we focused on PA.</td>
</tr>
<tr>
<td>Step 3: specify the target behavior and formulate intervention goals</td>
<td>Specify the target behavior by:</td>
<td>Cocreation session 1:</td>
</tr>
<tr>
<td>- Generating a nonexhaustive list of all potential barriers and facilitators that may be relevant to the target behavior</td>
<td>- Introduction session (ie, explanation of the project and its purpose, process of informed consent, and getting to know each other)</td>
<td></td>
</tr>
<tr>
<td>- Describing these barriers and facilitators as what needs to be targeted in the intervention (who needs to do it, what do they need to do differently to achieve change, where and when do they need to do it, and how often and with whom do they need to do it)</td>
<td>- Comapping barriers to and facilitators of PA</td>
<td></td>
</tr>
<tr>
<td>- Formulating 10 intervention goals based on the ranking by the cocreators</td>
<td>- What PAs are they currently performing? What do they like or dislike?</td>
<td></td>
</tr>
<tr>
<td>Step 4: link intervention goals to COM-B&lt;sup&gt;d&lt;/sup&gt; components and TDF&lt;sup&gt;e&lt;/sup&gt; domains</td>
<td>Select the components of the COM-B model and the theoretical domains of the TDF for each intervention goal</td>
<td>Cocreation session 3:</td>
</tr>
<tr>
<td>- Not applied as designers limited to a specific policy lever are directed immediately to step 7&lt;sup&gt;f&lt;/sup&gt; [33]</td>
<td>- Explore the most important barriers and facilitators on which the intervention should focus by voting and ranking them by importance</td>
<td></td>
</tr>
<tr>
<td><strong>Stage 2: identify intervention options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 5: select intervention functions</td>
<td>Select intervention functions using the APEASE&lt;sup&gt;f&lt;/sup&gt; criteria from the BCW guide [33]</td>
<td>No input was gathered from the participants with IDs in this step as their input (from step 3) only needed to be linked to the theoretical components of the COM-B and TDF.</td>
</tr>
<tr>
<td>Step 6: identify policy categories</td>
<td>Not applied as designers limited to a specific policy lever are directed immediately to step 7&lt;sup&gt;f&lt;/sup&gt; [33]</td>
<td>N/A&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Stage 3: identify content and implementation options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 7: identify BCTs&lt;sup&gt;h&lt;/sup&gt;</td>
<td>Choose the most appropriate BCT(s) based on the following:</td>
<td>Cocreation session 4:</td>
</tr>
<tr>
<td>- The BCW guide [33]</td>
<td>- Select BCTs and identify whether selected BCTs would suit the target group or how they can be redesigned to work for them</td>
<td></td>
</tr>
<tr>
<td>- Input from participants with IDs</td>
<td>- APEASE criteria (expert consultation)</td>
<td></td>
</tr>
<tr>
<td>- APEASE criteria (expert consultation)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Table 1. Merging the Behavior Change Wheel (BCW) with a cocreational approach.

<sup>b</sup> PA, physical activity.

<sup>c</sup> PI, principal investigator.

<sup>d</sup> COM-B, competences, opportunities, motivation, behavior.

<sup>e</sup> TDF, Theoretical Domains Framework.

<sup>f</sup> APEASE, affordability, practicality, effectiveness, acceptability, superiority, ease of use.

<sup>g</sup> N/A, not applicable.

<sup>h</sup> BCT, behaviour change technique.
**Ethical Considerations**

All participants and their parents or legal guardians received detailed and accessible information regarding the study design and purpose as well as data handling. To ensure privacy, the data were pseudonymized and only accessible to the researchers or their appointed representatives. Data confidentiality was always maintained. In consultation with the data protection officer of Ghent University (Belgium), the legal basis was changed from “active informed consent” of parents or legal guardians to “public interest,” although this did not exempt researchers from informing participants. This meant that parents or legal guardians needed to contact the researchers only if they disagreed with their child’s participation and, thus, researchers did not require active consent from parents or guardians to commence. The participants with IDs themselves were required to provide their active consent, which is why the first session involved a thorough, step-by-step review of the information and consent process with time for discussion. Young people with IDs participated voluntarily and could withdraw at any time. In appreciation of their participation, all participants received 2 cinema tickets, about which they were informed when they decided to participate. This study received approval from the Ethical Committee of the Faculty of Psychology and Educational Sciences at Ghent University, Belgium (2021/056 LM).

**Combining the BCW Development Process and a Cocreational Approach**

**Overview**

From April 2021 to June 2021, the 8 steps of the BCW were systematically followed for intervention development (Figure 1) [33]. In parallel, input was gathered from adolescents and young adults with IDs through 6 cocreation sessions (Table 1). The entire process was a mix of theoretical underpinnings (ie, COM-B), the domain expertise of the researcher, and the lived experiences of the target group (ie, cocreation sessions). The 6 sessions took place in their classrooms, each during 2 consecutive class hours. For a comprehensive explanation of the cocreation process, the methods used, and the participants’ experiences, we refer interested readers to our previously published paper [51].
Stage 1: Understand the Behavior

In step 1, the literature was reviewed by the principal investigator (PI; LM) to articulate the (health) problem in behavioral terms, this being “insufficient PA in people with IDs.” The next steps were then to select (step 2) and specify (step 3) the target behavior of the intervention, this being “increasing PA levels in adolescents and young adults with IDs,” by defining who needs to do it, what needs to be done differently to achieve change, where and when they need to do it, and how often and with whom they need to do it. This was done by generating a nonexhaustive list of possible barriers to and facilitators of PA for adolescents and young adults with IDs based on both the literature and information gathered in the second cocreation session (Table 1). Owing to the cocreative approach, insights from the literature (brought in by the PI) and input from the target group were intertwined (eg, visual cards of barriers and facilitators were developed by the PI inspired by the literature, which were brought up when the participants themselves could not come up with barriers and facilitators [anymore] [51]). In cocreation session 3, these barriers and facilitators were ranked according to importance by the target group. The most important barriers and facilitators were described as what needed to be targeted in the intervention and, consequently, formulated as the intervention goals. In step 4, these intervention goals were then assigned by the PI to the specific components of the COM-B model and theoretical domains of the TDF. No direct cocreation session was organized within this step as their input (from step 3) only needed to be linked to the theoretical components of the COM-B and TDF. However, this does not deviate from the essence of cocreation as the PI established these connections based on all the input provided by the participants.

Stage 2: Identify Intervention Options

In step 5, the BCW guide links COM-B components and TDF domains to 9 intervention functions [33]. Consequently, the broader research group of the PI (ie, the Physical Activity and Health research group) held expert meetings to decide which intervention functions were most suitable to work with based on the Affordability, Practicability, Effectiveness or Cost-Effectiveness, Acceptability, Side Effects or Safety, and Equity criteria [33]. These criteria are recommended by the BCW guide to make strategic judgments on the most appropriate intervention functions. No direct input from the cocreators was sought in this case, either. However, we approached this step with an open-minded perspective and only removed the intervention functions that were deemed not feasible by the project team. All other intervention functions were retained, allowing the cocreators to continue shaping the direction of development. The sixth step was to consider which policies would support the delivery of the intervention functions identified in step 5 [33]. However, as the researchers within this project did not have access to policy levers, step 6 was not applied. This is also described in the BCW guide by stating that “designers limited to a specific policy lever are directed to step 7 to identify BCTs” [33].

Stage 3: Identify Content and Implementation Options

In step 7, the BCW guide proposes the most appropriate BCTs for each intervention function (selected in step 5) [33]. In each of these, a distinction is made between “BCTs used most frequently and less frequently” [33]. For feasibility reasons, we focused primarily on the most frequently used BCTs during the development process. However, for the fourth cocreation session on BCTs, we also explored the less frequently used BCTs and selected relevant ones based on our expertise with the target group. The aim of this cocreation session was to find out which BCTs were understandable and feasible for adolescents and young adults with IDs and how BCTs could be adapted to meet these criteria. On the basis of the Affordability, Practicability, Effectiveness or Cost-Effectiveness, Acceptability, Side Effects or Safety, and Equity criteria and on input from the cocreators with IDs, a decision was made on which BCTs to include in the
Move it, Move ID! intervention. Finally, the eighth step was to identify the best way to deliver the intervention (ie, mode of delivery). As research has shown that the use of technology (ie, mHealth) is feasible and has high potential in adolescents and young adults with IDs [24,25,52], the target group was asked in the fifth cocreation session about their preferences and barriers to and facilitators of mHealth use.

Analysis

All the cocreation sessions were recorded and transcribed verbatim. A combination of a deductive (ie, 8 steps of the BCW) and inductive (ie, resonating the voice of the participants) analysis approach was applied specifically focusing on identifying and describing the findings from each of the steps of the BCW.

Results

Stage 1: Understand the Behavior

**Step 1: Define the Health Problem in Behavioral Terms**
Few people with IDs are sufficiently physically active [18,19].

**Step 2: Select the Target Behavior**
An increase in the total volume of PA should be targeted rather than aiming to meet the WHO guidelines regarding MVPA as even small positive changes in PA levels are associated with health benefits among people with IDs [53].

**Step 3: Specify the Target Behavior and Formulate Intervention Goals**

Multimedia Appendix 1 [32,54-63] provides an overview of 72 barriers to and 66 facilitators of PA for adolescents and young adults with IDs based on (1) a review of the literature by the PI in preparation for the cocreation sessions and (2) input from cocreators with IDs during these sessions. The appendix is divided into intrapersonal, interpersonal, and contextual factors, reflecting the multifaceted and complex nature of the influences on PA in this population. In the third cocreation session, participants ranked the barriers and facilitators according to their importance, providing guidance on which ones to address in the intervention. The 10 most important barriers (in the opposite direction, these would be facilitators) were identified: (1) the need for social connectedness, (2) the lack of practical support within the PA context, (3) the absence of a role model, (4) the need for others around them who also engage in PA, (5) the lack of confidence in their own abilities and body image, (6) the need for knowledge about the (health) benefits of PA, (7) the lack of knowledge about the different PA options available, (8) the low motivation to engage in PA, (9) the difficulty in setting goals, and (10) the need for help to incorporate PA into their existing schedules (ie, goal conflict) as they often depend on others for this. Evidently, this top list does not mean that the other barriers and facilitators were not relevant for some individuals at particular times, but in view of feasibility, it was decided to prioritize and primarily address those that were identified as the most important.

Previous studies have proposed schools as the ideal setting for PA promotion [8,32,64]. Participants with IDs in this study indicated that they are sufficiently encouraged at school to engage in PA via compulsory PE classes. However, they expressed difficulties in being physically active during leisure time. In the cocreation sessions, they expressed a preference for an intervention during their leisure time (ie, at home or in the community setting) rather than a school-based intervention:

> I think it’s best to go somewhere else. Then you have something separate from school. That you are really away. When you come back to school, that you can start again with a fresh head. [Cocreator 1; cocreation session 3; group A]

> [...] that you just keep your activities outside school and that you don’t keep it here between these four walls. [Cocreator 2; cocreation session 3; group A]

Textbox 1 summarizes the specifics of the target behavior gathered during the first 3 steps: who will perform the behavior; what needs to be done differently; and when, where, how often, and with whom it needs to be done.

Finally, the PI formulated 10 intervention goals targeting the most important barriers chosen by the cocreators (Table 2).
**Textbox 1.** Specify the target behavior (step 3 of the Behavior Change Wheel).

<table>
<thead>
<tr>
<th>Who needs to perform the behavior?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Flemish adolescents and young adults aged between 14 and 22 years with mild to moderate intellectual disabilities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What does the person need to do differently to achieve the desired change?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Address the 10 most important barriers or facilitators (described in Table 2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When do they need to do it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• During leisure time (weekdays+weekends)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Where do they need to do it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• In the community setting or at home</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How often do they need to do it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Not specified</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With whom do they need to do it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Together with someone (at this stage, it was not specified yet who this someone could be, but the need for social connectedness during physical activity did emerge as the main barrier or facilitator in both groups)</td>
</tr>
</tbody>
</table>
Table 2. Linking of intervention goals to Capability, Opportunity, and Motivation–Behavior (COM-B) components and Theoretical Domains Framework (TDF) domains (step 4 of the Behavior Change Wheel).

<table>
<thead>
<tr>
<th>COM-B component and relevant TDF domain</th>
<th>Most important barriers or facilitators</th>
<th>Intervention goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>• Insufficient knowledge about options for PA(^a), where, what suits the person best, and what are the barriers and how to counter them</td>
<td>Adolescents and young adults with IDs(^b) need a better understanding of where, when, and how to engage in PA; they need to be offered a range or variety of PA options they can choose from.</td>
</tr>
<tr>
<td>Behavioral regulation</td>
<td>• Difficulty in setting up PA goals (mostly because of a lack of knowledge about PA options)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Difficulty with planning PA (eg, mostly because of the dependency on others and goal conflict)</td>
<td>Adolescents and young adults with IDs need to be facilitated/supported in formulating specific PA goals.</td>
</tr>
<tr>
<td>Opportunity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social influences</td>
<td>• Lack of social connectedness; having no one to do PA with (eg, friends or loved ones)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Not having a role model (ie, seeing other people engage in PA as well)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No guidance during PA or no practical support</td>
<td>Adolescents and young adults with IDs need to have the opportunity to engage in PA together with someone (ie, social connectedness).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adolescents and young adults with IDs need a role model regarding PA.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adolescents and young adults with IDs need to have more (social and practical) support from others when engaging in PA.</td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentions</td>
<td>• No motivation to engage in PA</td>
<td>Adolescents and young adults with IDs need to be encouraged in feeling a sense of enjoyment when engaging in PA.</td>
</tr>
<tr>
<td>Beliefs about capabilities</td>
<td>• Insecure about own capabilities and skills (eg, afraid of doing something wrong, afraid of the reaction of others, afraid of PA being too difficult, or afraid of being laughed at)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Insecure or ashamed about weight or body shape</td>
<td>Adolescents and young adults with IDs need to build self-confidence regarding PA (ie, increase self-image and confidence in their ability to perform certain activities).</td>
</tr>
<tr>
<td>Beliefs about consequences</td>
<td>• Lack of awareness about the health consequences of physical inactivity</td>
<td>Adolescents and young adults with IDs need a better understanding of the benefits of PA.</td>
</tr>
</tbody>
</table>

\(^{a}\)PA: physical activity.  
\(^{b}\)ID: intellectual disability.

**Step 4: Link Intervention Goals to COM-B Components and TDF Domains**

In step 4, the PI assigned the 10 intervention goals to the specific COM-B and TDF components of the BCW (Table 2). From this behavioral analysis, it can be inferred that psychological capability, social opportunity, and reflective motivation would be the first targets to focus on for increasing PA levels in adolescents and young adults with IDs.

**Stage 2: Identify Intervention Options and Steps 5 and 6 (Identify Intervention Functions and Policy Categories)**

Linking the selected COM-B components and TDF domains from step 4 to the intervention functions proposed by the BCW guide, all 9 intervention functions could be applied. In total, 7 intervention functions were chosen to further focus on: education, persuasion, incentivization, training, environmental restructuring, modeling, and enablement (see the detailed argumentation in Multimedia Appendix 2). The 2 other intervention functions were not selected as they were deemed (1) less practicable to apply as a research team (ie, restriction).
and (2) less acceptable or unlikely to have an impact on adolescents or young adults with IDs (ie, coercion).

Stage 3: Identify Content and Implementation Options

Step 7: Identify BCTs

Overview

A total of 12 BCTs were selected to proceed with. We have outlined our selection and reasoning for each selected BCT within the specific intervention function in the following sections. Multimedia Appendix 3 provides a detailed explanation of all the BCTs that were considered for the 7 intervention functions that came out of step 5, along with the accompanying rationale for why they were chosen and others were not.

Education

Participants with IDs expressed a lack of knowledge about PA options (eg, what is out there, what suits the person best, and where can it be done). For this reason, it was considered valuable to provide information on various PA options. However, the only BCTs formulated within the taxonomy by Michie et al [41] related to providing information are pertaining to consequences (ie, social, emotional, environmental, and health). Although participants mentioned the value of information about the health benefits of PA in previous stages, we collectively decided not to place a direct emphasis on information provision within our intervention. Participants do not desire an intervention centered on “learning” or “teaching” (see also their preference for an intervention outside the school context). According to them, the focus should be on enjoyment. Nonetheless, we anticipate that the target audience may indirectly experience positive effects through the intervention. In this regard, the BCT “salience of consequences” (under the persuasion intervention function) seemed more applicable as it focuses on using methods to specifically emphasize the consequences of performing a behavior, making them more memorable, which goes beyond mere information provision about these consequences. “Feedback on behavior” was also selected as participants indicated that they would like to receive feedback on how well they are performing the behavior.

Persuasion

The BCT “credible source” was valued by participants, but opinions varied on its presentation. Some preferred health professionals using fun visual communication, whereas others liked animated movies. In this target group, experts or influencers explaining PA benefits in an engaging way were considered more appealing than scientific videos. Furthermore, the significance of “verbal persuasion about capability” was strongly emphasized. Given the low self-efficacy within this population, offering verbal persuasion to counteract self-doubt was deemed highly valuable for adolescents/young adults with IDs.

Incentivization

Owing to the prominent role of social factors, we observed that the BCT “social reward” would be highly motivating for this target group.

Environmental Restructuring

Cocreators highlighted the importance of social connectedness and support in encouraging PA. As a result, we expect the greatest impact from recognizing and meeting their social needs, which entails a “restructuring of their social environment.”

Modeling

Participants expressed that it would be motivating to witness others engaging in PA around them, whether in person or indirectly through influencers such as on TikTok (serving as role models). The cocreators showed enthusiasm for involving influencers they admired to encourage PA. Considering budget limitations, it would not be feasible for us to incorporate a well-known influencer into the intervention. However, this does indicate that “demonstration of the behavior” might be an interesting BCT to include.

Enablement

The entire development process highlighted a strong emphasis on the importance of social support and social connections, whether from friends or individuals with expertise in PA (ie, social support BCT—practical, emotional, and unspecified). Participants expressed increased confidence when they could openly discuss their goals and challenges with friends, and their motivation to engage in PA was significantly higher when they could do it with others rather than exercising alone. Peer support was generally preferred, although younger adolescents with IDs (aged 14 years) also mentioned the potential for support from family members. Furthermore, participants agreed that having a list of goals to choose from would make it easier for them rather than having to come up with their own goals (ie, goal setting BCT). Most participants recognized the importance of “action planning” as a valuable BCT. However, insights from teachers revealed that adolescents and young adults with IDs often struggle with tasks such as maintaining a personal agenda or planner, which is typically managed by parents or carers. Therefore, it would be crucial to offer guidance during action planning. Creating a detailed action plan independently, including specifics such as what, when, where, and with whom, seemed challenging and burdensome for this group. Simplicity and minimizing cognitive effort were emphasized as essential factors. Similarly, the collaborative review and adjustment of the behavioral goal with individuals with IDs based on their progress was seen as advantageous. It was considered feasible to engage in close negotiation with them to either retain the same goal, make minor adjustments, or establish a new goal if necessary (referred to as “reviewing behavior goals”). The primary focus in this case is on shared decision-making and active involvement.

Training

The only BCT that we considered including under the “training” intervention function is “demonstration of the behavior.” However, we view this as more related to modeling behavior rather than as actual behavior demonstration within a training context. In the course of our intervention development, it became evident that the primary focus should be on addressing social needs and creating enjoyable experiences rather than on formal training in activities. Therefore, the “training” intervention function was omitted from this phase onward.
Step 8: Identify Mode of Delivery

On the basis of the literature, an mHealth app appeared to be a good and feasible approach for adolescents and young adults with IDs and, therefore, was verified during the fifth cocreation session. The cocreators indicated that they preferred an mHealth app with a straightforward design that clearly indicated its purpose and functionality (eg, through an introductory video). They suggested that the app should be visually appealing, with minimal text, bright colors, and no foreign-language words. They also suggested that a game component or chat feature would be of added value. Cocreators would not use an app that they had to pay for, was childish, or looked rather old-fashioned. They mentioned preferring not to receive too many notifications (ie, no more than 1 notification per day). Finally, this is a group that often faces negative comments and experiences of failure. When talking about mHealth, this also emerged as an aspect to be considered (eg, by keeping the reactions that can be given to each other in an app controlled).

At the end of the fifth cocreation session, cocreators indicated that an app alone would not be sufficient to encourage them to engage in (more) PA. They suggested that an app could be integrated into a broader intervention but not be a stand-alone intervention. More specifically, the desire for social connection with peers and social integration in real life was found to be a more important theme in intervention development. Therefore, the decision was made to focus on a buddy system as many people with IDs reported a lack of friendships with peers outside school, resulting in decreased opportunities to engage in PA during leisure time. To facilitate this buddy partnership, we chose to work with a buddy without IDs who could offer practical support during the intervention period, which reduced the reliance on context alone (ie, parents or carers) to guide the intervention implementation.

Move It, Move ID! Intervention

On the basis of the systematic steps of the BCW combined with a cocreational approach, the Move it, Move ID! intervention ultimately consists of a buddy partnership with a supporting app (ie, dyadic intervention). Figure 2 illustrates the development process, showing how COM-B components, intervention functions, and behavior change techniques (BCTs) are intertwined with the selected intervention design. A more in-depth description can also be found in Multimedia Appendix 3.

Figure 2. Visualization of the development process showing how Capability, Opportunity, and Motivation–Behavior (COM-B) components; intervention functions; and behavior change techniques (BCTs) are intertwined with the selected intervention design. PA: physical activity; TDF: Theoretical Domains Framework.

During an intervention period of 3 months, adolescents and young adults with IDs will be paired with a buddy without IDs of the same age range and encouraged to try out weekly PAs in Ghent (Flanders, Belgium). Buddies without IDs will be students (aged 17-23 years) of the coauthors of this paper and will receive 3 short training sessions (ie, maximum of 1 hour per session) on their role and responsibilities as a buddy.

Although the buddy partnership forms the core of the intervention, a supporting app will also be provided in which buddies and participants with IDs will be in direct contact with each other (an explanation of the scope and screenshots of the supporting app can be found in Multimedia Appendix 4). The app is considered a private space between participants with IDs and their buddies without adding parents or carers to the app. The PI will add a range of activities (eg, walking a shelter dog, dancing, playing Kubb, and undertaking an altitude trail) to the app at the start of the intervention. Participants will have the autonomy to choose whether they want to try an activity by agreeing (swiping right) or disagreeing (swiping left) with a proposed activity. When both the participant with IDs and the
buddy agree with a certain proposed activity, they will receive a pop-up to a chat function to make arrangements and schedule this activity on their shared agenda. The buddy will take the lead in this process. During an activity, the buddy can provide feedback such as how well they perform the behavior or words of encouragement. On the app pinboard, pairs can share photos of the activity they performed together, give comments, and also rate the activity afterward. This allows them to keep track of successful activities and identify less enjoyable ones.

Discussion

Principal Findings

This paper describes the systematic, theory-driven development of a lifestyle intervention to promote PA in adolescents and young adults with IDs using the BCW planning model combined with cocreation sessions involving the target group. The purpose of this transparent and detailed description was 2-fold. First, it aimed to develop a PA promotion intervention by identifying intervention components and BCTs that address the specific needs of this target group. Second, it aimed to encourage future researchers and intervention developers interested in PA among adolescents and young adults with IDs to apply a theoretical planning model in combination with cocreation when designing similar interventions or take the insights described into account in their own intervention development. By transparently describing the theory and BCTs that underpin the intervention, researchers are facilitated in broader evaluations to explore their driving mechanisms. In doing so, we adhered to the Medical Research Council guidelines, which emphasize the importance of theorizing how an intervention works and what works in which setting and identifying its other impacts [65]. This discussion will first delve deeper into the key findings regarding the development of the Move it, Move ID! intervention followed by a reflection on the experience of the development process by combining the BCW and cocreation.

The development process underscored the essential importance of collaborating with the target group as the intervention looks different from what the research team had envisioned in the project proposal (ie, developing an mHealth app). Active collaboration with young people with IDs highlighted the urge for real-life social connectedness and social integration, which makes the social component as part of PA a cornerstone within our intervention development [54-58]. Although the importance of social interaction has emerged in qualitative studies with the target group [66,67], this correlate has surprisingly not been included in studies examining the correlates and determinants of PA levels among young people with IDs [16,68]. However, a study from 2004 conducted within the context of the Special Olympics has already articulated that social support may be particularly crucial for individuals with IDs as they likely have a more limited friendship network compared with individuals without IDs [67]. In total, 3 intervention goals within this development process were consequently directed toward emphasizing the importance of “social opportunity” within the COM-B model. In addition, “psychological capability” and “reflective motivation” emerged as important areas for PA interventions as young people with IDs indicated a lack of knowledge about their PA options, a need for assistance in setting and planning goals, a requirement to enhance their confidence in their own capabilities, and the need to experience genuine enjoyment during PA before they would be motivated to engage in it. Throughout the remainder of the development process, the appropriate intervention functions and BCTs were then linked to these 3 COM-B components. At the end of the process, the cocreators underscored that solely relying on an mHealth app would not meet their social connectedness needs. They preferred face-to-face interaction over distant delivery modes. In addition, they expressed a preference for an intervention targeting their leisure time rather than one connected to their school context. For these reasons, a dyadic intervention was chosen in which young individuals with IDs will be paired with a peer without IDs to explore various PAs together outside the school context. A dyadic intervention refers to an approach or program that involves 2 individuals, typically with a focus on the interaction, relationship, or dynamic between them. Dyadic behavior change has been proven to be a promising approach in previous research [69-71].

As such, by incorporating an extensive and collaborative development process within a project application, one could re-evaluate the initial project proposal (ie, develop an mHealth app for young people with IDs targeted at promoting PA) with a thorough argument that adaptations are necessary from the perspective of the target group itself. In that regard, the combination of actively involving the target audience and applying a clear and scientific planning model was crucial. The most prominent planning models that are currently proposed to guide the development of effective interventions are Intervention Mapping [72] and the BCW [33]. Intervention Mapping includes 6 different steps to rigorously select determinants, performance, and change objectives using appropriate methods and strategies [72]. Although Intervention Mapping is comprehensive, its level of detail makes it more complex and, thus, less feasible, especially in combination with cocreation [73]. The BCW, in contrast, is more open, practical, and flexible as it was developed with interdisciplinary application in mind [37]. However, in applying the BCW within this project, it was noticed that its openness and flexibility could also lead to variable interpretations, with judgments from the researchers often required throughout the development process (eg, step 5). The variations in intervention development mainly depended on the resources available to the project team (eg, affordability and practicability). Moreover, even within this small research team of the Move it, Move ID! intervention, different steps within the BCW were sometimes interpreted differently. Some researchers saw the formulation of barriers to and facilitators of PA as belonging to steps 2 and 3 (as it was described in this paper), whereas others ascribed this to step 4 [74]. In our opinion, assigning these aspects to a certain step will not differ much from the behavioral diagnosis one will eventually arrive at. We consider it more important to discuss the different steps thoroughly within the research team so that the decisions made are well informed and can be argued for. By going through the different steps of the BCW, we learned that interventions can look different depending on the choices made without necessarily making one intervention better than the other. Further research should subsequently indicate which
interventions prove to be effective and why (ie, identifying driving mechanisms [65]). This could potentially lead to the formulation of guidelines outlining the best possible choices that could be made during intervention development within a specific target group and setting.

Nevertheless, by applying the theoretical planning model, the PI had a clear goal in mind in setting up the structure and flow of the cocreation sessions. In doing so, the BCW was instrumental in identifying an informed behavioral diagnosis and choosing which BCTs would be most applicable to have an impact on PA behavior change within this target group and setting. Although the literature suggests that the use of theory in intervention development is key [2,20,33-36], a 2019 meta-analysis formulated that the effectiveness of interventions would be less influenced by whether they are theoretically developed than by the specific BCTs used [75]. In contrast, we believe that both (ie, theory and choice of BCTs) are intertwined. A 2017 systematic review found that lifestyle change interventions for people with IDs aimed at improving PA levels typically used 5.9 BCTs, with “provide information on consequences of behavior in general,” “plan social support/social change,” “provide instruction on how to perform the behavior,” and “goal setting (behavior)” being the most frequently used BCTs [31]. However, 73% of the studies did not use any theoretical framework for intervention development [31]. After completing the full behavioral diagnosis based on the BCW, we included 12 BCTs in our intervention. This is not to say that the inclusion of more BCTs would be better but, rather, that the transparent description of the BCW steps made more evident why these specific BCTs were chosen and how they are intertwined by means of the intervention design. This demonstrates why we believe that the use of theory and the selection of BCTs are strongly connected.

Linking cocreation to the BCW, our goal was to create an intervention that starts with the experiences of the target group. This approach was intended to enhance the effectiveness and sustainability of the intervention by making it more suitable and acceptable for the target audience [10,44-48]. Cocreation with the target audience extended well beyond the described cocreation sessions for intervention development in this project. As the project progressed toward the effect study, ongoing collaboration continued with 2 coresearchers with IDs (ie, inclusive research [48]). These coresearchers maintained regular meetings (every 2 weeks) with the PI (LM) at the Department of Movement and Sports Sciences (Ghent University), actively engaging in various facets of the project. Their responsibilities included assessing prototypes of the app; offering feedback on the training of buddies; testing measurement instruments (comprising questionnaires, interviews, and accelerometers); providing insights into the recruitment strategy; contributing to the development of promotional materials such as flyers, information letters, and informed consent forms; and participating in efforts to enhance the project’s visibility among their peers, classmates, and other stakeholders. This ongoing collaboration with the coresearchers was purposefully designed to ensure the continued accessibility of the project even beyond the initial phase of intervention blueprinting. To conclude, the described intervention development addresses an important and often overlooked population that experiences health disparities and is at higher risk of physical inactivity and related health issues. This study highlights the importance of considering the unique requirements of people with IDs to develop tailored interventions that effectively meet their needs.

Limitations and Strengths

This study has some limitations. First, a wide age range of adolescents and young adults with IDs was included, which might make us question whether this intervention is applicable to both an individual aged 13 years and one aged 22 years. Indeed, younger adolescents with IDs (ie, aged 14 and 15 years) did indicate that they would be open to involvement with parents as buddies within an intervention, whereas this was not the case for young adults (ie, aged 17-22 years old). Choosing a tighter age limit (eg, ages of 13-16 years or 17-22 years) is recommended in future intervention development. Second, of the 23 cocreators, 5 (22%) had a comorbidity with autism spectrum disorder, and 1 (4%) adolescent had attention-deficit/hyperactivity disorder. This is considered a limitation as previous research has found different effects on PA among youth who have IDs and youth who have other developmental disabilities in addition to IDs [8], suggesting that further comparison of PA experiences between these groups is warranted. Within the further intervention development, little weight was given to these comorbidities. In contrast, we can also conclude that their perspective was included from the start of intervention development as they also acted as cocreators and this was not an exclusion criterion. Third, following the prioritization of young people with IDs as cocreators in the initial stages of blueprinting an intervention idea, we were unable to gather input from buddies (peers intended to be matched with the participants with IDs) and consider the broader context of individuals with IDs in the actual development phase of the intervention. This constraint was due to the project’s timeline. In light of this constraint, we recommend that future intervention developers consider including these stakeholders in subsequent phases of intervention development. Their perspectives and insights are invaluable in creating interventions that are comprehensive, inclusive, and truly reflective of the needs and dynamics of the entire participant group. The greatest strength of this study was the fact that a theoretical planning model was used in combination with cocreation to develop a PA promotion intervention for this target group. In this way, it addressed the two main reasons why current interventions often prove to be ineffective: (1) a lack of use of theory and (2) a lack of population-specific research. To the best of our knowledge, this is the first study that describes the collaborative development of a PA promotion intervention for and with adolescents and young adults with IDs. Within the Move it, Move ID! project, the decision was made to work only with participants with mild to moderate IDs; consequently, the findings cannot be extended to the target group of severe or profound IDs. Although future research should focus on the representation of all people with IDs in health research, the fact that a specific group was chosen to truly tailor an intervention to their needs can also be seen as a strength.
Conclusions

The Move it, Move ID! intervention was developed based on the BCW in combination with cocreation. Going through this process was seen as an added value by the research team, which makes it highly recommended to allocate adequate time, budget, and experienced scientific staff for intervention development. By systematically identifying the needs of young people with IDs and linking them to theoretical concepts step by step, cocreators with IDs emphasized the importance of face-to-face interactions and social components in PA promotion interventions. They indicated that relying solely on an mHealth app would not fulfill their social needs. The intervention will consist of a dyadic approach in which young individuals with IDs are paired with a peer without IDs to engage in PAs together, with an app solely providing support within this partnership. The detailed and transparent development process described is a valuable blueprint for practitioners wanting to integrate the BCW and its associated BCTs, in combination with actively involving the target group, into their intervention development for people with IDs.

Acknowledgments

The authors wish to thank the adolescents or young adults and teachers who expressed their opinions during intervention development for this project. This work was supported by the Research Foundation–Flanders under grant 11F3621N (2020-2024).

Data Availability

The data sets generated and analyzed during this study are not publicly available to protect participants’ privacy and confidentiality because of the small number of participants but are available from the corresponding author upon reasonable request. In addition, the upcoming effect study has been preregistered at the Open Science Framework, where all other materials concerning this project can be found.

Authors’ Contributions

LM, SC, G Cardon, G Crombez, and GVH conceptualized the study. LM collected the data (ie, organized the cocreation sessions) and wrote the original draft. JL assisted in combining the Behavior Change Wheel planning model and a cocreational approach. SC, JL, G Cardon, G Crombez, and GVH edited the manuscript and provided feedback. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Barriers to and facilitators of physical activity for adolescents or young adults with intellectual disabilities.

[PDF File (Adobe PDF File), 170 KB - formative_v8i1e51693_app1.pdf ]

Multimedia Appendix 2

Selection of intervention functions using the Affordability, Practicality, Effectiveness and Cost-Effectiveness, Acceptability, Side Effects or Safety, and Equity criteria.

[PDF File (Adobe PDF File), 201 KB - formative_v8i1e51693_app2.pdf ]

Multimedia Appendix 3

Selection and reasoning for each selected and nonselected behavior change technique.

[PDF File (Adobe PDF File), 194 KB - formative_v8i1e51693_app3.pdf ]

Multimedia Appendix 4

Scope of the Move it, Move ID! app.

[PDF File (Adobe PDF File), 525 KB - formative_v8i1e51693_app4.pdf ]

References


Abbreviations

BCT: behavior change technique
BCW: Behavior Change Wheel
COM-B: Capability, Opportunity, and Motivation–Behavior
ID: intellectual disability
mHealth: mobile health
MVPA: moderate to vigorous physical activity
PA: physical activity
PE: physical education
PI: principal investigator
TDF: Theoretical Domains Framework
WHO: World Health Organization
Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study

Stephen Bonett¹, RN, PhD; Willey Lin¹, MB; Patrina Sexton Topper¹, BSN, MS, PhD; James Wolfe¹, MS; Jesse Golinkoff², MPH; Aayushi Deshpande², MPhil; Antonia Villarruel¹, RN, PhD; José Bauermeister¹, MPH, PhD

¹School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
²Department of Psychology, Ashoka University, Sonepat, India

Abstract

Background: Web-based surveys increase access to study participation and improve opportunities to reach diverse populations. However, web-based surveys are vulnerable to data quality threats, including fraudulent entries from automated bots and duplicative submissions. Widely used proprietary tools to identify fraud offer little transparency about the methods used, effectiveness, or representativeness of resulting data sets. Robust, reproducible, and context-specific methods of accurately detecting fraudulent responses are needed to ensure integrity and maximize the value of web-based survey research.

Objective: This study aims to describe a multilayered fraud detection system implemented in a large web-based survey about COVID-19 attitudes, beliefs, and behaviors; examine the agreement between this fraud detection system and a proprietary fraud detection system; and compare the resulting study samples from each of the 2 fraud detection methods.

Methods: The PhillyCEAL Common Survey is a cross-sectional web-based survey that remotely enrolled residents ages 13 years and older to assess how the COVID-19 pandemic impacted individuals, neighborhoods, and communities in Philadelphia, Pennsylvania. Two fraud detection methods are described and compared: (1) a multilayer fraud detection strategy developed by the research team that combined automated validation of response data and real-time verification of study entries by study personnel and (2) the proprietary fraud detection system used by the Qualtrics (Qualtrics) survey platform. Descriptive statistics were computed for the full sample and for responses classified as valid by 2 different fraud detection methods, and classification tables were created to assess agreement between the methods. The impact of fraud detection methods on the distribution of vaccine confidence by racial or ethnic group was assessed.

Results: Of 7950 completed surveys, our multilayer fraud detection system identified 3228 (40.60%) cases as valid, while the Qualtrics fraud detection system identified 4389 (55.21%) cases as valid. The 2 methods showed only “fair” or “minimal” agreement in their classifications (κ=0.25; 95% CI 0.23-0.27). The choice of fraud detection method impacted the distribution of vaccine confidence by racial or ethnic group.

Conclusions: The selection of a fraud detection method can affect the study’s sample composition. The findings of this study, while not conclusive, suggest that a multilayered approach to fraud detection that includes conservative use of automated fraud detection and integration of human review of entries tailored to the study’s specific context and its participants may be warranted for future survey research.

(JMIR Form Res 2024;8:e47091) doi:10.2196/47091

KEYWORDS
web-based survey; data quality; fraud; survey methodology; COVID-19; survey; fraud detection; Philadelphia; data privacy; data protection; privacy; security; data; information security; data validation; cross-sectional; web-based
Introduction

Web-based survey research has become increasingly common in recent years, particularly because of its ability to reach broad populations efficiently and economically [1]. Web-based surveys involve inviting potential respondents to complete questionnaires through digital platforms that manage how questions are presented and how data are collected and stored [2,3]. These research methods have been used in response to the difficulties faced in traditional survey methods (i.e., recruiting participants using flyers, newspaper or radio or television advertisements or spreading by word of mouth and collecting data in person using computer-assisted survey instruments or over the phone), especially in reaching underrepresented populations [4,5]. The advantages of web-based surveys include eliminating the requirement for face-to-face interaction, offering flexible access to surveys, removing transportation and logistical barriers, and preserving anonymity. In recent years, COVID-19 pandemic restrictions limited opportunities for in-person research and provided additional justification for researchers to adopt web-based study designs while leveraging social media recruitment methods to reach diverse populations [6-9].

As web-based recruitment and survey methods in health research have become more ubiquitous and refined, so too have methods of web-based research fraud [10,11]. Fraud can manifest in multiple ways. For example, individuals may misrepresent themselves in order to appear eligible for a study or may submit duplicate surveys in order to receive multiple incentive payments. Additionally, fraudulent data may also come from automated operations enacting fraud at a large scale, often referred to as “bots” [11,12]. These methods are often used to target surveys offering participation compensation payments and can be lucrative when aimed at large web-based surveys, even those offering small payments [9,13]. Such fraud poses risks not only to research resources but also, importantly, to the integrity of research findings, as fraudulent data can distort results and undermine data quality. Specifically, fraudulent responses can introduce additional random noise or potentially add systematic bias to the data [14-16].

In response, researchers, companies operating in the digital research space (e.g., Qualtrics) [17], and organizations interested in digital data integrity (e.g., Google) [18] have developed methods to address fraudulent activity. The research community has crafted recommendations for fraudulent data identification and participant identity verification protocols [13,19,20]. Platforms specializing in web-based survey research such as Qualtrics [17] and Amazon Mechanical Turk [21] have also developed fraud detection features that accompany their services. While these proprietary systems for fraud detection offer a simple, automated approach to improving data quality, little information is available about the mechanisms they use [22]. Fraud detection systems often obscure details about how their validation process functions as an important strategy to protect the integrity of the fraud detection system, making it more difficult for fraudulent participants to circumvent protections. However, obfuscation also introduces questions about how fraud detection algorithms alter study samples and whether they introduce bias into analyses [20].

Little research has compared how fraud detection strategies impact study sample composition or examined their comparative effectiveness in correctly identifying fraud [22-25]. By accurately identifying and removing fraudulent responses to web-based surveys, research can improve data quality and strengthen the overall rigor of their methods. Robust, reproducible, and context-specific methods of accurately detecting fraudulent responses are needed to ensure integrity and maximize the value of web-based survey research. This paper aims to (1) describe the multilayer fraud detection techniques we developed and implemented in a large web-based survey collecting data about attitudes, beliefs, and behaviors related to COVID-19; (2) examine the degree of agreement between our multilayer fraud detection strategy and the proprietary fraud detection system used by Qualtrics; and (3) compare the study samples that resulted when using each of the 2 fraud detection methods.

Methods

Study Design

We collected data from November 2021 through February 2022 for the PhillyCEAL Common Survey, a cross-sectional study using a web-based survey to assess how the COVID-19 pandemic and response have impacted individuals, neighborhoods, and communities across the city of Philadelphia, Pennsylvania. The Checklist for Reporting Results on Internet E-Surveys was used to guide the reporting of our methods and results (Multimedia Appendix 1) [26]. The Qualtrics web-based survey platform was used to design the survey and automatically capture responses in a database. The usability and technical functionality of the survey were tested by the study team before launching the survey. Individuals were eligible to participate if they (1) resided within Philadelphia County (coterminous with the city limits) and (2) were at least 13 years of age. We recruited participants through advertisements on social media platforms (i.e., Facebook, Instagram, Twitter, and Reddit) and referrals from community partners (including one partner that provided study recruitment materials to individuals via door-to-door canvassing). The recruitment process directed individuals to a voluntary, open web-based survey, where they completed a screener consent form and answered a series of questions to determine eligibility and record basic demographic information. We did not allow participants to change their answers through a back button feature.

Eligible participants were automatically directed to the full study consent form. Following consent, participants were asked to complete a 20-minute questionnaire about their experiences, behaviors, and beliefs about COVID-19, risk of infection, testing, vaccination, treatment, and knowledge and beliefs about COVID-19 clinical trials. Participants completed 1 of 3 slight variations of the survey (i.e., the adult survey, the parent survey, and the youth survey), where additional questions or slight changes to wording were used on the parent and youth surveys. Participants completed only 1 of these 3 potential variations (i.e., survey groups were mutually exclusive), with participants of any age who reported having minor children completing the parent survey (103 questions), participants ages 25 years or
younger and not having minor children completing the youth survey (126 questions), and all other participants completing the adult survey (92 questions). To reduce participant burden, we used adaptive questioning to reduce the number of questions displayed based on their answers. Participants saw an average of 6 questions per page.

Participants were excluded if they did not complete the entire survey (n=2930) or did not provide a residential zip code matching 1 of the 48 zip codes of Philadelphia County (n=647). Participants confirmed eligible and marked as complete by both Qualtrics automation and our manual review of the data were compensated with a US $15 electronic gift card. To protect participant data, the web survey data were downloaded to a secure university server, deidentified by replacing contact information with unique IDs, and stored in a restricted folder with password protection.

Ethical Considerations

All participants completed an informed consent process before proceeding to the survey. Ethics approval was obtained from the institutional review board at the University of Pennsylvania (protocol 848650).

Preliminary Fraud Protection

Given the prevalence of fraud and duplicate responses in web-based survey research, we used several strategies to prevent fraudulent participants from accessing and completing the survey, serving as a baseline level of fraud protection for both our multilayer and the Qualtrics detection methods. Our preliminary line of defense against fraud was targeted toward nonhuman interferences such asbots. To proceed with the survey, all respondents had to pass a built-in Turing test provided by Qualtrics using Google’s reCAPTCHA (version 2) antifraud technology [27]. Depending on the respondent’s on-device saved data such as browser cookies, they either had to click a checkbox or solve a simple image challenge to pass the reCAPTCHA (version 2) test.

Since sophisticated bots can trick the reCAPTCHA (version 2) test [28], we added a honeypot question as a second line of defense against bots. Honeypots are survey questions hidden from rendering on the screen using custom JavaScript code [11]. They are, therefore, invisible to human respondents but accessible to bots that do not rely on what is rendered on screen. Since the honeypot is not visible to human respondents, any responses to the honeypot would immediately disqualify the entire survey response and end the survey.

We also created unique URLs for each recruitment source and advertisement campaign for the study. The unique URLs enabled us to identify the origin (recruitment source and ad campaign) of each survey response. This allowed us to individually monitor and suspend links that became the target of fraudulent survey responses.

Multilayer Fraud Detection Methods

Real-Time Exclusion of Fraudulent Responses

We implemented a suite near real-time data verification procedures to assess the veracity of data as they were collected, including techniques based on recommendations from prior research as well as several manual checks that were developed specifically for this study. A research team member would individually inspect submitted responses and label responses as fraudulent using the following criteria: (1) participants were asked to provide their residential address and the colloquial name of their neighborhood. Responses were marked as fraud if the neighborhood name provided did not match a standard Philadelphia neighborhood name corresponding to the residential address provided or any adjacent neighborhood; (2) the residential address provided did not match an existing address in Philadelphia County [20,23,29]; (3) the survey had the same start times and stop times plus or minus 1 minute as 2 or more other submitted surveys (rapid survey submission) [11,29]; (4) the respondent’s email address matched a previously enrolled participant’s email address [11,20]; (5) the zip code provided as part of the residential address was nonstandard (ie, a post office box code or a unique code) [20,23]; (6) the residential address had already been reported by at least 2 other respondents [20,23]; or (7) the URL from which the response was referred did not match any of the URL links distributed by the study team during recruitment. Responses labeled as fraud during real-time validation were not eligible to receive survey compensation.

Automated Post Hoc Identification of Fraudulent Responses

We developed a set of automated post hoc techniques designed to detect fraud that our real-time procedures may not have captured. Three criteria were developed for this post hoc fraud identification based on recommendations from prior research [11,20,22,23,29-32]. Since these criteria only identify suspicious entries and do not definitively prove that an entry is fraudulent, responses were labeled as fraud only if they satisfied 2 or more criteria. We settled on using a threshold of 2 criteria (rather than 1 or 3) in order to balance concerns about the potential of each of our 3 criteria to incorrectly label a participant as fraud with the necessity to exclude causes that showed strong evidence of fraud. The criteria were as follows:

1. In response to a free text item at the end of the survey soliciting additional comments or questions from the participant, the submitted survey included text that was identical to text submitted by other respondents. We considered a free text response an identical match if it was among free text entries of 1 word or greater that were repeated 100 or more times, free text entries of 2 words or greater that were repeated 10 or more times, or free text entries of 3 words or greater that were repeated 3 or more times (see Table S1 in Multimedia Appendix 2 for list of unique text strings excluded and their frequency in the full set of responses).
2. The IP address of a response belonged to a virtual private network or data center originated from outside the United States, as determined by using a security service for proxy and virtual private network detection and IP location information [33]; and
3. Responses provided in the main survey were inconsistent with responses to the same items in the screener for one or more key items that would not be expected to be variable
(ie, age, zip code, number of adults living at home with the participant, number of minors living at home with the participant, Hispanic or Latinx ethnicity, “Have you ever been tested for COVID-19?,” and “Have you received at least 1 dose of the COVID-19 vaccine?”).

Qualtrics Fraud Detection Methods

Qualtrics is a widely used web-based survey platform that allows users to create surveys with complex flow logic and customizable visual design. Qualtrics surveys are easily optimized for use on mobile devices and can display a wide variety of question types on both computer and mobile phone interfaces. Another key strength of the Qualtrics platform is its integration of 1-click translation, allowing users to quickly switch between various languages. This was crucial for our study, which recruited participants from diverse populations across Philadelphia and was available in English, Spanish, and Mandarin. In addition to these valuable features, Qualtrics also offers tools for detecting fraudulent survey responses. This automated and user-friendly system for fraud detection has the potential to help researchers improve data quality in their web-based surveys. Given the lack of research exploring how these consumer tools compare to existing published protocols for fraud detection, we sought to compare our multilayer fraud detection methods to the system used by Qualtrics.

The Qualtrics fraud detection system relies on Google’s reCAPTCHA (version 3) and Imperium’s RelevantID antifraud technologies. Both tools rely on proprietary machine learning models that analyze passive and behavioral data, browser interactions, and respondent metadata to identify abuse and fraud [18,34,35]. Unlike the reCAPTCHA (version 2) test respondents had to solve at the start of the survey, but detection using reCAPTCHA (version 3) does not present respondents with an image challenge nor block respondents and bots from proceeding with the survey. Instead, it returns a score (Q_RecaptchaScore) between 0.0 and 1.0 that Qualtrics records as part of the survey response. We used the recommended 0.5 score as the threshold for fraud, where a score under 0.5 is deemed likely to be a bot [17,18].

Like reCAPTCHA (version 3), RelevantID does not prevent bots from completing the survey. Instead, it attaches a score (Q_RelevantIDFraudScore) between 0 and 130 to each survey response. We followed Qualtrics’ recommendation in interpreting a score ≥30 as fraudulent and likely a bot [17]. In addition to bot detection, RelevantID identifies duplicate responses through digital fingerprinting and proprietary detection algorithms [34]. Qualtrics then attaches another score (Q_RelevantIDDuplicateScore) between 0 and 100 to the survey response. We followed the suggested score threshold where any score ≥75 is considered a duplicate [17].

Statistical Analysis

Agreement and Comparative Performance

The classification tables were created to display the degree of agreement between the 2 fraud detection methods for the full sample and for each of the 3 survey-type categories (ie, adult, parent, and youth).

Impact of Fraud Detection Method on Sample Characteristics

Descriptive statistics were computed for the full sample of responses, the subset classified as valid by our multilayer fraud detection method, and the subset classified as valid by the Qualtrics fraud detection method. As these 3 sets of responses are not mutually exclusive, we did not directly compare them statistically.

To test for differences between fraudulent and valid responses as classified by each fraud detection method, statistical comparisons were conducted for key study variables between the mutually exclusive sets of responses classified as fraudulent or valid within each method. Specifically, we used chi-square tests for categorical variables, 2-tailed t tests for normally distributed continuous variables, and Mann-Whitney U tests for continuous variables that were not normally distributed. The results of these analyses are presented in Table S2 in Multimedia Appendix 2 for the multilayer method and Table S3 in Multimedia Appendix 2 for Qualtrics.

To assess the degree to which the 2 fraud detection methods would impact the distribution of a key study variable, the point estimate and 95% CI were calculated for vaccine confidence by racial or ethnic group for each fraud detection method and for the entire sample without any fraud mitigation.

Variations in Survey Responses During Study Period

A time-series plot was created to show the cumulative responses to the study survey over time and their fraud classification by each of the 2 fraud detection methods. This plot highlights the periods in which social media recruitment campaigns are active and can also shed light on how the 2 fraud detection methods diverge in their classification of responses during different periods of high survey response. Additionally, we present a time-series plot showing the proportion of responses classified as fraud across the study period, including smooth locally weighted smoothing lines to visualize the trends over time. All analyses were performed with R (version 4.1.0; R Foundation for Statistical Computing).

Results

Multilayer Fraud Detection Methods

A total of 7950 completed survey responses were received. See Figures 1 and 2 for an overview of fraud detection results from our multilayer fraud detection methods. Using the real-time exclusion criteria of the multilayer fraud detection method, 4207 (52.92%) entries were classified as fraud. Of those classified as fraud, 1242 (29.52%) reported a neighborhood name that did not match their residential address, 648 (15.4%) provided an invalid residential address, 1397 (33.21%) displayed rapid survey submission, 42 (1%) used a repeated email address, 77 (1.83%) reported a nonstandard zip code, 398 (9.46%) reported a residential address that was used more than twice, and 403 (9.58%) did not have a valid recruitment URL. After the real-time exclusion, 3743 (47.08%) cases remained classified as valid.
Our automated post hoc fraud detection criteria identified additional cases as fraud. Of the remaining 3743 initially valid cases, 1561 (41.70%) cases had a duplicate response in the free text entry item, 394 (10.53%) cases had an IP address from outside the United States or from a virtual private network, and 619 (16.54%) had inconsistencies between the screener and main survey on at least 1 key item. Using our “2-strike” rule, we classified an additional 515 (13.76%) responses as fraud for meeting at least 2 of the above criteria. Thus, our multilayer fraud detection strategy classified a total of 4722 (59.40%) entries as fraud and 3228 (40.60%) entries as valid.

### Qualtrics Fraud Detection Methods

The Qualtrics fraud detection methods identified 498 (6.26%) cases as fraud by the RelevantID FraudScore, and 938 (11.80%) cases as duplicates by the RelevantID DuplicateScore. The Qualtrics fraud detection strategy classified a total of 3561 (44.79%) entries as fraud (ie, meeting one or more of the 3 criteria above) and 4389 (55.21%) entries as valid.

### Agreement and Comparative Performance

Table 1 presents confusion matrices showing the degree of agreement between our multilayer fraud detection method and the Qualtrics fraud detection method for the full sample and each survey-type category. The interrater reliability indicated “fair” or “minimal” agreement between the 2 methods for the full sample ($\kappa=0.25$; 95% CI 0.23-0.27), “moderate” or “weak” agreement for the adult ($\kappa=0.48$; 95% CI 0.43-0.53) and youth ($\kappa=0.50$; 95% CI 0.43-0.58) surveys, and “slight” or “none” agreement for the parent survey ($\kappa=0.13$; 95% CI 0.10-0.15) [36,37].
Table 1. Confusion matrix and interrater reliability ($\kappa$) between our multilayer fraud detection system and the Qualtrics fraud detection system for the full sample, only adult surveys, only parent surveys, and only youth surveys.

<table>
<thead>
<tr>
<th></th>
<th>Fraud (multilayer)</th>
<th>Valid (multilayer)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full sample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>2627$^b$</td>
<td>934$^c$</td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>2095$^c$</td>
<td>2294</td>
</tr>
<tr>
<td><strong>Adult survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>299</td>
<td>166$^c$</td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>174$^c$</td>
<td>904</td>
</tr>
<tr>
<td><strong>Parent survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>2184</td>
<td>710$^c$</td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>1848$^c$</td>
<td>1102</td>
</tr>
<tr>
<td><strong>Youth survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>144</td>
<td>58$^c$</td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>73$^c$</td>
<td>288</td>
</tr>
</tbody>
</table>

$^a$$\kappa$=0.25; 95% CI 0.23-0.27.

$^b$Values in italics represent agreement between the 2 methods.

$^c$Values represent disagreement between the 2 methods.

$^d$$\kappa$=0.48; 95% CI 0.43-0.53.

$^e$$\kappa$=0.13; 95% CI 0.10-0.15.

$^f$$\kappa$=0.50; 95% CI 0.43-0.58.

We conducted sensitivity analyses to assess the impact of choosing a “2-strike rule” for our post hoc fraud detection rather than a “1-strike rule” or a “3-strike rule.” Compared to the “2-strike rule,” which resulted in 515 additional cases being classified as fraud during the post hoc phase of fraud detection, the “1-strike rule” would have classified 2047 additional cases as fraud, and the “3-strike rule” would have classified 12 additional cases as fraud. In terms of agreement with Qualtrics’ fraud detection methods, the “1-strike rule” would have resulted in a $\kappa$ of 0.20 (95% CI 0.19-0.22) for the full sample, and the “3-strike rule” would have resulted in a $\kappa$ of 0.24 (95% CI 0.22-0.26) for the full sample.

Additionally, we explored how the 2 fraud detection strategies compared in their ability to classify cases with validated email addresses as valid entries. Validated email addresses were defined as email addresses ending in “.edu” or “.gov,” indicating an institutional affiliation. Of the 168 cases with validated emails, the multilayer fraud detection system correctly classified 166 (98.81%) as valid, while the Qualtrics fraud detection system correctly classified only 126 (75%) as valid.

**Impact of Fraud Detection Method on Sample Characteristics**

Decisions about which fraud detection strategies to use can impact the results of web-based survey research. Table 2 presents the descriptive statistics for sociodemographic variables, survey metric variables, and key study outcome variables on 3 versions of the data set: the full data set with no fraud detection (n=7950), the cases identified as valid by our multilayer fraud detection methods (n=3228), and the cases identified as valid by the Qualtrics fraud detection methods (n=4389). As these sets are not mutually exclusive, we cannot compare them directly; however, there are clear differences in the distributions of many study variables between the 3 sets. When comparing entries classified as fraud to those classified as valid for each of the 2 fraud detection methods (ie, mutually exclusive sets), all study variables, except for lifetime COVID-19 testing for the multilayer fraud detection, were found to be significantly different for both methods (Tables S2 and S3 in Multimedia Appendix 2).

Table 3 showcases in detail how a key variable of interest to researchers may be affected by using different fraud detection methods. In this data set, vaccine confidence among White respondents was greater when using our multilayer fraud detection ($\mu$=0.867; 95% CI 0.851-0.882) when compared to Qualtrics fraud detection ($\mu$=0.782; 95% CI 0.766-0.798). A similar pattern is seen for Hispanic or Latinx respondents and Black or African American respondents.
Table 2. Demographics, survey metrics, and key study responses in overall sample, multilayer valid set, and Qualtrics valid set.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Full sample (N=7950)</th>
<th>Multilayer valid set (n=3228)</th>
<th>Qualtrics valid set (n=4389)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>35.54 (9.70)</td>
<td>38.09 (12.15)</td>
<td>37.01 (10.81)</td>
</tr>
<tr>
<td>Race or ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>1188 (14.9)</td>
<td>254 (7.9)</td>
<td>571 (13)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>135 (1.7)</td>
<td>7 (0.2)</td>
<td>34 (0.8)</td>
</tr>
<tr>
<td>Asian</td>
<td>311 (3.9)</td>
<td>219 (6.8)</td>
<td>221 (5)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1856 (23.3)</td>
<td>728 (22.6)</td>
<td>853 (19.4)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>42 (0.5)</td>
<td>11 (0.3)</td>
<td>10 (0.2)</td>
</tr>
<tr>
<td>White</td>
<td>4272 (53.7)</td>
<td>1889 (58.5)</td>
<td>2600 (59.2)</td>
</tr>
<tr>
<td>Multiracial or others</td>
<td>146 (1.8)</td>
<td>120 (3.7)</td>
<td>100 (2.3)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>4253 (53.5)</td>
<td>2028 (62.8)</td>
<td>2645 (60.3)</td>
</tr>
<tr>
<td>Man</td>
<td>3571 (44.9)</td>
<td>1108 (34.3)</td>
<td>1663 (37.9)</td>
</tr>
<tr>
<td>Transgender or gender diverse</td>
<td>105 (1.3)</td>
<td>76 (2.4)</td>
<td>64 (1.5)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>21 (0.3)</td>
<td>16 (0.5)</td>
<td>17 (0.4)</td>
</tr>
<tr>
<td>Sexual orientation, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bisexual</td>
<td>361 (4.5)</td>
<td>242 (7.5)</td>
<td>262 (6)</td>
</tr>
<tr>
<td>Gay</td>
<td>231 (2.9)</td>
<td>101 (3.1)</td>
<td>147 (3.3)</td>
</tr>
<tr>
<td>Lesbian</td>
<td>142 (1.8)</td>
<td>69 (2.1)</td>
<td>65 (1.5)</td>
</tr>
<tr>
<td>Straight (i.e., not gay, lesbian, or bisexual)</td>
<td>7039 (88.5)</td>
<td>2682 (83.1)</td>
<td>3791 (86.4)</td>
</tr>
<tr>
<td>Others</td>
<td>94 (1.2)</td>
<td>84 (2.6)</td>
<td>71 (1.6)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>83 (1)</td>
<td>50 (1.5)</td>
<td>53 (1.2)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>215 (2.7)</td>
<td>49 (1.5)</td>
<td>65 (1.5)</td>
</tr>
<tr>
<td>High school or equivalent</td>
<td>1013 (12.7)</td>
<td>299 (9.3)</td>
<td>479 (10.9)</td>
</tr>
<tr>
<td>Some college</td>
<td>1890 (23.8)</td>
<td>579 (17.9)</td>
<td>964 (22)</td>
</tr>
<tr>
<td>College graduate</td>
<td>3935 (49.5)</td>
<td>1672 (51.8)</td>
<td>2253 (51.3)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>882 (11.1)</td>
<td>620 (19.2)</td>
<td>617 (14.1)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>15 (0.2)</td>
<td>9 (0.3)</td>
<td>11 (0.3)</td>
</tr>
<tr>
<td>Survey type, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>1543 (19.4)</td>
<td>1070 (33.1)</td>
<td>1078 (24.6)</td>
</tr>
<tr>
<td>Parent</td>
<td>5844 (73.5)</td>
<td>1812 (56.1)</td>
<td>2950 (67.2)</td>
</tr>
<tr>
<td>Youth</td>
<td>563 (7.1)</td>
<td>346 (10.7)</td>
<td>361 (8.2)</td>
</tr>
<tr>
<td>Survey metrics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey duration (minutes), median (IQR)</td>
<td>23.46 (18.38-38.10)</td>
<td>22.02 (18.13-32.57)</td>
<td>22.82 (18.52-35.13)</td>
</tr>
<tr>
<td>User language=Spanish, n (%)</td>
<td>127 (1.6)</td>
<td>22 (0.7)</td>
<td>53 (1.2)</td>
</tr>
<tr>
<td>Key study variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever tested for COVID-19=yes, n (%)</td>
<td>6968 (87.6)</td>
<td>2840 (88)</td>
<td>3903 (88.9)</td>
</tr>
<tr>
<td>Ever COVID-19–positive, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>5836 (83)</td>
<td>2498 (87.3)</td>
<td>3294 (83.8)</td>
</tr>
</tbody>
</table>
Table 3. COVID-19 vaccine confidence (somewhat confident or very confident) grouped by race compared across the 2 fraud detection methods.

<table>
<thead>
<tr>
<th>Race or ethnicity</th>
<th>Multilayer valid set (n=3228)</th>
<th>Qualtrics valid set (n=4389)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>254 (7.9)</td>
<td>0.87 (0.34)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>7 (0.2)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Asian</td>
<td>219 (6.8)</td>
<td>0.92 (0.28)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>728 (22.6)</td>
<td>0.80 (0.40)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>11 (0.3)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>White</td>
<td>1889 (58.5)</td>
<td>0.87 (0.34)</td>
</tr>
<tr>
<td>Multiracial or others</td>
<td>120 (3.7)</td>
<td>0.76 (0.43)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Variations in Survey Responses During Study Period

The response rate varied throughout the study and was correlated to several social media advertising campaigns and an extended holiday break where no responses were accepted. Figure 3 shows the cumulative number of responses over time, differentiated by fraud detection method and fraud classification. Time periods when social media advertising campaigns were active are highlighted on these plots. Responses tend to increase during social media campaigns. Notably, between January 13 and February 7, 2022, no social media campaign was active, yet a significant number of responses were received (n=766). These responses were largely classified as fraud by our multilayer fraud detection method (n=716, 93.47% classified as fraud) but were often classified as valid by the Qualtrics fraud detection method (n=296, 38.64% classified as fraud). Figure 4 shows the proportion of responses classified as fraud across the study period. The comparative fraud detection between the 2 methods was similar during the first half of the study period (November to December 2021), while the multilayer fraud detection method consistently identified a higher proportion of responses as fraud during the later portion of the study period (January to February 2022).
Figure 3. Fraud detection by multilayer fraud detection system and Qualtrics fraud detection system during the study period (November 2021 to February 2022). Highlighted regions indicate periods when social media campaigns were active (November 3-18, 2021; November 29-December 22, 2021; and February 7-10, 2022). Data collection was paused during an extended winter break from December 23, 2021, until January 12, 2022. The dotted vertical line represents January 12, 2022, when data collection was resumed.

Figure 4. Proportion of responses classified as fraud by multilayer fraud detection system and Qualtrics fraud detection system during the study period (November 2021 to February 2022). Smooth locally weighted smoothing lines are included to help visualize the trends over time.

Discussion

Principal Findings

Our multilayer fraud detection methods identified a substantial number of fraudulent cases. However, when comparing our fraud detection methods to proprietary fraud detection systems provided by web-based survey software, we saw low levels of agreement between the 2 methods. Our results highlight how the choice of fraud detection method can alter the distribution of key study variables.

Both our multilayer fraud detection methods and the Qualtrics fraud detection system identified significant levels of fraud; however, the 2 methods differed in which cases they identified as fraudulent and in which they identified as valid. Notably, these differences were most pronounced for participants who administered the survey specifically for parents. It is possible that fraudulent participants made assumptions about eligibility or compensation (e.g., parents are a more specific demographic that may be of interest to researchers, and thus, fraudulent entries that claim to be parents may be more likely to screen as eligible and receive compensation) and responded to eligibility questions in ways that guided them to the parent survey. With a greater
number, and perhaps a greater variety, of fraudulent participants, we then may have seen greater variability in the 2 methods’ ability to consistently identify the fraud.

The disagreement between our multilayer fraud detection and the Qualtrics fraud detection suggests that there are important differences in the methodologies being used by the 2 systems, which resulted in differences in classification. To fully understand and compare the relative performance of the 2 systems, detailed information is needed about the methods used by the Qualtrics fraud detection system. It is likely that the features of the RelevantID proprietary fraud detection method used by Qualtrics are intentionally obfuscated to prevent fraudulent participants from undermining its effectiveness. The trade-off for this black box tactic is that researchers who use the Qualtrics platform cannot ascertain how Qualtrics’ fraud detection algorithms function and how these methods compare to alternative fraud detection strategies. There is an inherent tension between transparency (ie, publishing the features of a fraud detection method improves scientific rigor) and defending against fraud (ie, making comprehensive information about a fraud detection method available may enable fraudulent participants to avoid detection) [22]. Additional research is needed to evaluate the effectiveness of proprietary fraud detection systems and compare them to published techniques used by researchers.

For both fraud detection methods, the cases identified as fraud differed in nonrandom ways from the cases classified as valid on key study variables. These differences could have implications for the interpretation of study results; if legitimate survey participants are classified as fraudulent, critical data are lost, and potential bias could be introduced. In addition, many automated fraud detection tools turn to proprietary machine learning data and predictive modeling for fraud detection [22,38]. This could disproportionately affect those with low literacy or barriers to internet access, as fraud detection tools may be more likely to flag them as fraudulent [19]. Given the existing digital divide across racial or ethnic groups [39,40], this may result in the further exclusion of racial or ethnic minorities in research if they are more likely to use older technologies and shared devices at home or in public spaces, such as computers at libraries or community centers (which may trigger the threshold for fraud due to user behavior that is atypical of a single-user device). One approach to overcoming this challenge is to integrate manual inspection of survey entries in place of or in addition to automatic processes that could result in bias [41,42]. However, human inspection of each survey entry can be time-consuming, more variable than automated processes, and could also result in bias. Future work should aim to better characterize subgroups that may be disproportionately flagged by fraud detection systems and develop ensemble approaches that integrate manual and automatic fraud detection while balancing fraud detection accuracy with protections against excluding valid participants.

While overly sensitive fraud detection could result in bias, fraud detection methods that are not sensitive enough to detect fraudulent entries could also add random noise or systematic bias to the data and threaten the integrity of the research [14-16]. It is important to note that we do not have insight into fraudulent participants’ techniques for responding to survey questions. Fraudulent participants may deliberately select specific demographic options (characteristics they believe will be more likely to result in their entry into the study), randomly select their responses, or use some combination of those techniques [9]. Additionally, rapid developments in machine learning and artificial intelligence have increasingly allowed bots to mimic human behavior [11,22], which could contribute to the seemingly human selection of responses on these surveys, including entries into free text fields [43]. Regardless, this analysis demonstrated the importance of developing study-specific fraud detection methods to supplant or supplement the proprietary fraud detection methods of web-based survey platforms.

Another point of note is the decreased effectiveness of fraud detection tools in determining user legitimacy, as major technology companies take increasing measures to protect user privacy. For example, it is common for fraud detection tools to rely on device fingerprinting and browser cookies to help determine the legitimacy of an individual [44]. While these 2 methods are regularly used by advertisers and marketers to track individuals and deliver targeted advertisements, they also provide a way for fraud detection tools to flag known bad actors and differentiate between legitimate and fraudulent responses. However, the invasive and comprehensive nature of device and browser fingerprinting has raised privacy concerns from users and privacy advocates alike [45,46]. Technology companies, such as Apple, Mozilla, and Brave, have in turn introduced measures to hide users’ identities and activity in a bid to protect user privacy. For instance, Apple’s Safari browser on the macOS desktop operating system now strips all unique identifiers from a user’s device profile, so they appear no different from millions of other Safari users [47]. These privacy-protecting measures, while helpful in safeguarding an individual’s digital presence, make it more difficult for fraud detection tools to differentiate between a legitimate human and a bot. This could partially explain the discrepancy we found between the fraud detection by Qualtrics using reCAPTCHA and RelevantID and our multilayer fraud detection.

Without a method to make a conclusive determination regarding which entries are truly fraudulent and which entries are genuinely valid, it is difficult to compare the relative performance of our multilayer fraud detection methods with the Qualtrics fraud detection methods. However, several pieces of evidence suggest that our fraud detection methods have advantages over Qualtrics in this study context. First, we saw that for email addresses that had an institutional affiliation (ie, “.edu” or “.gov,” which require identity confirmation and cannot be generated en masse) and thus were presumed to be valid, our fraud detection methods correctly validated 98% (n=166) of cases. In comparison, Qualtrics only validated 75% (n=126) of cases. Second, we saw an unusually large discrepancy between the 2 fraud detection methods during a period when the survey link was open, but no advertising or recruitment had recently been active. During this time when we did not expect to receive legitimate responses, we received hundreds of responses that were largely classified as fraud by our fraud detection methods but were generally classified as valid by the Qualtrics system.
While it is possible that valid participants were still able to find and access this survey in the absence of active recruiting, we believe this pattern is evidence of noneligible actors using automated systems in an attempt to gain additional compensation payments from the survey. Taken together, these 2 observations are indirect evidence that our multilayer fraud detection method may have better specificity (ie, can correctly identify valid entries) as well as better sensitivity (ie, can accurately detect fraudulent entries) when compared to the Qualtrics system in this study. While we are unable to conclude that the approach we developed for this study is more or less accurate in identifying fraud when compared to the system used by Qualtrics, we believe these pieces of indirect evidence suggest that using an automated system, such as the one available through Qualtrics, alone may be suboptimal. A multilayered approach was recommended to fraud detection that includes conservative use of automated fraud detection and integration of human review of entries that is tailored to the study’s specific context and its participants.

Limitations
This study is subject to several limitations. First, our comparison of fraud detection methods is limited by the fact that we are unable to definitively determine which entries are valid and which are fraudulent. We selected fraud detection criteria specifically intended to identify repeat respondents (eg, multiple responses providing identical information), fraudulent submissions from outside the Philadelphia region (eg, location verification using IP addresses), and submissions from bots or bot-assisted fraudulent participants (eg, requiring responses that would be difficult to generate via algorithm such as local, colloquial neighborhood names). Applying these criteria may still have resulted in the inclusion of illegitimate responses and the exclusion of legitimate ones. Second, because we cannot know for certain the true fraud status of participants, we are unable to calculate metrics like precision and recall for the fraud detection methods. Future research should aim to establish gold-standard indicators for fraud that could then be used to directly compare the efficacy of the different methods for fraud detection. Third, without knowing how Qualtrics detects fraud, we are unable to determine which components of our strategy may overlap with the Qualtrics strategy. This limits the conclusions we can draw about the comparative effectiveness of these fraud detection methods. Fourth, while we choose to compare our fraud detection methods with the automated systems used by the Qualtrics platform, we believe that similar comparisons and research are also needed with other proprietary fraud detection systems.

Recommendations
The following recommendations are offered for improving data integrity in web-based survey research based on the findings from this study:

- Use a multilayered approach to fraud detection that combines different techniques like bot detection, location verification, consistency checks, and manual review. Relying solely on one method may miss certain types of fraud.
- Carefully evaluate proprietary fraud detection systems and request details on their methodology if possible. Black box methods make it difficult to fully assess their impact on sample composition.
- Avoid overly strict fraud detection rules that may disproportionately exclude valid respondents from vulnerable groups. Balance rigor with inclusion.
- Continuously monitor survey responses over time to identify changes in fraud patterns that may require adjustments to detection methods.

Conclusions
Web-based research and recruitment through social media platforms offer powerful flexibility for researchers to collect large, diverse samples. Web-based surveys, however, are vulnerable to low-quality data from fraud and duplicate entries. Researchers must actively design their web-based studies with this vulnerability in mind and adopt active and adaptable methods of detecting and responding to fraudulent survey responses. Automated, proprietary fraud detection systems offered by web-based survey software may be an important tool in combating fraud, but additional research is needed to evaluate their effectiveness. Human verification of survey entries, while time-consuming, can add another layer of protection and enhance the rigor of web-based survey research. We believe a multilayered strategy that includes a combination of automated fraud detection tools, data enrichment, and human intelligence is the best approach for combating fraud.

Acknowledgments
The authors would like to recognize the contributions of the community partners who helped to distribute the survey as well as the research participants who shared their experiences and insights as part of this research. The authors would also like to thank Gared Harbison for his contributions to validating survey responses and cleaning and managing the data. This work was supported by funding from the National Institutes of Health Agreement OT2HL16156 as part of the Community Engagement Alliance Against COVID-19 Disparities, PhillyCEAL.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.
Authors' Contributions
SB, PST, JW, AD, AV, and JB conceptualized the research project; SB, PST, JW, and AD cleaned and managed the data; SB conducted formal statistical analysis of the data; AV and JB acquired financial support for the project; SB, PST, JW, JG, and JB developed the methodology for the analysis; SB, PST, and JW managed and coordinated day to day operations for the project; AV and JB provided oversight and leadership for the research; SB and WL created visualizations and data presentations; and SB, WL, PST, JW, and JG wrote the original draft of the paper. All authors reviewed and revised the paper and approved the final version.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 22 KB - formative_v8i1e47091_app1.docx ]

Multimedia Appendix 2
Fraud detection analysis.
[DOCX File , 39 KB - formative_v8i1e47091_app2.docx ]

References


27. reCAPTCHA v2. Google. 2023. URL: https://developers.google.com/recaptcha/docs/display [accessed 2023-12-14]


33. Proxy and VPN detection API. proxyclick. 2022. URL: https://www.proxyclick.io [accessed 2023-12-14]


Virtual and Interprofessional Objective Structured Clinical Examination in Dentistry and Dental Technology: Development and User Evaluations

MengWei Pang1,2,3,4*, MDS; YanLing Dong1,2,3,4*, MDS; XiaoHan Zhao5*, MD; JiaWu Wan6, BA; Li Jiang1,2,3,4, DT; JinLin Song1,2,3,4, PhD; Ping Ji1,2,3,4, PhD; Lin Jiang1,2,3,4*, DDS

1 Stomatological Hospital of Chongqing Medical University, Chongqing, China
2 Chongqing Key Laboratory of Oral Diseases and Biomedical Sciences, Chongqing, China
3 Chongqing Municipal Key Laboratory of Oral Biomedical Engineering of Higher Education, Chongqing, China
4 College of Stomatology, Chongqing Medical University, Chongqing, China
5 State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China
6 Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd, Beijing, China

* these authors contributed equally

Corresponding Author:
Lin Jiang, DDS
Stomatological Hospital of Chongqing Medical University
426# Songshibei Road
Yubei District
Chongqing, 401147
China
Phone: 86 15922650133
Email: jianglin@hospital.cqmu.edu.cn

Abstract

Background: Interprofessional education (IPE) facilitates interprofessional collaborative practice (IPCP) to encourage teamwork among dental care professionals and is increasingly becoming a part of training programs for dental and dental technology students. However, the focus of previous IPE and IPCP studies has largely been on subjective student and instructor perceptions without including objective assessments of collaborative practice as an outcome measure.

Objective: The purposes of this study were to develop the framework for a novel virtual and interprofessional objective structured clinical examination (viOSCE) applicable to dental and dental technology students, to assess the effectiveness of the framework as a tool for measuring the outcomes of IPE, and to promote IPCP among dental and dental technology students.

Methods: The framework of the proposed novel viOSCE was developed using the modified Delphi method and then piloted. The lead researcher and a group of experts determined the content and scoring system. Subjective data were collected using the Readiness for Interprofessional Learning Scale and a self-made scale, and objective data were collected using examiner ratings. Data were analyzed using nonparametric tests.

Results: We successfully developed a viOSCE framework applicable to dental and dental technology students. Of 50 students, 32 (64%) participated in the pilot study and completed the questionnaires. On the basis of the Readiness for Interprofessional Learning Scale, the subjective evaluation indicated that teamwork skills were improved, and the only statistically significant difference in participant motivation between the 2 professional groups was in the mutual evaluation scale ($P=0.004$). For the viOSCE evaluation scale, the difference between the professional groups in removable prosthodontics was statistically significant, and a trend for negative correlation between subjective and objective scores was noted, but it was not statistically significant.

Conclusions: The results confirm that viOSCE can be used as an objective evaluation tool to assess the outcomes of IPE and IPCP. This study also revealed an interesting relationship between mutual evaluation and IPCP results, further demonstrating that the IPE and IPCP results urgently need to be supplemented with objective evaluation tools. Therefore, the implementation of viOSCE as part of a large and more complete objective structured clinical examination to test the ability of students to meet undergraduate graduation requirements will be the focus of our future studies.
Introduction

Interprofessional Collaboration Between Dentists and Dental Technicians

Conflicts are part of the life of any organization, and the dental professions are not spared. Jurisdictional battles and supremacy struggles are not alien to dentistry [1]. Unfortunately, despite the obvious reported benefits of interprofessional education (IPE) for interprofessional collaborative practice (IPCP) [2], there is a paucity of data about IPE to promote IPCP among dental professionals. Dentists and dental technicians need to communicate effectively and contribute their professional skills to ensure that they make decisions that are in the best interests of their patients [3]. A clear understanding of the interactions of the dental care team can promote teamwork [4], establish cooperative goals [5], encourage mutual respect [6], and promote IPCP between dental students and dental technology students [7].

IPE encourages teamwork among dental care professionals [8-11] and is increasingly becoming a part of the training programs for dental and dental technology students [12-17]; however, gaps still remain. Perhaps, the largest gaps are owing to the predominant focus of previous studies regarding IPE and IPCP on student and instructor perceptions and a lack of objective assessment of collaborative practice as an outcome measure [18,19]. This marked gap has necessitated the development of a conceptual framework to evaluate the impact of IPE on IPCP to strengthen the evidence for IPE as a tool to improve IPCP between dental and dental technology students [20].

The Objective Structured Clinical Examination

The objective structured clinical examination (OSCE) is an assessment tool based on the principles of objectivity and standardization, in which individual students move through a series of time-limited stations in a circuit for the purpose of assessment of professional performance in a simulated environment. At each station, the student is assessed and marked against standardized scoring rubrics by trained assessors [21]. OSCE has been widely adopted as a summative assessment in the medical undergraduate curriculum and is universally accepted as the gold standard for assessing clinical competence in dental education [22,23]; furthermore, its effectiveness has been confirmed by several studies [24-26]. On the basis of the extensive application of OSCE, the interprofessional OSCE (iOSCE) was initially developed to simulate IPCP [27]. Unlike conventional OSCE, iOSCE involves students from different professions, encourages students to work as a team, and requires the entire team to participate in all tasks [28]. This is performed to objectively evaluate the results of IPE. Within this framework, several variations of iOSCE have been developed to accommodate the training needs of health care teams built to address different disease categories (team OSCE [29,30], group OSCE [31,32], interprofessional team OSCE [28,33], etc). These iOSCE variants can be roughly divided into synchronous [33-36] and asynchronous [29,37] task-based variants. A team working in an operating room typically works synchronously, whereas health care teams of dentists and dental technicians typically work asynchronously. Although the use of iOSCE in medical education has been extensively reported [27-30,38-42], to the best of our knowledge, the use of iOSCE for asynchronous work, especially within dentistry and dental technology cross-professional education, has not been reported.

Although iOSCE may provide an ideal solution for dental and dental technology students to perform IPCP simulation based on real patient cases, the COVID-19 pandemic [43] highlighted the limitations of this traditional approach. For example, a plaster model generated from a clinical case and passed multiple times among students and examiners may pose a risk of infection. In addition, diagnostic stations are usually set up to facilitate OSCE. A station is typically equipped with a trained, standardized patient, and the students complete the diagnosis by asking questions and examining this standardized patient. The risk of infection at this type of station was heightened during the pandemic. Nevertheless, compared with the traditional OSCE, iOSCEs are more time consuming and resource intensive [44,45], especially in dental education; hence, a virtual approach, as developed and piloted in this study, is justified [46,47]. Notably, the conventional virtual OSCE (vOSCE) has been described as a method of performing OSCE using internet technology in medicine [47-49]. The major reason for this technological approach was the scattered nature of the locations of students requiring assessment. However, this approach does not fully leverage virtual technology in dentistry. The integration of digital dental technologies and cloud-based dental laboratory workflows could be practiced within the vOSCE framework [50], which now also forms a professional core course in dental technology education [51-53]. The development of iOSCE based on virtual technology could facilitate the inclusion of digital dental technology in the blueprint design of examination stations. This combination could simulate the critical needs of present-day dental laboratories and promote students’ improved perception about the current demands of the profession.

Objective

To address these research gaps, this study presented a new virtual iOSCE (viOSCE) to objectively assess the effectiveness of IPE as a tool to promote IPCP among dental and dental technology students. We have described the development and piloting of a viOSCE framework and its virtual techniques to validate the user-friendliness of IPE and document its effect on IPCP among dental and dental technology students. Data from both subjective and objective evaluations were collected, and their correlation was assessed.
Methods

Development of viOSCE

The principal investigator (PI) first limited the viOSCE knowledge to content related to the prosthodontics course. Content related to implantology and orthodontics was excluded because it is not part of the core undergraduate coursework for dental or dental technology students. On the basis of the Association for Medical Education in Europe guide [54], a modified Delphi method was used to generate content for viOSCE. The Delphi method is a decision-making process that uses expert opinion, gathered in the form of a survey, under the guidance and direction of the PI to reach group consensus through collaboration, independent analysis, and iteration [55]; this process is the most frequently used method to generate content for OSCEs [54]. The panel of experts in this study consisted of 9 instructors (including the PI) from the College of Stomatology, Chongqing Medical University. All 9 instructors had prosthodontics teaching experience and digital technology practical teaching experience with undergraduate dental and dental technology students. They had also participated in the design and examiner training for traditional OSCE, but only the PI had experience in IPE and vOSCE design.

In this study, there were 4 iterations (rounds) before the viOSCE station design was finalized. In the first round, the PI identified 10 potential topics for viOSCE based on the syllabus of the prosthodontics course for dentistry and dental technology students, gave initial suggestions for the station design, and created a manuscript that was emailed to the panel of experts. Each expert independently gave their opinion and selected 5 topics that they considered as the most important in the syllabus and the most suitable for assessment using viOSCE. In the second round, the PI identified 3 topics with the highest selection rate based on the expert feedback and designed draft blueprints for 20 stations based on the top 3 selected topics using existing virtual technology support. These were sent to the expert panel via email. The expert panel commented about the potential effectiveness of interprofessional collaboration at the stations, made necessary corrections, and returned the design drafts to the PI. In the third round, the PI summarized all the changes made by the expert panel and, finally, decided on 7 stations based on the availability of virtual technology and the time to be spent on the stations within the allotted time frame of the examination. Stations consuming a lot of time, requiring multiple devices for support, or requiring very large spaces were rejected. Next, the selected viOSCE station blueprint design was completed, the virtual technical support was finalized, and the PI sent the final viOSCE station blueprint to the expert panel via email. The expert panel created the scoring rubrics based on the final viOSCE station blueprint, and these were returned to the PI for finalization. In the final round, the PI compiled all the information and met with the group to get a consensus regarding the viOSCE station blueprint and scoring rubrics. Once all the experts approved the viOSCE test station blueprint and scoring rubrics, the PI declared the viOSCE design as complete and declared the panel of experts the viOSCE examiner panel (Figure 1).
**The viOSCE Framework**

The developed viOSCE framework consisted of 3 topics, namely, fixed prosthodontics, removable prosthodontics, and clinical diagnostics. There were 7 collaborative examination stations consisting of 4 asynchronous and 3 synchronous stations. All these stations were designed and developed using the Delphi method (Figure 2).
Figure 2. The framework of the viOSCE. CAD: computer-aided design; RPD: removable partial denture; viOSCE: virtual and interprofessional objective structured clinical examination.

At the fixed prosthodontic stations, the dental student prepared tooth 8 (maxillary right central incisor) on the simulator (Nissan Dental Products) and then worked with the dental technology student to scan the preparations using an intraoral scanner (Panda P2; Freqty Technology). The dental and dental technology students at the intraoral scanning station worked collaboratively. The dental student performed an intraoral scan task, and the dental technology student observed the scan results to determine whether they could be used for the computer-aided design (CAD) wax pattern station. After obtaining a digital model, the dental technology student used a CAD system (Dental system; 3shape) to design a single crown on the digital model of the preparation. Individual scoring rubrics were designed for tooth preparation, intraoral scan, and CAD wax pattern. The 3 examiners scored each of the 3 stations (Figure 3 and Tables 1-3).
Figure 3. The fixed prosthodontics stations of the viOSCE. (A) A dental student prepared tooth 8 on the simulator. (B) Dental and dental technology students scanned the preparation using an intraoral scanner. (C) A dental technology student created a digital wax pattern using the computer-aided design system. (D) A viOSCE examiner scored the preparation process and results. (E) A viOSCE examiner scored the intraoral scanning process and results. (F) A viOSCE examiner scored the digital wax patterns. viOSCE: virtual and interprofessional objective structured clinical examination.
Table 1. The scoring rubric used to assess the tooth preparation stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation before operation</td>
<td>5</td>
<td>• Infection control was correctly performed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct adjustment of phantom head position and lighting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The operating position is correct</td>
</tr>
<tr>
<td>Fine motor skills</td>
<td>15</td>
<td>• Holds the handpiece correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fulcrum stability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct use of the mouth mirror to reflect areas to be operated under indirect vision</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Accurate application of burs</td>
</tr>
<tr>
<td>Preparation during operation</td>
<td>15</td>
<td>• Operation sequence correctly performed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Placement of depth orientation grooves</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Labial surface prepared in 2 planes</td>
</tr>
<tr>
<td>Incisal reduction</td>
<td>10</td>
<td>• 1.5-2 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Formed a small bevel inclined 45° to the lingual side</td>
</tr>
<tr>
<td>Axial reduction</td>
<td>15</td>
<td>• 2 mm for the labial surface</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 1 mm for the proximal surface</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 0.7-1 mm for the lingual surface</td>
</tr>
<tr>
<td>2-plane reduction</td>
<td>5</td>
<td>• Labial surface forms 2 planes and has rounded line angles and point angles</td>
</tr>
<tr>
<td>Taper</td>
<td>5</td>
<td>• Retentive walls: 6°-10°</td>
</tr>
<tr>
<td>Margin placement</td>
<td>10</td>
<td>• Margins extended to a specified target (1 mm supragingivally)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 0.8-1 mm for the shoulder, modified form of the shoulder, and small radius internal angle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with a 90° cavosurface margin</td>
</tr>
<tr>
<td>Details</td>
<td>20</td>
<td>• Adjacent teeth and gingiva are unaffected by the preparation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No undercut areas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Margins and walls are smooth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Margins are continuous and well defined</td>
</tr>
</tbody>
</table>

Table 2. The scoring rubric used to assess the intraoral scan stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning preparation</td>
<td>25</td>
<td>• Order creation is correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The tip is held smoothly and stable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No saliva interference during scanning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cleans the lens and waits for 10 s to preheat the lens</td>
</tr>
<tr>
<td>Scanning operation</td>
<td>35</td>
<td>• Continuous operation of the standard scanning sequence without pauses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• During the scanning, the lip and other soft tissues are pulled to expand</td>
</tr>
<tr>
<td></td>
<td></td>
<td>the scanning field</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scanning should be completed in 6 min (upper and lower jaws)</td>
</tr>
<tr>
<td>Scanning integrity</td>
<td>35</td>
<td>• Mesial and distal interproximal surfaces are intact with no missing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>red-blue data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The scan width of the gingival area is at least 2 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scanning of the occlusal surface or incisal edge is complete and clear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The bite registration is correct</td>
</tr>
<tr>
<td>Software tool selection</td>
<td>5</td>
<td>• Ability to use the software tools accurately</td>
</tr>
</tbody>
</table>
Table 3. The scoring rubric was used to assess the computer-aided design wax pattern stations (crowns).

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order creation</td>
<td>5</td>
<td>- Selects preparation in the teeth overview correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Selects the category correctly (anatomy, wax, and zirconia)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Selects the import option correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Selects the relevant import type correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Imports the intraoral scan data correctly</td>
</tr>
<tr>
<td>Margin</td>
<td>10</td>
<td>- Places the margin line correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Sets the insertion direction correctly</td>
</tr>
<tr>
<td>Occlusion</td>
<td>15</td>
<td>- Normal overlap and overbite</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Accurate restoration of the occlusal vertical dimension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- The occlusion can be checked by dynamic virtual articulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Balanced occlusal forces and no premature contacts</td>
</tr>
<tr>
<td>Proximal contact area</td>
<td>10</td>
<td>- Correct position and shape of the proximal contact area</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Correct contact relationship between adjacent teeth</td>
</tr>
<tr>
<td>Shape</td>
<td>35</td>
<td>- Tooth position: long axis is correctly aligned with the lip and tongue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>direction, correct proximal and distal orientation, tooth is correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>positioned in the dental arch, and ratio of the tooth length to width</td>
</tr>
<tr>
<td></td>
<td></td>
<td>is coordinated with that of the adjacent teeth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Thickness: the thinnest thickness is not &lt;0.5 mm, and the axial surface</td>
</tr>
<tr>
<td></td>
<td></td>
<td>thickness is not &lt;1 mm and not &gt;1.5 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Gingival embrasures are correctly designed and coordinated with</td>
</tr>
<tr>
<td></td>
<td></td>
<td>those of the adjacent teeth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Tooth length: the incisal position is in harmony with that of the adja-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cent teeth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Detailed structure of the surface, such as developmental grooves and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ridges</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Lingual morphology: lingual fossa and marginal ridge morphology</td>
</tr>
<tr>
<td>Cement space</td>
<td>5</td>
<td>- Acceptable cement space</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acceptable extra cement space</td>
</tr>
<tr>
<td>Restoration effect</td>
<td>20</td>
<td>- Acceptable functionality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acceptable esthetics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acceptable visual harmony</td>
</tr>
</tbody>
</table>

At the removable prosthodontics station, real patient cases and intraoral digital models were selected and prepared by the PI, followed by approval by the expert panel. The intraoral digital model was a clinical plaster model scanned using Lab Scanner (E4; 3shape). Each dental student used our previously developed Objective Manipulative Skill Examination of Dental Technicians (OMEDT) system [56] to observe the intraoral digital model and to design a removable partial denture (RPD) framework. At the end of the design task, the dental student submitted the design and then discussed the design with the dental technology student; the dental student could make modifications if they wanted to. Next, each dental technology student used a CAD system (Dental system; 3shape) to design the framework of an RPD on the intraoral digital model based on the final design. A viOSCE examiner scored the first RPD design using the OMEDT system. Next, the viOSCE examiner scored the final RPD design and the digital framework of the RPD. The design discussion station was not scored by a separate examiner (Figure 4 and Tables 4 and 5).
Figure 4. The removable prosthodontics stations of the viOSCE. (A) A dental student designed the framework of a RPD using the Objective Manipulative Skill Examination of Dental Technicians system. (B) Dental and dental technology students discussed the RPD design. (C) A dental technology student created a digital framework of an RPD using the computer-aided design system. (D) A viOSCE examiner scored the first RPD design using the Objective Manipulative Skill Examination of Dental Technician system. (E) A viOSCE examiner scored the final RPD design and the digital framework of the RPD. RPD: removable partial denture; viOSCE: virtual and interprofessional objective structured clinical examination.

Table 4. The scoring rubric used to assess the removable partial denture design stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case observation</td>
<td>20</td>
<td>• The missing tooth position is identified accurately and marked correctly on the drawing</td>
</tr>
<tr>
<td>Design choices</td>
<td>40</td>
<td>• No missing component</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Indirect retainer is present in the optimal position</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Design choices do not violate biological principles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clasp choice is optimal for the case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The major connector is selected properly with reasonable extension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Justified use of clasps and rests</td>
</tr>
<tr>
<td>Drawing</td>
<td>20</td>
<td>• Ideal drawing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Metal components are painted in blue, and resin bases are painted in red</td>
</tr>
<tr>
<td>Consistency with task description</td>
<td>10</td>
<td>• Exactly as described in the task description</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clearly presents the requirements implied in the description, and the design is well aligned with the corresponding description</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Gives consideration to both esthetics and functions</td>
</tr>
<tr>
<td>Neatness and accuracy in presentation</td>
<td>10</td>
<td>• Neat and accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No inconsistencies between the table and drawing</td>
</tr>
</tbody>
</table>
Table 5. The scoring rubric used to assess the computer-aided design wax pattern stations (removable partial denture [RPD] framework).

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order creation</td>
<td>5</td>
<td>• Selects artificial teeth in the teeth overview correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the category correctly (removable—RPD frame)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the import option correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the relevant import type correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Imports the laboratory scan data correctly</td>
</tr>
<tr>
<td>Surveying</td>
<td>10</td>
<td>• Insertion direction is correctly chosen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Undercuts are correctly identified</td>
</tr>
<tr>
<td>Virtual cast preparation</td>
<td>20</td>
<td>• Correct paralleled blockout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct shaped blockout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct arbitrary blockout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct relief setting</td>
</tr>
<tr>
<td>Framework design</td>
<td>40</td>
<td>• Reasonable position and shape of clasp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of rest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of major connector</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of retention grid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of finishing line</td>
</tr>
<tr>
<td>Form</td>
<td>25</td>
<td>• All parts are connected as a whole</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The thickness and strength of the framework meet the requirements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The thickness is uniform, and the surface is smooth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Esthetics are acceptable</td>
</tr>
</tbody>
</table>

The clinical diagnostics station used a virtual standardized patient (VSP) with the haptic device (UniDental, Unidraw). The VSP hardware does not have an anthropomorphic shape, but it interacts through vocal, visual, and haptic devices. On the basis of the novel oral knowledge graph and the coupled, pretrained Bert models, the VSP can accurately interact with a dentist’s underlying intention and express the symptom characteristics in a natural style [57]. On the basis of this algorithm, the PI adjusted and entered the real patient case details, allowing the dental technology student to work with the dental student as a chairside dental technician to make a diagnosis based on the information obtained from the interactions with the VSP. In this study, the clinical case designed on the VSP was a patient who required root canal treatment and full crown restoration. At the end of the dental student’s diagnosis and simulation, the dental technology student was required to assist the dental student in designing the restoration plan and help the patient in choosing the materials for crown restoration (this often determines the price of the treatment). Thus, dental and dental technology students finalized the prostodontic treatment plan collaboratively. The visual device built a virtual dental clinic environment and VSP model, allowing the students to view the VSP from global, extraoral, and intraoral perspectives. The haptic device allows dental students to perform intraoral and extraoral examinations using essential tools to explore the diagnostic evidence.

Owing to the complexity of collaborative diagnosis, the station was manually scored by 2 examiners independently based on the previously developed scoring rubrics, whereas the UniDental output machine provided an additional score according to the previously developed scoring rubrics. The average of the 3 scores formed the final score for the station. To ensure the relative independence and internal consistency of all scores, the examiners were not informed about the existence of the machine score. The PI exported the machine score data from the VSP at the end of the experiment (Figure 5 and Table 6).
Figure 5. The clinical diagnostics station of the viOSCE. (A) The VSP with the haptic device, UniDental. (B) Dental and dental technology students performed intraoral palpation on the VSP using the haptic device. (C) Then, 2 viOSCE examiners scored the process and clinical diagnostic results. viOSCE: virtual and interprofessional objective structured clinical examination; VSP: virtual standardized patient.
Table 6. The scoring rubric used to assess the clinical diagnostic stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>History taking</td>
<td>25</td>
<td>• The content of the inquiry is accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Few questions unrelated to the disease or clinical situation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Inquiries are made sequentially, purposefully, and hierarchically</td>
</tr>
<tr>
<td>Intraoral examination</td>
<td>25</td>
<td>• Tool selection is accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Humanistic care is reflected during the examination</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Appropriate oral examination items are performed based on the case information</td>
</tr>
<tr>
<td>Auxiliary examination</td>
<td>10</td>
<td>• Correct auxiliary examination items are selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct interpretation of auxiliary examination results</td>
</tr>
<tr>
<td>Case analysis</td>
<td>20</td>
<td>• Correct diagnosis of the case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct selection of the diagnostic criteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct differential diagnosis of the case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct selection of the basis for differential diagnosis</td>
</tr>
<tr>
<td>Plan design</td>
<td>20</td>
<td>• Correct treatment plan design according to the disease condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Provides advice about material selection based on patient request</td>
</tr>
</tbody>
</table>

**Performance Evaluation of viOSCE**

In this study, fourth-year undergraduate dental students and third-year undergraduate dental technology students participated in viOSCE because students at this stage of education had completed preclinical professional training. Overall, 50 students who met these requirements, including 25 (50%) dental students and 25 (50%) dental technology students, were recruited into the viOSCE user evaluation experiment and were divided into groups of 2 comprising 1 dental student and 1 dental technology student. The PI and examiner teams did not influence or determine the team-formation process. All participating students were informed that as this viOSCE was in the experimental phase, it was conducted as a small extracurricular skills competition, thus allowing for self-evaluation without a final examination situation, as previously reported [58]. This approach allowed for the simulation of an examination situation without affecting the final examination grade of the students. A month before commencing the experiment, the PI led an web meeting for students to explain the viOSCE, the relevant knowledge points, and the need to practice fully during the upcoming month. At the end of the meeting, the students completed the Readiness for Interprofessional Learning Scale (RIPLS) pretest questionnaire, which is a 19-item 5-point Likert-scale questionnaire; this type of questionnaire is the most frequently used method for the subjective evaluation of IPE and IPCP [18].

viOSCE was piloted after the 1-month preparation period. The panel of examiners marked points according to the previously prepared scoring rubrics, whereas some of the points were automatically scored by a machine. After this step, the participating students completed the posttest self-made questionnaire, to which a mutual evaluation scale and a viOSCE evaluation scale were added. The mutual evaluation scale asked the students to score the performance of their partner, whereas the viOSCE evaluation scale asked the students to score viOSCE. In total, 6 items were included in the mutual evaluation scale, and 7 items were included in the viOSCE evaluation scale (Textboxes 1 and 2). All items in both questionnaires were set to a maximum score of 100. Before issuing the questionnaire, the panel first reviewed all the questions, clarified ambiguities, and removed any double-barreled questions [59,60]. At the end of the experiment, one-on-one interviews were conducted with all the students to determine their perceptions about viOSCE.

**Textbox 1.** The mutual evaluation scale administered to dental and dental technology student groups who participated in the virtual and interprofessional objective structured clinical examination (viOSCE).
The virtual and interprofessional objective structured clinical examination (viOSCE) evaluation scale administered to dental and dental technology student groups who participated in viOSCE.

**Items**

- Evaluation of viOSCE effectiveness
- Evaluation of equipment, network operation and maintenance
- Evaluation of viOSCE examiners
- Evaluation of viOSCE staff
- Rationality of the clinical diagnostic design
- Rationality of the fixed prosthodontics design
- Rationality of the removable prosthodontics design

**Statistical Analysis**

Data were tabulated in a Microsoft Excel spreadsheet and imported into IBM SPSS Statistics for Windows (version 26.0; IBM Corp) for descriptive analysis. GraphPad PRISM 8.0 software (GraphPad Software) was used to create the graphs. Responses were summarized, and comparisons were made. Output data were presented as percentages and in graphical format. The Shapiro-Wilk test was used to test for normal distribution. Specific data analysis tests performed included descriptive statistics, 2-tailed paired t tests, and correlation analyses.

**Ethical Considerations**

The research ethics committee of the Affiliated Hospital of Stomatology, Chongqing Medical University, approved this study protocol (COHS-REC-2022; LS number: 096). All participants provided written informed consent before participation in the study.

**Results**

Of the 50 students, 32 (64%) completed the experiment. Interviews were conducted with the students who dropped out of further participation in the study. The main reasons for dropping out included the students’ belief that they or their collaborating partners had not practiced sufficiently to perform well in the experiment. A group had a verbal confrontation approximately an hour before the experiment began. The main reason for the conflict was that the dental technology student accused the dental student of not practicing sufficiently before the experiment. According to the study protocol, at the end of the experiment, the conflict was resolved by the PI. Both parties were counseled, mediated by the PI, and the 2 parties reconciled.

Data from the RIPLS, mutual evaluation scale, and viOSCE evaluation scale were first analyzed to determine the impact of viOSCE on the subjective evaluation of IPCP. All students (32/32, 100%) who completed the experiment were administered the RIPLS questionnaire before and after the experiment. The Cronbach $\alpha$ values were .835 for the pretest data and .731 for the posttest data, suggesting that the reliability and internal consistency were acceptable. The results failed the Shapiro-Wilk test for normality; therefore, the data were analyzed using the Wilcoxon signed rank test. The teamwork and collaboration subscale scores were significantly increased after the experiment ($P=.004$). In addition, there was a nonsignificant decrease in the negative professional identity subscale scores ($P=.21$). There was also an insignificant increase in the scores on the positive identity subscale and on the roles and responsibilities subscale ($P=.13$ and $P=.96$, respectively). Figure 6 depicts the RIPLS data before and after the viOSCE pilot.
After the experiment, the mutual evaluation scale was administered to all participating students (32/32, 100%) who completed the experiment. The Cronbach $\alpha$ value was .873, suggesting good reliability and internal consistency. Comparison of the results of the dental and dental technology students revealed that only the mutual evaluation scores for competition motivation were significantly different between the 2 groups ($P=.04$). The dentistry and dental technology students evaluated each other’s motivation to participate in the competition (competition motivation), and the dental students had higher scores than the dental technology students. Figure 7 depicts the mutual evaluation scale scores of the dental and dental technology students.
Similarly, after the experiment, the viOSCE evaluation scale was administered to all students (32/32, 100%). The Cronbach \( \alpha \) value was .706, suggesting acceptable reliability and internal consistency. Comparison of the viOSCE evaluation scale results of the dental and dental technology students with the Wilcoxon signed rank test results revealed that only the evaluation scores for the removable prosthodontics design were statistically significant \((P=.01)\) among the 7 items. Figure 8 depicts the viOSCE evaluation scale scores of the dental and dental technology students.

**Figure 8.** viOSCE evaluation scale data for dental and dental technology students. viOSCE: virtual and interprofessional objective structured clinical examination.
To explore the validity of the examiner panel scores in viOSCE, correlation analysis was conducted on the scores of each station under the 3 topics. Using Spearman correlation coefficient, for the fixed prosthodontics topic, a strong positive correlation between the scores of the tooth preparation station and the CAD wax pattern station was noted, and it was statistically significant ($r=0.67; P=.005$). Positive correlations between the scores of the intraoral scan station and the CAD wax pattern station and between the intraoral scan station and the tooth preparation station were not statistically significant ($r=0.179; P=.51$ and $r=0.387; P=.14$, respectively). For the removable prosthodontics topic, 11 (69%) of the 16 student groups finally decided to modify the RPD design initially made by the dental students. A negative but statistically insignificant correlation between the scores of the RPD design station and the CAD wax pattern station was noted ($r=-0.111; P=.68$). For the clinical diagnostics topic, the correlation analysis was conducted primarily for the machine scores and the examiner scores to determine the usability of the VSP in viOSCE and the consistency of machine scoring and examiner scoring. The results revealed a significant positive correlation between the scores of the 2 examiners, and the positive correlation between the machine scores and the 2 examiners’ scores was also significant. The results are shown in Table 7 and Figure 9.

Table 7. Spearman correlation analysis of the virtual and interprofessional objective structured clinical examination scores.

<table>
<thead>
<tr>
<th>Topic and station</th>
<th>Correlation coefficient</th>
<th>$P$ value (2-tailed)</th>
<th>Participants (n=16), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed prosthodontics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tooth preparation vs CAD$^a$ wax pattern</td>
<td>0.670</td>
<td>.005$^b$</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Intraoral scan vs tooth preparation</td>
<td>0.387</td>
<td>.14</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Intraoral scan vs CAD wax pattern</td>
<td>0.179</td>
<td>.51</td>
<td>16 (100)</td>
</tr>
<tr>
<td><strong>Removable prosthodontics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPD$^b$ design vs CAD wax pattern</td>
<td>$-0.111$</td>
<td>.68</td>
<td>16 (100)</td>
</tr>
<tr>
<td><strong>Clinical diagnostics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine score vs examiner-1 score</td>
<td>0.601</td>
<td>.01$^d$</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Machine score vs examiner-2 score</td>
<td>0.629</td>
<td>.009$^b$</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Examiner 1 score vs examiner-2 score</td>
<td>0.855</td>
<td>$&lt;.001^e$</td>
<td>16 (100)</td>
</tr>
</tbody>
</table>

$^a$CAD: computer-aided design.
$^b$RPD: removable partial denture.
To explore the relationship between the objective and subjective evaluations, correlation analysis was conducted between the viOSCE scores and the RIPLS scores as well as between the viOSCE scores and the mutual evaluation scale scores. Insignificant negative correlations were noted between the subjective evaluation scores presented by RIPLS and viOSCE. Similarly, the correlation of the mutual evaluation scale score with the viOSCE scores was not significant. The SD of the scores on the mutual evaluation scale showed a decreasing trend among students with higher viOSCE scores and those with lower scores, but an increasing trend was observed among those with median scores (Table 8 and Figure 10).
Table 8. Spearman correlation analysis between the subjective and objective evaluations presented by the Readiness for Interprofessional Learning Scale (RIPLS) and mutual evaluation scale.

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>P value (2-tailed)</th>
<th>Participants (n=16), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The examiner panel scores of viOSCE&lt;sup&gt;a&lt;/sup&gt; vs the intragroup mean score of RIPLS (before the test)</td>
<td>−0.272</td>
<td>.15</td>
</tr>
<tr>
<td>The examiner panel scores of viOSCE vs the intragroup mean score of RIPLS (after the test)</td>
<td>−0.302</td>
<td>.13</td>
</tr>
<tr>
<td>The examiner panel scores of viOSCE vs the intragroup mean score of the mutual evaluation scale</td>
<td>−0.038</td>
<td>.44</td>
</tr>
</tbody>
</table>

<sup>a</sup>viOSCE: virtual and interprofessional objective structured clinical examination.

Figure 10. Correlation analysis of the viOSCE scores and the mutual evaluation scale. (A) Correlation analysis of the viOSCE scores and the SD of the scores for each item on the mutual evaluation scale. (B) Correlation analysis of the viOSCE scores and the SD of the mean scores on the mutual evaluation scale. viOSCE: virtual and interprofessional objective structured clinical examination.

In the one-on-one interviews, 29 (91%) of the 32 students approved of the effectiveness of viOSCE and wanted to use it to assess their IPCP ability in the graduation examination. At the fixed prosthodontics station, 56% (9/16) of the dental technology students complained about the lack of lingual space prepared by their partners at the tooth preparation station, which made it difficult to design crown wax patterns, and the corresponding dental students reported not being aware of the condition before viOSCE. At the removable prosthodontics station, almost all the dental students (15/16, 94%) reported that the advice given by the dental technology students was effective in helping them complete the RPD design and considered their
design practice to be insufficient. In contrast, the dental technology students reported that helping the dental students complete the RPD design made them feel satisfied with their professional competence and felt that they were truly part of the team during the collaboration. At the clinical diagnosis station, the dental students felt that their clinical practice experience was not sufficient, especially when the dental technology students could provide a diagnostic plan faster than themselves.

In terms of positive feedback, the students believed that viOSCE promoted the friendship between themselves and their partners, helped them realize the continuity and relevance between their own work and the work of their partners, and enabled them to acquire a deep understanding of IPCP. The negative feedback mainly focused on their lack of clinical knowledge, inadequate preparation, and long waiting time at some stations.

**Discussion**

**Principal Findings**

The IPCP results of dentists and dental technicians reflect the quality of their IPE, skill training, and clinical experience. The results contribute to the much-needed IPE assessment literature and suggest that teamwork skills can be improved by IPCP and effectively assessed using this new evaluation scale. We used a modified Delphi process in this study. This is in accordance with Simmons et al [27], who found that the modified Delphi process is an effective tool to obtain consensus among professionals for the foundational work required. In addition, our study demonstrated the effectiveness of viOSCE in asynchronous and synchronous collaboration scenarios, while providing a methodological reference for developing a new iOSCE for dental health care professionals. As the collaboration scenario between the dentist and the dental technician may be both asynchronously applied through prescriptions and synchronously conducted in chairside discussions [61-63], it was deemed appropriate for the viOSCE framework to consider both synchronous and asynchronous scenarios.

The viOSCE scores in this study also reflect the effectiveness of the framework design. From the viOSCE examiner scores in the fixed prosthodontics section, a significant positive correlation between the scores of the tooth preparation station and the CAD wax pattern station was evident. This finding is consistent with the actual clinical asynchronous delivery scenario, where the dentist’s preparation largely determines the quality of the dental technician’s crown wax pattern. Qualitative evaluations extracted from the one-on-one interviews also supported this result. For the removable prosthodontics section, the negative correlation between the scores of the RPD design station and the CAD wax pattern station was not statistically significant, which might be owing to the fact that more than half of the groups (15/16, 94%) worked collaboratively to modify the RPD design to possibly compensate for the lack of training, which is consistent with the findings about dentists’ inadequate competence in RPD design reported in other studies [64,65].

As OSCE is essentially a simulated scenario-based examination, the use of virtual technology to build simulated scenarios has become an important direction for OSCE-related studies, especially in the field of dental education [43]. The COVID-19 pandemic has further contributed to dental educators’ interest in this area, as dental clinical practice typically occurs in a virus-laden aerosolized environment [66]. Therefore, providing a safe and robust learning environment in the simulation clinic is also critical to help students compensate for lost educational time. The virtual technologies used to construct the simulated clinical environment in this study include VSP and CAD. Previously, Janda et al [67] developed a virtual patient as a supplement to standard instruction in the diagnosis and treatment planning of periodontal disease. However, it could not fully understand complex or ambiguous questions, and the students felt frustrated during the practice [67].

Tanzawa et al [68] developed a robot patient that could reproduce an authentic clinical situation and introduced it into OSCE. However, the dialogue recognition of the robot patients was prespecified; the robot was unable to identify subjective patient descriptions or the dentist’s interrogation intention and could not support intraoral or extraoral examinations to obtain diagnostic evidence [68]. To fill these gaps, our study used VSP with intention recognition and haptic feedback to construct virtual dental clinical practice and diagnosis scenarios more realistically. As the diagnostic evidence collected by students through interrogation, inspection, and palpation was automatically summarized for the final differential diagnosis, and omissions in the examination process eventually led to a misdiagnosis, the system simulated a high-fidelity clinical environment. In addition, the results showed that 1 (6%) of the 16 student groups misdiagnosed their VSP because of incomplete interrogation and palpation. The correlations between the scores of the 2 examiners and the machine scores were statistically significant, thus confirming the robustness of the high-fidelity simulation scenarios constructed by the VSP and the machine scores. On the basis of these results, the use of VSP should be expanded and integrated into daily teaching to give students more opportunities for clinical practice training.

Consistent with the results of previous OMEDE studies [56], the use of CAD technology in viOSCE significantly reduced the time spent at each station for the dental technology students. Some dental technology students complained about the slowness of the CAD program. Upon further investigation, it was found that they imported both impressions at the same time. In dental laboratory practice, dental technicians usually import the impressions separately to prevent computational issues. This finding exposes the lack of virtual dental laboratory practice skills in teaching, which needs to be addressed.

The results showed that the teamwork and collaboration subscale scores were significantly increased at the end of the study (P<.004), suggesting that viOSCE can improve students’ teamwork skills. The increase in the other 3 subscale scores, although not statistically significant, can be explained by the choice of timing of viOSCE. The optimal time to expose medical students to IPE is still subject to debate [18]. viOSCE, as a clinical IPCP intervention introduced during the clinical year, had no significant effect on the promotion of negative or positive
identity or roles and responsibilities. This finding may be due to the fact that the students’ professional cognition had been stereotyped at this time, making it difficult to affect significant changes through IPE or IPCP intervention. This conclusion is supported by a previous study [69].

The results of the mutual evaluation scale showed statistically significant difference in participant motivation between the 2 professional groups, which could be explained by the results of the roles and responsibilities subscale. Of the 16 dental technology students, 4 (25%) expressed that they would not practice as dental technicians in the future because they wanted to choose other careers. Differences in the scores of the other items were not statistically significant, thus showing the effectiveness of viOSCE in the development of teamwork spirit. This result confirms that the OSCE design is well suited as a final evaluation of IPE and IPCP. In addition, the average score of each item of viOSCE was >60, indicating that the students were satisfied with the design and operation of viOSCE. The differences in scores between the 2 types of professionals were not statistically significant, except at the removable prosthodontics station, which was probably caused by the dental technology students’ unfamiliarity with the CAD program.

Overall, the internal consistency of all subjective evaluations was acceptable, and the results met expectations. Interesting observations were also made regarding the correlation between the subjective and objective evaluations. The SD of the scores on the mutual evaluation scale showed a decreasing trend among the dental and dental technology students with higher viOSCE scores and those with lower scores, but an increasing trend in the median score was observed. Although this trend was not statistically significant due to sample size limitations, this early finding provides data support for a summary of clinical experience published previously by Preston [70], who reported that the intensity of the relationship between dentists and dental technicians is determined by the difference in their professional skills. If the professional skills of both parties are high, there will be few problems in their cooperative relationship. The more discriminating and demanding the technician or dentist becomes, the more the relationship is strained when either fails to perform up to the other’s standards. This result suggests that in the study of IPE and IPCP for dentists and dental technicians, it is not sufficient to explore the improvement of the traditional assessment dimensions such as team collaboration skills and identity. The final quality of the output must be included in the assessment dimension. This also reaffirms the effectiveness of viOSCE as an objective, quantitative evaluation tool for IPE and IPCP.

Limitations and Future Studies
The main limitation of our study is the small convenience sample of participating students, which could have led to self-selection bias. The sample size should be expanded in the future to obtain more data and to further verify the robustness of the viOSCE framework. In addition, whether viOSCE should be made a part of the large and more complete OSCE to test the ability of students to meet undergraduate graduation requirements will also be the focus of our next study. Moreover, the independent application of the novel VSP in the education of dental students is an interesting topic that will be explored in the next step of this study.

Recommendations
On the basis of our results, we provide the following recommendations:

1. All dental health professionals should be educated to deliver patient-centered care as members of an interdisciplinary team [16].
2. IPE intervention–related skills should be introduced as preclinical skills.
3. The cooperation of the dental care team is complex, and the training for improving the cooperation ability of the dental care team should include both subjective and objective assessments.
4. viOSCE and scale assessment should be introduced for the assessment of IPE and IPCP at the clinical stage of training.

Conclusions
In this study, a novel viOSCE framework was developed and piloted. Data based on subjective evaluation scales and objective examiner scores were collected and analyzed, confirming the effectiveness of viOSCE as an objective evaluation tool for IPE and IPCP. The experimental design should be expanded to include more randomly selected students with a scientifically determined sample size to further develop studies focused on IPE and IPCP in dentistry and dental technology, ultimately promoting quality in dental clinical practice.

Acknowledgments
This study was funded by a major project of Chongqing Higher Education and Teaching Reform of the People’s Republic of China (reference number 201019 to JS and reference number 193070 to MP) and the Teaching Reformation Fund of the College of Stomatology, Chongqing Medical University (KQJ202105 to YD and KQJ202003 to MP). All authors thank Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd for assisting in the pilot study and the maintenance of the virtual standardized patient and Objective Manipulative Skill Examination of Dental Technicians systems.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.
Authors' Contributions

MP was the principal investigator who organized the group of experts to develop the framework of the virtual and interprofessional objective structured clinical examination (viOSCE); designed the validity experiment; collected and analyzed the data; and drafted the manuscript. YD organized the viOSCE pilot study and assisted with fundraising and distribution, expert panel recruitment, and data collection. XZ and JW led an engineer team from Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd to complete the development of the virtual standardized patient and Objective Manipulative Skill Examination of Dental Technicians systems and assisted in the maintenance of the virtual standardized patient and Objective Manipulative Skill Examination of Dental Technicians systems during the viOSCE pilot study. Li Jiang assisted in the preparation of clinical cases and equipment related to viOSCE, checked all the details, and was responsible for maintaining order at the viOSCE facility. PJ and JS supervised and directed the project. Lin Jiang supervised the advancement of the project and assisted in recruiting the participating students. All authors approved the manuscript.

Conflicts of Interest

None declared.

References


61. Mykhaylyuk N, Mykhaylyuk B, Dias NS, Blatz MB. Interdisciplinary esthetic restorative dentistry: the digital way. Compendium 2021 Nov;42(10) [FREE Full text]


Abbreviations

CAD: computer-aided design
IPCP: interprofessional collaborative practice
IPE: interprofessional education
OMEDT: Objective Manipulative Skill Examination of Dental Technicians
OSCE: objective structured clinical examination
PI: principal investigator
RIPLS: Readiness for Interprofessional Learning Scale
RPD: removable partial denture
vOSCE: virtual and interprofessional objective structured clinical examination
vOSCE: virtual objective structured clinical examination
VSP: virtual standardized patient

Edited by A Mavragani; submitted 28.11.22; peer-reviewed by C Fonseca, Y Zhou, N Ab Malik; comments to author 16.03.23; revised version received 26.04.23; accepted 13.12.23; published 17.01.24.

Please cite as:
Pang M, Dong Y, Zhao X, Wen J, Jiang L, Song J, Ji P, Jiang L. Virtual and Interprofessional Objective Structured Clinical Examination in Dentistry and Dental Technology: Development and User Evaluations. JMIR Form Res 2024;8:e44653
URL: https://formative.jmir.org/2024/1/e44653
doi:10.2196/44653
PMID:38231556

©MengWei Pang, YanLing Dong, XiaoHan Zhao, JiaWu Wan, Li Jiang, JinLin Song, Ping Ji, Lin Jiang. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Use of Machine Learning Tools in Evidence Synthesis of Tobacco Use Among Sexual and Gender Diverse Populations: Algorithm Development and Validation

Shaoying Ma1*, PhD; Shuning Jiang2*, BS; Olivia Yang2; Xuanzhi Zhang2, BS; Yu Fu2, BS; Yusen Zhang2, BS; Aadeeba Kaareen1, BSoCSci; Meng Ling2, PhD; Jian Chen2, PhD; Ce Shang1, PhD

1Center for Tobacco Research, The Ohio State University Comprehensive Cancer Center, Columbus, OH, United States
2Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, United States
*these authors contributed equally

Corresponding Author:
Shaoying Ma, PhD
Center for Tobacco Research
The Ohio State University Comprehensive Cancer Center
3650 Olentangy River Road
1st Floor, Suite 110
Columbus, OH, 43214
United States
Phone: 1 6148976063
Email: shaoying.ma@osumc.edu

Abstract

Background: From 2016 to 2021, the volume of peer-reviewed publications related to tobacco has experienced a significant increase. This presents a considerable challenge in efficiently summarizing, synthesizing, and disseminating research findings, especially when it comes to addressing specific target populations, such as the LGBTQ+ (lesbian, gay, bisexual, transgender, queer, intersex, asexual, Two Spirit, and other persons who identify as part of this community) populations.

Objective: In order to expedite evidence synthesis and research gap discoveries, this pilot study has the following three aims: (1) to compile a specialized semantic database for tobacco policy research to extract information from journal article abstracts, (2) to develop natural language processing (NLP) algorithms that comprehend the literature on nicotine and tobacco product use among sexual and gender diverse populations, and (3) to compare the discoveries of the NLP algorithms with an ongoing systematic review of tobacco policy research among LGBTQ+ populations.

Methods: We built a tobacco research domain–specific semantic database using data from 2993 paper abstracts from 4 leading tobacco-specific journals, with enrichment from other publicly available sources. We then trained an NLP model to extract named entities after learning patterns and relationships between words and their context in text, which further enriched the semantic database. Using this iterative process, we extracted and assessed studies relevant to LGBTQ+ tobacco control issues, further comparing our findings with an ongoing systematic review that also focuses on evidence synthesis for this demographic group.

Results: In total, 33 studies were identified as relevant to sexual and gender diverse individuals’ nicotine and tobacco product use. Consistent with the ongoing systematic review, the NLP results showed that there is a scarcity of studies assessing policy impact on this demographic using causal inference methods. In addition, the literature is dominated by US data. We found that the product drawing the most attention in the body of existing research is cigarettes or cigarette smoking and that the number of studies of various age groups is almost evenly distributed between youth or young adults and adults, consistent with the research needs identified by the US health agencies.

Conclusions: Our pilot study serves as a compelling demonstration of the capabilities of NLP tools in expediting the processes of evidence synthesis and the identification of research gaps. While future research is needed to statistically test the NLP tool’s performance, there is potential for NLP tools to fundamentally transform the approach to evidence synthesis.

(JMIR Form Res 2024;8:e49031) doi:10.2196/49031
KEYWORDS

machine learning; natural language processing; tobacco control; sexual and gender diverse populations; lesbian; gay; bisexual; transgender; queer; LGBTQ+; evidence synthesis

Introduction

The use of nicotine or tobacco products is a leading preventable cause of cancer, heart diseases, and lung diseases in the United States [1], with cigarette smoking alone responsible for the death of half a million Americans each year [2]. Notably, sexual and gender diverse individuals, often referred to as the LGBTQ+ (lesbian, gay, bisexual, transgender, queer, intersex, asexual, Two Spirit, and other persons who identify as part of this community) populations, are particularly vulnerable to nicotine and tobacco product use [3]. Both the National Cancer Institute and the Centers for Disease Control and Prevention have recognized the LGBTQ+ populations as a critical target in their efforts to combat tobacco use disparities [4–10].

In response to the pressing need for tobacco control and the rapidly evolving landscape of the tobacco market, the National Institutes of Health (NIH) and other health foundations, including the American Cancer Society, have made substantial investments in tobacco control research and tobacco regulatory science [11,12]. According to our calculations using data from the NIH era reporter, funding for tobacco research has shown a remarkable increase, growing from US $7.7 billion in 2016 to US $11.2 billion in 2021 (Multimedia Appendix 1 [13]). Consequently, the volume of peer-reviewed publications related to tobacco has experienced a significant increase. This presents a considerable challenge in efficiently summarizing, synthesizing, and disseminating research findings, especially when it comes to addressing specific target populations, such as the LGBTQ+ populations.

One promising pathway to rapidly assessing the expanding body of literature is the use of natural language processing (NLP) models. NLP is dedicated to deciphering and comprehending how computers interpret human language, equipping them to analyze extensive data sets of natural language [14–16]. While NLP tools have garnered considerable recognition in biomedical research [4–10], aiding in tasks such as disease surveillance (eg, COVID-19) and diagnosing using medical records [17–23], their potential to expedite near real-time synthesis of evidence in tobacco control research remains untapped [24].

Another gap in existing NLP tools is the lack of applications in synthesizing social science research and modeling. A noteworthy example in the domain of tobacco research is the evaluation of the effectiveness of tobacco control policies, which are often assessed using complex statistical modelling and large-scale survey data. These methods demand a specialized semantic database for labelling studies and interpreting results. However, to the best of knowledge, such a semantic database has not been developed yet. Considering that policy interventions at federal, state, and local levels are designed to reach a large number of populations, the lack of a database to facilitate NLP applications may significantly undermine evidence synthesis and thereby the timely adoption of effective policies [25].

Furthermore, in light of the calls from entities such as the NIH and other health agencies to address tobacco use disparities within priority populations, including LGBTQ+ populations, the development NLP tools to aid in the discovery of effective policies tailored to these special populations remains uncharted territory [26–31]. There is an urgent demand for the development of NLP tools (eg, semantic database, NLP algorithms) in tobacco research that have the abilities to synthesize evidence in social science and assist in research gap discovery for priority populations.

In this pilot study, we aimed to achieve the following goals to address the identified research and application gaps: (1) compile a specialized semantic database for tobacco policy research to extract information from journal article abstracts, (2) develop NLP algorithms that comprehend the literature on nicotine and tobacco product use among sexual and gender diverse populations, and (3) compare the discoveries of the NLP algorithms with an ongoing systematic review of tobacco policy research among LGBTQ+ populations [32]. While this pilot study does not fully address the gaps by developing a comprehensive evidence synthesis or discovery tool for tobacco research, the outcomes may pave the road for future tools that can achieve this goal. Our vision is that NLP tools may be able to assist academic scholars and policy makers in prescribing public health policies, such as tobacco control policies, and addressing public health needs, such as reducing health disparities.

Methods

Development of a Tobacco Research Domain–Specific Semantic Database

Overview

To generate a tobacco research domain-specific semantic database, we used an iterative process that combines expert opinions and the reading of tobacco research papers in 4 leading tobacco journals (Tobacco Control, Nicotine and Tobacco Research, Tobacco Induced Diseases, and Tobacco Prevention and Cessation). The main categories of keywords were the followings: (1) tobacco use behaviors, prevalence, and outcomes; (2) population characteristics; (3) geographic locations; (4) method and inference; (5) policy; (6) tobacco products; (7) relation statement; and (8) tobacco characteristics. Under each main category, there were one or more subcategories, and each subcategory contained a list of named entities. Table 1 presents the categories of named entities in a domain-specific semantic database that were used for training and improving a language model for tobacco research on sexual and gender diverse populations. These categories are based on journal articles’ keywords, further guided by existing literature on how to use NLP methods to synthesize public health evidence [25,33]. These categories are important components of a study, encompassing measures, methods, results, conclusions, and hypothesis testing.
### Table 1. Main categories and subcategories of named entities.

<table>
<thead>
<tr>
<th>Main categories</th>
<th>Subcategories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobacco use behavioral outcomes</td>
<td>Tobacco cessation</td>
</tr>
<tr>
<td></td>
<td>Exposure to tobacco-related or antitobacco content, or exposure to secondhand or thirdhand smoking</td>
</tr>
<tr>
<td></td>
<td>Health and disease</td>
</tr>
<tr>
<td></td>
<td>Perception and belief</td>
</tr>
<tr>
<td></td>
<td>Tobacco use prevalence</td>
</tr>
<tr>
<td></td>
<td>Time period</td>
</tr>
<tr>
<td>Population characteristics</td>
<td>Age groups</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
</tr>
<tr>
<td></td>
<td>Sexual and gender diverse populations</td>
</tr>
<tr>
<td></td>
<td>Racial and ethnic minoritized groups</td>
</tr>
<tr>
<td></td>
<td>Socioeconomic status</td>
</tr>
<tr>
<td>Geographic locations</td>
<td>Countries, states, provinces, or cities</td>
</tr>
<tr>
<td>Method and inference</td>
<td>Data</td>
</tr>
<tr>
<td></td>
<td>Methodology</td>
</tr>
<tr>
<td></td>
<td>Statistics</td>
</tr>
<tr>
<td>Policy</td>
<td>Marketing</td>
</tr>
<tr>
<td></td>
<td>Law, policy, and regulation</td>
</tr>
<tr>
<td></td>
<td>Regulation body</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
</tr>
<tr>
<td>Tobacco products</td>
<td>Combustible tobacco products</td>
</tr>
<tr>
<td></td>
<td>Noncombustible tobacco products</td>
</tr>
<tr>
<td>Relation statement</td>
<td>Relation terms</td>
</tr>
<tr>
<td>Tobacco characteristics</td>
<td>Chemical</td>
</tr>
<tr>
<td></td>
<td>Flavor</td>
</tr>
</tbody>
</table>

### Journal Selection

We chose 4 peer-reviewed tobacco-specific multidisciplinary journals, namely, *Tobacco Control*, *Nicotine and Tobacco Research*, *Tobacco Induced Diseases*, and *Tobacco Prevention and Cessation*, to extract articles and compile keywords at the initial stage. The first 2 are among the journals that have the highest impact factors in addiction research; in 2022, *Tobacco Control* had an impact factor of 5.2 and a 5-year impact factor of 5.7 [34], and *Nicotine and Tobacco Research* had an impact factor of 4.7 and a 5-year impact factor of 4.2 [35]. *Tobacco Induced Diseases* [36] and *Tobacco Prevention and Cessation* [37] are 2 other peer-reviewed journals that specifically publish research on nicotine and tobacco products but are not as highly ranked as the other 2 journals. The textual data from the 4 peer-reviewed journal articles contained a total of 2993 abstracts from published papers from 2015 to early 2021.

While the 2993 articles extracted from these journals do not represent the full body of tobacco research, they cover a significant share of tobacco studies and integrate evidence across the 5 translational research stages: basic research, preclinical research, clinical research, clinical implementation, and public health. These journals also ask authors to specify how the research reported contributes to tobacco control objectives, which have policy implications. Alternatively, a random sampling from PubMed searches using tobacco related terms may not yield studies that are necessarily translational in nature. Therefore, we focused on the articles published in the 4 journals in our study.

### Iterative Process to Expand Terms (Named Entities) in the Database

The general process included the following iterative steps: (1) to generate initial annotation data, we first compiled key terms from extracted articles and allocated key terms to categories using group discussions; (2) we enriched the database using various sources and group discussions (more specific descriptions below); (3) we fine-tuned the spaCy en_core_web_lg model with the initial annotation and following iterative versions of data (the en_core_web_lg model is a pretrained large language model that can extract multiple general named entities); (4) we expanded the list of named entities to include more keywords of similar meanings using SeedNER [38,39], that is, a small set of initial labeled examples or patterns that was used as a starting point for training a model; (5) we searched the occurrence of each keyword in the 2993 paper abstracts and kept those with high frequency; (6) during this process, named entities that were too generic to yield meaningful relations were removed from the database; and (7) we repeated steps 3 to 6 until the set of entities reached our satisfaction during group discussions.
Specific approaches were used for conducting step 2. For categories including “tobacco use behavioral outcomes,” “tobacco products,” and “tobacco characteristics,” the iterative process involved four steps: (1) discussions to determine whether to include newly identified key terms and how to allocate them into additional subcategories (Table 1); (2) using a named entity recognition (NER) model to extract named entities from 2993 paper abstracts from the 4 specific journals; (3) randomly sampling and reviewing the output of the NER model, correcting identified errors, and adding missed NERs; and (4) repeating steps 1 to 3 until we were satisfied with the model output.

The categories “population characteristics,” “geographic locations,” and “relation terms” are commonly used concepts in real life and not specific to tobacco control. We used Google searches, Wikipedia, and WordNet to enrich the key terms. In addition, for the “method and inference” category, we used the glossary of an econometrics methodology textbook by Cameron and Trividi to enrich the terms [40]. This textbook is widely used in economics and social science and its glossary should provide sufficient terms for this category.

For the “policy” category, we drew named entities from 2 sources that comprehensively summarize available tobacco control policies in the regulatory space. The first source was a peer-reviewed journal article by McDaniel et al [41] that conducted an intensive policy scan of all possible regulations that can contribute to tobacco endgame. The second source was the World Health Organization’s report on the global progress in implementing tobacco control policies, as recommended by the World Health Organization’s Framework Convention on Tobacco Control [42], which is the largest public health treaty signed by 182 countries and prescribes a comprehensive set of tobacco control policies. These policies are classified into 5 groups: M (monitor tobacco use and prevention policies), P (protect people from tobacco smoke), O (offer help to quit tobacco use), W (warn about the dangers of tobacco), E (enforce bans on tobacco advertising, promotion, and sponsorship), and R (raise taxes on tobacco) [42]. These sources cover policy key terms related to both national and international contexts and together create the most comprehensive policy terms to our knowledge.

Development of NLP Algorithms That Comprehend the Literature on Nicotine and Tobacco Product Use Among Sexual and Gender Diverse Populations

We used RoBERTa, an optimized BERT (bidirectional encoder representations from transformers)-based language model [43], to perform NER tasks. BERT is a state-of-the-art language model that excels at tasks such as sentiment analysis and text summarization. By learning patterns and relationships between words and their context in text, BERT can extract named entities that it has learned during training and potentially discover new ones.

We developed an NER model based on RoBERTa using the Python (Python Software Foundation) programming language and the spaCy library [44]. We began by defining 36 labels of categories (main and subcategories; Table 1) and extracting 1582 named entities using the existing NER model RoBERTa. Next, those named entities were used to tag abstracts and create a training set, using the annotation tool Prodigy [45]. A subset of the abstracts with labeled named entities was reviewed by 2 domain experts to identify key terms that were missing in our semantic database, which were added to the lists of named entities.

The RoBERTa model was then updated based on the richer database and further trained for a maximum of 20,000 steps, with early stopping implemented if no improvement was observed for 1600 consecutive steps. With a series of iterations, we used the updated RoBERTa model to assess the 2993 abstracts and labeled them with the categories.

When identifying studies related to LGBTQ+ populations, it is important to understand that this community is heterogeneous [46,47]. Given that LGBTQ+ key terms are included in the “population characteristics” categories, we were able to identify LGBTQ+ populations based on categorization. There were 111 LGBTQ+-related named entities in our database.

Comparison of the Discoveries of the NLP Algorithms With an Ongoing Systematic Review of Tobacco Policy Research Among LGBTQ+ Populations

Ideally, we would like to compare the results from our tools with those from systematic reviews and meta-analyses of studies related to tobacco control issues among LGBTQ+ populations. Systematic reviews and meta-analyses are state-of-the-art evidence synthesis methods that can provide the ground truth [48-50]. While we are currently conducting a separate systematic review of the effectiveness of tobacco control policies among LGBTQ+ populations, this review has not been finalized yet [32]. Nonetheless, the ongoing systematic review does provide some data points for comparisons, including the number of studies extracted from the 4 journals and presence of policy assessment. Therefore, we conducted comparisons of these 2 domains.

Ethical Considerations

This study does not involve human subjects, as it synthesizes data from research articles published at peer-reviewed journals. The Ohio State University Institutional Review Board has determined that it contains no human subjects and thus no further review is needed (study number: 2021E0776).

Results

In total, we identified 33 articles relevant to sexual and gender diverse populations from the 2993 abstracts. Our trained model successfully extracted 773 named entities (181 unique named entities) from the 33 paper abstracts to describe the themes of these articles. Among the 773 extracted named entities, 688 were already learned by the model during training, while 70 were new time- or age-related words (eg, 18 years, 2013), 9 were new statistical terms (eg, N=20), and 6 were newly discovered and labeled within other categories. We did not observe any newly discovered policy-related terms.

In Figures 1-3, we present the hierarchy of named entities extracted from abstracts in published papers that studied nicotine...
or tobacco product use among sexual and gender diverse individuals. Each number on the right is the frequency of the corresponding named entity by paper abstract. Named entities with the same color belong to the same main category.

Figure 1. Hierarchy and frequency counts of named entities extracted from published research in tobacco-specific journals from 2015 to early 2021 in 4 main categories: tobacco use, products, characteristics, and relation statement. Numbers represent the frequency of the corresponding named entity by paper abstract.
Figure 2. Hierarchy and frequency counts of named entities extracted from published research in tobacco-specific journals from 2015 to early 2021 in the main category of population characteristics. Numbers represent the frequency of the corresponding named entity by paper abstract.
According to our tool, among the 33 tobacco studies related to LGBTQ+ populations, the most frequent use outcomes were “cigarette smoking” (n=17), “substance use” (n=16), “prevalence” (n=16), and “risk” perception (n=14). Also, for these populations, “cigarettes” (n=15) were the most frequently mentioned combustible tobacco product and “e-cigarettes are” (n=8) was the most frequently mentioned noncombustible tobacco product. In addition, for tobacco characteristics, “alcohol” (n=7) and “nicotine” (n=5) were the most mentioned attributes among LGBTQ+ tobacco research papers.

The relation statement findings suggest that a majority of the studies examined “comparison” (n=26), “association” (n=23), and “correlation” (n=6). We found no studies that explicitly used the term “causal” or “causality” in the studies.
The population characteristics mentioned in the studies illustrated that among socioeconomic status terms, the most frequently included were “demographics” (n=8) and “SES factors” (n=8). Among sex and sexual and gender minority terms, the most frequent ones were “bisexual” (n=21), “lesbian” (n=19), and “gay” (n=19). Among racial and ethnic minority group terms, the most frequent ones were “minority groups” (n=8) and “Race/ethnicity” (n=6). For age group terms, the terms included “adult” (n=14), “young adult” (n=11), “adolescent” (n=5), “students” (n=3), and “adolescents and young adults” (n=2).

The policy category showed that in these studies, the most mentioned term was “intervention” (n=16). In addition, while the general term “tobacco control” was mentioned in 6 studies, only 1 study contained any specific policy term (“smoke free air law”). As such, there was a significant gap in policy research among the published articles in the 4 leading tobacco journals between 2015 and early 2021, since only 1 study mentions specific policies when it comes to tobacco research among the LGBTQ+ populations. The statistics and methodology terms further indicated that the most used terms included “survey” (n=18) and “logistic regression” (n=10), and relatively fewer studies mentioned terms related to causal inferences, such as “experimental research” (n=4), “randomization” (n=3), and “clinical trial” (n=1). The studies mentioning “US” also dominated in the numbers, with 12 studies in total. Several studies that assessed countries with multilevel governing levels, such as Canada and the United States, also appeared to have mentioned “state,” “city,” and “province,” suggesting that attention was paid to these defined areas.

We next compared our results using the NLP tools with our ongoing systematic review. Similar to the conclusions of the ongoing systematic review, we found very few studies that yielded specific policy recommendations. This finding was further corroborated by the lack of causal inference methods labeled by the NLP tool. While our NLP tool cannot replace systematic reviews just yet, it does show potential to complement the existing methods and requires less human supervision (systematic reviews usually require at least 2 human coders).

**Discussion**

This pilot study builds a semantic database dedicated to tobacco research and developed NLP algorithms to automatically identify, extract, and summarize textual data from published tobacco studies. We further demonstrated a user case wherein we assessed LGBTQ+ tobacco research by labeling key components of a tobacco study: tobacco use outcomes, tobacco characteristics, population characteristics, geographic locations, method and inference, and policy relevance.

It is worth noting that the components we categorized, such as “method and inference,” align with the typical sections found in scientific articles in social science, including measures, methods, results, conclusions, and hypothesis testing. As a result, our tool extracts text segments that are frequently assessed in evidence synthesis, thereby showing the potential of using NLP tools to enhance systematic reviews and facilitate meta-analyses.

Additionally, we leveraged the NLP algorithms we created to identify gaps in tobacco research concerning the LGBTQ+ populations and concluded that there is a scarcity of studies assessing impact policies on this demographic using causal inference methods. This finding is consistent with our ongoing systematic review, highlighting how NLPs have the capacity to aid in both evidence synthesis and research gap discoveries. This, in turn, has the potential to streamline research efforts, reduce labor costs, and influence the trajectories of future research directions.

Using the NLP tool, we further found some interesting patterns in tobacco research involving LGBTQ+ populations. It appears that the product drawing the most attention in the field is cigarettes or cigarette smoking and that the number of studies of various age groups is almost evenly distributed between youth or young adults and adults. Moreover, the existing evidence body is dominated by studies coming from the United States. These patterns are consistent with the research needs to reduce cigarette smoking among LGBTQ+ populations in the United States, where 16.1% of LGBTQ+ adults and 17.4% of LGBTQ+ high schooler students smoke cigarettes—this is 4% to 6% higher than their heterosexual counterparts. Therefore, our findings align with the ongoing research needs and the financial investments made by the US health agencies like the NIH, thereby bolstering the confidence in the NLP tool we developed.

Finally, while the semantic database and language model in this pilot study are designed to extract and summarize key components of tobacco research, many of the terms and labeling categories are broad and applicable to public health and social science research in general, such as “methods and inference” and “relation terms.” Therefore, our tool has the potential to transform the evidence synthesis paradigm in tobacco control and public health at large by enabling more efficient and effective analyses of large volumes of textual data. Future tool development may extend its reach to other public health domains, fostering the real-time translation of research findings into evidence-based policymaking, thereby contributing significantly to the advancement of public health initiatives.

Our study has several limitations. First, for the development of keywords and the application of the NLP, we focused on 4 peer-reviewed tobacco-specific research journals, which were not representative of the entire tobacco control literature. However, considering the prominence and extensive content covered by these journals, we believe that this selection is unlikely to introduce significant selection bias or result in the omission of crucial keywords. Second, although we used our ongoing systematic review as a benchmark for the qualitative assessment of the results obtained in this pilot study, we did not perform a quantitative comparison of our findings with the ground truth derived from the systematic review. This quantitative evaluation, which might include measures like Cohen kappa, was not conducted because the systematic review has not yet been finalized. Consequently, future research endeavors are required to undertake a thorough quantitative evaluation.
comparison between the training data and the established ground truth using statistical testing for a more comprehensive assessment of the NLP tool’s performance.

Despite the limitations, our pilot study serves as a compelling demonstration of the capabilities of NLP tools in expediting the processes of evidence synthesis and the identification of research gaps. Expanding the scope of this pilot research to encompass other public health disciplines, extending beyond the realm of tobacco control, holds the promise of fundamentally transforming the approach to evidence synthesis. Such expansion has the potential to play a pivotal role in shaping policy development across a wide spectrum of public health domains.

Acknowledgments
This study was supported by the President’s Research Excellence (PRE) Accelerator Grant from The Ohio State University (principal investigator: CS). CS was funded by the National Cancer Institute (R21CA249757). SM was supported by the Pelotonia Fellowship from The Ohio State University Comprehensive Cancer Center.

During the preparation of this work, the authors used ChatGPT 3.5 in order to check grammar errors and improve language flow. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability
The data sets generated and analyzed during this study are available in the GitHub repository [55].

Authors’ Contributions
CS and SM conceptualized the study. JC, ML, SJ, CS, and SM designed the methodology. SJ and ML were responsible for the software. CS and JC validated the data. SM and SJ performed the formal analysis. ML, SJ, and SM carried out the investigation. CS and JC provided resources. ML, OY, XZ, YF, YZ, SJ, and SM performed data curation. SM and SJ wrote the original draft. SM, SJ, JC, and CS reviewed and edited the manuscript. CS and JC supervised the study. ML and SJ were responsible for project administration. CS acquired funding. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Tobacco-related funding from the National Institutes of Health (NIH), 2010-2022. Data was obtained from the National Institutes of Health [13].

References


32. Ma S, Mirza M, Schuster A, Bridges J, Shang C. A systematic review of the effects of tobacco control policies on tobacco use among LGBTQI+ populations. PROSPERO 2022 CRD42022360559. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022360559 [accessed 2024-01-09]


34. About. Tobacco Control. URL: https://tobaccocontrol.bmj.com/pages/about [accessed 2023-12-01]

35. About the journal. Nicotine & Tobacco Research. URL: https://academic.oup.com/ntr/pages/About [accessed 2023-12-01]

36. Aims and scope. Tobacco Induced Diseases. URL: http://www.tobaccoinduceddiseases.org/Aims-and-Scope,386.html [accessed 2023-12-01]


Abbreviations

BERT: bidirectional encoder representations from transformers
LGBTQ+ : lesbian, gay, bisexual, transgender, queer, intersex, asexual, Two Spirit, and other persons who identify as part of this community
NER: named entity recognition
NIH: National Institutes of Health
NLP: natural language processing
Use of Machine Learning Tools in Evidence Synthesis of Tobacco Use Among Sexual and Gender Diverse Populations: Algorithm Development and Validation


URL: https://formative.jmir.org/2024/1/e49031
doi:10.2196/49031
PMID:38265858

©Shaoying Ma, Shuning Jiang, Olivia Yang, Xuanzhi Zhang, Yu Fu, Yusen Zhang, Aadeeba Kaareen, Meng Ling, Jian Chen, Ce Shang. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Designing and Developing Online Training for Diabetes Prevention Program Coaches Using an Integrated Knowledge Translation Approach: Development and Usability Study

Kaela D Cranston\(^1\), BSc, MSc; Natalie J Grieve\(^1\), BA, MSc; Tineke E Dineen\(^1\), BSc, MA, PhD; Mary E Jung\(^1\), BHK, MSc, PhD
Faculty of Health and Social Development, University of British Columbia, Kelowna, BC, Canada

Corresponding Author:
Mary E Jung, BHK, MSc, PhD
Faculty of Health and Social Development
University of British Columbia
1238 Discovery Road
Kelowna, BC, V1V 1V7
Canada
Phone: 1 2508079670
Email: mary.jung@ubc.ca

Abstract

**Background:** e-Learning has rapidly become a popular alternative to in-person learning due to its flexibility, convenience, and wide reach. Using a systematic and partnered process to transfer in-person training to an e-learning platform helps to ensure the training will be effective and acceptable to learners.

**Objective:** This study aimed to develop an e-learning platform for Small Steps for Big Changes (SSBC) type 2 diabetes prevention program coaches to improve the viability of coach training.

**Methods:** An integrated knowledge translation approach was used in the first 3 stages of the technology-enhanced learning (TEL) evaluation framework to address the study objective. This included three steps: (1) conducting a needs analysis based on focus groups with previously trained SSBC coaches, meetings with the SSBC research team, and a review of research results on the effectiveness of the previous in-person version of the training; (2) documenting processes and decisions in the design and development of the e-learning training platform; and (3) performing usability testing. Previously trained SSBC coaches and the SSBC research team were included in all stages of this study.

**Results:** Step 1 identified components from the in-person training that should be maintained in the e-learning training (ie, a focus on motivational interviewing), additional components to be added to the e-learning training (ie, how to deliver culturally safe and inclusive care), and mode of delivery (videos and opportunities to synchronously practice skills). Step 2 documented the processes and decisions made in the design and development of the e-learning training, including the resources (ie, time and finances) used, the content of the training modules, and how coaches would flow through the training process. The design and development process consisted of creating a blueprint of the training. The training included 7 e-learning modules, the learning modalities of which included narrated demonstration videos and user-engaging activities, a mock session with feedback from the research team, and a final knowledge test. Step 3, usability testing, demonstrated high levels of learnability, efficiency, memorability, and satisfaction, with minor bugs documented and resolved.

**Conclusions:** Using an integrated knowledge translation approach to the technology-enhanced learning evaluation framework was successful in developing an e-learning training platform for SSBC coaches. Incorporating end users in this process can increase the chances that the e-learning training platform is usable, engaging, and acceptable. Future research will include examining the satisfaction of coaches using the SSBC coach e-learning training platform, assessing coach learning outcomes (ie, knowledge and behavior), and estimating the cost and viability of implementing this training.

(JMIR Form Res 2024;8:e50942) doi:10.2196/50942

**KEYWORDS**
program evaluation; prediabetic state; e-learning education; e-learning; platform; usability; diabetes; prevention; knowledge translation; end user; type 2 diabetes; framework

https://formative.jmir.org/2024/1/e50942
**Introduction**

e-Learning has emerged as a popular alternative to traditional in-person education in recent years, particularly with the advancement of technology and widespread availability of internet connectivity. e-Learning has revolutionized the way people learn by providing flexible, wide-reaching, and convenient access to educational resources, as well as personalized learning experiences that cater to individual needs and preferences. e-Learning has been adopted across various fields, including health care, and is effective in enhancing learning outcomes and improving overall learner engagement [1-3].

There is a paucity of research on how e-learning platforms are designed and developed. Partnering with end users can help guide e-learning platform development. An integrated knowledge translation (IKT) approach meaningfully engages the right research users at the right time throughout the research process and can increase the chances that the e-learning platform is usable, engaging, and acceptable [4-7]. Such approaches require in-depth qualitative methods to understand the perspectives of end users throughout the entire research process.

Frameworks can help guide the process of platform conception, design, development, implementation, and evaluation. This study was guided by the technology-enhanced learning (TEL) evaluation framework [8]. The TEL evaluation framework was informed by commonly used learning models, including the Kirkpatrick model [9]. The seven TEL evaluation framework stages are (1) conduct needs analysis and environmental scan; (2) document processes, decisions, and final product (3) test usability; (4) document key events during implementation and final product; (5) assess participant experience and satisfaction; (6) assess learning outcomes; and (7) estimate cost, reusability, and sustainability. The TEL lacks specific guidelines, which can be seen as both a strength and a limitation, as it permits flexibility for different users and contexts, but it can also be used haphazardly and may result in a product that has not considered end user needs.

This paper offers a guide for including an IKT approach to the needs analysis, design, development, and usability testing of an e-learning platform using the first 3 stages of the TEL evaluation framework. The proposed partnered stages of the TEL evaluation framework are described using the example of developing an e-learning platform for type 2 diabetes (T2D) prevention program coaches.

**Methods**

**Context**

Small Steps for Big Changes (SSBC) is a community-based T2D prevention program that aims to empower individuals with prediabetes to increase physical activity and improve their diet to reduce their risk of developing T2D. SSBC is delivered in and by staff and volunteers of the YMCA of Southern Interior British Columbia. Trained SSBC coaches (YMCA staff and volunteers) deliver 6 one-on-one sessions to clients over 4 weeks using a motivational interviewing (MI)-informed approach.

In-person 3-day workshops were previously used to train coaches to deliver SSBC. After completing the in-person training, coaches delivered the SSBC program with good levels of fidelity [10,11]. However, in-person coach training is not a viable format for future coach training due to difficulty in scheduling training time and the inability to revisit training components. Additionally, the in-person training was not developed using an IKT approach. To ensure usability and acceptability, the development of the SSBC coach e-learning platform used an IKT approach (ie, partnering with current SSBC coaches).

**Needs Analysis**

**Overview**

The information was gathered on the program needs and capacities (eg, the specific knowledge and skills required to become an SSBC coach) and individual needs of SSBC coaches (eg, preference for specific content and instructional approach). The various components of the needs analysis stage were assessed through meetings with the SSBC research team, focus groups with previously trained SSBC coaches, and analysis of previous research.

**Previous Research**

Dineen et al [10] evaluated the effectiveness of the SSBC in-person training. Data from this study helped determine coaches’ baseline levels of knowledge and skills prior to any training.

**SSBC Research Team Meetings**

The SSBC research team included the founder and director of SSBC (the last author) and the 2 in-person training facilitators (first and third author). The first meeting was held to determine whether the in-person training was a feasible and viable mode to continue to train SSBC coaches. The main reasons discussed for transitioning to e-learning were difficulties scheduling a 3-day in-person SSBC coach training for YMCA staff and volunteers, and concerns for practicality in the future expansion of SSBC to other Canadian and global facilities. Additional meetings with the research team were used to determine what knowledge and skills would be required of new SSBC coaches, what content from the in-person training would be included and excluded, and what new concepts could be added. Finally, SSBC research team meetings included discussions on measures to determine the effectiveness of the e-learning platform. In-depth meeting notes and minutes were recorded during meetings for later analysis.

**Focus Groups With SSBC Coaches**

Incorporating end users into the development of the e-learning platform (a main tenet of IKT) was done throughout the needs analysis, design, and development of the SSBC coach e-learning platform. Within the need analysis, focus groups with previously trained SSBC coaches were conducted. SSBC coaches who were trained through the in-person workshop were recruited via email to participate in a 2-hour focus group. Focus group questions centered on what content should or should not be included in the e-learning platform. Coaches also answered questions about motivation to learn, learning preferences (ie,
synchronicity, length of time, interactivity, and choice), and potential barriers for new coaches. Focus group data were analyzed using conventional content analysis [12].

**Design and Development**

To facilitate the design and development phase of this project, the research team hired a third-party digital health solutions company (3C Institute) that specializes in developing e-learning platforms. The first and last author had meetings with 3C Institute every 2 weeks to discuss project progress and make decisions. Throughout this time, the research team created a blueprint and storyboard of the detailed topics to be taught, discussed options for teaching formats (ie, a variety of didactic and user-engaging methods), and defined the learning objectives. The research team wrote the content for the module scripts, which were edited by 3C Institute to ensure they were clear and succinct. Modules were filmed by the research team, and 3C Institute developed graphics and compiled the modules.

SSBC coaches from the needs analysis phase were invited to participate in the design and development phase, specifically to review the overview of modules and methods of instruction and provide feedback.

All processes and decisions were documented in detailed meeting notes (eg, attendees, discussion points, and decisions). Calendar events and contracts captured the resources (eg, financial and time) required to develop the e-learning platform.

**Test Usability**

Usability evaluation helps to ensure that a product is usable, efficient, effective, and satisfying to its end users. Nielsen [13] identified five key components of usability: (1) learnability—how easily end users can accomplish basic tasks the first time they enter the e-learning platform; (2) efficiency—how well end users can perform tasks on the e-learning platform; (3) memorability—how easily end users can reestablish proficiency when they return to the e-learning platform after a period of not using it; (4) errors—how many errors the end users make, the severity of the errors, and their ease of recovery from the errors; and (5) satisfaction—how pleasant it is to use the e-learning platform.

Initial usability testing was conducted by the SSBC research team and the SSBC coaches involved in the e-learning platform design and development. Users were asked to navigate through the modules and learning activities and complete various tasks (eg, log in to the training platform, move through modules, and access the resource center). Written and verbal feedback on issues and errors were collected. A summary of concerns and usability issues was documented and sent to 3C Institute to be addressed. The level of usability was deemed acceptable by the first and last authors once all errors were resolved. In the true spirit of IKT, further usability testing will be examined through qualitative and quantitative methods upon release and implementation of the e-learning platform to ensure that the platform is usable by the end users.

**Ethical Considerations**

This study was approved by the behavioral research ethics board of the University of British Columbia (H21-01800). The participants provided informed consent for the focus groups and were involved in the further design and development of the e-learning platform. All data were deidentified upon data collection. Individuals were renumerated for their participation via e-transfer. Participants each received $50 CAD (US $36.83) for their participation in a focus group, and participants were given $25 CAD (US $18.42) for each module storyboard that they reviewed.

**Results**

**Needs Analysis**

**Previous Research**

Dineen et al [10] demonstrated that coaches had little knowledge about MI and SSBC program content prior to taking the in-person training. Coaches also reported high levels of satisfaction with the in-person training topics.

**SSBC Research Team Meetings**

The SSBC research team made the following decisions based on 3 meetings. SSBC coaches needed to have basic knowledge about T2D prevention, specific SSBC session content (ie, Canada’s physical activity guidelines, daily added sugar limits, carbohydrates, and the talk test), what MI is, the spirit of MI, MI skills (ie, open-ended questions, affirmations, reflections, summaries and ask-tell-ask), the 4 processes of MI, how to deliver SSBC content using MI, and how to deliver culturally safe and inclusive care. Coaches would need to demonstrate skills associated with delivering the SSBC program to clients, which includes appropriately delivering the correct information using an MI-informed approach.

Once a coach completed the e-learning modules, the coach would demonstrate their skills through a mock session, providing an opportunity for assessment and provision of feedback. Sufficient knowledge would be determined by scoring a minimum of 70% on a knowledge test on T2D prevention, SSBC-specific content, and MI at the end of the training. The research team would develop the knowledge tests, drawing from content directly covered in the SSBC coach e-learning training. Finally, the coaches’ skills would be monitored and assessed by coding a random selection of audio recordings from sessions with clients using session-content checklists and the Motivational Interviewing Competency Assessment tool [14].

**Focus Groups With SSBC Coaches**

Invitation emails were sent to 15 SSBC coaches to participate in a focus group, including information about the purpose of the focus group and remuneration of time. A total of 9 SSBC coaches volunteered to take part in a focus group, and they were split into 3 equal-numbered focus groups based on their availability. See Table 1 for focus group participants’ demographics and experience as a coach. On average, each focus group was 62 minutes.

The results from the conventional content analysis showed that coaches wanted more information on the evoking process of MI, on transitioning from sustain talk to change talk, and an understanding of the effectiveness of MI. The coaches spoke about the importance of knowledge checks throughout the
e-learning training, a collection of resources and commonly used documents (ie, a resource center), role-play videos demonstrating the skills being taught, and live sessions with a workshop facilitator to practice skills and receive feedback. The coaches suggested that modules be short in duration and for training to span over 2 weeks. The coaches’ recommendations also included the integration of a variety of learning methods through the use of didactic and user-engaging components. Additional resources suggested by coaches included video examples of delivering SSBC content using MI and various SSBC program delivery styles and scenarios. The participants desired a blended approach of asynchronous and synchronous learning, with an opportunity for practice and feedback.

Table 1. Focus group participants’ demographics and experience (N=9).

<table>
<thead>
<tr>
<th>Individual-level factor</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>3 (33)</td>
</tr>
<tr>
<td>25-34</td>
<td>3 (33)</td>
</tr>
<tr>
<td>35-44</td>
<td>1 (11)</td>
</tr>
<tr>
<td>45-54</td>
<td>2 (22)</td>
</tr>
<tr>
<td>Gender identity</td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>8 (89)</td>
</tr>
<tr>
<td>Man</td>
<td>1 (11)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>8 (89)</td>
</tr>
<tr>
<td>White and Indigenous</td>
<td>1 (11)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>University certificate, diploma, or degree</td>
<td>4 (44)</td>
</tr>
<tr>
<td>Apprenticeships, trades services, or diploma</td>
<td>1 (11)</td>
</tr>
<tr>
<td>College, college d’enseignement general et professionnel, or other nonuniversity certificate or diploma</td>
<td>1 (11)</td>
</tr>
<tr>
<td>High school</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Years of experience working in customer service</td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>1 (11)</td>
</tr>
<tr>
<td>Greater than 5</td>
<td>8 (89)</td>
</tr>
</tbody>
</table>

Design and Development

The research team worked with 3C Institute and SSBC coaches over a series of 13 months (21 meetings) to design and develop the e-learning platform. Correspondence with the involved SSBC coaches to collect feedback was done over email. The flow of the entire design and development phases can be seen in Figure 1, showing each group’s contributions to development. Information collected in the needs analysis stage was used to develop a blueprint for the e-learning platform, which was then reviewed by the 3C Institute. After consultations with 3C Institute and SSBC coaches, it was determined that the e-learning training would include seven modules—1 introducing SSBC and inclusivity, 1 covering SSBC-specific content, 4 covering MI, and 1 putting everything together. Additionally, 3C Institute provided suggestions for learning activities (eg, narrated videos, demonstration videos, and user-engaging activities) based on their expertise in developing e-learning platforms. All modules were designed to be asynchronous so coaches could move through them at their own pace, with user-engaging components and knowledge checks throughout the modules, and at the end of each module. The e-learning platform includes a resource center for coaches to access resources (eg, PDFs, video examples, and SSBC session paperwork).

The research team drafted 1 module script at a time. Scripts included what the narrator would say, and the content for the user-engaging activities and knowledge. The script for each module underwent numerous editing cycles between the 3C Institute and the research team before being sent to the 3C Institute multimedia artist. The multimedia artist developed storyboard proofs for each module. Storyboard proofs included still images representing what would be on the screen with the associated script. Feedback was collected from both the research team and SSBC coaches from the focus groups. For each module storyboard proof, 2 to 4 coaches provided feedback (dependent on their schedules and availability) and were remunerated for their time. After the storyboard proofs were approved by the SSBC research team and coaches, the narrators for the videos (first and last author) worked with a University of British Columbia Studios Okanagan producer and the 3C Institute editor to film the 7 modules. Talent was hired to act as coaches and clients in the role-play videos for the modules and the resource center. After each module was filmed, the 3C Institute
multimedia artist and video director worked to edit and finalize the modules. An overview of the SSBC e-learning training platform is provided in Table 2 and a selection of screenshots from the online training platform are provided in Multimedia Appendix 1.

Based on decisions made following the needs analysis, the entire training process for new SSBC coaches includes 3 stages. First, the coaches must complete the 7 modules. Next, the coaches are required to schedule a 1-hour mock session with an SSBC research team member and are then provided with written feedback within 2 weeks of completing the mock session. Following the mock session, coaches are given access to the final knowledge test. After passing the final knowledge check, coaches are certified to take on SSBC clients independently.

Overall, the authors spent a total of 192 hours on this project (e.g., preparing for and leading SSBC research team meetings and focus groups, analyzing data from needs analysis, designing blueprints, writing and editing scripts, developing knowledge check questions, emails, correspondence with SSBC coaches and talent, meetings with 3C Institute, filming, and usability testing). The final cost of developing the SSBC coach e-learning platform was ~US $80,000 (cost to hire 3C Institute and professional talent for filming).

Figure 1. e-Learning platform design and development process with each group’s contributions.
Table 2. Overview of SSBC\textsuperscript{a} e-learning platform.

<table>
<thead>
<tr>
<th>e-Learning component</th>
<th>Topics covered</th>
<th>Learning objectives</th>
<th>Module features</th>
</tr>
</thead>
</table>
| Module 1             | Introduction to SSBC, prediabetes and T2D\textsuperscript{b}, cultural safety, and inclusivity | • Broadly describe T2D and prediabetes  
• Explain how diet and exercise changes can help prevent T2D  
• Describe what the SSBC program includes  
• Understand why having an inclusive mindset is integral to being a SSBC coach  
• Understand the need to learn and act with empathy when working with clients, regardless of their background and life experiences | Narrated video, role-play videos, and knowledge check questions |
| Module 2             | SSBC session content, taking client measurements, tools associated with the program (eg, health tracking mobile app), and exercise protocols | • Describe the content that comprises each of the 6 SSBC counseling sessions  
• Understand the program tools and session documentation requirements  
• Know the difference between SSBC’s moderate-intensity continuous training and high-intensity interval training exercise sessions and the weekly progressions for each  
• Know which measurements to perform on your clients and when to take them | Narrated videos, demonstration videos, and knowledge check questions |
| Module 3             | Introduction to MI, definition of MI, and spirit of MI | • Define MI and identify where it falls on the spectrum of counseling styles  
• Describe each of the 4 elements of the spirit of MI  
• Understand how to implement the spirit of MI in SSBC sessions | Narrated videos, role-play videos, and knowledge check questions |
| Module 4             | MI skills: open-ended questions, affirmations, reflections, summaries (OARS\textsuperscript{d}), and ask-tell-ask | • Define and use the OARS skills  
• Describe and use the method “ask-tell-ask” to provide information or advice to your clients | Narrated videos, role-play videos, and knowledge check questions |
| Module 5             | Listening and responding to change talk, sustain talk, and ambivalence | • Define change talk and sustain talk  
• Define ambivalence  
• Describe preparatory change talk and mobilizing change talk  
• Understand which OARS skill will work best when a client engages in change talk, sustain talk, or ambivalence | Narrated videos, role-play videos, and knowledge check questions |
| Module 6             | 4 processes of MI: engaging, focusing, evoking, and planning | • Describe the 4 processes of MI  
• Understand how to use them in sessions with your clients | Narrated videos, role-play videos, and knowledge check questions |
| Module 7             | Putting it all together: how to deliver the SSBC content using MI | • Describe the elements of an SSBC session  
• Understand how to provide SSBC content in a way that embodies the spirit and skills of MI | Narrated videos, role-play videos, and knowledge check questions |
| Resource Centre      | Additional information on topics covered in modules, sessions scripts, session checklists, and video role-plays | • N/A\textsuperscript{e} | PDFs and role-play videos |

\textsuperscript{a}SSBC: Small Steps for Big Changes.  
\textsuperscript{b}T2D: type 2 diabetes.  
\textsuperscript{c}MI: motivational interviewing.  
\textsuperscript{d}OARS: open-ended questions, affirmations, reflections, summaries.  
\textsuperscript{e}N/A: not applicable.

**Test Usability**

Results from the initial usability tests conducted by the research team and the SSBC coach partners showed high levels of learnability, efficiency, memorability, and satisfaction within the provided feedback. Some errors were recorded including technical bugs, which were corrected by 3C Institute.
Discussion

Principal Results

This paper provides a clear and transparent outline for the needs analysis, design, development, and initial usability testing of the SSBC coach training e-learning platform. Importantly, this process demonstrates how an IKT approach can be used with the TEL evaluation framework. To our knowledge, this is one of the first e-learning platforms that has been developed using an IKT approach. Engaging end users through this process increases the chances that they will find the e-learning platform useful, acceptable, and appropriate [15,16].

The development and use of e-learning platforms for SSBC coaches will improve the viability of the program as new coaches will be able to take the training from any location and at a time that is convenient for them. This process can be used to inform other T2D prevention and health programs currently training coaches in person. e-Learning has become a popular education tool; however, there is still a lack of meaningful engagement with end users and research outlining the development process.

A preliminary study demonstrated high levels of satisfaction and gain in coach knowledge after completing this SSBC e-learning platform [17]. The next steps include implementing and evaluating the SSBC coach e-learning platform to examine the real-world applicability and effects of the platform using the TEL evaluation framework phases 4-7. This will include documenting key events during implementation, assessing new coaches’ experience and satisfaction with the e-learning platform (ie, user-friendliness and appropriateness of the platform), assessing learning outcomes (Kirkpatrick levels 2-4) through changes in knowledge, behavior, SSBC client outcomes, and finally, estimating the total costs to develop and implement the e-learning platform, as well as determining viability.

Beyond the use of SSBC, other programs should consider an IKT approach in developing e-learning platforms and transparently report the process. Once training programs have been developed and implemented, it is crucial that research teams monitor coaches’ knowledge, skills, and behaviors and report on the fidelity of their coaches [18].

Limitations

Only a small number of previous SSBC coaches were involved in the design and development of this e-learning platform, and their preferences for content and delivery style might not be reflective of that of all coaches who will take this e-learning training. However, working with 3C Institute allowed us to incorporate their expertise in e-learning platforms, content, and learning styles. This study does not examine the effectiveness of this training, and at this point, we cannot be certain that this training will improve coaches’ knowledge, skills, and behaviors. Future research will look at these learning outcomes.

Conclusions

This is the first paper to outline the needs assessment, design, and development of an e-learning platform for a T2D prevention program. One of the main limitations of the current T2D prevention programs is the limited reporting of coach training modes and training fidelity. This paper demonstrates how an e-learning platform can be designed and developed for T2D prevention program coaches using an IKT approach combined with the TEL evaluation framework. It is expected that this training will be acceptable and effective because of the methods and approaches used in this study.

Acknowledgments

The authors would like to acknowledge the Small Steps for Big Changes coaches for their important input on this project and the 3C Institute as the software developer and host of the e-learning training platform. This work was supported by a Canadian Institute for Health Research Grant (018647). This work was also supported by a Social Sciences and Humanities Research Council of Canada doctoral fellowship and the WorkSafe BC Doctoral Research Training Award.

Data Availability

The data sets generated and analyzed during this study are not publicly available to protect the intellectual property of SSBC but are available from the corresponding author on reasonable request.

Authors’ Contributions

KDC conceived the study, conducted the focus groups, led analysis, and wrote the paper. NJG assisted with the focus groups, data analysis, development, and usability testing. TED assisted in study conception and development. MEJ assisted with study conception and oversaw the project. All authors reviewed and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots from the e-learning platform.

[DOCX File, 4863 KB - formative_v811e50942_app1.docx]
References

Abbreviations
IKT: integrated knowledge translation
MI: motivational interviewing
SSBC: Small Steps for Big Changes
tel: technology-enhanced learning
2D: type 2 diabetes

https://formative.jmir.org/2024/1/e50942
JMIR Form Res 2024 | vol. 8 | e50942 | p.2350
(page number not for citation purposes)
Evaluation of Eligibility Criteria Relevance for the Purpose of IT-Supported Trial Recruitment: Descriptive Quantitative Analysis

Romina Blasini1, MSc; Cosima Strantz2, MSc; Christian Gulden2, PhD; Sven Helfer3, MSc, MD; Jakub Lidke4, MSc; Hans-Ulrich Prokosch5, PhD; Keywan Sohrabi5, PhD; Henning Schneider1,5, MD

1Institute of Medical Informatics, Justus Liebig University, Giessen, Germany
2Department of Medical Informatics, Biometrics and Epidemiology, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany
3Department of Pediatrics, Medical Faculty and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany
4Data Integration Center, Medical Faculty, Philipps University of Marburg, Marburg, Germany
5Faculty of Health Sciences, Technische Hochschule Mittelhessen University of Applied Sciences, Giessen, Germany

Abstract

Background: Clinical trials (CTs) are crucial for medical research; however, they frequently fall short of the requisite number of participants who meet all eligibility criteria (EC). A clinical trial recruitment support system (CTRSS) is developed to help identify potential participants by performing a search on a specific data pool. The accuracy of the search results is directly related to the quality of the data used for comparison. Data accessibility can present challenges, making it crucial to identify the necessary data for a CTRSS to query. Prior research has examined the data elements frequently used in CT EC but has not evaluated which criteria are actually used to search for participants. Although all EC must be met to enroll a person in a CT, not all criteria have the same importance when searching for potential participants in an existing data pool, such as an electronic health record, because some of the criteria are only relevant at the time of enrollment.

Objective: In this study, we investigated which groups of data elements are relevant in practice for finding suitable participants and whether there are typical elements that are not relevant and can therefore be omitted.

Methods: We asked trial experts and CTRSS developers to first categorize the EC of their CTs according to data element groups and then to classify them into 1 of 3 categories: necessary, complementary, and irrelevant. In addition, the experts assessed whether a criterion was documented (on paper or digitally) or whether it was information known only to the treating physicians or patients.

Results: We reviewed 82 CTs with 1132 unique EC. Of these 1132 EC, 350 (30.9%) were considered necessary, 224 (19.8%) complementary, and 341 (30.1%) total irrelevant. To identify the most relevant data elements, we introduced the data element relevance index (DERI). This describes the percentage of studies in which the corresponding data element occurs and is also classified as necessary or supplementary. We found that the query of “diagnosis” was relevant for finding participants in 79 (96.3%) of the CTs. This group was followed by “date of birth/age” with a DERI of 85.4% (n=70) and “procedure” with a DERI of 35.4% (n=29).

Conclusions: The distribution of data element groups in CTs has been heterogeneously described in previous works. Therefore, we recommend identifying the percentage of CTs in which data element groups can be found as a more reliable way to determine the relevance of EC. Only necessary and complementary criteria should be included in this DERI.

(JMIR Form Res 2024;8:e49347) doi: 10.2196/49347
KEYWORDS
CTRSS; clinical trial recruitment support system; PRS; patient recruitment system; clinical trials; classifications; data groups; data elements; data classification; criteria; relevance; automated clinical trials; participants; clinical trial

Introduction

Background

Clinical trials (CTs) are key to medical progress as they are used to implement a new therapy, a medical device, a diagnostic procedure, or a preventive measure [1,2]. CTs thus form an essential component of “translation,” the transfer of findings from basic medical research to clinical application [3].

An important part of planning a CT is to define criteria that all participants must meet. The main goal of these inclusion—and exclusion—or eligibility criteria (EC) is to specify the CT’s target population (ie, patients who have specific conditions and might benefit from the studied therapy). EC are also used to minimize disruptive factors that are under suspicion to interfere with the CT objectives. Additionally, EC consider individuals for whom a CT could pose a health risk, such as pregnant women. Finally, some EC are necessary for legal or organizational reasons [4,5].

The successful implementation of CTs depends on the recruitment of a suitable number of participants who fulfill all EC. Insufficient participant recruitment is the foremost reason for the premature discontinuation of CTs, which raises ethical concerns because participants are exposed to risk, without potential benefits. Furthermore, the extension of the recruitment period is also associated with significant financial costs and is consequently inefficient [6-10].

Clinical Trial Recruitment Support Systems

Identifying individuals who fulfill all the EC of a study and are willing to enroll is challenging and time-consuming. In many cases, trial personnel manually search for suitable candidates in electronic health records (EHRs) [9,11]. A clinical trial recruitment support system (CTRSS), also called a patient recruitment system (PRS), can assist in increasing participant numbers [12-14]. These systems simplify study participant identification with secondary use of data already collected for care and billing purposes in clinics [15] and work by comparing EHR data with the specified EC of CTs. Most CTRSSs described in the recent literature are implemented for only 1 specific trial, medical department, or clinic, but there are also some approaches to develop a CTRSS that can be used for a wide range of CTs [14,16-22].

Several key considerations are necessary for the successful deployment of a CTRSS. First, to implement a CTRSS, it is necessary to have patient records as well as EC in machine-readable format to perform a comparison of both and create a list of potential participants [23,24]. The formatting of EC in machine-readable form depends on the underlying technology used. For instance, in a database-oriented system, the criteria can be developed using Structured Query Language (SQL). Previous research efforts have used ATLAS software (Observational Health Data Sciences and Informatics [OHDSI]) for this purpose. Artificial intelligence (AI)–based methods exist for translating ethical considerations from study protocols into ATLAS software, which can partially automate the process [25].

As previously stated, medical data from hospitals or medical centers are used to compare a CTRSS with the EC and shortlist individuals from the data pool who meet the EC for the study [26]. Consequently, a CTRSS can only specify search criteria that correspond to the existing data pool. Hence, to ensure efficient prefiltering of the data pool, the CTRSS must contain the maximum number of desired search criteria.

The consolidation of medical data in a centralized format remains a significant challenge in many systems due to limited data availability [27]. Additionally, many hospitals lack a retrievable centralized system for all accumulating medical data. Despite ongoing efforts to harmonize data, these approaches are not yet widely accessible. Therefore, establishing a connection to a comprehensive data repository that facilitates all the necessary search criteria is crucial in implementing a CTRSS, and a predetermined list of search criteria is essential.

Eligibility Criteria

To characterize the necessary data for CTRSS implementation, multiple studies have examined the prevalence of data element groups in the EC of CTs. Even though the results of these examinations are heterogeneous, it is evident that diagnosis is the most frequent data element used in official study protocols, followed by data about therapies, medications, and diagnostic results [25,28-30].

When searching for potential study participants, there is often a need to manually search through a large number of patient files. Study personnel typically start by making a preselection. Initially, they check the EC that they consider most important, as this helps narrow down the pool of potential CT participants effectively. Some EC can only be assessed right before including participants in the study or necessitate a personal evaluation by the trial staff. One such criterion is the consent form that participants are required to sign during the inclusion process. Not all the EC specified in the study protocol are likely used for preselection in the context of a CTRSS [11]. When implementing a CTRSS, it is sufficient that only relevant data element groups be queried to obtain appropriate suggestions [30].

Objectives

In previous work, we identified which data element groups are most commonly used in CT EC [30]. In this study, we investigated which of the data element groups identified in the previous studies, mentioned in the Eligibility Criteria section, are relevant in practice [25,28-30].

Another objective was to categorize these EC according to their underlying data element in order to identify the element groups, such as diagnosis, laboratory values, or demographics, that are most commonly used for patient recruitment, as well as those
that are mostly irrelevant to a CTRSS. Since the use of different data elements and search algorithms is strongly influenced by the availability of this information in EHRs, we also investigated when a data element needs to be checked but is not available in the patient’s EHR.

The overall goal was to determine the relevance of data element groups for use in a CTRSS. Therefore, we wanted to find out how often different data element groups occur in CTs and how often the groups are considered relevant. We also wanted to determine how many studies these data element groups occur in.

### Methods

#### Participants

Two groups of participants were enrolled in the study. The first group included 1 or more project participants (PPs) at each site who were responsible for data collection. These individuals had deep knowledge of medical informatics in general and were also involved in the development of a CTRSS.

The second group of participants were trial staff actively working in the field of study recruitment. They were referred to as trial professionals.

#### Data Collection Sheet

To assess the relevance of data element groups, we first developed a data collection sheet to capture relevant information from the trial professionals. We wanted to capture all EC of the selected CTs in their original format, the underlying data element, and the assessment of study personnel in terms of relevance to patient recruitment. The development of the data collection sheet was an iterative process in which the categories were first discussed in a group of 10 CTRSS developers and then tested by 2 persons of this group on 3 randomly selected studies. In the next step, the group discussed any issues that arose. After 3 iterations, we achieved full agreement among all testers and developers.

For simple but unambiguous categorization, we used the 40 most common data element groups of EC from our previous work [30] and combined some rarely used groups into broader categories (eg, special laboratory information into a category laboratory value). If none of the given data element groups were appropriate, it was also possible to select a broader category, such as other procedural information, and provide a more specific description as a comment. These categories are then strongly linked to the data in EHRs and can therefore provide more insight into the possibilities of accessing elements of EC in clinical systems. For a complete list of all data element groups and other information composed on the data collection sheet, see Multimedia Appendix 1.

EC were classified into 3 CTRSS relevance categories: necessary, complementary, and irrelevant. Necessary items are those that determine the main selection of the desired cohort, complementary criteria in a manual process are mostly used in a second step to obtain more precise results, and irrelevant criteria are not used at all. Because there are criteria that may be important for participant selection but are not regularly documented in the EHRs and therefore must either be known by the treating physician or verified by direct questioning of the patient, we added the categories “necessary, not documented” and “complementary, not documented.” In addition, there are various reasons criteria may be irrelevant. Therefore, we decided to add the categories “irrelevant, redundant” and “irrelevant, recorded at the time of enrollment.” The category descriptions are summarized in Table 1.

<table>
<thead>
<tr>
<th>Relevance category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necessary</td>
<td>Used for the main filter criterion, for example, the main diagnosis under investigation</td>
</tr>
<tr>
<td>Necessary, not documented</td>
<td>Used for the main filter criterion but cannot be checked on paper or digitally</td>
</tr>
<tr>
<td>Complementary</td>
<td>Used for the secondary filter criteria used to achieve more precise search results</td>
</tr>
<tr>
<td>Complementary, not documented</td>
<td>Used for the secondary filter criteria but cannot be checked on paper or digitally</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>Not relevant for the participant search</td>
</tr>
<tr>
<td>Irrelevant, recorded at the time of enrollment</td>
<td>Not relevant, because it is only relevant after the initial participant search during the process of enrollment</td>
</tr>
<tr>
<td>Irrelevant, redundant</td>
<td>Used for criteria listed twice (duplicated), one time marked as redundant</td>
</tr>
</tbody>
</table>

In addition, we had a field to note surrogate data element groups that could be used if a criterion was not documented in the EHR system. For example, if a particular condition is not documented in a timely manner and there is a lab result that is indicative of that condition, that lab result can potentially be used as an alternative data element.

Taking an example CT on diabetes with the inclusion criteria of a diagnosis of diabetes, an elevated laboratory value, and consent to participate in the CT, as well as the exclusion of drug abuse, this can be classified as shown in Table 2. In this example, the presence of diabetes is the main inclusion criterion, so it is classified as “necessary” and the laboratory value is classified as “complementary,” since this criterion usually applies to all patients with diabetes. Alcohol abuse can be diagnosed in the medical history but is usually not documented in the patient’s record, and therefore, it is an additional filter criterion but cannot be verified by inspection of the file. Alternatively,
it is possible to find notes about possible alcohol abuse in the records of the patient’s medical history.

### Table 2. Sample representation of a completed data extraction sheet.

<table>
<thead>
<tr>
<th>Original description</th>
<th>Content (simplified)</th>
<th>Data element</th>
<th>Relevance assessment</th>
<th>Surrogate parameter</th>
</tr>
</thead>
</table>
| Participants with a confirmed diagnosis of type 2 or type 1 diabetes | Type 1 or 2 diabetes | Diagnosis | Necessary | —
| Patients with controlled diabetes (HbA1c<9%) | HbA1c<9% | Laboratory result | Complementary | —
| Willing to take part in the trial | Consent | Informed consent | Irrelevant, recorded at the time of enrollment | —
| History of drug abuse within 1 year prior to screening | Drug abuse | Diagnosis | Complementary, not documented | Other medical history

*aNot applicable.

*bHbA1c: glycated hemoglobin.

### Selection of Trials

We collected CT information from 8 university hospitals in Germany between December 2021 and March 2022 using the data collection sheet described in the Data Collection Sheet section.

The participating centers conducted the CT selection, including only CTs that recruited participants prospectively. This means that the CTs actively searched for individuals and asked them to participate. Animal, biomaterial, and case-control CTs were excluded. Psychiatric and oncological CTs were also excluded due to organizational reasons. There were no further restrictions on the selection of medical specialties.

Trial centers were contacted by the PPs to inquire about participation in the study. If a positive response was received, the CTs recommended by the trial personnel were incorporated into the analysis. The trial professionals were also asked to participate in this study to obtain a real-world view of the CT recruitment process. The PPs had either face-to-face or video meetings with the trial professionals to discuss the process of identifying potential participants and to categorize each individual criterion based on both parties’ experiences. If a positive response was received, the CTs recommended by the trial professionals were discussed and incorporated into the analysis.

### Data Validation and Analysis

All PPs were trained in a common training session where the data collection sheet was presented and tested. To ensure the correct use of the data element groups and the CTRSS relevance categories, we performed an additional step to validate the collected data: 2 authors went through all records and checked for consistency and face validity. If the entries were not fully understandable, they contacted the responsible PPs and discussed the case until agreement was reached. Despite the validation step, the distribution of relevance categories remained unchanged. However, the distribution of data element groups changed. All steps are shown in Figure 1.

### Statistical Analysis

Statistical analysis and graphing were performed using R (R Core Team and the R Foundation for Statistical Computing) [31]. Due to a high number of items in the collection categories such as “Other” or “Other medical history,” we decided to group the corresponding items into new categories. This work was performed by 2 authors to avoid errors.

### Ethical Considerations

Data capture and analysis only included individuals associated with the project. No Institutional Review Board approval was requested, as this would only be appropriate for studies including direct contact with patients [32], but we did not capture or process any patient-related data, so informed consent was not required.

### Results

### Studies Investigated

Figure 2 shows the departments to which the reviewed studies belong. Although no prior decision was made as to which departments to select, there was an overrepresentation of neurological studies.
Figure 2. Distribution of the medical specialties among the analyzed studies, sorted in descending order: neurology (n=30, 36.6%), cardiology (n=8, 9.8%), urology (n=7, 8.5%), general surgery (n=7, 8.5%), dermatology (n=6, 7.3%), gastroenterology (n=5, 6.1%), orthopedics (n=3, 3.7%), pulmonology (n=2, 2.4%), neurosurgery (n=2, 2.4%), gynecology (n=2, 2.4%), diabetics (n=2, 2.4%), anesthesiology (n=2, 2.4%), trauma surgery (n=1, 1.2%), rheumatology (n=1, 1.2%), and other (n=1, 1.2%).

Data Element Groups

In total, we included 82 CTs from 8 different university hospitals, and at each site, 3-25 (4%-30%) CTs were processed. In total, we identified 1157 EC, of which 1132 (97.8%) unique criteria remained after the removal of duplicates, which were classified as “irrelevant, redundant.”

Figure 3 shows the frequency of data element groups in all examined EC. With 28.4% (321/1132) of all EC, diagnosis was, by far, the most common data element, followed by informed consent and date of birth/age in second and third places, respectively.
Relevance Categories

In terms of EC relevance, 350 (30.9%) of the 1132 EC were categorized as “necessary,” 224 (19.8%) as “complementary,” 217 (19.2%) as “not documented,” 52 (4.6%) as “irrelevant,” and 289 (25.5%) as “irrelevant, recorded at the time of enrollment.” The overall percentages are shown in Figure 4.

Figure 4. Distribution of relevance categories of all EC: 350 (30.9%) “necessary” criteria, 109 (9.6%) “necessary, not documented,” 224 (19.8%) “complementary,” 108 (9.5%) “complementary, not documented,” 52 (4.6%) “irrelevant,” and 289 (25.5%) “irrelevant, recorded at the time of enrollment.” EC: eligibility criteria.

Total Irrelevant Criteria

In total, 341 (30.1%) EC were categorized as “irrelevant” and “irrelevant, recorded at the time of enrollment.” These were mainly patient consent information (n=90, 26.4%). In addition, the EC “other” (n=42, 12.3%) and “other medical history” (n=26, 7.6%) were often marked as not relevant for patient screening. Diagnosis was only a small part (n=24, 7%) of all irrelevant EC, which is low compared to the general data element distribution shown in Figure 3, where this category makes up the largest part. A visualization of the complete list and frequencies can be found in Multimedia Appendix 2.

Nondocumented Criteria

In total, 217 (19.2%) necessary and complementary EC were categorized as not documented and were most commonly the data element groups diagnosis (n=121, 55.8%), examination results (n=36, 16.6%), and procedures (n=28, 12.9%). As noted in the example in Table 2, it is possible that data element groups such as diagnosis were sometimes categorized as documented and sometimes with other categories, which is valid because it may depend on the medical context of what information is documented. All data element groups and their frequencies can be found in Multimedia Appendix 3.

All Relevant Criteria

As defined before, both complementary and necessary EC were relevant for an automated search of potential participants, which were 574 in number (Multimedia Appendix 4). The most frequently used data element groups were diagnosis code, which accounted for 44.4% (n=255) of all necessary EC, followed by date of birth (n=74, 12.9%) and procedure codes (n=48, 8.4%).

Relevance Distribution by Data Element Group

Since each criterion was individually classified into a data element group and a relevance category, it is possible that common data element groups play a role in several of the categories. For this reason, the categorization between data element groups was heterogeneous. For example, the data element group diagnosis was mostly categorized as necessary or complementary (n=255, 79.4%), while the data element group informed consent was mostly categorized as “irrelevant, recorded at the time of enrollment” (n=90, 92.8%). In addition, the data element groups pregnancy, contraception, and lactation were mostly either not documented or documented at the time of enrollment. Figure 5 shows the relevance distribution for all data element groups.
Figure 5. Distribution of relevance categories in percentage by data element group; the groups were recorded at least 5 times in our data set, ordered by the proportion of necessary EC: date of birth/age, other details of encounter, diagnosis, and diagnosis date showed the highest proportion of necessary EC. EC: eligibility criteria.

Data Element Relevance Index

More important than the absolute frequency of a data element is the question of how many CTs the element is used for in patient screening and not how often a group is represented in a CT. This parameter, the data element relevance index (DERI), can be calculated without considering the frequency of data element groups in CTs.

For the calculation, we determined the number of CTs in which the data element was used at least once and removed all entries marked as undocumented or irrelevant. The results showed that the data element groups diagnosis and date of birth/age were present in more than 50% of the CTs: diagnosis, n=79 (96.3%); date of birth/age (85.4%). All other DERI values are shown in Figure 6.
Surrogate Parameters

For the case of unavailability of an original data element, 92 (8.1%) of the EC were documented surrogate parameters to use. Surrogate parameters were often used for the grouping categories “Other medical history” (n=14, 15.2%) and “Other diagnosis information” (n=13, 14.1%). Diagnosis had the highest frequency in both original elements (n=20, 20.7%) and surrogates (n=52, 56.5%). As a surrogate data element for diagnosis, procedures (n=5, 25%) and laboratory results (n=5, 25%) were most often used.

Discussion

Principal Findings

EC have been analyzed and categorized repeatedly in recent years to measure their prevalence in CTs. These studies have mostly categorized EC into semantic categories [33]. Comparing the studies, we saw that the frequency of semantic categories shows some overlap but also varies to some extent [28]. This could be due to the selection process of the studies or a different way of categorization by the diverse researchers.

Data Element Groups

Although the categorization by Luo et al [33] focuses on the semantic categories of EC rather than individual data elements, as in this paper, parallel categories exist that we used to find similarities and differences. Previous studies have measured the prevalence of data element groups as a percentage of all EC examined [25,28-30,33].

Our measured prevalence of similar data element groups differed by no more than 0.5 from the minimum and maximum of the measured frequency percentages of comparable studies. This comparison using the semantic categories described by Luo et al [33] is provided in Multimedia Appendix 5. An exception is the consent category, which we used more often. In our review, consent was often used more than once for a CT, not only to describe consent to the CT itself, but also when a participant was asked to consent to specific procedures or circumstances of the CT. This included, for example, consent to use adequate contraception for the duration of the CT. The high deviation from other studies may be due to the fact that other studies have only used this data category for specific CT consent [25,28,29].

Relevance Categories

The results show that about 70% of the examined EC are relevant for the selection of potential CT participants. About 19% are not usually documented in EHRs and therefore cannot be used for filtering, even if the trial professional searches all patient records. Whether a criterion is classified as undocumented depends, in part, on the capabilities of hospital information systems and therefore varies from hospital to hospital. In addition, certain information may only be collected depending on the context of care and the severity of illness.

About 51% of all EC are relevant for the electronic prefiltering of patients. Since the implementation of EC is one of the obstacles in the development of a CTRSS, the realization that only about half of the EC are relevant can mean a simplification in the implementation of EC in the systems.

Figure 5 shows that it is not possible to link data element groups only to 1 relevance category. Instead, depending on the context of the CT, different relevance categories were assigned to the data element groups. In some cases, the same group was sometimes categorized as documented and sometimes as undocumented. This can be explained by the fact that whether information is documented depends not only on the data element group but also on several other factors. For example, often, only diagnoses that are considered important and investigated in the context of therapy are documented. In these cases, the trial
professionals may be able to make an assessment because they are either part of the patient care or are in close contact with the treatment physicians and are therefore aware of these details for documentation purposes. Figure 5 can provide some guidance as to which data element groups are typically not needed, but an assessment should always be made directly by the trial professionals.

In previously published studies on the prevalence of EC, this was usually also referred to as their relevance. Our results show that a significant proportion of EC is not used to search for participants. Since the classification according to relevance categories varies between the data element groups, as shown in Figure 5, it is apparent that filtering according to relevance criteria affects not only the number of EC but also the frequency distribution of the data element groups.

For this reason, we compared the general frequency of data element groups (Figure 3) with those categorized as necessary or complementary (Multimedia Appendix 4). From this comparison, we saw that the percentage prevalence of all data element groups is sometimes different from the prevalence of the relevant data element groups. Diagnosis and date of birth/age as well as procedures show a particularly strong increase in frequency. However, consent, other medical history, and pregnancy are significantly less frequent among the data elements classified as relevant.

Data Element Relevance Index
The frequency order of data element groups depends on how they are viewed. If the EC criteria are filtered based on their relevance in prefiltering possible participants, the frequency order changes.

The introduction of a new DERI measurement value changes the perspective. The occurrence frequency of data element groups, such as various laboratory values, in a study is no longer relevant. The consideration now lies only in the frequency of the respective group’s use in studies. Studies with a significant number of exceptional cases exert less influence on the general outcome.

Furthermore, by using the DERI value, it is possible to determine the number of studies reliant on a specific data element group. Consequently, a direct inference can be made about its impact within a CTRSS, which is unachievable through pure frequency data.

Data Completeness
In 2018, Vass et al [34] examined the data quality of data element groups commonly used in CT EC in 10 university hospitals in Germany. They found that the data completeness is partly heterogeneous and that elements are rarely collected in a structured way in daily clinical practice, which hinders automated retrieval [34].

Comparing the DERI values of the data element groups and the completeness of data in the EHRs (Figure 7), we saw that more than 80% of the information is available for the 3 most relevant data element groups (diagnoses, demographics, and procedures). The less relevant categories of laboratory results and other diagnostic information are still available in more than 65% of cases. However, medication information (medication history and current medication) is problematic, with availability ranging from 13% to 61%. Nevertheless, medication history and current medication are relevant groups, with DERI values of 13% and 26%, respectively. Scores are also poorly documented, with a data completeness of 18%.

Figure 7. Data completeness measured by Vass et al [34] in comparison with DERI. It can be seen that although diagnosis and date of birth/age both have a high data completeness rate and DERI, problems arise with scores, medication anamnesis, and current medication, which have a low data completeness rate. OPS: Operationen- und Prozedurenschlüssel.
To reliably use these data in a CTRSS, a higher level of completeness should be present for DERI values higher than 10%. Poor data quality of data from EHRs, when matched with EC from CTs, can lead to high false-negative rates for inclusion criteria and high false-positive rates for exclusion criteria. The former, in particular, is fatal to the use of a CTRSS, as matching individuals are thus overlooked. High false-negative rates, in turn, lead to increased workload for trial personnel.

The study by Vass et al [34] showed that some data element groups are well captured in EHRs but other elements are problematic. This generates a data gap between EC and EHRs, which was analyzed by Butler et al [35]. They found that about 40% of all EC are not captured in EHRs. Since a structured collection of clinical patient-related information is important not only for the implementation of a CTRSS but also for billing purposes and patient safety, hospitals as well as governments are pushing the digitalization of patient records. In recent years, many initiatives and laws have been implemented to make data from EHRs available for research purposes [36,37]. Therefore, it is likely that the accessibility of medical information has recently improved or will improve in the future.

The timeliness of data element accessibility was not considered here but should be in future studies to assess not only whether a data element is documented in EHRs but also whether this data element is accessible in time. We should also examine whether the results of Butler et al [35] are reproducible when only relevant criteria are examined.

Implementation of Patient Recruitment Systems

In previous publications, the relevance of data element groups has been described as the frequency of a group relative to the number of all data elements found in all EC, which depends only on the distribution of data element groups. Since the prevalence of EC is defined heterogeneously in the literature, it is not useful to assess the relevance of data element groups based on the frequency distribution alone.

Especially for the implementation of a CTRSS, it is useful to evaluate how many CTs use a data element group. Using the DERI to determine the relevance of data element groups can be helpful here, as it additionally includes only groups that are used to search for participants.

The core functionality of a CTRSS is to compare EC with clinical data to identify potential participants. For the system to be useful to trial staff, the suggestions must be as accurate as possible, with minimal false-positive or false-negative rates. Since the comparison between the clinical database and EC is the critical factor at this point, it is particularly important that as many EC as possible be checked automatically. When implementing a CTRSS for different types of CTs, it is recommended to first identify important data element groups by determining a local DERI for all data element groups. Therefore, a set of CTs should be analyzed in cooperation with the trial sites, and all data element groups with a high DERI should be accessible to the CTRSS. For this purpose, it is useful to set a threshold value for the DERI determined and to make available all data element groups that exceed this value. For this purpose, we could start with a value of 10, since this means that a data element group is used in at least 10% of the CTs, and gradually increase this value. Further studies are needed to determine the quantitative relationship of the DERI to the results of a CTRSS.

In contrast to the determination of local DERI values, there is the additional task of determining the data quality of the available data sources to be used for matching with the EC of CTs. Again, data completeness and timeliness should be determined locally to ensure that the necessary data element groups identified are available in the highest-possible data quality.

Limitations

Since the PPs at the study sites were allowed to select participants themselves, the selection of participants was through the convenience sampling method. Randomization was not possible; instead, all participants who were currently supervising at least 1 CT and who were willing to participate in the study were included. Additionally, we could see an overrepresentation of neurological CTs in our sample and had to exclude oncological and psychiatric CTs.

The categorization of EC was performed with the cooperation of CTRSS experts and trial professionals at each site. To minimize the bias of different categorization methods, all the categorization of data element groups was validated by 2 experts.

The results presented here are dependent on the study sites selected. The clusters we identified may be prone to local variation, and their applicability to other sites is unclear.

Conclusion

In this study, we demonstrated that automated recruitment support of CT personnel requires only roughly 50% of the EC indicated in the CT protocols. Since the frequency of EC in CTs is described differently in the literature, exclusively focusing on the frequency of EC is misleading. Instead, we propose to define the relevance of EC as the proportion of CTs in which a criterion occurs. In addition, only EC considered relevant (necessary or complementary) for patient recruitment should be included for this determination. This DERI can be used to quantify the relevance of EC data element groups.

Further examination is necessary to find out whether the relevant data element groups are documented in EHRs in a structured way and accessible in time for the implementation of a CTRSS.

Acknowledgments

This study was conducted in the Medical Informatics in Research and Care in University Medicine (MIRACUM) Consortium. MIRACUM is funded by the German Ministry for Education and Research (funding number FKZ 01ZZ1801A/B/C/D/G/L/M).
We greatly appreciate the work of Ines Reinecke, Susanne Zabka, Phillip Macho, Alexandra Stein, and Frederike Elisabeth Euchner for their support in the development of data extraction sheets and in data collection. We also thank Fabian Wälldchen for his help with writing and proofreading.

Data Availability

We added the parameters of our data collection sheet, as well as all analysis and results, either in this paper or in the appendices. Additionally, all data our analysis is based on can be found in Multimedia Appendix 6. We only had to remove original trial descriptions, as they would allow a conclusion on concrete clinical trials and locations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Captured information.

[DOCX File, 14 KB - formative_v8i1e49347_app1.docx]

Multimedia Appendix 2
Irrelevant data element groups with frequency.

[ PNG File, 50 KB - formative_v8i1e49347_app2.png]

Multimedia Appendix 3
Not-documented data element groups with frequency.

[ PNG File, 48 KB - formative_v8i1e49347_app3.png]

Multimedia Appendix 4
Relevant data element groups with frequency.

[ PNG File, 45 KB - formative_v8i1e49347_app4.png]

Multimedia Appendix 5
Comparison of data element groups' distribution with other studies.

[ PNG File, 51 KB - formative_v8i1e49347_app5.png]

Multimedia Appendix 6
All analyzed eligibility criteria of all included trials.

[XLSX File (Microsoft Excel File), 38 KB - formative_v8i1e49347_app6.xlsx]

References


37. Hendolin M. Towards the European health data space: from diversity to a common framework. Eurohealth 2021;27(2):15-17 [FREE Full text]

Abbreviations

CT: clinical trial
DERI: data element relevance index
EC: eligibility criteria
EHR: electronic health record
OPS: Operationen- und Prozedurenschlüssel
PP: project participant

Edited by A Mavragani; submitted 25.05.23; peer-reviewed by W Wei, V Barber; comments to author 07.08.23; revised version received 28.09.23; accepted 22.11.23; published 31.01.24.

Please cite as:
Evaluation of Eligibility Criteria Relevance for the Purpose of IT-Supported Trial Recruitment: Descriptive Quantitative Analysis
JMIR Form Res 2024;8:e49347
URL: https://formative.jmir.org/2024/1/e49347
doi:10.2196/49347
PMID:38294862

©Romina Blasini, Cosima Strantz, Christian Gulden, Sven Helfer, Jakub Lidke, Hans-Ulrich Prokosch, Keywan Sohrabi, Henning Schneider. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

User Requirements in Developing a Novel Dietary Assessment Tool for Children: Mixed Methods Study

Zoë van der Heijden¹, MSc; Femke de Gooijer¹, MSc; Guido Camps¹, PhD; Desiree Lucassen¹, PhD; Edith Feskens¹, Prof Dr; Marlou Lasschuijt¹, PhD; Elske Brouwer-Brolsma¹, PhD

Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, Netherlands

Corresponding Author:
Zoë van der Heijden, MSc
Division of Human Nutrition and Health
Wageningen University & Research
Stippeneng 4
Wageningen, 6700 AA
Netherlands
Phone: 31 7480 100
Email: zoe.vanderheijden@wur.nl

Abstract

Background: The prevalence of childhood obesity and comorbidities is rising alarmingly, and diet is an important modifiable determinant. Numerous dietary interventions in children have been developed to reduce childhood obesity and overweight rates, but their long-term effects are unsatisfactory. Stakeholders call for more personalized approaches, which require detailed dietary intake data. In the case of primary school children, caregivers are key to providing such dietary information. However, as school-aged children are not under the full supervision of one specific caregiver anymore, data are likely to be biased. Recent technological advancements provide opportunities for the role of children themselves, which would serve the overall quality of the obtained dietary data.

Objective: This study aims to conduct a child-centered exploratory sequential mixed methods study to identify user requirements for a dietary assessment tool for children aged 5 to 6 years.

Methods: Formative, nonsystematic narrative literature research was undertaken to delineate initial user requirements and inform prototype ideation in an expert panel workshop (n=11). This yielded 3 prototype dietary assessment tools: FoodBear (tangible piggy bank), myBear (smartphone or tablet app), and FoodCam (physical camera). All 3 prototypes were tested for usability by means of a usability task (video analyses) and user experience (This or That method) among 14 Dutch children aged 5 to 6 years (n=8, 57% boys and n=6, 43% girls).

Results: Most children were able to complete FoodBear’s (11/14, 79%), myBear’s (10/14, 71%), and FoodCam’s (9/14, 64%) usability tasks, but all children required assistance (14/14, 100%) and most of the children encountered usability problems (13/14, 93%). Usability issues were related to food group categorization and recognition, frustrations owing to unsatisfactory functioning of (parts) of the prototypes, recall of food products, and the distinction between eating moments. No short-term differences in product preference between the 3 prototypes were observed, but autonomy, challenge, gaming elements, being tablet based, appearance, social elements, and time frame were identified as determinants of liking the product.

Conclusions: Our results suggest that children can play a complementary role in dietary data collection to enhance the data collected by their parents. Incorporation of a training program, auditory or visual prompts, reminders and feedback, a user-friendly and intuitive interaction design, child-friendly food groups or icons, and room for children’s autonomy were identified as requirements for the future development of a novel and usable dietary assessment tool for children aged 5 to 6 years. Our findings can serve as valuable guidance for ongoing innovations in the field of children’s dietary assessment and the provision of personalized dietary support.

(TMIR Form Res 2024;8:e47850) doi:10.2196/47850

KEYWORDS
diet; children; dietary assessment; recall; technological innovation; mobile health; mHealth; mobile phone

https://formative.jmir.org/2024/1/e47850

(JMIR Form Res 2024;8:e47850) doi:10.2196/47850

(keyword not for citation purposes)
Introduction

In the Netherlands, childhood overweight and obesity have reached an alarming prevalence of 15% in 2021, indicating a pressing public health concern [1]. Childhood obesity has a major impact on psychological health and later-life risks of developing noncommunicable diseases, mortality, and morbidity [2,3]. As a healthy diet is known to play a vital role in the prevention of obesity [4], interventions encouraging children toward healthier food choices receive high priority globally. However, the long-lasting effects of past and ongoing dietary interventions are rather limited, mainly related to their “one-size-fits-all” approach, which calls for more personalized interventions [5].

Personalized dietary interventions require accurate individual-level dietary assessment and monitoring [6] to facilitate realistic personalized dietary feedback. However, dietary assessment methods in young children are extremely challenging owing to their limited literacy, writing skills, food knowledge, and interest [7]. Consequently, caregivers currently serve as the primary sources of (surrogate) dietary information for their children. However, as primary school children gain independence in their food choices, the likelihood of misreporting increases [8]. Recent technological advancements now provide opportunities for a role for children themselves in the dietary assessment, which may serve the overall obtained dietary data quality.

Accordingly, there is growing interest in the development of innovative tools to assess dietary intake in children, with a particular emphasis on more effective and engaging technology-based solutions [9,10]. However, most of the developed tools thus far lack proper validation and are not tailored to the Dutch context, including Dutch food databases [11,12]. Country-specific dietary assessment tools are essential to accurately capture dietary information while considering cultural, regional, and nutritional differences. Moreover, given children’s rapid cognitive development, there is also a need for dietary assessment tools that align with their age-specific developmental stages. Current research has mostly focused on tools for children aged ≥8 years [11], as children tend to be better at independently reporting their food intake from this age onward [7]. However, considering the high level of technology readiness in today’s generation of children and the continuously changing technological possibilities, it is worth exploring the development of dietary assessment tools for younger children.

A “child-centered approach,” which places the user in the center of the design and development process, can effectively address young children’s age-specific cognitive needs for innovative dietary assessment. This approach enables researchers to understand the context, needs, and preferences of the tool’s intended end users [13,14] by engaging children in identifying challenges and finding solutions [15]. As a result, a child-centered approach can enhance design outcomes, improve user experience [16], and potentially improve data collection procedures and accuracy. In the decades marked by an increasing prevalence of childhood health issues, understanding the dietary behaviors and needs of young children is vital for informing effective interventions.

Therefore, as a first step, this study aimed to reveal user requirements for a novel child-friendly food intake registration tool designed for Dutch children aged 5 to 6 years. In pursuit of this goal, we developed and evaluated 3 distinct prototypes specifically created for children aged 5 to 6 years while considering age-specific cognitive and developmental characteristics to serve as valuable guidance for advancing the field of dietary assessment tools for children across a broader age range.

Methods

This study applied an exploratory sequential mixed methods study design, combining qualitative and quantitative measures [17]. Phases included a formative research phase (qualitative), a developmental phase (qualitative), and an evaluation phase (mixed methods, but with qualitative emphasis; Figure 1).

Figure 1. Flowchart illustrating the design process of the dietary assessment tool prototypes FoodBear, myBear, and FoodCam.

Formative Research

We performed a nonsystematic narrative literature research to identify existing dietary intake assessment tools for young children, their validity, and age-specific developmental considerations. This information was then used in an expert workshop to probe prototype idea generation. This nonsystematic search followed an abductive approach [18], focusing on extracting valuable insights to inform the design process effectively (eg, formulation of a list of user requirements and wishes for designing a dietary assessment tool for children aged 5-6 years; Textbox 1), rather than aiming for an exhaustive review of the available literature. The search query included a combination of the following Medical Subject Headings terms: dietary assessment, food intake, nutritional assessment, child centered design, child*, kid*, preschool*, child computer interaction, and eHealth. To our knowledge, this study is the first to focus on self-reported tool development among children.
aged 5 to 6 years. Therefore, our requirements are based on heterogeneous literature, including studies related to design for children [19-21], dietary assessment tools for older children outside the Netherlands [22-27], or child development [7,28].

In the context of this study, requirements were defined as being vital for usability and wishes as being desirable for enhancing usability and motivation among children. Our requirements were assessed in terms of perceived importance (ranging from 1 to 5) based on close consultation and consensus within our research team (Textbox 1). In the weighted decision matrix (WDM), decisions were rated on a scale of 1 to 5, with 1 indicating low importance and 5 signifying high importance for successful use.

Textbox 1. List of user requirements for a novel dietary assessment tool for children resulting from formative research and their importance (score ranging from 1 to 5). The list includes aspects that the dietary assessment tool should have (ie, requirements) or could have (ie, wishes).

<table>
<thead>
<tr>
<th>Requirements (importance score)</th>
<th>Wishes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Collect accurate and useful data on dietary intake in children (importance score 5) [7,22-26]</td>
<td>• Incorporate photography [22,27]</td>
</tr>
<tr>
<td>• Be understandable (ie, simple, easy-to-use, and intuitive; importance score 5) [26]</td>
<td>• Incorporate an avatar [21,25,26]</td>
</tr>
<tr>
<td>• Be fast paced (ie, completed in a short time; importance score 4) [26,28]</td>
<td>• Incorporate gamification [21,26]</td>
</tr>
<tr>
<td>• Give feedback and context-specific help (eg, auditory or visual; importance score 3) [19,20]</td>
<td>• Include a storyline [21,25,26]</td>
</tr>
<tr>
<td>• Be motivating and encouraging to use (importance score 3) [19,26,27]</td>
<td>• Include rewards [21]</td>
</tr>
<tr>
<td>• Be social (importance score 3) [19]</td>
<td>• Incorporate learning and/or repetitive elements [19]</td>
</tr>
<tr>
<td>• Be challenging (importance score 3) [19]</td>
<td></td>
</tr>
</tbody>
</table>

Development

Idea Generation

Concepts for dietary assessment tools were developed through a 1-hour web-based expert panel workshop hosting nutrition (n=4), design (n=4), behavior (n=1), and technology researchers (n=2). Experts were carefully selected from various universities in the Netherlands and came together on the web-based Miro whiteboard platform. The workshop unfolded in 3 key stages. First, the experts immersed themselves in the world of our target audience by engaging with emotional image prompts. Next, experts were presented with our comprehensive list of requirements, as detailed in Textbox 1, and tasked with generating prototype ideas that could address these requirements. In the third and final phase of the workshop, the experts collaborated in pairs to refine these ideas and transform them into feasible prototypes. The results of this collaborative effort produced a wide range of innovative concepts, including a food piggy bank, food camera, digital plate, Tamagotchi, smartwatch, and a food diary.

Prototype Development

The results of the expert panel workshop were scored and evaluated against the list of requirements and subsequently multiplied by their importance (ranging from 1 to 5) in a WDM [18] (Multimedia Appendix 1), a decision-making tool that can be used to evaluate a set of options against critical factors and compare design concepts based on the overall value of each design concept. Three researchers from Wageningen University and Research (WUR) completed the WDM individually to ensure objectivity. The 3 concepts were considered to align most closely with the requirements and subsequently further advanced, which resulted in the 3 functional prototypes, FoodBear (average WDM score: 144.1), myBear (average WDM score: 148.4), and FoodCam (average WDM score: 144.7; Figure 2; Multimedia Appendix 2).
Figure 2. From left to right, FoodBear, myBear, and FoodCam. myBear (in Dutch) shows the entry screen for lunch, where children can report their intake during lunch based on the food groups depicted in Textbox 1.

FoodBear can be used as a food recall or food record [9] and serves as a physical eating buddy in the shape of a bear. FoodBear can be fed with coins that match the foods consumed by the child (Table 1). A coin represents 1 item of the food group eaten (e.g., 1 slice of bread) and children are challenged to put the number of coins corresponding to the number of items eaten in the bear’s belly as an estimate of portion size. This allows the assessment of food group diversity and provides a rough estimate of dietary intake.

Table 1. Included food groups in myBear and FoodBear and their contribution to the lunch of Dutch children aged 5 to 6 years [24].

<table>
<thead>
<tr>
<th>Food group</th>
<th>Contribution to lunch (%)</th>
<th>Child-friendly categories in prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereals and grain products</td>
<td>43</td>
<td>Bread</td>
</tr>
<tr>
<td>Eggs</td>
<td>28</td>
<td>Eggs</td>
</tr>
<tr>
<td>Milk products</td>
<td>26</td>
<td>Milk, yogurt, and cheese</td>
</tr>
<tr>
<td>Soups, broth</td>
<td>23</td>
<td>Soup</td>
</tr>
<tr>
<td>Meat and meat products</td>
<td>20</td>
<td>Sausage or meat</td>
</tr>
<tr>
<td>Sugar and confectionery</td>
<td>15</td>
<td>Sweet toppings and candy</td>
</tr>
<tr>
<td>Fruits, nuts, and olives</td>
<td>14</td>
<td>Fruit</td>
</tr>
<tr>
<td>Different</td>
<td>13</td>
<td>Something different</td>
</tr>
<tr>
<td>Fish, crustaceans, and shellfish</td>
<td>12</td>
<td>Fish</td>
</tr>
</tbody>
</table>

myBear can be used as a food recall or food record [9] but in the form of a tablet-based app. The user interface design was developed with Adobe XD software, and the prototype app can be displayed on a smartphone or tablet. Children feed the bear with the same foods as they ate themselves. On the home screen of the app, children select the food groups they consumed, which then appear in the belly of myBear. The child uses plus and minus buttons to indicate the quantities of items from the food group they have consumed. Children receive a sticker on a digital sticker sheet after completing all entries. Similar to FoodBear, myBear can be used to provide a rough estimate of dietary intake and to track food group diversity.

FoodCam is based on the food record methodology [9] and consists of a camera and a “cookbook template.” The camera is specifically designed for children, and for this specific purpose, it was used to take pictures of food items. As the camera immediately prints the captured photo, it also provides immediate feedback to the child. Subsequently, the printed photos can be used by children to create their own cookbook, which offers the opportunity for children to draw and express their creativity. FoodCam can be used to assess daily food intake. FoodCam is accompanied by a cookbook stencil (in Dutch) on which children can put a picture of their lunch, draw it, indicate how full they feel [29], and how much they enjoyed it [30].

Prototype Content

The first prototypes were developed to assess lunchtime and focused on the most frequently consumed foods during lunchtime by young Dutch children according to the Dutch Food Consumption Survey [31] (Table 1). Only food groups that contributed for >10% to children’s lunch were included.

Recruitment

Overview

We recruited 14 Dutch (n=6, 43% girls and n=8, 57% boys) children aged 5 to 6 years through purposive sampling, as 10 (±2) participants were considered sufficient for usability evaluation [32]. Data saturation was assumed to be acquired within 6 to 12 sessions [33], which in this study was reached after 11 sessions. Children were recruited via colleagues within the Division of Human Nutrition and Health at WUR and through personal networks.
**User Testing**

The functional prototypes were evaluated on usability and user experience by combining qualitative and quantitative measures [17]. To make the child feel comfortable, the researcher visited the child at home in the presence of a parent or caregiver. Before the start of the session, the parent or caregiver was instructed to introduce the researcher as a toy inventor and that their child would act as an assistant inventor. Moreover, parents were instructed to interfere as little as possible during the user tests. The entire procedure took approximately 30 minutes and took place after lunch, between 1:30 PM and 3:00 PM. The 30-minute time frame was selected to align with the attention span of young children. We conducted 3 streamlined usability tasks (lasting a maximum of 140 seconds) and a user experience task and interview within this time frame to minimize any potential loss of interest or fatigue among participating children. To ensure that all tasks would fit within our proposed time frame, the procedures were piloted twice. All parents completed a demographic questionnaire on the child’s age (y), gender (boys or girls), number of siblings, siblings’ age (y), interactive screen time (h/d), foods and portion sizes eaten during lunch, and time of lunch. Parental lunch dietary intake data gathered in the questionnaire were used as the criterion to assess successful recall by the children, instead of more objective direct lunch observation, to be able to create a study environment that is comfortable and engaging for children. Testing order for FoodBear and myBear was alternated across participants. To assess the usability of FoodCam, parents were instructed to prepare a duplicate (ie, an “identical meal to what their child had eaten for lunch on the day of testing”) of the children’s previously consumed lunch during the test. This prototype was always tested last because FoodCam is the only prototype that does not rely on children memorizing their lunch.

The procedure consisted of 3 usability tests (steps 3, 5, and 6; Textbox 2) and a user experience test (steps 7–9; Textbox 2). To assess usability, children performed a task with every prototype while measuring the completion rate and task completion time. Completion rate was defined as the proportion of children that successfully completed the usability tasks, and completion time was defined as the time needed to complete the task. The time required for the researcher to explain or draw attention to the task was subtracted and the number of interruptions required to complete the tasks were registered. Behavioral observations and field notes were evaluated to identify the usability issues. A detailed description of the study procedures is provided in Textbox 2.

Textbox 2. Within-participant procedures consisting of 9 steps (maximum 30 min in total). Steps 3 and 5 are alternated across participants.

**Step 1: Introduction**
- The researcher engaged in a small talk with the child to build trust. Study procedures were explained in a child-friendly manner, and it was emphasized that the child could say anything.

**Step 2: Lunch recall**
- As the usability task of myBear and FoodBear required recall of the lunch, the child was asked to do this before starting these tasks. Lunch was considered as being correctly recalled when it resembled the lunch written down in the questionnaire by caregivers, in terms of food items and numbers. If the child was unable to recall his or her lunch independently, standardized help questions were asked: (1) “Did you eat bread?” (2) “How much bread did you eat?” (3) “What kind of topping did you eat? Cheese, meat or something sweet?” (4) “How much bread did you eat with this topping?” If the child answered a question with “no,” the following questions were prepared: (5) “Did you eat a salad, pasta or rice for lunch?” (6) “Did you drink something with your lunch? Milk, tea, or water?” (7) “Did you eat anything else, such as candy, fruit or soup?” The number of questions required was noted.

**Step 3: Usability task: myBear**
- The child was asked to provide myBear with the same foods as recalled in step 2 or 4. The session started by clicking the “lunch” button (ie, as one of 5 different eating moments), which started the time measurement. The time measurement ended once the last food item was entered in myBear. The assignment was completed when all food groups and amounts were correctly entered.

**Step 4: Lunch recall**
- To mitigate potential effects arising from the passage of time between recalling the lunch and subsequent assessment of prototype’s usability, the child was asked to recall the lunch again.

**Step 5: Usability task: FoodBear**
- The child was asked to give the same lunch to FoodBear as recalled in step 2 or 4. The time measurement started once the researcher asked the child to start feeding FoodBear and ended when the child put the last coin into its belly. The assignment was completed when all food groups and amounts were correctly entered.

**Step 6: Usability task: FoodCam**
- The child was asked to take a picture of his or her (duplicate) lunch with FoodCam. Time measurement started once the researcher handed over the camera to the child and ended when the child took the picture. The assignment was completed when the (1) photo was sharp and (2) included all consumed foods in a recognizable way. To assure objectivity, photos were assessed by 3 researchers.

**Step 7: This or That method**
- As a response to 5 “This or That” questions, children indicated which prototype they liked best [34]. The original This or That method uses pairwise comparison, but this study compared 3 prototypes. One of the original questions was considered irrelevant and excluded: “Which of these three would you most like to take home?” The following This or That questions were included: “Which of these three was most fun?” “Which of these three was a bit stupid?” “Which of these three was a little boring?” “Show me which of these three you would like to play again?” “Show me which of these three you would like to receive as a gift?” Children could indicate more than 1 prototype but were not told beforehand to facilitate decision-making.

**Step 8: Reward**
- The child received a biscuit and a stamp set to express gratitude for time investment and participation in the study.

**Step 9: Behavioral choice selection**
- The researcher told the child “that there was some time left,” and that he or she could select a prototype to play with again. The researcher ensured that the child ate the biscuit first to prevent the child from automatically choosing the prototype they played with last.

**Data Analysis**

**Usability**

All sessions were audio- and video-recorded and transcribed verbatim. The researcher watched the videos and documented the examples of interest. Using the qualitative data analysis software ATLAS.ti (ATLAS.ti Scientific Software Development GmbH), examples of interest were grouped into themes to identify the most important usability issues, by means of using a reflexive and inductive approach, allowing for the emergence of unexpected insights, and understanding of prototype usability. Specific attention was paid to behaviors that hindered the completion of the usability tasks. As this study was exploratory in nature, our goal was to generate a foundation for further research. Therefore, data were coded by a single coder to gain a deeper understanding of the research objectives and context. To determine usability task effectiveness and efficiency for each prototype, average time and corresponding SDs were calculated for the usability tasks. By integrating qualitative and quantitative usability in the discussion and interpreting our results, we aimed to provide a more comprehensive assessment of the prototypes’ usability among the target group.
**User Experience**

The quantitative results for This or That method were coded dichotomously for each of the 5 questions. A preference was scored 1 in case of the 3 positive questions. In contrast, a preference was scored -1 in case of the 2 negative questions. Consequently, the total score for the 3 prototypes ranged from a minimum of -2 to a maximum of 3, for which a mean score and corresponding SD were calculated. One-way ANOVA was performed to test for significant differences using SPSS Statistics (version 25; IBM Corp). A P value of ≤0.05 was considered statistically significant. Furthermore, qualitative data were analyzed by using a combination of deductive and inductive thematic coding in ATLAS.ti. Qualitative and quantitative data were integrated to gain a deeper understanding of prototype user experience.

**Ethical Considerations**

All parents provided written informed consent and children gave their verbal consent. When the child did not fully understand the study procedures, these were explained again. Participation was voluntary and participants could withdraw from the study at any time, without stating a reason. The study protocol was reviewed and deemed not subject to the Medical Research Involving Human Subjects Act (2021-13199). Subsequently, the protocol was reviewed and approved by the Social Sciences Ethics Committee of the WUR. The organization conducting this study established procedures for data management and data protection. Participants were not financially compensated for participating in this study, but children received a small gift as a thank you (a cookie and a stamp set).

**Results**

**Sample Characteristics**

Table 2 presents the descriptive characteristics of the study sample. A total of 14 children participated in the study and evaluated the 3 prototypes. Most children had at least 1 highly educated parent (12/14, 86%) and actively used technology for ≤1 hour per day (11/14, 79%). Moreover, most of the children had no older siblings (11/14, 79%).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>8 (57)</td>
</tr>
<tr>
<td>Girls</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>5.9 (0.8)</td>
</tr>
<tr>
<td>Girls</td>
<td>6.0 (0.6)</td>
</tr>
<tr>
<td>Older siblings, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3 (21)</td>
</tr>
<tr>
<td>No</td>
<td>11 (79)</td>
</tr>
<tr>
<td>Interactive screen time(^a), n (%)</td>
<td></td>
</tr>
<tr>
<td>≤1 h daily</td>
<td>11 (79)</td>
</tr>
<tr>
<td>2 h daily</td>
<td>3 (21)</td>
</tr>
<tr>
<td>3 h daily</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Education level of caregiver or caregivers(^b), n (%)</td>
<td></td>
</tr>
<tr>
<td>No highly educated caregiver</td>
<td>2 (14)</td>
</tr>
<tr>
<td>1 highly educated caregiver</td>
<td>2 (14)</td>
</tr>
<tr>
<td>2 highly educated caregivers</td>
<td>10 (72)</td>
</tr>
</tbody>
</table>

\(^a\)Interactive screen time use is defined as the time a child spends actively interacting with a device (eg, tablet, PC, or smartphone).

\(^b\)“Highly educated” is defined as having completed a degree at a university or a university of applied sciences.

**Usability Testing**

**Quantitative Results**

Of 14 children, 10 (71%) correctly recalled their lunch; half of the boys (4/8, 50%) and all the girls (6/6, 100%) succeeded in recalling their lunch. None of the successful children were able to do so without the assistance of standardized recall questions. Overall, 1 (7%) child needed the maximum number of 5 recall questions and 4 (40%) children needed one recall question; successful children needed an average of 2.2 (SD 1.4) recall questions. FoodBear’s usability task had the highest completion rate (n=11, 79%), followed by myBear (n=10, 71%) and FoodCam (n=19, 64%). The mean completion time (s) was faster and the number of interruptions (n) was lower for FoodCam (mean 9, SD 6; n=0.9) followed by myBear (mean 51, SD 17; n=3.3), and FoodBear (mean 65, SD 43; n=3.9).
Children who failed to complete either FoodBear’s (n=3, 21%) or myBear’s usability task (n=4, 29%) were all boys, but 1 (20%) girl failed to complete FoodCam’s usability task (n=5, 36%). Among these children, approximately all children had no highly educated parents (100%) or one highly educated parent (50%). When visually inspecting our data, no differences in age, interactive screen time (≤1 hour per day), or having older siblings (no) were observed across the children who did not complete their tasks. Finally, the order in which FoodBear and myBear were tested alternated. Children who successfully completed both FoodBear’s and myBear’s usability task (n=10, 71%) performed their second task on average 26.1 (SD 27.1) seconds faster.

**Table 3.** Descriptive data (task effectiveness and task efficiency) of usability tasks of FoodBear, myBear, and FoodCam.

<table>
<thead>
<tr>
<th>Task</th>
<th>Completion rate, n (%)</th>
<th>Completion time (seconds), mean (SD; range)</th>
<th>Interruptions for help, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FoodBear: “Give FoodBear the same lunch as you ate”</td>
<td>11 (79)</td>
<td>65 (43; 16-140)</td>
<td>4 (4)</td>
</tr>
<tr>
<td>myBear: “Give myBear the same lunch as you ate”</td>
<td>10 (71)</td>
<td>51 (17; 15-70)</td>
<td>3 (2)</td>
</tr>
<tr>
<td>FoodCam: “Photograph your lunch with FoodCam”</td>
<td>9 (64)</td>
<td>9 (6; 3-24)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

**Observational Results**

Thematic analysis of the video and field notes obtained during usability testing revealed several key usability issues that were mainly related to 4 themes: *food groups, frustrations related to unsatisfactory functioning of (parts of) the prototype, recall of food products, and distinction between eating moments* (Figure 3).

**Figure 3.** Thematic map of usability issues revealed by user testing of the dietary assessment tool prototypes FoodBear, myBear, and FoodCam.

**Theme 1: Food Groups**

In terms of food group-related usability issues, 93% (13/14) of the children reported problems related to the *recognition of icon/images and classification of food products into food groups*. When using FoodBear and myBear, approximately all children (13/14, 93%) experienced usability problems related to the depicted icons or images, which were mainly related to recognizing the food group represented by the icon, resulting in incorrect food group reporting. To illustrate, most children (n=10, 71%) became confused when the consumed product did not exactly resemble the depicted icon: “Cheese spread and jam. Which one is the jam?” (Participant 11). Some children (n=8, 57%) solved the problem by choosing the icon that they thought most resembled the product they ate. In some cases (n=5, 36%),
the children eventually classified food products into the correct food group: "I don’t see baguette, but I do see bread, so I press the bread" (Participant 10), but other participants failed to do so: “Which one is gingerbread? This one? [points to icon for fish]” (Participant 14). Moreover, 21% (3/14) of the children ate >1 product of the same food group during their lunch and reported problems with categorizing these different food products into the same food group. One participant asked the researcher for help and the other 2 chose different icons for different products. Finally, it was also observed that some icons were not chosen at all even though the icon was applicable for several of the children, for example, the icon for “something else.”

**Theme 2: Frustrations Related to Unsatisfactory Functioning of (Parts of) the Prototypes**

Overall, 93% (13/14) of the children encountered ≥1 usability problem related to unsatisfactory functioning of (parts of) the prototypes, that is, prototype-specific issues. For FoodBear, several children encountered problems with *where and how to put coins in FoodBear’s belly* (n=4, 29%) and *that coins did not fit properly* through the intended opening (n=6, 43%). The latter led to visible frustrations among participants: “Stupid bear!” (Participant 6). When using myBear, the main usability issues were caused by *interaction design*. Most children (n=9, 64%) struggled with the plus and minus button to indicate the amount of the product eaten; children either pressed the button again, causing the product to disappear from myBear’s belly or asked the researcher for help. FoodCam’s usability issues were mainly related to *finding the photo button* (n=5, 36%) and *quality of the printed photos* (n=6, 43%): “It’s annoying that you cannot choose colors when you print it, so you can see the right colors” (Participant 6).

**Theme 3: Distinction Between Eating Moments**

More than half of the participants (n=8, 57%) had difficulties with *distinguishing eating moments throughout the day*. Some children (n=4, 29%) pointed out that they did not fully understand the concept of lunch. Other children (n=2, 14%) tended to also name products they consumed earlier that day: “But I had also yogurt today!” (Participant 6).

**Theme 4: Recall**

In the category of recall-related usability issues, the subcategories *forgetfulness, portion size*, and *fantasy* issues were distinguished. Overall, 57% (8/14) of the children reported usability issues related to ≥1 of these subcategories. For FoodBear and myBear, 29% (4/14) of the children failed to perform the usability tasks because of *forgetfulness*. For example, children forgot to feed certain products to myBear or FoodBear: “O, I just completely forgot about that one!” (Participant 6). Moreover, 36% (5/14) of the children had problems indicating consumed *quantities*, which resulted in underreporting by all 5 children, that is, an insufficient number of icons or coins for a specific food group in FoodBear’s or myBear’s belly. Finally, 21% (3/14) of the children were unable to complete the usability tasks because of their *fantasy*, that is, children fed FoodBear or myBear products that were not consumed (n=2, 14%) or expressed unrealistic amounts (n=1, 7%). To illustrate the latter, one participant reported: “I think ten breads. In every lunchbox” (Participant 2).

**User Experience**

**Quantitative Results**

No significant difference in mean preference scores for FoodBear, myBear, and FoodCam was observed based on the This or That method, that is, 0.57 (SD 1.94), 0.50 (SD 1.45), and 0.86 (SD 1.65; F=0.18; P=.80), respectively (Table 4).

<table>
<thead>
<tr>
<th></th>
<th>FoodBear</th>
<th>myBear</th>
<th>FoodCam</th>
</tr>
</thead>
<tbody>
<tr>
<td>First choice, n (%)</td>
<td>5 (36)</td>
<td>5 (36)</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Last choice, n (%)</td>
<td>7 (50)</td>
<td>7 (50)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>Score, mean (SD)</td>
<td>0.57 (1.94)</td>
<td>0.50 (1.45)</td>
<td>0.86 (1.65)</td>
</tr>
</tbody>
</table>

*aPercentages are calculated based on the number of times participants selected each prototype. As children were allowed to choose more than one prototype, the sum of percentages may exceed 100%.

**Observational Results**

During the behavioral choice selection (Textbox 2; step 9), all children (14/14, 100%) chose (one of) the prototypes they scored best during This or That.

**Qualitative Results**

Although some children (n=5, 36%) experienced difficulties with answering the “why-questions” that followed the 5 This or That questions, several determinants for product liking could be identified. First, an important reason for children to choose one prototype over the others was *autonomy*, because they liked being able to do it “themselves” (n=5, 36%): “Because I can put the coins in there myself!” (Participant 13). Moreover, children (n=5, 36%) referred to the *reward*, such as the printed photo, as being a determinant for product liking: “I like that one because you can print your taken picture and keep it as a memory!” (Participant 6). In addition, 29% (4/14) of the children indicated that they liked the prototype being *challenging*: “I like that one because you can do a lot with that one. The camera is a bit stupid because you can only take a picture with it” (Participant 4). Such a challenge could be presented in the form of a game; the *gaming element* was emphasized by some children (n=4, 29%) as a fun element of one of the prototypes: “I like myBear because you can play games on it” (Participant 9). The shape of the prototype was also mentioned by several...
children; 29% (4/14) of the children indicated that they liked the prototype because it was tablet based or because they liked its appearance (n=2, 14%). Other determinants of preference included its social aspect (n=2, 14%) and the time frame (n=2, 14%): “I didn’t like this one very much because this one took too long and that one took too short” (Participant 13).

Discussion

Principal Findings

This study provided several insights related to usability and user experience that can be used to inform the development of dietary assessment tools for use by children. At the first encounter, most children were able to use FoodBear, myBear, and FoodCam and fulfill the accompanying usability tasks. However, all children required assistance from the researcher to succeed, and most of the children encountered several usability problems. The most important usability issues included problems related to food groups, frustrations related to the unsatisfactory functioning of (parts of) the prototypes, recall of food products, and distinction between eating moments. These issues, along with the queries needed to accomplish usability tasks, may suggest that dietary assessment tools may not be independently usable by children aged 5 to 6 years. However, the completion rates suggest that children can play a complementary role in dietary data collection to enhance data collected by their parents. No differences in product liking were observed when comparing the 3 prototypes. However, it is notable that all children selected one of the prototypes they scored best with This or That for the behavioral choice selection to play with again. The qualitative part of This or That revealed several determinants for liking a product, including autonomy, challenge, gaming elements, being tablet based, appearance, social elements, and time frame.

Usability

Overview

The 3 prototypes differed in terms of usability rate, time needed to complete the assessment, and required number of interruptions by the researcher, which may be partly explained by the fact that prototypes are based on different dietary assessment methodologies, that is, food recall (FoodBear and myBear) and food record (FoodCam) [9]. As a recall requires additional memory-based cognitive capacities compared with a food record, FoodCam was expected to yield the best results in this population. However, a lower number of children successfully completed FoodCam’s (9/14, 64%) usability task compared with FoodBear’s (11/14, 79%) and myBear’s (10/14, 71%) tasks, meaning that a lower number of children were able to take a sharp photo on which all consumed products were recognizable. On the other hand, if completed, the time and help needed with FoodCam’s usability task was substantially less than the other 2 tools. Aflague et al [27] showed that using the Mobile Food Record for capturing eating occasions could be a feasible method for use by children aged ≥3 years. In contrast to our study, participants in the study by Aflague et al [27] were allowed to practice and use a tablet or smartphone, whereas FoodCam is based on a more old-fashioned camera. Additional research is required to determine whether there is a difference in usability between traditional cameras and the cameras on smartphones or tablets. However, this research should also address the current challenges related to automatically extracting dietary information from real-world, user-generated images. As our prototypes were tested at the first encounter, it is likely that usability will increase with practice or a training module, but further studies are required to test this. Moreover, this study primarily evaluated the usability of 3 prototypes designed for independent use, thereby revealing some inherent challenges. Nonetheless, adopting an approach that combines children’s data with those collected by parents can potentially enrich the comprehensiveness of a child’s daily dietary intake assessment. Such a combined method would offer the possibility of gaining more detailed insights into foods consumed outside the home, ultimately enhancing the reliability of the dietary intake data. Further research is needed to investigate the potential bias of this approach. Moreover, consistent with previous findings [23,35], girls performed better than boys for all 3 usability and lunch recall tasks in terms of completion rate. This difference may be explained by girls’ higher attentional and memory performance compared with boys and emphasizes the need to consider sex differences in further development of the tools [36].

Strategies to Increase Usability

To address these usability issues, we identified several strategies for further improvement. To increase usability, the tools might benefit from a training module providing practice runs on estimating quantities and portion sizes, recognizing food categories, handling FoodCam, or the interaction design of myBear. Practical effects were already observed in this study, that is, all children who successfully completed myBear’s or FoodBear’s usability task performed their second usability task at a faster pace. Similar training effects have been observed in other studies [27,37]. Auditory or visual prompts, reminders, and feedback may also improve the usability of updated versions of the prototypes, that is, to remind participants to report their dietary intake throughout the day, or help with the correct use of the tool, for example, by checking whether all products have been reported in the correct amount, or send reminders when photos are incomplete or unsharp. Reminders and help with the tasks were now verbally performed by the researcher (eg, with recall questions), but should be automated in the next versions of the prototypes to facilitate independent use by the target group. Integrating multiple reminders is commonly used in other methods as well, for example, in Compl-eat [38], and is used to trigger the report of often forgotten products, such as cooking fats or drinks. In addition, myBear could particularly benefit from a more user-friendly and intuitive interaction design. Improvements in the interaction design should among others focus on simplifying consumed portion sizes. For example, using a slider to indicate portion size or pressing the button twice for the specific product may be more intuitive than using a plus or minus button. The use of age-appropriate interactions and images could also contribute to a better understanding of the different eating moments, for example, by using a clock model to capture mealtimes throughout the day. The direct effect of improving the interaction design has proven to be effective
in another study on the adolescent dietary assessment tool myfood24, where they compared the usability and acceptability of myfood24 among adolescents before and after making amendments [39]. Finally, the usability tasks of FoodBear and myBear illustrated that children experienced difficulties in understanding or interpreting the food group icons and categorizing their consumed products into food groups, asking for a more child-friendly approach. Therefore, further research is needed to identify child-friendly food groups and icons [25].

User Experience

No preferences were observed for one prototype over the others when using This or That method. However, it is notable that all children selected one of the prototypes they scored best with This or That to play again. This consistency suggests that the reported This or That choice is a good predictor for short-term preference in this sample. However, as the children only used the prototypes for a short time, it should be emphasized that This or That may not reflect the long-term preference. To gain insight into long-term engagement, more research is needed where children use the prototypes for a longer period in a home-use setting handling different mealtimes throughout the day.

As This or That determines preference relatively, it does not offer the opportunity to determine the magnitude of preference [34]. Although the This or That scores revealed no differences between the prototypes, it is important to consider this relativity when interpreting the qualitative results. What stands out is that most of the determinants for product liking pointed out by the children were in line with our list of requirements (Textbox 1), except for the determinant autonomy. Children in the preoperational phase, including our target group, have a strong curiosity and are interested in learning [40]. Therefore, the finding of autonomy being a determinant for product liking by children is not unexpected and should be included in the updated version of the program of requirements. Strategies to increase this feeling of autonomy within young children’s dietary assessment could include, for example, making the design accessible for children’s independent navigation (eg, by using navigation without text and making it real-time responsive), focus on children’s decision-making (eg, by including options for personalization and customization in the design), or encouraging their initiatives (eg, by including a reward system) [41].

Strengths and Limitations

Although this exploratory study contributes to the body of knowledge in several areas, it has some limitations. First, our first list of user requirements (Textbox 1) is based on literature only, which ideally would have included expert interviews as well, as conducted by de Gooijer et al [42]. As this study is the first to explore self-reported dietary assessment among children aged 5 to 6 years, this first list may have been insufficient. Second, as the usability tasks in this study were performed under favorable circumstances, the usability for FoodBear and myBear may have been overestimated. More specifically, the dietary assessment tasks took place shortly after lunch (a maximum of 3 hours after lunch). As other studies showed that meals with shorter retention intervals are in general easier to accurately recall and report compared with meals with longer retention intervals, results may become less accurate when measurements were performed after a longer period [43,44]. Moreover, prototypes were only evaluated for lunch and longer interaction (eg, over the course of a day) with the prototypes is needed to evaluate the accuracy of dietary intake data collected through our prototypes. In addition, it is worth noting that Dutch children typically have bread with spreads or toppings for lunch [45]. This was also reflected in our study sample, where all participating children ate bread for lunch. As previous studies showed that children struggle with identifying components within mixed meals [46], it is important to consider the relative simplicity of the Dutch lunch when interpreting our results. Considering these favorable conditions in this study, it raises questions about the ability of children aged 5 to 6 years to accurately use FoodBear and myBear without parental assistance for more complex meals consumed over an extended time frame in future research. Therefore, FoodBear and myBear might have more potential for use with caregivers. Another important limitation that should be considered when interpreting our findings is related to our sample. The sample size of this study was small and not representative of the Dutch population, primarily owing to the high proportion of highly educated parents among the participants. This demographic bias limits our ability to generalize our findings to a more diverse population. To gain a better understanding of the application of such tools in populations of lower socioeconomic status, further studies are necessary. Moreover, considering the qualitative focus of our research, the small sample size of our study underscores the necessity for caution when interpreting our quantitative results. This is particularly relevant in terms of statistical power and generalizability. Future research efforts could focus on studying the quantitative aspects of our study in more detail by recruiting a larger sample size.

Conclusions

This exploratory study identified essential user requirements for a novel dietary assessment tool designed for children aged 5 to 6 years, including (1) a comprehensive training program, (2) incorporation of auditory or visual prompts, (3) implementing reminders and feedback mechanisms, (4) a focus on a user-friendly and intuitive interaction design, (5) use of child-friendly food groups or icons, and (6) allowing room for children to exercise autonomy. By addressing these identified user requirements in the development of new dietary assessment tools, we can significantly enhance the quality of dietary intake data collected among children. Furthermore, these findings can serve as valuable guidance for ongoing innovations in the field of children’s dietary assessment and the provision of personalized dietary support. This, in turn, can inform strategies aimed at guiding children toward healthier food choices.
Acknowledgments
The authors would like to thank Eline Chin for assisting in completing the weighted decision matrix and photo quality assessment.

Data Availability
The data sets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions
ZvdH, FdG, ML, GC, DL, and EB-B participated in concept development. ZvdH, FdG, and EB-B cooperated in the development of the prototypes. ZvdH collected and analyzed the data and wrote the first draft. All the authors reviewed and commented on the drafts of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Summary of weighted decision matrixes.
[DOCX File, 27 KB - formative_v8i1e47850_app1.docx ]

Multimedia Appendix 2
The functionality of myBear and FoodCam.
[DOCX File, 1409 KB - formative_v8i1e47850_app2.docx ]

References


Abbreviations

WDM: weighted decision matrix
WUR: Wageningen University and Research

Edited by A Mavragani; submitted 26.04.23; peer-reviewed by T Baranowski, L Lafave; comments to author 23.10.23; revised version received 10.11.23; accepted 08.12.23; published 01.02.24.

Please cite as:

URL: https://formative.jmir.org/2024/1/e47850
doi: 10.2196/47850
PMID: 38300689

©Zoë van der Heijden, Femke de Gooijer, Guido Camps, Desiree Lucassen, Edith Feskens, Marlou Lasschuijt, Elske Brouwer-Brolsma. Originally published in JMIR Formative Research (https://formative.jmir.org). 01.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
The Development and Use of a New Visual Tool (REVISIT) to Support Participant Recall: Web-Based Interview Study Among Older Adults

Eileen M Dryden¹, PhD; Chitra Anwar¹*, MA; Jennifer Conti¹, MPH; Jacqueline H Boudreau¹, MPH; Meaghan A Kennedy²,³, MPH, MD; William W Hung⁴,⁵, MPH, MD; Kathryn A Nearing⁶,⁷, MA, PhD; Camilla B Pimentel¹,²,⁸, MPH, PhD; Lauren Moo¹,²,⁹, MD

¹Center for Healthcare Organization and Implementation Research, VA Bedford Healthcare System, Veterans Health Administration, Bedford, MA, United States
²New England Geriatric Research, Education, and Clinical Center, VA Bedford Healthcare System, Veterans Health Administration, Bedford, MA, United States
³Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States
⁴Bronx Geriatric Research, Education, and Clinical Center, James J. Peters VA Medical Center, Veterans Health Administration, Bronx, NY, United States
⁵Icahn School of Medicine, New York, NY, United States
⁶Eastern Colorado Geriatric Research, Education, and Clinical Center, Rocky Mountain Regional VA Medical Center, Veterans Health Administration, Aurora, CO, United States
⁷Division of Geriatric Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
⁸Department of Public Health, Zuckerberg College of Health Sciences, University of Massachusetts Lowell, Lowell, MA, United States
⁹Harvard Medical School, Boston, MA, United States
*these authors contributed equally

Corresponding Author:
Eileen M Dryden, PhD
Center for Healthcare Organization and Implementation Research
VA Bedford Healthcare System
200 Springs Road
Bedford, MA, 01730
United States
Phone: 1 781 506 2369
Email: eileen.dryden@va.gov

Abstract

Background: Qualitative health services research often relies on semistructured or in-depth interviews to develop a deeper understanding of patient experiences, motivations, and perspectives. The quality of data gathered is contingent upon a patient’s recall capacity; yet, studies have shown that recall of medical information is low. Threats to generating rich and detailed interview data may be more prevalent when interviewing older adults.

Objective: We developed and studied the feasibility of using a tool, Remembering Healthcare Encounters Visually and Interactively (REVISIT), which has been created to aid the recall of a specific telemedicine encounter to provide health services research teams with a visual tool, to improve qualitative interviews with older adults.

Methods: The REVISIT visual appointment summary was developed to facilitate web-based interviews with our participants as part of an evaluation of a geriatric telemedicine program. Our primary aims were to aid participant recall, maintain focus on the index visit, and establish a shared understanding of the visit between participants and interviewers. The authors’ experiences and observations developing REVISIT and using it during videoconference interviews (N=16) were systematically documented and synthesized. We discuss these experiences with REVISIT and suggest considerations for broader implementation and future research to expand upon this preliminary work.

Results: REVISIT enhanced the interview process by providing a focus and catalyst for discussion and supporting rapport-building with participants. REVISIT appeared to support older patients’ and caregivers’ recollection of a clinical visit, helping them to
share additional details about their experience. REVISIT was difficult to read for some participants, however, and could not be used for phone interviews.

**Conclusions:** REVISIT is a promising tool to enhance the quality of data collected during interviews with older, rural adults and caregivers about a health care encounter. This novel tool may aid recall of health care experiences for those groups for whom it may be more challenging to collect accurate, rich qualitative data (eg, those with cognitive impairment or complex medical care), allowing health services research to include more diverse patient experiences.

**KEYWORDS**

qualitative interviews; visual recall aid; older adults; health services research; web-based methods; visual tool; recall; qualitative interview; experience; perspective; motivation; patient; recall capacity; medical information; visual appointment; geriatric; older people; teledmedicine; videoconference; e-consultation; e-medicine; internet medicine; REVISIT; Remembering Healthcare Encounters Visually and Interactively; mobile phone

**Introduction**

Qualitative health services research often relies on semistructured or in-depth interviews to develop a deeper understanding of patient experiences, motivations, and perspectives. The quality of data gathered is contingent upon a patient’s recall capacity. Studies consistently show recall of medical information is low. Patients remember between 20% and 60% of the information provided by health care practitioners immediately after an encounter [1], dropping to 12.8% a month later [2]. In a seminal study of patient recall in a routine clinical setting by Anderson et al [3], of the 40% of medical information recalled by patients, 48% of it was misconstrued. Various practitioner- and patient-related factors pose threats to recall: practitioner-related factors include the use of complicated medical terminology, high volume of information relayed, and mode of information presentation (eg, verbal vs visual), while patient-related factors include low education level and emotional state during the visit [1,4].

Threats to generating rich and detailed interview data may be more prevalent when interviewing older adults. Aging is associated with a decline in sensory and cognitive function, making it difficult to understand and remember medical information [5]. Compared to younger individuals, older adults have more difficulty recalling details of health care experiences that researchers may be interested in exploring, including medication regimens [6], treatment recommendations [7], and appointment reminder telephone messages [8]. Routine recurring visits are also more poorly recalled than nonrecurring ones—patients tend to collapse recurring visits into a single, generic memory instead of separate, specific occurrences [9]. Older adults may be especially prone to do so as they are estimated to have an average of 7 medical visits per year [10].

To ensure qualitative data are accurate, researchers must carefully consider how to plan and conduct qualitative interviews with older adults. Visual methodologies have been used to mitigate the threats to validity resulting from recall bias in qualitative health services research [11-13]. These methods invite participants to tap into memories through nonverbal ways of thinking, improving participant recall and allowing researchers to access participant perspectives that can be difficult to articulate through conversation alone. Commonly used strategies include viewing and discussing photographs, video elicitation, drawing, chart-stimulated recall, and mapping and timelines exercises [11,14,15]. In our review of the literature, we found no documented cases of using visual recall aids with older adults, a group for whom such tools may be particularly useful, given known challenges with medical information recall [1].

In this paper, we explore the development and use of a new visual tool, Remembering Healthcare Encounters Visually and Interactively (REVISIT), created to aid recall of a specific telemedicine encounter among older adult interview participants. In spring 2021, a team of Veterans Health Administration (VA) qualitative researchers interviewed 30 rural, older (65 years of age and older) veterans and their caregivers remotely as part of an evaluation of GRECC Connect, a program that uses telemedicine to connect rural veterans with complex care needs to geriatric specialty care at 15 urban VA medical center hub sites. GRECC Connect hub teams are comprised of interprofessional care teams affiliated with Geriatric Research, Education, and Clinical Centers, VA centers of excellence focused on aging. Given the focus of many GRECC Connect sites on treating cognitive impairment, we anticipated that interviewees might experience challenges recalling details of their most recent GRECC Connect appointment (the “index visit”), posing a risk to the completeness and validity of interview data. We also anticipated challenges isolating information about their most recent GRECC Connect appointment from other appointments due to the increase in telemedicine visits during the COVID-19 pandemic. The REVISIT visual appointment summary was developed to better facilitate interviews conducted remotely with our participants. Our primary aims were to aid participant recall, maintain focus on the index visit, and establish a shared understanding of the visit between participants and interviewers. In this paper, we describe the development of REVISIT and interviewer experiences with the tool and suggest considerations for broader implementation and future research to expand upon this preliminary work.

**Methods**

**Evaluation Team**

A multidisciplinary VA project team contributed to the evaluation. Team members included physicians with expertise...
in primary care, geriatrics, and dementia; a veteran consultant; GRECC Connect leadership; and researchers with expertise in qualitative methods and project coordination.

**Developing REVISIT**

REVISIT was designed as a template to be populated with data from the veteran’s electronic health record (EHR). A member of the team with a background in media design drafted template options on Canva (Canva), a free web-based graphic design platform. REVISIT’s design drew upon VA’s Patient Experience Journey Map, a visual representation of commonly experienced moments before, during, and after a veteran’s health care visit [16]. Draft REVISIT templates were presented to the full multidisciplinary team for review, resulting in 3 iterative rounds of feedback and refinement.

The information included in the final iteration of REVISIT focused on aspects of the index visit we sought to confirm and explore, which were separated into three groupings: (1) the referral, including the reason and referring provider; (2) the index visit, including individuals present and main topics discussed; and (3) changes in the veteran’s health and health care resulting from the visit, including changes in diagnoses, medications, and referrals.

The overall structure of the first iteration of REVISIT (Figure 1) contained 3 columns, with each column representing a step in the GRECC Connect visit (before, during, and after the visit). For the first iteration, initial refinements suggested by the team focused on simplifying the template to include only information pertinent to the interview. Team members also felt REVISIT should focus more on the “Post-Visit” section to better aid participants’ recall of what worked well about the visit and what health needs remained unmet.

**Figure 1.** First iteration: the first iteration of REVISIT included elements subsequently omitted, such as the sections containing questions at the bottom of each column. REVISIT: Remembering Healthcare Encounters Visually and Interactively.

```
<table>
<thead>
<tr>
<th>Pre-Visit</th>
<th>GRECC Connect Visit</th>
<th>Post-Visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Services Used</td>
<td>CBG Participants: Participant, Participant 2, Participant 3</td>
<td>Services Used</td>
</tr>
<tr>
<td>Active health issues</td>
<td>Hub Participants: Participant, Participant 2, Participant 3</td>
<td>Active health issues</td>
</tr>
<tr>
<td>Medications Used</td>
<td>Dec. 6, 2020: 1:01 PM</td>
<td>Medications Used</td>
</tr>
<tr>
<td>What kinds of medical care do you get outside the VA? And why?</td>
<td>Diagnosis: Details, Details</td>
<td>What, if anything, has changed in your life since your virtual visit?</td>
</tr>
<tr>
<td></td>
<td>Chief Complaint: Details, Details</td>
<td></td>
</tr>
</tbody>
</table>
```

The second iteration (Figure 2) incorporated the aforementioned feedback for simplification. For example, the “Pre-Visit” section was changed completely to include only GRECC Connect referral information, and the heading for this section was edited to “Referral” to reflect this change. The “GRECC Connect Visit” section still included persons present during the index visit, but the other sections were collapsed into one summary of the visit details. The “Post-Visit” section was expanded to take up more of the page, emphasizing this section as the focus of the interview.
After reviewing the second iteration, the team encouraged further simplification of the template’s design to reflect REVISIT’s primary goal of helping participants recall details about their experience of the telemedicine visit. Team review of the second iteration also focused on possible modifications to the included language. EHRs contain medical jargon that is likely unfamiliar to interview participants. Team members suggested translation of these terms into more lay language for the last iteration, a process that relied heavily on input from the physician team members.

Design considerations for the final REVISIT iteration (Figure 3) included using boxes with rounded edges, as the team felt this connoted friendliness compared to the sharp edges shown in Figure 1. Icons were included alongside text descriptions wherever possible to increase ease of understanding. Arrows showed flow from one section to the next, green “+” symbols signified newly prescribed medications, and red “x” symbols signified deprecribed medications. Calibri font was used in accordance with VA’s graphic design standards.
The colors of each section were specifically chosen based on color-in-context theory [17], which posits color meanings are grounded in learned associations that develop from repeated pairings of colors with particular messages, concepts, or experiences. The color motif was loosely based on the 3 phases of a traffic light—the initial referral to GRECC Connect was yellow (to symbolize a transition) and the postvisit was green (to symbolize moving forward). Blue, as opposed to red, was chosen to represent the index visit because of its generally accepted calming effects [18]. The team felt this was a more suitable color choice, given the potentially sensitive topics that may surface during discussion of the index visit during interviews. The team also opted for colors with lighter versus darker hues, as these were felt to be easier on the eyes.

Language edits were incorporated into the final iteration. Potentially sensitive medical issues such as dementia diagnoses or cognitive decline were instead referred to as “memory changes.” We deemed this step necessary as it was sometimes unclear, based on the medical notes, what was explicitly discussed with the participant versus only documented in the clinic note. The inclusion of more neutral language helped ensure that REVISIT would avoid distressing a participant with potentially new information about their health.

Participants

REVISIT was used in the context of a health care evaluation with a sample of 30 rural veterans attending specialty geriatric telemedicine visits at 6 geographically diverse GRECC Connect hub sites. We defined “telemedicine” as medical appointments conducted through one of three modalities: (1) video appointments from a veteran’s home or other location to a remote specialist using VA Video Connect (VVC), (2) video appointments from a VA outpatient clinic near the veteran’s home to a remote specialist using Clinical Video Telehealth (CVT), or (3) telephone. An option to participate as a veteran-caregiver dyad was offered in cases where veteran participants had some degree of cognitive impairment or where caregivers were substantially involved in care. Due to the impact of dementia, some dyads were primarily represented by the caregiver.

Prior to initial contact, 3 team members briefly reviewed each veteran’s EHR to confirm the most recent telemedicine visit date and modality (VVC, CVT, or phone), veteran location, initial reason for referral, the presence of a caregiver, and any cognitive or other health concerns that would preclude study participation (eg, a veteran in hospice or deceased). Veterans were considered eligible for participation if they were 65 years and older of age, resided in a rural area (rural-urban commuting...
Interview Preparation

Once participants agreed to participate in an interview, a team member performed a detailed chart abstraction of the veteran’s EHR 6 months prior to the index appointment using a structured data abstraction template (Multimedia Appendix 1). This data abstraction template helped team members extract only information relevant to the GRECC Connect visit from the participant’s EHR, which can contain many notes from numerous clinicians. GRECC physicians on the multidisciplinary team helped to develop the data abstraction template and interpret EHR data when questions arose. These data were then used to populate REVISIT to create an individualized visual appointment summary and tailored interview guide for each participant. To protect participants’ health information, each populated REVISIT was saved in a password-protected participant-specific folder on a secure server. On average, team members spent 2 to 4 hours abstracting data and creating the visual summary. Time varied depending on the extensiveness and clarity of the participant’s medical chart.

Data Collection

Four experienced qualitative researchers conducted semistructured qualitative interviews with veterans and their caregivers who agreed to participate in the evaluation. We asked participants about their GRECC Connect telemedicine visit, including support received, what worked and did not work well, preferred modality for medical care, impact of visit, satisfaction, and recommendations. Interviews took place approximately a month after the index visit and were conducted via VVC or by phone depending on participant preference and ability. In total, 16 interviews were conducted via VVC on various devices (eg, smartphone, tablet, and laptop), and 14 were conducted by phone.

REVISIT was shared with participants who were interviewed via VVC using its screen-sharing feature. The use of REVISIT was incorporated into the GRECC Connect interview guide. The interviewer shared REVISIT when beginning to discuss the index visit following initial rapport-building questions. In at least one case, REVISIT was shown earlier in the interview because the participant needed more recall support.

Following the data collection process, team members who conducted interviews debriefed their experience using REVISIT, sharing the benefits and challenges of using the tool. Elements of the debrief were recorded on digital sticky notes, which were then grouped together by theme along the project timeline. Team members’ perspectives were informed by observations of participants when REVISIT was shared onscreen. We reviewed participant interview transcripts to find relevant excerpts to illustrate our observations.

Ethical Considerations

The VA Bedford Healthcare System Institutional Review Board determined this work was undertaken to inform VA operations as part of program evaluation and quality improvement activities and was not human subjects research.

Results

Use of REVISIT was limited to the 16 participants with whom interviews were conducted via VVC. Given the focus on our development of and initial experience with REVISIT, patient perspectives are only included insofar as their observed reactions influenced the team’s experiences and perspectives.

Interviewers used REVISIT to familiarize themselves with relevant details of the index visit prior to conducting interviews. This was particularly helpful when the interviewer did not complete the detailed chart review and was therefore less immersed in the details of each participant’s case or care. REVISIT provided the most salient information at a glance, so interviewers felt it was easier to review than the longer summary extracted from the chart review. With REVISIT, team members also felt better prepared to tailor interview questions to each participant. Additionally, the process of creating or reviewing each participant’s REVISIT visual encouraged the team to consider appropriate language to use during the interviews, such as using “changes in memory” versus “cognitive impairment.”

During the interview, team members used REVISIT as a shared reference point with participants, providing a focus and catalyst for discussion and prompt for further questioning. In one example, REVISIT allowed an interviewer to probe about other aspects of the index visit that were not brought up by the participant organically:

Interviewer: So this [REVISIT] is what we saw as sort of the summary of the visit that you and Mr. XXXX had. We’ve talked about a lot of this. We’ve talked about the changes in diagnoses, the memory changes. It did look like … they referred you to Audiology to check his hearing. Do you remember that referral at all?

In this way, REVISIT allowed interviewers to bring up contextual details about the visit, which helped to confirm that participants were discussing the index visit. This was particularly important for those who had numerous health care encounters.

Using REVISIT also helped interviewers cross-reference EHR data with participant accounts in real time, confirming congruence or revealing discrepancies between participant recollection and EHR data:

Interviewer: So you mentioned that there were some suggestions for medication changes in the future if anything progresses. We also noticed that there was a recommendation to consider using B12 supplements.

Participant: I don’t recall hearing the recommendation of the B12 supplements.

Interviewer: Okay.
Caregiver: So I didn’t have that in my notes. He has previously been given B12 but it has been quite some time since he’s been given B12 so I don’t remember her mentioning that.

In several cases, interviewers felt that REVISIT seemed to help both veterans and caregivers remember details about the index visit that they otherwise did not bring up:

Caregiver: So, I remember now those conversations and that I had requested – his [the veteran’s] mood was so angry, rage-y – that I had requested Dr. XXXX to increase his quetiapine, get to add one more during the morning and one at noon in addition to the three at night and so your visual helps me remember that.

This contributed to rapport-building by alleviating participants from the onus of remembering every detail of their visit. REVISIT was appreciated by participants, at least one of whom expressed challenges with information recall:

Caregiver: ... it was hard for me to remember what all we talked about that day and that [REVISIT] was very helpful.

However, the use of the tool was not without challenges. Interviewers noted that several participants expressed difficulty seeing REVISIT when shared over the videoconference platform. At least one participant felt that the visual was too light, while several others noted it was too small to read. Most participants who had difficulties with the size of REVISIT viewed it through their cell phone, so the issue of size may partially have to do with the device used:

Participant: I think I can say, for me, it was too small a screen, and you could probably mention, you know, it’s better if you’ve got a tablet or a laptop.

However, one participant on a larger tablet still had issues with font size and readability. It is thus unclear whether these challenges can be attributed to the visual alone or other computer-related factors (eg, whether the VVC window was maximized on the screen and the device’s brightness display). Researchers could not assess or control participants’ computer settings during interviews.

See Table 1 for further organization of interviewer experiences with REVISIT into relevant benefits and challenges.

### Table 1. Benefits and challenges of using Remembering Healthcare Encounters Visually and Interactively throughout the interview process.

<table>
<thead>
<tr>
<th>Benefits</th>
<th></th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Promoted use of sensitive language (eg, describing symptoms discussed with doctor vs displaying sensitive diagnoses like dementia)</td>
<td>- Contributed to rapport-building with participants through creation of shared understanding of events</td>
<td>- Required careful thought about displaying sensitive information that may be upsetting to the participants, for example, a new or sensitive diagnosis</td>
</tr>
<tr>
<td>- Supported organization of participant index visit data from participant’s health records</td>
<td>- Helped catalyze and focus discussion, providing a basis from which the interviewers could probe</td>
<td>- Some participants had difficulty seeing visual due to the size or low contrast of the document</td>
</tr>
<tr>
<td>- Served as a succinct preinterview refresher for interviewers</td>
<td>- Appeared to help participants recall and share details about their experiences</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Allowed interviewers to contrast recall with health record information in real time</td>
<td></td>
</tr>
</tbody>
</table>

### Discussion

**Principal Results**

REVISIT is a promising visual tool for enhancing the quality of data collected during interviews with older, rural veterans and caregivers. REVISIT enhanced the interview process by providing a focus and catalyst for discussion and supporting rapport-building with participants. Interviewers felt that the tool supported participants’ recollection of the clinical visit, as many participants noted this while sharing additional details about their experience. Our findings demonstrate that the novel use of visual methodologies during videoconference interviews with older adults is feasible and may be useful in supporting the overall success of qualitative evaluations.

**Comparison With Prior Work**

Visual methods, combined with in-depth interviews, have been shown to increase data quality, relevance, and trustworthiness [11,13]. Using REVISIT in our evaluation of GRECC Connect appeared to lead to similar enhancements in data quality by aiding participant recall during interviews, resulting in additional disclosure from participants. This finding is consistent with neuroscience principles that demonstrate how visual stimuli evoke brain regions involved in nonverbal information processing and memory [19].

Our experience is also consistent with other studies that demonstrate visual methods support interviewers in facilitating discussions with participants by prompting further questioning by interviewers, providing direction for discussion, streamlining topic transitions, and promoting increased attention [20-24]. Maintaining participant focus on relevant topics during an interview is essential to generating valid data [25]. Yet, researchers have argued that inhibition, or the ability to direct attention away from irrelevant information, declines with age [5]. Using a simplified visual aid such as REVISIT, which we found to provide a focus for discussion, may be particularly useful when interviewing older adults.

Using REVISIT highlighted, for interviewers, difficulties participants experienced recalling details of their index visit. A real-world implication of this insight is that patients may not remember the health and health care information shared by clinicians to properly care for themselves after the visit. For
some evaluation participants, viewing REVISIT was the first time they saw any written information about their visit from a practitioner. There are a number of provider-focused information-giving interventions that have been shown to positively influence patient recall [26,27], including intentional specific structuring of written postdischarge information [28]. At the very least, then, as our experience also suggests, older telemedicine patients may benefit from an after-visit summary outlining pertinent details about their health and health care discussed during the visit.

Consistent with previous studies [23], interviewers felt that using a visual memory aid contributed to rapport-building by providing a shared focus with which to interact and reflect upon. Rapport building is an important dimension of interviewing older adults with communication or cognitive barriers, as these challenges may lead them to view interviewers as threatening and increase feelings of powerlessness or a desire to withdraw from study participation [25,29]. Kirkevold and Bergland [30] suggest allocating more time over the course of a project to establish rapport with older interviewees, which can be challenging for research and evaluation projects with strict time constraints. REVISIT addresses this challenge by providing an accelerated rapport-building option for use directly within participant interviews.

The main challenge of using REVISIT as expressed by participants was the inability of some to see the visual due to its light color and font size. Age-related changes in visual acuity and contrast sensitivity can make it more difficult for older adults to read [5]. Therefore, images should have a high degree of contrast and use a large font. Further, participants in this study noted they or others might benefit from viewing images from a larger screen (eg, a tablet or laptop vs a cell phone). Additionally, although our evaluation of GRECC Connect showed it is possible to use visual tools during video interviews with older adults, it does not address potential barriers to the use of technology among this population; limited knowledge, comfort, or experience with technology, challenges with internet access, and existing cognitive and sensory impairments may hinder participants in studies conducted over videoconferencing platforms [31].

**Additional Considerations**

Future users of REVISIT and other similar recall aids should be mindful of how to introduce such tools and integrate them into the interview process. REVISIT may diminish rapport if interviewers share the visual at the wrong time; doing so may inadvertently disrupt the flow of the conversation, distracting participants from the interview as they try to make sense of the visual tool. Using REVISIT also reduces the capacity for nonverbal communication when shared on screen, since this action usually minimizes the window of the participant and researcher across videoconferencing platforms. Additionally, if a participant disagrees with the information presented on the visual, it may create confusion, discomfort, or distrust. Another consideration is the substantial amount of time it takes to prepare the visit summaries and subsequent REVISIT visuals for each interviewee. While the preparation time reduced as the team members gained experience with the methodology, given the time investment needed, this method may not be practical for studies with considerably larger sample sizes.

More research is needed to optimize REVISIT’s usability and understand its acceptability among older adults with cognitive impairments. Future research should also explore the extent to which the visual tool affects recall by systematically comparing appointment recall using REVISIT with interview discussion alone. REVISIT may be useful for understanding the experiences of other patient populations with cognitive impairment (eg, traumatic brain injury and posttraumatic stress disorder) or complex medical care (eg, cancer treatment) and adaptable to in-person use (eg, on an iPad or paper). Further research and evaluation are needed to ensure the efficacy of REVISIT with different populations and settings.

**Limitations**

This is a preliminary study. Observations were limited to our sample of 16 veterans and veteran-caregiver dyads, most of whom had some degree of cognitive impairment and were interviewed over VVC, as REVISIT use was only possible through its screen-sharing feature. Further, participants were not systematically asked about their experience of viewing and using REVISIT during the interview. Because of this, we only included patient experiences that directly influenced team members’ own experience with and perceptions of the tool. Additionally, though REVISIT appeared to support recall in this study, it does not guarantee that a participant will truly recall relevant details as opposed to simply agreeing with what they are seeing.

**Conclusions**

REVISIT is a novel visual tool that aids the recall of health care encounters by tapping into memories through nonverbal ways of thinking. The use of REVISIT, a carefully curated visual representation of one particular health care encounter, helps to address a number of threats to generating rich, detailed interview data that may be more prevalent when interviewing older adults. As health services research seeks to understand more diverse patient experiences within health care, a tool such as REVISIT may aid recall of health care experiences for those groups for whom it may be more challenging to collect accurate, rich qualitative data. Further research is needed to understand its usefulness with different populations and settings.

**Acknowledgments**

The authors wish to acknowledge Mr Robert Bruno, a veteran consultant on our project team, for providing valuable input and feedback on our work. This work was supported by funding from the US Department of Veterans Affairs, Office of Rural Health, Veteran Rural Health Resource Center-Salt Lake City for portfolio project FY21-16035. MAK is supported by Career Development Award (IK2 RX003930) from the VA Rehabilitation Research and Development Service. CBP is supported by Career Development Award (IK2 RX003930).
References


provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Testing Behavioral Messages to Increase Recruitment to Health Research When Embedded Within Social Media Campaigns on Twitter: Web-Based Experimental Study

Sandro T Stoffel¹,², MSc, PhD; Jing Hui Law³, MSc, MRes; Robert Kerrison⁴, MSc, PhD; Hannah R Brewer⁵, MSc, PhD; James M Flanagan⁵, BSc, PhD; Yasemin Hirst¹,⁶,⁷, MSc, PhD

¹Department of Behavioural Science and Health, University College London, London, United Kingdom
²Institute of Pharmaceutical Medicine, University of Basel, Basel, Switzerland
³Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom
⁴School of Health Sciences, University of Surrey, Guildford, United Kingdom
⁵Department of Surgery and Cancer, Imperial College London, London, United Kingdom
⁶Lancaster Medical School, Lancaster University, Lancaster, United Kingdom
⁷Applied Health Research Hub, University of Central Lancashire, Preston, United Kingdom

Corresponding Author:
Sandro T Stoffel, MSc, PhD
Department of Behavioural Science and Health
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 2076792000
Email: s.stoffel@ucl.ac.uk

Abstract

Background: Social media is rapidly becoming the primary source to disseminate invitations to the public to consider taking part in research studies. There is, however, little information on how the contents of the advertisement can be communicated to facilitate engagement and subsequently promote intentions to participate in research.

Objective: This paper describes an experimental study that tested different behavioral messages for recruiting study participants for a real-life observational case-control study.

Methods: We included 1060 women in a web-based experiment and randomized them to 1 of 3 experimental conditions: standard advertisement (n=360), patient endorsement advertisement (n=345), and social norms advertisement (n=355). After seeing 1 of the 3 advertisements, participants were asked to state (1) their intention to take part in the advertised case-control study, (2) the ease of understanding the message and study aims, and (3) their willingness to be redirected to the website of the case-control study after completing the survey. Individuals were further asked to suggest ways to improve the messages. Intentions were compared between groups using ordinal logistic regression, reported in percentages, adjusted odds ratio (aOR), and 95% CIs.

Results: Those who were in the patient endorsement and social norms–based advertisement groups had significantly lower intentions to take part in the advertised study compared with those in the standard advertisement group (aOR 0.73, 95% CI 0.55-0.97; P= .03 and aOR 0.69, 95% CI 0.52-0.92; P=.009, respectively). The patient endorsement advertisement was perceived to be more difficult to understand (aOR 0.65, 95% CI 0.48-0.87; P=.004) and to communicate the study aims less clearly (aOR 0.72, 95% CI 0.55-0.95; P=.01). While the patient endorsement advertisement had no impact on intention to visit the main study website, the social norms advertisement decreased willingness compared with the standard advertisement group (157/355, 44.2% vs 191/360, 53.1%; aOR 0.74, 95% CI 0.54-0.99; P= .02). The majority of participants (395/609, 64.8%) stated that the messages did not require changes, but some preferred clearer (75/609, 12.3%) and shorter (59/609, 9.7%) messages.

Conclusions: The results of this study indicate that adding normative behavioral messages to simulated tweets decreased participant intention to take part in our web-based case-control study, as this made the tweet harder to understand. This suggests that simple messages should be used for participant recruitment through Twitter (subsequently rebranded X).
Introduction

For researchers, increasing use of the internet has opened up new ways to investigate society and behavior at lower costs [1], producing higher data quality (Kongsved et al [2]), faster return rates, and lower data entry times [3]. Web-based participant recruitment (eg, websites and apps, such as social media) has also been applied in the social and behavioral sciences.

The new internet, or social media, also known as Web 2.0, enables people to connect with friends, family, companies, and other entities to produce and share content on the web [4,5]. The benefits of using social media for participant recruitment, over and above Internet 1.0, have been increasingly demonstrated [6,7]. For example, at the height of the COVID-19 pandemic in 2020, social media allowed many social scientists to connect with thousands of people and include populations with unmet needs in their research [8-10]. In addition, compared to paid panels, using paid social media advertisements to recruit research participants has been shown to cost less [11,12] and be more time-efficient [13,14]. Furthermore, it is also suggested that social media can reach larger pools of participants and access hard-to-reach populations [15]. Social media allow researchers to recruit participants quickly and cost-effectively, as they are accessible through various devices at any time and allow the creation of specific digital content to target specific populations, increasing the likelihood of achieving the required sample size [15,16].

Facebook, Twitter (subsequently rebranded X), and Instagram are some of the most popular social media platforms for health research [13,17-20]. In 2022, more than 70% of internet users in the United Kingdom reported using Facebook and 42.8% Twitter [21]. Social media has become so deeply embedded in our daily lives that people rely on them for every need, such as entertainment, information, purchases, social connections, and work [22]. As new social media platforms enter the market, it is expected that the number of social media users will continue to grow [22].

So far, only a few studies have investigated the effectiveness of social media for health research. Most studies have investigated their effectiveness concerning recruitment in offline studies [13,16,17,23], with only a few making comparisons between different web-based recruitment methods for web-based studies [24]. While there is some guidance for the ethical use of social media for recruiting participants for health research [25], little is known about optimizing the use of platforms such as Twitter and Facebook for health-related research recruitment [19,26], such as the contents of the message to be used for targeting eligible individuals.

However, we can infer potential key components of a social media message from other areas using behavioral sciences. For example, marketing research has shown that credibility and trust in the source are important factors for clicking on advertisements [17,27-30]. In relation to this, studies have suggested that web-based health information from an expert source is viewed as more experienced and credible [31,32]. These findings demonstrate that aspects of endorsement and credibility when creating and disseminating messages are important. Similarly, several studies have shown that messages containing descriptive and normative social norms can be effective methods of engaging with the public [33-35]. In these studies, individuals receive information about socially desired (normative norms) or most frequently observed behavior (descriptive norms). Social norm messages provide individuals with a standard against which they can compare their intentions [36]. To our knowledge, no previous studies have tested whether social norms or patient endorsement messages on social media posts increase engagement with target audiences. The primary aim of this web-based experimental study, therefore, was to design and test the use of tailored Twitter posts, which integrate elements of patient endorsement and social norms, for the recruitment of participants into an observational case-control study.

Methods

Setting and Context

In 2020, a simulated randomized web-based experiment was programed on SurveyMonkey (SurveyMonkey Inc). The experiment was designed to test the effectiveness of targeted social media messages to increase intentions to participate in a real-world observational case-control study called the Cancer Loyalty Card Study (CLOCS) [37]. CLOCS is an observational case-control study that aims to investigate the self-care behaviors of patients with ovarian cancer before their cancer diagnosis. It seeks to do this by investigating differences in transactional data (such as medication purchasing) between women with and without ovarian cancer (the transactional data are collected through the loyalty cards of 2 UK-based high street retailers). Cases (ie, women with ovarian cancer) were recruited through participating National Health Service sites, while controls were recruited through the study website. Thus, those who were eligible to take part as control participants were recruited through social media and other internet-based sources.

Study Eligibility and Recruitment

The study sample comprised women aged between 18 and 70 years living in the United Kingdom without an ovarian cancer diagnosis who were potentially eligible for the real-world observational case-control study. Study participants were recruited through a web-based survey vendor, Dynata (Dynata Global UK Ltd).
Procedure
At the beginning of the experiment, those who were interested in taking part in the web-based experiment were presented with information about the study, including a brief description of CLOCS as well as a consent form. If participants consented and were eligible, they were randomized (in a 1:1:1 ratio) to receive 1 of 3 simulated Twitter posts: a standard advertisement (control condition), an advertisement with patient endorsement (patient endorsement condition), or an advertisement with a descriptive social norms message (social norms condition; Table 1). To generate authentic Twitter messages, real tweets were posted on a dummy Twitter account, alongside an infographic detailing information about CLOCS. Screenshots were taken of these posts for use in the experiment (the messages were immediately deleted after each one was posted; Figures S1, S2, and S3 in Multimedia Appendix 1 contain screenshots of the messages).

Table 1. Messages used in the experimental study with readability scores and character count. The Flesch-Kincaid readability score ranges from 0 (“extremely difficult to read, best understood by university graduates”) to 100 (“very easy to read, easily understood by an average 11-year-old student”).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Content</th>
<th>Readability score without special symbols or URL</th>
<th>Readability score with special symbols and URL</th>
<th>Number of characters with special symbols or URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control condition</td>
<td><strong>NEW RESEARCH RECRUITING WOMEN WITHOUT OVARIANCANCER</strong> We are recruiting women in the UK, aged 18+ without ovarian cancer to an online survey about potential symptoms, shopping and self-care behaviours. Take part @ clocsproject.org.uk/participants</td>
<td>23.8</td>
<td>15</td>
<td>222</td>
</tr>
<tr>
<td>Patient endorsement</td>
<td><strong>NEW RESEARCH RECRUITING WOMEN WITHOUT OVARIANCANCER</strong> Fiona (CLOCS patient representative): #ParticipatedinCLOCS because I bought medication for my symptoms from retailers before my cancer diagnosis.” Take part @ clocsproject.org.uk/participants</td>
<td>24</td>
<td>7.9</td>
<td>224</td>
</tr>
<tr>
<td>Descriptive norms</td>
<td><strong>NEW RESEARCH RECRUITING WOMEN WITHOUT OVARIANCANCER</strong> Most women with #ovariancancer are happy to take part in CLOCS. You can help us better understand their illness and symptoms by taking part as a healthy volunteer. Take part @ clocsproject.org.uk/participants</td>
<td>57.8</td>
<td>37</td>
<td>232</td>
</tr>
</tbody>
</table>

The content of the messages is presented in Table 1, along with a Flesch-Kincaid readability score calculated using the web-based software Grammarly (Max Lytvyn, Dmytro Lider, and Alex Shevchenko). This was done to ensure that the message was understandable to the target audience [38]. The standard and patient endorsement messages had the lowest readability scores (15 and 7.9, respectively) and were the easiest to understand (Table 1).

After being presented with the Twitter messages, participants were asked 2 comprehension questions on whether CLOCS only recruits women with ovarian cancer and what kind of data CLOCS are analyzing. Participants could only continue in the survey if they answered the questions correctly [35,39,40]. The primary outcome was participants’ intention to take part in CLOCS, and we asked individuals whether they would participate in the advertised study, adapted from previous literature [34,35,39,41,42]. It featured a fully labeled 4-point response scale (“definitely not,” “probably not,” “yes probably,” and “yes definitely”).

To explore how the messages were perceived by the participants, we included 2 questions on how easy the message was to understand (“very difficult,” “fairly difficult,” “fairly easy,” or “very easy”) and how clearly the aims of the study were communicated (“not at all,” “a little,” “very,” or “extremely”).

In the next step, participants were asked about their past participation in health care research (“yes” or “no”) and whether they had loyalty cards from UK-based high-street retailers. Sociodemographic questions covered age (“18-24,” “25-34,” “35-44,” “45-54,” or “55-70”), education (“no college degree” or “college degree, equivalent, or higher”), employment status (“yes” or “no”), marital status (“single”, “married or living with a partner”, “divorced or separated or widowed”), self-reported health (“poor,” “fair,” “good,” or “excellent”), and history of cancer in themselves, family, or close friends (“yes” or “no”).

Individuals were then given the opportunity to state their thoughts on improving social media messages for the recruitment of study participants in an open-ended question.

The survey concluded with an active interest question on whether participants would be interested in being redirected to the CLOCS website for more information on how to participate [34,40-42]. Those who responded yes were provided with a link to the CLOCS website on the final page of the survey. The website opened in a new tab for participants who clicked on the link. No further data associated with their direct participation in CLOCS were collected in this experiment. The web-based experiment took, on average, 5 minutes to complete.

Ethical Considerations
Ethics approval for this study was obtained from the University College London Research Ethics Committee (17813/001). All participants provided consent to take part in the study. All the data collected as part of the study were anonymized, meaning no identifiable information were collected. Eligible participants who completed the questionnaire received a small financial incentive from Dynata, as per their panelist agreements.
Data Analysis

A pilot study was conducted beforehand for sample size calculations. Based on the findings from the initial sample of 359 participants, with a 10 percentage point difference in the intention to take part ("yes, definitely" or "yes, probably" versus "definitely no" or "probably no"), we determined that the number of participants needed to achieve 95% CI and 80% power was 350 per trial arm. Data from participants in both the pilot and final samples were combined for analysis.

Sample characteristics were assessed using descriptive statistics (Table 2). Differences in participants’ intention to take part in CLOCS and perception of the messages were assessed using univariate and multivariate ordinal logistic regression. Willingness to visit the actual website was assessed between groups using univariate and multivariate binary logistic regressions. Adjusted odds ratios (aORs), 95% CIs, and \( P \) values are presented in the results, with \( P \) values below .05 regarded as statistically significant.

The responses to the open-ended feedback question were categorized into main themes through content analysis [43].

Table 2. Sociodemographic characteristics of study participants.

| Demographic categories | Control condition (n=360), n (%) | Patient endorsement condition (n=345), n (%) | Social norms condition (n=355), n (%) | Overall (N=1060), n (%) | Chi-square test (df) | \( P \) value
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>48 (13.3)</td>
<td>45 (13.0)</td>
<td>51 (14.4)</td>
<td>144 (13.6)</td>
<td>8.54 (8)</td>
<td>.38</td>
</tr>
<tr>
<td>25-34</td>
<td>71 (19.7)</td>
<td>74 (21.4)</td>
<td>76 (21.4)</td>
<td>221 (20.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>80 (22.2)</td>
<td>87 (25.2)</td>
<td>86 (24.2)</td>
<td>253 (23.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>76 (21.1)</td>
<td>74 (21.4)</td>
<td>87 (24.5)</td>
<td>237 (22.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-70</td>
<td>85 (23.6)</td>
<td>65 (18.8)</td>
<td>55 (15.5)</td>
<td>205 (19.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>15 (4.2)</td>
<td>17 (4.9)</td>
<td>25 (7.0)</td>
<td>57 (5.4)</td>
<td>12.40 (6)</td>
<td>.05</td>
</tr>
<tr>
<td>Fair</td>
<td>94 (26.1)</td>
<td>118 (34.2)</td>
<td>90 (25.4)</td>
<td>302 (28.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>201 (55.8)</td>
<td>175 (50.7)</td>
<td>190 (53.5)</td>
<td>566 (53.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>50 (13.9)</td>
<td>35 (10.1)</td>
<td>50 (14.1)</td>
<td>135 (12.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower than a college degree</td>
<td>171 (47.5)</td>
<td>181 (52.5)</td>
<td>197 (55.5)</td>
<td>549 (51.2)</td>
<td>4.67 (2)</td>
<td>10</td>
</tr>
<tr>
<td>College degree, equivalent, or higher</td>
<td>189 (52.5)</td>
<td>164 (47.5)</td>
<td>158 (44.5)</td>
<td>511 (48.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paid employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>119 (33.1)</td>
<td>115 (33.3)</td>
<td>121 (34.1)</td>
<td>355 (33.5)</td>
<td>0.09 (2)</td>
<td>.96</td>
</tr>
<tr>
<td>No</td>
<td>241 (66.9)</td>
<td>230 (66.7)</td>
<td>234 (65.9)</td>
<td>705 (66.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>164 (45.6)</td>
<td>145 (42.0)</td>
<td>146 (41.1)</td>
<td>455 (42.9)</td>
<td>0.450</td>
<td>.45</td>
</tr>
<tr>
<td>Married or living with a partner</td>
<td>196 (54.4)</td>
<td>200 (58.0)</td>
<td>209 (58.9)</td>
<td>605 (57.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experienced cancer closely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>263 (73.1)</td>
<td>255 (73.9)</td>
<td>254 (71.5)</td>
<td>772 (72.8)</td>
<td>0.776</td>
<td>.78</td>
</tr>
<tr>
<td>No</td>
<td>97 (26.9)</td>
<td>90 (26.1)</td>
<td>101 (28.5)</td>
<td>288 (27.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacy loyalty card</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>264 (73.3)</td>
<td>263 (76.2)</td>
<td>256 (72.1)</td>
<td>783 (73.9)</td>
<td>0.445</td>
<td>.45</td>
</tr>
<tr>
<td>No</td>
<td>96 (26.7)</td>
<td>82 (23.8)</td>
<td>99 (27.9)</td>
<td>277 (26.1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a^{Chi-square test.\)
Results

Study Sample

Figure 1 demonstrates the flow of participants through the study. In total, 2500 invitations were sent out on behalf of University College London researchers to women registered on a survey panel (Dynata), and 47.8% (1195/2500) responded to the invitation. Out of these potential participants, 92.6% (1107/1195) were eligible for the study.

Eligible participants were randomized to the experimental conditions: 376 to the control condition, 358 to the patient endorsement condition, and 373 to the social norms condition. Across conditions, 4.2% (47/1107) did not finish the survey after randomization, leaving a final sample of 1060, who were all included in the analysis: 34% (360/1060) in the control condition, 32.5% (345/1060) in the patient endorsement condition, and 33.5% (355/1060) in the social norms condition. Most women in the analytical sample were in paid employment (705/1060, 66.5%), married or cohabiting (605/1060, 57.1%), did not have a college degree (549/1060, 51.2%), owned at least 1 loyalty card from a pharmacy (783/1060, 73.9%), experienced cancer closely (ie, either themselves or with family or close friends) (772/1060, 72.8%), and reported good or excellent health (701/1060, 66.1%). Post hoc comparisons revealed that sociodemographic variables did not vary significantly across the experimental conditions (Table 2).

Figure 1. Flow through the study.

<table>
<thead>
<tr>
<th>Invited to the survey (n=2500)</th>
<th>Started the survey (n=1195)</th>
<th>Excluded as they did not meet the inclusion criteria (n=88)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized to experimental conditions (n=1107)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control condition (n=376)</td>
<td>Patient endorsement condition (n=358)</td>
<td>Social norms condition (n=373)</td>
</tr>
<tr>
<td>Finished survey (n=360)</td>
<td>Finished survey (n=345)</td>
<td>Finished survey (n=355)</td>
</tr>
</tbody>
</table>

Intention to Take Part in CLOCS

Overall, the intention to take part in CLOCS was high, with 60% (636/1060) of women stating that they would probably or definitely participate. Table 3 shows the distribution of intentions after reading the Twitter messages. The ordered logistic regressions in Table S1 in Multimedia Appendix 1 show that the behavioral messages, both patient endorsement (odds ratio [OR] 0.74, 95% CI 0.56-0.98; \( P = .03 \)) and social norms (OR 0.74, 95% CI 0.54-0.93; \( P = .015 \) and aOR 0.69, 95% CI 0.52-0.92; \( P = .009 \)), decreased intention to take part in CLOCS. None of the sociodemographic variables were significantly associated with the intention to participate in CLOCS.

Table 3. Intention to take part in the case-control study.

<table>
<thead>
<tr>
<th>Control (n=360), n (%)</th>
<th>Patient endorsement (n=345), n (%)</th>
<th>Social norms (n=355), n (%)</th>
<th>Overall (N=1060), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely not(^{a,b})</td>
<td>24 (6.7)</td>
<td>27 (7.8)</td>
<td>29 (8.2)</td>
</tr>
<tr>
<td>Probably not(^{a,b})</td>
<td>101 (28.1)</td>
<td>125 (36.2)</td>
<td>118 (33.2)</td>
</tr>
<tr>
<td>Probably yes(^{a,b})</td>
<td>183 (50.8)</td>
<td>146 (42.3)</td>
<td>179 (50.4)</td>
</tr>
<tr>
<td>Definitely yes(^{a,b})</td>
<td>52 (14.4)</td>
<td>47 (13.6)</td>
<td>29 (8.2)</td>
</tr>
</tbody>
</table>

\(^a\chi^2_{6}=14.52.\)

\(^bP=.02.\)

Perception of the Messages

Table 4 shows that most study participants stated that the messages were fairly or very easy to understand (796/1060, 75.1%) and that the aims of the study were very or extremely clearly communicated (594/1060, 56%). However, the ordered logistic regression results in Table S2 in Multimedia Appendix 1 show that individuals in the patient endorsement condition perceived the message as more difficult to understand (OR 0.63, 95% CI 0.47-0.84; \( P = .002 \)) and the study aims as less clear (OR 0.70, 95% CI 0.53-0.92; \( P = .01 \) and aOR 0.72, 95% CI 0.55-0.95; \( P = .02 \)) than those in the control condition. There were no statistically significant differences in the perceptions of those in the social norms condition and those in the control condition.
Table 4. Perception of the messages.

<table>
<thead>
<tr>
<th>Understanding the message</th>
<th>Control (n=360), n (%)</th>
<th>Patient endorsement (n=345), n (%)</th>
<th>Social norms (n=355), n (%)</th>
<th>Overall (N=1060), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very difficult</td>
<td>6 (1.7)</td>
<td>10 (2.9)</td>
<td>8 (2.3)</td>
<td>24 (2.3)</td>
</tr>
<tr>
<td>Fairly difficult</td>
<td>72 (20.0)</td>
<td>94 (27.3)</td>
<td>74 (20.8)</td>
<td>240 (22.6)</td>
</tr>
<tr>
<td>Fairly easy</td>
<td>213 (59.2)</td>
<td>196 (56.8)</td>
<td>218 (61.4)</td>
<td>627 (59.2)</td>
</tr>
<tr>
<td>Very easy</td>
<td>69 (19.2)</td>
<td>45 (13.0)</td>
<td>55 (15.5)</td>
<td>169 (15.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communication of study aims</th>
<th>Overall (N=609), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all clear</td>
<td>9 (2.5)</td>
</tr>
<tr>
<td>A little clear</td>
<td>139 (38.6)</td>
</tr>
<tr>
<td>Very clear</td>
<td>145 (40.3)</td>
</tr>
<tr>
<td>Extremely clear</td>
<td>67 (18.6)</td>
</tr>
</tbody>
</table>

Active Interest in CLOCS

Almost half of the study participants (526/1060, 49.6%) indicated that they would like to be redirected to the CLOCS website after the survey. The binary logistic regression in Table S1 in Multimedia Appendix 1 shows that participants who were presented with the social norms message were less interested in being redirected than those in the control condition (157/355, 44.2% vs 191/360, 53.1%; OR 0.70, 95% CI 0.52-0.94; P=.02 and aOR 0.74, 95% CI 0.54-0.99; P=.05). While there were no significant differences between the patient endorsement and control conditions (178/345, 51.6% vs 191/360, 53.1%; OR 0.94, 95% CI 0.70-1.27; P=.70 and aOR 0.96, 95% CI 0.71-1.30; P=.78), women with a loyalty card (117/205, 57.1% vs n/N, 40.3%; aOR 1.47, 95% CI 1.10-1.95; P=.008), excellent health (70/135, 51.8% vs 22/57, 38.6%; aOR 1.94, 95% CI 1.00-3.76; P=.05), aged between 55 and 70 years (117/205, 57.1% vs n/N, 40.3%; aOR 1.89, 95% CI 1.19-3.00; P=.007) and those who had experienced cancer closely (404/772, 52.34% vs 122/288, 42.4%; aOR 1.37, 95% CI 1.03-1.83; P=.03) were more interested in visiting the study website. Those who had previously participated in health research were less likely to want to be redirected (348/734, 47.4% vs 178/326, 54.6%; aOR 0.75, 95% CI 0.58-0.99; P=.04).

Feedback Question

Table 5 shows the main themes of the content analysis per message. While 57.5% (609/1060) of the study participants were willing to provide some feedback, the majority (395/609, 64.8%) stated that the messages did not require changes. Another common theme was clarity (75/609, 12.3%), where participants thought there was too much jargon, that the message should be shorter, adding the hashtags at the end would make it more readable, and message format (59/609, 9.7%), where participants recommended using brighter colors or adding more infographics or a video instead of text.

Moreover, some participants (237/609, 3.8%) stated that the messages were unclear on how the advertised study uses loyalty cards to help with an ovarian cancer diagnosis. To increase the credibility of the message, 3.3% (20/609) of participants suggested including the university’s or sponsor’s logo at the beginning of the message. Some participants also suggested posting the messages on several social media platforms (16/609, 2.6%), as well as using patient or celebrity endorsement (9/609, 1.5%) or advertising an incentive (7/609, 1.1%).
Discussion

Overview
This randomized web-based experiment examined the effectiveness of adding behavioral messages to Twitter advertisements for participant recruitment in a real-world case-control study (CLOCS). The results show that the standard messages yielded the highest intentions compared to the 2 normative behavioral messages. Furthermore, the social norms message decreased the willingness to visit the real study website after the survey. The vast majority of participants stated that the messages did not require changes, but some preferred clearer and shorter advertisements.

Comparison With Previous Literature
Our findings contrast with previous research, which has shown that using behavioral messages, such as social norms [34,35] and patient endorsement [31,32] can be effective methods to engage with the public on the web. The negative effect we found can partially be explained by the reduced readability of the messages, as individuals in the social norms condition perceived the message to be more difficult to understand—and the aims of the study were less clear than those in the standard advertisement.

Individuals who have had experience with a cancer diagnosis, either themselves, with family, or with close friends, were found to be more interested in CLOCS. This is in line with research reporting that familial history of cancer is associated with increased breast and ovarian cancer screenings due to individuals’ increased awareness of cancer-related complications [44]. In this study, participants’ awareness of ovarian cancer and its risks—partially informed by their close experiences with cancer—may explain their increased interest in CLOCS. Therefore, having a personal awareness of or connection to the proposed project can increase individuals’ interest in health-related research. We also found that women aged between 55 and 70 years had increased interest in visiting the CLOCS website. Ovarian cancer is rare in women younger than 30 years, but the risk increases with age, drastically spiking after 50 years—with the average age of diagnosis being between the ages of 50 and 70 years [45]. Thus, the saliency of the risk for ovarian cancer in these age groups may, in part, explain their interest. Final, while individuals recommended including a video in the message, a recent experimental study did not find any effect of adding animated decision aids to a website with the intention to participate in a case-control study [42].

Strengths and Limitations
One strength of this study was the use of a randomized experimental design to evaluate the effectiveness of adding behavioral messages to Twitter messages. Additionally, the study used validated questions on intentions and active interest. A final strength of this study is that the statistical analysis included a large number of covariates known to influence participation in health research.

This study has some important limitations, which call for follow-up research. First, the 2 messages were grounded in social norms and patient endorsement, which have mixed and limited evidence supporting the efficacy of these messages in influencing participation in clinical research [46] and may not have been the right theoretical basis for the content of the recruitment messages. This limitation is further exacerbated by the paucity of experimental research testing and reporting different messages on digital and social media platforms and their effectiveness on research recruitment. More theory-based formative research using social media marketing techniques, field experiments, and co-design approaches is needed to improve our understanding of the evidence-based application of social influence on research participation for recruiting participants to health research using social media.

Second, throughout the design and testing of both the social norms and patient endorsement messages, the authors considered whether the messages were suboptimally designed despite having contributions from patient representatives who reviewed the messages, and these have undergone various iterations. This is due to 2 reasons. While previous studies have shown that proximal social norms are more effective in different contexts [30-36], it was not possible to use them in our experiment due to the lack of data supporting the claim at the time of the CLOCS recruitment [37] and the novelty of this case-control study. Additionally, the social norm message had to use a vague verbal quantifier, “many women,” and focus on satisfaction with participation rather than the participation rate to ensure messages were ethical and not coercive. Similarly, the patient endorsement message may not have highlighted the link between motivation and action because the message only referred to the patient representative buying medication for symptoms from retailers before a cancer diagnosis. It is possible that future studies focusing on barriers and facilitators of health research participation in the design of the recruitment messages rather than normative behaviors may demonstrate different outcomes.

The authors aimed to address the aforementioned issues with feedback from the participants. However, this exercise did not lead to clear future recommendations other than the use of the factual message used in the control condition. Nevertheless, the outcomes of this experiment informed the recruitment of the CLOCS participants. The authors gained further understanding of the potential limitations of recruiting participants to CLOCS and successfully recruited 249 participants using Facebook advertisements with the control message [47]. The cost per participant recruited was between US $12 and $19, which is comparable to and less than other health-related studies with a targeted population [48]. This hypothetical experimental study demonstrates the importance of testing messages to be used in internet-based recruitment strategies, the potential limitations, and biases, and not relying only on consensus methods. Embedding process evaluations and pre- and postresearch data collection could have a significant impact on the resources allocated to recruitment as well as whether they reach their intended outcomes. Based on the outcomes of the CLOCS study [37], future studies could emphasize how the participation of women without ovarian cancer in the case-control study can help better early diagnosis of ovarian cancer, use their response rates, and further explore why the existing participants took part in this research to develop effective messages.

https://formative.jmir.org/2024/1/e48538
Furthermore, in line with previous literature, we measured attitudes toward the simulated Twitter message and CLOCS to capture individuals’ potential reactions to the website [42], which has its limitations, as several studies have reported on the intention-behavior gap [49]. As such, motivational interventions are necessary but often not sufficient to change behavior. It is possible that the hypothetical nature of the web-based experiment may have introduced a potential response bias. Similarly, there might have been a social desirability or agreement bias, where study participants tended to overestimate their intentions.

Last, we did not account for participants’ familiarity with Twitter. Our sample may have contained women who were not used to reading messages containing hashtags (Twitter use was not verified). Moreover, while we tried to include a behavioral outcome by including an option for participants to visit the CLOCS website, our experiment did not formally assess the analytics of the website or investigate how the messages influenced click behavior. Final, this study may have been affected by selection bias, as factors shaping computer use (age, gender, socio-economic status, etc) tend to influence the demographics of the sample in web-based studies [50]. For instance, there are usually similar demographic patterns across social media platforms, where users are primarily made up of young, female, and urban individuals [51,52].

Implications for Policy and Future Research

Our findings suggest that researchers conducting health-related studies should focus on using simple messages for participant recruitment through Twitter. To increase engagement with potential participants through social media, recruitment messages should be easy to read, transparent, and appropriately targeted to an audience that could have experience related to or an interest in the proposed study. Future research could test messages involving social proofing, such as sharing the experiences of study participants. Additionally, messages could be tested in field experiments by controlling the date, time, hashtags, and images used.

Conclusion

In conclusion, our results indicate that adding behavioral messages containing patient endorsement or social norms to simulated recruitment messages on Twitter decreased participants’ intention to take part in a real-world case-control study. The social norms message also decreased participant interest in visiting the actual study website. These results can be partially explained by difficulties in reading and understanding the message content, with the addition of normative behavioral components. Future research should continue exploring and optimizing methods that can effectively leverage social media platforms for the engagement of potential participants in health-related research.

Acknowledgments

This study was funded by Cancer Research UK’s Early Detection and Diagnosis Project (grant C38463/A26726 to YH and HRB) with support from the Peter Sowerby Foundation. The authors would like to acknowledge Emily Kuipers, MSc, who worked on this project as part of her MSc health psychology dissertation project at University College London. We did not use generative artificial intelligence for this project or for its write-up.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request

Authors’ Contributions

STS, RK, and YH developed the study concept and design. YH, HRB, and JMF designed and developed the Twitter messages for the Cancer Loyalty Card Study. STS and YH performed the data analysis and interpretation. STS and YH drafted the manuscript, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.

[DOCX File, 2367 KB - formative_v8i1e48538_app1.docx ]

References


6. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMRI Res Protoc 2016;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]


21. Most popular social media platforms in the United Kingdom (UK) as of the third quarter 2022, by usage reach. Statista.


24. Stoffel et al. JMIR FORMATIVE RESEARCH


Abbreviations

aOR: adjusted odds ratio
CLOCS: Cancer Loyalty Card Study
OR: odds ratio

©Sandro T Stoffel, Jing Hui Law, Robert Kerrison, Hannah R Brewer, James M Flanagan, Yasemin Hirst. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Clinical Informatics Team Members’ Perspectives on Health Information Technology Safety After Experiential Learning and Safety Process Development: Qualitative Descriptive Study

Chantelle Recsky¹, PhD; Kathy L Rush², PhD; Maura MacPhee¹, PhD; Megan Stowe³, MSN; Lorraine Blackburn⁴, MSN; Allison Muniaik⁵, MASc; Leanne M Currie¹, PhD

¹School of Nursing, University of British Columbia, Vancouver, BC, Canada
²School of Nursing, University of British Columbia Okanagan, Kelowna, BC, Canada
³Digital Health, Provincial Health Services Authority, Vancouver, BC, Canada
⁴Vancouver Coastal Health, Vancouver, BC, Canada
⁵Health Quality BC, Vancouver, BC, Canada

Corresponding Author:
Chantelle Recsky, PhD
School of Nursing
University of British Columbia
T201-2211 Wesbrook Mall
Vancouver, BC, V6T 2B5
Canada
Phone: 1 604 822 7417
Email: chantelle.recsky@ubc.ca

Abstract

Background: Although intended to support improvement, the rapid adoption and evolution of technologies in healthcare can also bring about unintended consequences related to safety. In this project, an embedded researcher with expertise in patient safety and clinical education worked with a clinical informatics team to examine safety and harm related to health information technologies (HITs) in primary and community care settings. The clinical informatics team participated in learning activities around relevant topics (e.g., human factors, high reliability organizations, and sociotechnical systems) and cocreated a process to address safety events related to technology (i.e., safety huddles and sociotechnical analysis of safety events).

Objective: This study aimed to explore clinical informaticians’ experiences of incorporating safety practices into their work.

Methods: We used a qualitative descriptive design and conducted web-based focus groups with clinical informaticians. Thematic analysis was used to analyze the data.

Results: A total of 10 informants participated. Barriers to addressing safety and harm in their context included limited prior knowledge of HIT safety, previous assumptions and perspectives, competing priorities and organizational barriers, difficulty with the reporting system and processes, and a limited number of reports for learning. Enablers to promoting safety and mitigating harm included participating in learning sessions, gaining experience analyzing reported events, participating in safety huddles, and role modeling and leadership from the embedded researcher. Individual outcomes included increased ownership and interest in HIT safety, the development of a sociotechnical systems perspective, thinking differently about safety, and increased consideration for user perspectives. Team outcomes included enhanced communication within the team, using safety events to inform future work and strategic planning, and an overall promotion of a culture of safety.

Conclusions: As HITs are integrated into care delivery, it is important for clinical informaticians to recognize the risks related to safety. Experiential learning activities, including reviewing safety event reports and participating in safety huddles, were identified as particularly impactful. An HIT safety learning initiative is a feasible approach for clinical informaticians to become more knowledgeable and engaged in HIT safety issues in their work.

(JMIR Form Res 2024;8:e53302) doi:10.2196/53302

KEYWORDS
informatics; community health services; knowledge translation; qualitative research; patient safety
Introduction

Background

Health care delivery is increasingly dependent on technology, and health care organizations are heavily investing in technological infrastructure [1]. Health information technologies (HITs), such as electronic health records (EHRs), computerized provider order entry, and mobile devices, play an ever-increasing role in clinical practice, and it is largely thought that these technologies have the potential to promote safe care and contribute to better patient outcomes [2,3]. However, a growing body of research [4-7] has also identified the potential for HITs to contribute to harm, where harm is defined as something “that should not have happened and that you don’t want to happen again” [8]. In this study, the acronym HIT refers to technologies used in the health care system for health information management. This study focused on HIT safety concerns, more specifically, unintended harm or potential harm that involved an HIT-based system.

An example of harm related to HIT is an overdose event that occurred when a patient aged 16 years received 39 antibiotic pills when they should have received only 1 pill [9]. The overdose event occurred after a series of computer and human failures, including a confusing computer interface design that forced weight-based dosing for pediatrics, an automated “robot” system in the pharmacy that performed a “double check” rather than human verification, a hidden curriculum for prescribing doctors to “ignore all computer alerts,” and a novice nurse who trusted the computer recommendations because “the computer had been right” in the past [9]. The design of technology and overreliance on the accuracy of the information presented can lead to unintended consequences, and there is increasing recognition that events such as this are occurring alongside the increasing uptake of technologies in health care [4-6,10-13]. There is also a growing identification that well-designed and well-deployed systems may help to mitigate some of these issues [14,15]. Clinical informatics teams, with expertise in the effective use of HITs, are ideally positioned to recognize and respond to HIT-related harms and contribute toward enhanced quality and safety in health care delivery. In this study, we explored clinical informaticians’ experiences in incorporating safety practices into their work.

Defining HIT Safety and Harm

The field of patient safety has conventionally focused on hospital and inpatient settings, with the measurement and monitoring of adverse events (including near misses) well established as a widespread practice designed to reduce harm from activities in the acute care setting (eg, medications, surgical procedures, falls, and diagnostics) [16-19]. For this study, focused on primary and community care settings, a broader view of safety was taken up, with consideration for all 6 interconnected dimensions of quality health care (safety, effectiveness, patient centeredness, timeliness, efficiency, and equity) [20]. Using the broad definition of harm as “something happened that you did not want to happen,” harm is not simply the opposite of safety; instead, the definition makes space for recognizing the potential for harm related to other circumstances, such as inequity, inaccessibility, and poor patient experiences. Recognizing the interconnectedness of the 6 dimensions of quality, an expansive definition of harm is useful because it places a greater emphasis on the complex nature of the health care system and can thus help identify latent problems. A focus on latent problems that contribute to harm (as opposed to a focus on human error) has the potential to yield more systems-focused solutions to mitigate against recurrence and thereby improve the quality of care. This is particularly relevant with HIT, where latent errors may impact people who are several degrees away from an HIT origin of error, such as the numerous clinicians involved in the overdose example described in the preceding section.

Embedded Research Context

This study was conducted as part of an embedded researcher project supported by an innovative program and funding model designed to maximize the impact of research by supporting formal partnerships between emerging academics and health system leaders to address pressing challenges in health service delivery [21]. The funding program provided support for a doctoral student researcher to conduct their dissertation research while holding a position within the health service organization. The program is designed to help researchers develop professional skills to support evidence-informed improvement within the health system alongside conventional research outputs [22]. The embedded researcher, the researcher’s supporting academic committee, and leaders from the organization collaborated to design an applied research project that met usual academic requirements and simultaneously was relevant and useful to the organization. In project conceptualization, the health care organization prioritized patient safety and a focus on HIT in primary and community care settings. Upon funding, the researcher was embedded within the organization’s community clinical informatics team for a 2-year period (2019 to 2021). The embedded researcher was experienced in clinical education and had experience in the areas of clinical nursing, clinical education, informatics, patient safety, and quality improvement. Having a role within the organization allowed the researcher extraordinary insights into the role of the team within the organization, the team’s learning needs related to HIT safety, and opportunities to address learning needs. Fostering daily working relationships was an intentional aspect of the project to support engagement and build capacity in the team, apply learning to existing processes, and promote sustainable practices [23]. The organization’s quality and safety leaders were also key contributors to the project, collaborating closely with the embedded researcher to guide the course of the project.

Codevelopment of the HIT Safety Process

In the early stages of the project, the embedded researcher worked with a clinical informatics leader to conduct a retrospective sociotechnical analysis of reported HIT safety concerns [24]. This part of the project was possible because the organization had added a question (“Was a computer involved in the incident?”) to their web-based voluntary incident reporting system in 2016. The findings from the incident analysis study [25] provided the foundation for the clinical informatics team and the embedded researcher to cocreate a new process that
uses sociotechnical systems analysis for identifying, analyzing, and responding to HIT safety events. The new process used safety huddles [26,27] and was aligned with the concept of a learning health system in which theoretical frameworks and scientific evidence are integrated with internal data to inform continuous improvements [28,29]. Throughout the project, the researcher was positioned within the clinical informatics team with a constant focus on facilitating learning using adult learning principles [30-32], drawing on a variety of fields such as patient safety [17,18,20,33,34], quality improvement [35-37], and learning health systems [38-40] as well as HIT safety and harm specifically [5,41-45]. The aim of this study was to examine the clinical informaticians’ experiences in learning about HIT safety and to understand their experiences in codeveloping the new process to address HIT safety concerns in their work.

Methods

Setting
The study was carried out in a large health care organization in western Canada that provides acute, primary, and community care for >1.25 million people and includes both densely populated urban areas as well as rural and remote communities. The study focused on members of a clinical informatics team assigned to support services delivered in non–acute care settings (ie, primary care, home care, population and public health, long-term care, mental health, and substance use services). The clinical informatics team comprised 15 to 20 multidisciplinary staff members as part of the organization’s efforts to strengthen primary and community care delivery. The team was established at the beginning of the research project, which proved to be serendipitous for the research project because the embedded researcher was able to enter a newly formed group. The team was responsible for the clinical integration and operation of all HIT systems in primary and community care settings. There were different roles and responsibilities among the members of the team. For example, the educators were responsible for supporting clinical staff in using the clinical software systems. The specialists were responsible for working with the HIT software development team to communicate the changes that might be required for the HIT system (eg, practice policy changes). In addition, team members were dedicated to specific clinical service areas, such as mental health or home care. The clinical informatics team did not include any prescribing clinicians, pharmacists, or medical office assistants; however, the embedded researcher did consult and obtain input from members of these groups to inform them of the learning experiences and codevelopment of the HIT safety process. An organizational chart of the team is available in Figure 1.

Figure 1. Team chart and research participation.

Study Design
This study used qualitative description methods to explore the clinical informaticians’ perspectives and reflections on their experiences of learning about HIT safety and the codevelopment of the process to manage HIT safety events. Qualitative description is useful for capturing the meanings and interpretations that informants ascribe to their experiences [46,47].Aligned with a constructivist paradigm, which posits that human knowledge is subjective and socially constructed, the study aimed to capture both the participants’ perspectives and acknowledge the subjectivity and involvement of the embedded researcher in constructing interpretations of the data [48]. In this study, having been embedded within the team for an extended period, the researcher was able to glean an in-depth understanding of the context surrounding the participants’ accounts.
During data collection and analysis, the embedded researcher engaged in ongoing reflexivity by attending to their unique positioning, the circumstances surrounding the study, and the potential influences on knowledge construction [48]. From a postpositivist perspective, the close relationship between the researcher and participant is conventionally thought to perpetuate bias and prevent the attainment of rigor in research [49,50]. However, we contend that it was the strength of the relationship between the embedded researcher and the clinical informatics team that allowed the researcher to fully explore how HIT safety was in alignment (or not) with the work of the team. The intention of situating an embedded researcher with the team was to encourage relationships between the researcher and the clinical informatics team members to support the meaningful, effective, and sustainable integration of knowledge into practice [51-54]. Indeed, as team members began to apply their learning in their work (ie, using the sociotechnical framework to analyze a problem), the researcher was available for guidance and consultation as needed, and as the team’s capabilities developed, less support from the researcher was required.

Ethical Considerations
Before conducting this study, ethics approval was obtained from the University of British Columbia Research Ethics Board (H18-02677), and all participants signed a consent form. The conduct and reporting of the study followed the Consolidated Criteria for Reporting Qualitative Studies guidelines for qualitative research reporting.

Data Collection
We used purposive sampling, targeting clinical informatics team members who supported primary and community care and who had participated in the HIT safety initiatives over the previous 12 to 24 months. The embedded researcher emailed all team members (N=16), inviting them to participate in 1 of 3 web-based focus groups, with a clear statement that participation was optional and in no way related to their job. The study was carried out during the COVID-19 pandemic when face-to-face meetings were discouraged; thus, the embedded researcher facilitated focus groups over Zoom (Zoom Video Communications, Inc) [55]. A semistructured interview guide was used (Multimedia Appendix 1), and the focus group sessions were recorded using Zoom video capture, downloaded onto the private secure computer of the researcher, and manually transcribed verbatim by the researcher. The informants reviewed the transcript of their comments, and all participants approved the transcripts with no revisions.

Data Analysis
Thematic analysis was conducted [46,56,57] using NVivo (version 12; Lumivero). Thematic analysis is well suited to address broad research questions and provides a flexible approach to remain “data-near” [47], searching across the data for patterns and allowing for both inductive and deductive approaches to the analysis. An overarching framework of 3 categories—barriers, enablers, and outcomes (Textbox 1)—was used to provide an initial structure for the analysis.

After completing the deductive coding to classify the data as barriers, enablers, or outcomes, the data were re-examined to inductively generate subcategories. Subcategories were developed and refined over several iterations (between CR and LMC), and a member-checking session was conducted with the clinical informatics team to support the descriptive and interpretive validity of the findings [46,48].

Textbox 1. Operational definitions of high-level analytic categories.

- Barriers to understanding and applying methods to address health information technology (HIT) safety: Activities or conditions that may have impeded learning
- Enablers to learning about safety and mobilizing their knowledge: Activities or experiences that facilitated or promoted learning
- Outcomes of the HIT safety project: A product or result from engaging in learning activities

Results

Overview
Three 1-hour web-based focus groups were held with 10 informants. Of the 10 informants, half were in clinical informatics educator roles, and the other half held roles such as clinical informatics team leader or clinical informatics project manager. A total of 50% (5/10) of the informants had been in their current role for <2 years, 40% (4/10) for 2 to 5 years, and 10% (1/10) for >10 years. In total, 90% (9/10) of the informants had a clinical background, including 6 nurses, 1 physiotherapist, 1 occupational therapist, and 1 social worker. The team member, who did not have a clinical background, had been working in the health care sector for >10 years. The informants’ previous work experience in their respective clinical roles before taking on an informatics-focused role ranged from 0 to 24 years, with an average of 9.7 years.

The informants shared several barriers and enablers related to their experiences of learning about and developing strategies to address safety. They also described outcomes such as new learning and capabilities. Figure 2 displays the categories and subcategories. A description of each item follows, including excerpts from the data. Characteristics of the individual informants are limited to preserve their anonymity.
Barriers to Understanding and Applying Methods to Address HIT Safety

The informants identified several barriers to understanding and applying methods to address HIT safety, including a lack of knowledge, previously held perspectives, organizational pressures, and challenges related to event reporting before the project began.

Limited Prior Knowledge of HIT Safety

A common barrier described by the informants was initially having little or no foundational knowledge related to safety principles and their application to HIT. As an informant reported, “My experience with technology safety was pretty limited prior to all of this... I didn’t really know much at all.” More specifically, informants shared that they had a limited understanding of any negative impact HIT systems could have on patient care. For example, one of the most experienced informants stated, “I had no idea what was going on for people at the front-line level with whatever they consider a computer incident.” This limited initial understanding exemplifies the learning needs within the team that needed to be addressed to begin to incorporate HIT safety practices into their work.

Previous Assumptions and Perspectives

The informants also shared different assumptions and perspectives about safety and harm in health care that they had previously held. Some were unaware that HIT could introduce risks to safety, were under the impression that safety concerns were beyond the scope of their role, or attributed safety concerns to user mistakes. An informant stated:

Technology was always supposed to be something that would make things safer, right? You streamline some processes, you make some things maybe, more automatic, take out some of the human element—that should be safer.

Others expressed that they were initially uncertain whether this topic was relevant. For example, an informant who had been with the team for 2 years stated, “I felt like it wasn’t really my role or responsibility to be identifying them [safety concerns].” Another shared that they had previously assumed another department of “internal auditors” was responsible for addressing problems related to HIT and now they saw the value in having a clinical perspective, “focusing on solutions, rather than, it’s just a number.”

Competing Priorities and Organizational Barriers

The informants indicated that their capacity and capabilities to address HIT-related safety concerns were challenged by competing priorities and organizational barriers. An informant who had worked within the organization for 4 years perceived this as a systems issue: “The whole organization gets caught up in all these new initiatives, all these new projects, all these things...”

Outcomes of the HIT safety project

- Limited prior knowledge of HIT safety
- Previous assumptions and perspectives
- Competing priorities and organizational barriers
- Safety culture and reporting

Enablers to learning about safety and mobilizing their knowledge

- Participating in learning sessions
- Gaining experience in analyzing reported events
- Participating in safety huddles
- Role modeling and leadership from embedded researcher

Individual outcomes

- Expanding ownership and interest in HIT safety
- Developing a sociotechnical systems perspective
- Thinking differently about safety
- Developing ability to identify and analyze reported events
- Applying learning to mitigate against harm
- Growing consideration for user perspectives

Team outcomes

- Enhanced communication within the team
- Informed future team work and strategic planning
- Promoted a culture of safety
you know to make things better, but learning from the past and all these safety incidents, I feel like they get brushed aside.” Another team member described recognizing that there was a need for a focus on HIT safety but struggled with knowing how to proceed: “It just seemed too big to wrap our hands around without some support.”

Safety Culture and Reporting
A further challenge was related to the internal patient safety reporting system and processes, including concerns that it is onerous to use and, therefore, underutilized. A person who had extensive clinical experience questioned the usefulness of the reporting system, “Coming from the mindset of a clinician, you’re busy you know, am I going to take the time to do a [report], what’s the value there?”

On a broader scale, there were also knowledge gaps related to using the existing reporting system. As 1 person commented, “There’s still a lot to do in developing the reporting culture around technology needs in community.” More specifically, an informant noted the point-of-care staff’s lack of understanding of what constituted a computer-related safety event, “People don’t really understand what [an HIT] system-related error really is, and so a lot gets put into that [reporting] system that may not be appropriate for our eyes.”

The informants were also concerned that underuse of the patient safety reporting system meant that issues reported in the event reports were just the tip of the iceberg: “There’s so many safety issues that we don’t know about, things that are actually happening that aren’t being reported.” The challenges with the reporting system impeded their ability to address HIT issues because fewer reports meant that they had fewer opportunities for learning by analyzing events. Another informant made an analogy to the concurrent COVID-19 pandemic:

“It’s just like the COVID out there right now, there’s probably more cases than there actually are, we just don’t know about them...what we don’t know, is how to actually properly capture that all, and encourage people to come forward when they have an issue.

Enablers to Learning About HIT Safety and Mobilizing Their Knowledge
The informants reported some key enablers to learning about HIT safety and mobilizing their knowledge, including making space to participate in learning sessions and safety huddles, the hands-on experience of analyzing reports, and their observations of the embedded researcher as a role model.

Participating in Learning Sessions
Team members identified short education sessions led by the embedded researcher as supportive of their learning, having “collapsed all the salient points into a quick, easy-to-understand, salient presentation.” The evidence and resources referenced in the learning sessions were also identified as helpful: “I don’t know that there’s often opportunity to bring in scientific literature into our day-to-day jobs so I think that was a great opportunity, to hear and to see what’s happening in the academic realm and consider its application to practice.”

The informants specifically highlighted the case study about the antibiotic overdose (described earlier in this paper) [9] from the learning sessions as an effective tool to understand complexity and sociotechnical systems. A clinical informatics educator who had been with the organization for 3.5 years stated: “That [case study] was very, very engaging and very interesting, and you could see how like, just to see the breakdown like that...and it again, made me aware of all these little things that can go wrong or have to go wrong to lead to something like this, and how the [HIT] system played into it at each step.”

Gaining Experience Analyzing Reported Events
The informants also shared that the experience of analyzing safety events was valuable in learning how to apply what was gleaned from the analyses to make improvements. An informant who was newer to the team explained, “Actually receiving the [reports] and actually doing the investigative work teaches you a lot about the [HIT] systems... I enjoyed doing that and then thinking about how it could be better in the long run.” They went on to explain how this experience provided an opportunity to consider how different sociotechnical dimensions may be related to safety concerns: “You’re not just thinking about the actual documentation system, but you’re thinking about all the systems around that, like whether it be a workflow, or chaos, or whatever it is that contributed to that scenario.”

Participating in Safety Huddles
The informants also explained that the experience of sharing the event analyses with other members of the team and participating in the safety huddles was supportive of their learning around HIT safety in that this activity provided a peer learning experience. One of the educators stated, “I found [safety huddles] really informative, especially having other people there that use other [HIT] systems, and again, it’s someone else’s perspective and how they’re reading the situation and what I can learn from that other person that I’m working with.”

Another team member expressed an appreciation for safety huddles as a venue for communication among the team members and found them worth the time and effort for the team:

“I found that making space for us, like dedicated time and focus, to talk about these concerns that we have, or the patient safety events that have occurred... I’d never had that experience before, and I found it was so helpful talking as a group about what we found or what those problems that were being reported were about... I found that it was easy to actually make the time and spend the effort to do that. You know we are all busy, but I think in the long run it’s all going to do us well as an organization and as a team to continue that [the safety huddles]”

Role Modeling and Leadership From the Embedded Researcher
The informants also indicated that the role of the embedded researcher supported their learning and facilitated changes within the team’s practices. An informant who had been with the team for 2 years described the key function of the embedded researcher as initiating a focus on HIT safety: “I think we needed
someone to come in and really help to set the tone and set that framework...and so it’s been learning. I think the last year has been learning across the board.” Building on this, the informants also recognized the embedded researcher as an expert and champion for HIT safety, as a different informant explained: “It’s helped really mobilize the team in that direction and create more of a team sort of focus on working with these teams committed to safety-related issues.” An educator from the team described the embedded researcher as a role model for how to approach analyzing safety events, having “instilled curiosity” in the team. Finally, there was also a recognition of how to integrate HIT safety into the work of the informatics team; a long-standing member of the team noted, “what I didn’t realize before is really how well this conversation fits within an organizational structure and within a team structure.”

**Outcomes**

A variety of outcomes from the HIT safety project surfaced in the focus group discussions, some of which were individual outcomes and some of which were team outcomes. Individual outcomes included increased ownership and interest in HIT safety, the development of a sociotechnical systems perspective, and increased consideration of end-user perspectives. Team outcomes included increased team communication and the ability to use the processes to guide strategic planning.

**Individual Outcomes**

**Expanding Ownership and Interest in HIT Safety**

Several informants shared an increased sense of personal interest in the topic of HIT safety. A team member explained, “It’s given me an appreciation and actual interest in safety and how that pertains to design of [HIT] systems and how we interact with the [HIT] systems.” Furthermore, an informant with 8 years at the organization described having an increased sense of both personal and team ownership in relation to HIT safety: “Not only is it my role and responsibility, but we’re really well positioned to identify and sort of bridge between practice and workflows.” The informants recognized the role their team plays within the organization in supporting the delivery of care, as one person with extensive clinical and informatics experience noted, “Having a good understanding of why we exist as informatics...it’s not just for the users, although that’s important, it’s also for the patients and reducing risk...so it all ties together...that kind of holistic view.”

**Developing a Sociotechnical Systems Perspective**

The informants shared new insights into their work based on learning about sociotechnical systems theory as it applies to informatics. They expressed an increased appreciation for the relationships among the technology, the users, and the context in which these are situated. An educator with 24 years of experience explained:

> I sort of think that technology doesn’t take into account the human being. It’s just, technology is a set of algorithms, it’s a set of stuff that’s written by a developer who tries to take in all the considerations possible. But you can’t take in all the considerations of a human being, and how a human being will respond to certain situations, or certain pieces of technology.

This was echoed by another educator with 3.5 years of experience, who stated, “Our technology is only as good as how people understand it, and so the education piece around it and you know, understanding the workflow and how to actually apply it and use our [HIT] systems is so important...because the [HIT] system could be working as designed, but if people don’t know how to use it properly, then it just leads to a lot of problems.”

Related to this, the notion of human factors as it applies to safe HIT use was also a new concept for many team members. An informant explained, “I never knew this existed, human factors—and now I am seeing that there is a whole theory behind it, there’s a lot to learn about, there’s best practices in design, there’s all these things that I had no idea even existed.”

**Thinking Differently About Safety**

Building on their knowledge of the sociotechnical perspective, the informants demonstrated an increased awareness of the factors and circumstances that may increase the risks of harm. As one educator highlighted, they previously “had just kind of considered the obvious errors, like with a malfunction or with a bug or something like that.” However, this educator went on to explain how their awareness of safety risks had grown beyond the technical aspects of HIT, noting, “Technology’s not infallible. There’s so many factors, and it’s quite complex, and it’s given me kind of an appreciation for...the whole topic, and it makes me think about problems in a different way.” Similarly, another educator described how their view of safety had expanded beyond just focusing on the HIT end user:

> It’s not simply, one person did something wrong. There [are] so many different things—there’s the workflow, the human factor, was it the actual user interface—all those different subcategories.

**Developing the Ability to Identify and Analyze Reported Events**

The informants shared how they applied what they learned to their work and developed the ability to identify and analyze HIT-related safety events. First, among the informants, there was a greater awareness of “just what is a safety event,” as one of the more experienced members of the team put it. Several people described adopting the practice of systematically analyzing reported events. For example, how they learned to “think of [HIT safety events] in those different dimensions that we learnt about with the sociotechnical model, just being able to, like, think of it in a framework like that, in sort of a structured way, to help break it down.” Another informant shared how taking a systematic approach had changed their thinking:

> It does help us see sort of where perhaps a gap was with a reported [event]. So, before we just knew there was an issue, but...I wasn’t looking at it as all these different sort of levels.

Similarly, another informant explained:

> What I’ve learned is how to break it up. Was it the [HIT] system?... Was it the workflow? The process?
Applying Learning to Mitigate Against Harm

Going a step further, informants also articulated how they learned to be more proactive and tried to prevent harm from occurring in the future. An educator explained that they have observed patterns over time from the analyses of events, and this has increased their awareness of the potential for future risks:

I’m actually already starting to see some patterns and starting to think about the complexities of some of these [HIT] systems. Just this one tiny little move can make a big change, can put someone at risk, and it shouldn’t be that easy to put a client [patient] at risk.

Another educator described having an increased awareness of patterns as well, and how this informed their thinking around mitigating future issues: “I think being aware of the patterns and being aware of the common sort of issues is really helpful in terms of thinking about future solutions and making sure that those problems can’t be easily replicated.” Furthermore, the informants described how they had been able to communicate concerns about HIT safety in the context of their work. One of the more experienced team members explained, “I didn’t always have the language to describe why something in the [HIT] system was a problem, but that information with systems thinking really helped me frame those conversations.” Another informant described how the clinical informatics team has begun to take a more proactive approach to safety:

I think it’s also changed the conversation around; just when we’re discussing [HIT] system changes or potential projects that we may undertake, is just the safety risks factors. Having more general dialogue around that...just being more proactive.

Growing Consideration for User Perspectives

Consideration for clinicians and HIT users was a commonly expressed sentiment. Team members reflected on their previous experiences as clinical care providers and reflected on this when considering the functions and dysfunctions of HIT in a clinical context. An educator commented:

When we get these reports now, I’m trying to think like the clinician. I’m not working as a clinician anymore, but I am trying to put myself in their shoes—what are all these other surrounding factors, what led them to report about this?

Another informant in the role of educator elaborated on this idea:

But it’s probably easier if you have the open mind to actually really understand, again the empathy factor of it, understanding what had happened in terms of if it’s a system error or whatever, and yeah, it’s just having that understanding that it’s not always the person at fault, it’s not always the system at fault, it could be a combination of everything. And again, what do we do next? It’s how do we learn from this.

Pairing this consideration with their knowledge of HIT safety, another informant demonstrated new insight into the users’ experiences with HIT:

And so they think it’s one way, and then an error happens because they misread something or they didn’t know to check somewhere, and...I just felt like I never clued into how much the design can really impact that front end user experience.

Furthermore, another informant contextualized the challenges clinicians and HIT users may face in using HIT:

I’m just thinking about, really, the environment that people are working in and the complexity of that...the environment might be...very chaotic, and then we’re asking them to do something very complex in the [HIT] system. I think that is a safety concern.... I think that’s where it would be very easy to have errors, obviously.

Team-Level Outcomes

Enhanced Communication Within the Team

The informants described how sharing their experiences about HIT safety worked to enhance communication within the team and further expand their awareness of potential HIT safety concerns. As described earlier, there were separate groups within the team that focused on different HIT systems and clinical areas in primary and community care, which could sometimes create siloed communication. Team members expressed that participating in the HIT safety activities opened up new communication channels and supported an exchange of learning across the different teams. An educator, whose work was focused on a particular clinical area, shared the following:

It was great awareness to hear what was happening in other places...kind of raising that awareness so that if we see something similar in our clinical area, it just kind of alert you to look out for things that you maybe would’ve never considered...and all of a sudden, your level of awareness is there.

Another informant highlighted how increasing communication across the smaller teams within the larger clinical informatics team helps provide better support to the clinicians or HIT users:

“Clinicians are interacting with many systems, and many applications, and many types of technologies, that in fact, we need those opportunities to speak with our colleagues who lead or support other [HIT] programs so that we really get a sense of what those safety events meant.”

Informed Future Teamwork and Strategic Planning

On a broader scale, the informants also gleaned new insights into the role of their team within the organization. One of the more senior team members asserted:

I think that our team is perfectly positioned to handle, be handlers of technology related [safety] reports, and make sure we close the loop. I think it needs to be part of our work and just have it as a regular ongoing piece of work that we do and...service that we provide to the organization.
In addition, the informants shared insights into how their knowledge of HIT safety relates to organizational decision-making and strategic planning. An informant in an educator role considered how their team can contribute to future decisions about HIT: “Whether it be just the organization, or operationally within a clinic, and they make a request for a change in a [HIT] system, or it’s a bigger change, like made at a higher level, even at the [executive] level, maybe they’re not making the best decision because we’re not providing them with the best information about our clinical [HIT] systems.”

**Promoted a Culture of Safety**

From a patient safety perspective, the informants shared an increased emphasis on a culture of safety and leveraged learning from their analyses to mitigate future concerns and make improvements. An informant explained:

> I think as a team member in clinical informatics, instead of focusing on the mistakes...thinking of the next time. What can be done better? How can it be resolved much better? It’s always the next time, it’s always learning, and it’s always a different situation. But it’s probably easier if you have the open mind to actually really understand...it’s just having that understanding that it’s not always the person at fault, it’s not always the system at fault, it could be a combination of everything. And again, what do we do next? How do we learn from this?

**Discussion**

**Principal Findings**

**Overview**

This study examined clinical informaticians’ perspectives on learning about HIT safety and the co-creation of a process to manage HIT-related safety reports. To our knowledge, this is a novel examination of the topic. The findings from this study provide valuable new insights into the barriers and facilitators to developing HIT safety within clinical informaticians’ practices, as well as the potential outcomes that a robust, evidence-based approach to knowledge mobilization can have.

**Barriers to the Uptake of HIT Safety**

From the clinical informaticians’ perspectives, some of the barriers that were initially challenging included the lack of knowledge about HIT safety and previous assumptions they carried at the outset of the project. For patient safety in general, a lack of knowledge about incident reporting and assumptions about the value or repercussions of event reporting are known barriers to initiating reports [58]. The embedded researcher was able to spend time at the beginning of the project to build relationships and assess the team’s gaps in knowledge. Recognizing the team’s initial limited understanding of HIT safety, learning activities were focused on fundamental topics in HIT safety and contextualizing learning within their existing work. Another identified barrier was competing priorities and organizational barriers, a challenge echoed in the literature about evidence-based practice [51,59]. While competing priorities may be a perennial challenge in health care, strategically aligning and incorporating learning activities into the current priorities of the clinical informatics team was crucial because mutual learning and appreciation of others’ perspectives and contributions may lead to better processes and outcomes by generating more relevant and applicable knowledge [60].

The findings surfaced challenges with the internal safety reporting systems, the processes, and the underuse of the safety system. The informants expressed concern over the value of reporting if there is not a proper follow-up, stressing the value of “closing the loop” to ensure that the person who reported the event is aware of the implications of their report from an informatics perspective. The costs and benefits of safety reporting are debated in the literature. Insufficient action following a safety report is thought to negatively affect clinicians’ commitment to the reporting process [61]. Macrae [62] argues that “we collect too much and do too little” [63], explaining that although the technical infrastructure for safety reporting has been established in many health care organizations, the requisite processes of investigation and improvement have been underemphasized. Other research suggests that low rates of safety reporting derive from clinicians being prone to applying quick fixes or workarounds to system failures rather than reporting issues to trigger more in-depth analysis and sustainable solutions [64,65]. In any case, it seems that safety reporting is yet to achieve its full potential [61], and the latency of safety issues related to HIT may pose further challenges, with near misses and errors being dismissed or going undetected [66]. However, the findings of this study suggest a way to establish practices to identify and mitigate latent errors.

**Supportive Learning Environment and Openness to New Ideas**

The enabling elements identified by the clinical informaticians focused primarily on participatory and experiential activities. Facilitating informal, locally owned processes for clinical informaticians’ learning around safety has been shown to enable the staff to raise concerns and actively contribute to improvement [58]. Although there is a growing collection of research studies that have applied the Sittig and Singh [24] sociotechnical model in analyzing safety concerns [27,67-69], no literature was identified where this framework was incorporated into experiential learning activities or embedded into clinical informatics work processes in real time. Safety huddles are thought to be transformational in shifting attitudes and practices related to safety, providing a “reliable framework for interdisciplinary communication and action” [26]; however, the evidence to support this is largely anecdotal [70]. Menon et al [27] used safety huddles to address EHR safety concerns in a hospital setting, and although the format differed from this project, the huddles promoted a culture of safety for clinical informaticians, providing a venue for open communication about safety concerns and facilitating learning and improvement, which was also found in this study.

The informants also noted the role the embedded researcher played on the team to “set the tone,” “mobilize the team,” and “instill curiosity,” which supported their learning in the project. A concerted effort was made in the initial stages of the project to develop strong, collaborative partnerships at all levels of the...
The findings indicated that the clinical informatics team exhibited the knowledge, skills, and attitudes of an effective team [72], which in turn support the notion of a high reliability team [73]. For example, the participants found the safety huddles a “good use of their time.” Furthermore, formal financial and organizational support for the embedded researcher throughout the duration of the project created a fertile environment for learning, with dedicated resources, endorsement and collaboration from leaders, and multiple opportunities for interactions between researchers and knowledge users [51,60].

**Effective Knowledge Mobilization**

The findings of this study include positive outcomes in terms of moving knowledge into action. The focused approach to supporting HIT safety seemed to support a group-level identity transformation, incorporating different professional perspectives, adding value, and acting as a lever for system-wide, evidence-informed sustainable change [74]. The informants described being more knowledgeable and engaged in HIT safety issues in their work. They developed their knowledge base in clinical informatics, with an increased recognition of some perennial problems related to HIT [60,75], and the contextual issues that surround the use of HIT in health care settings [76,77]. The informants also expressed greater ownership regarding safety. This is echoed in the literature in which EHR safety is ascribed as being a shared responsibility among key stakeholders including EHR developers, health care organizations and users, and government regulators [42]. Clinical informatics, with its emphasis on bridging technical and clinical perspectives, can play a central role in facilitating efforts to improve safety [78,79].

The cocreated sociotechnical analysis process for addressing HIT safety events produced immediate and ongoing insights to inform operational decision-making within the organization. The current findings provide additional evidence that a clinically focused informatics team is well positioned to take on this work of “closing the loop” with the end user using the system to report an event. Similarly, by leveraging voluntary safety reporting for quality assurance, Williams et al [54] identified 242 EHR-related safety events analyzed by nurse informaticians, 30 of which led to specific system changes to improve usability. In this study, informants expressed an appreciation for the structure that the Sittig and Singh [24] analytic framework brought to conducting an analysis and that they learned to look for patterns in reviewing reported events. The process developed by the team was an adaptive approach based on experiential cycles of learning, from which they gradually developed new insights and expanded their collective expertise on HIT safety [80]. Demonstrating a thoughtful approach to safety, the findings indicate that informants’ perspectives moved beyond a reactive “find and fix” approach and instead they were embracing complexity to “enable things to go right more often” [81].

The findings of our study also indicate that enhanced communication helped team members develop a more empathetic approach to supporting clinicians using HIT. Specifically, safety huddles were thought to have improved communication within the team as well as informed their perspectives on all aspects of their work, including planning for future HIT-related needs of the organization [27,82]. Although safety event reporting is not without its limitations [83], reporting can effectively contribute to participatory learning, improve practice, and promote safer care [84].

**Limitations**

A possible limitation of this study is that the embedded researcher functioned as both the lead of the initiative and the interviewer for this project. It is possible that the informants’ responses were influenced by social desirability [48]. However, the participants’ responses also showed vulnerability. For example, several respondents indicated that they had been judgmental about end users’ errors in the past. This level of candor suggests that social response bias may have been minimal and possibly was overcome by prolonged engagement, given the long duration that the embedded researcher participated with the team. This study also assessed only the experiences of the clinical informaticians in retrospect. It did not account for the impact on the knowledge users in the same way that a longitudinal design may have. Tracking the knowledge user’s experiences over the course of the project may have offered a more precise account of the impact of the various approaches to supporting HIT safety and the progression of the partnerships within the research [52]. Future research should focus on assessing the mechanisms by which the impact is achieved to articulate an optimal, replicable approach to knowledge mobilization. In addition, this study was situated in a nonacute setting, and therefore, the applicability of our findings is limited as such. However, given the adaptable and codeveloped nature of the processes for learning, it is possible that other health care settings may benefit from using similar approaches [45,85].

**Conclusions**

Overall, the findings of this study indicate that the evidence-based, experiential learning model used in this instance was an impactful approach to supporting HIT safety in the context of clinical informatics. Furthermore, the intensive focus on HIT safety resulted in increased knowledge and some evidence of group-level identity transformation related to clinical informaticians’ management of HIT safety events. An embedded researcher model can be an effective mechanism to support clinical informaticians in learning and applying HIT safety practices in their work.

**Acknowledgments**

The Canadian Institute of Health Research Health System Impact Fellowship funded this study.
Authors' Contributions

CR contributed to the conceptualization, investigation, formal analysis, and wrote the original draft. KLR was involved in the conceptualization, methodology, and participated in writing, reviewing, and editing. MM contributed to the conceptualization, methodology, and also participated in reviewing and editing. MS, LB, and AM each contributed to the conceptualization, provided resources, and were involved in the writing, reviewing, and editing process. LMC contributed to the conceptualization, supervised the project, was involved in formal analysis, and participated in writing the original draft, reviewing, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Semistructured interview guide.

References


Abbreviations

EHR: electronic health record
HIT: health information technology
Attitudes, Barriers, and Motivators Toward Daily Walking and a Mobile App to Increase Walking Among Women: Web-Based Anonymous Survey

Catherine Jones1*, MPH; Shikha Chandarana1*, MPH; Amita Vyas1*, PhD; Melissa Napolitano1*, PhD
Milken Institute School of Public Health, The George Washington University, Washington, DC, United States
*all authors contributed equally

Abstract

Background: There are disparities in the prevalence of physical activity (PA) with women engaging in less PA than men, a gap which widens during midlife. Walking is a generally accepted form of PA among women and should be encouraged. Motivations, barriers, and attitudes to engaging in walking change with age, but the influencing factors are not well understood nor are the features of mobile apps that facilitate daily walking.

Objective: This study explores the relationship between age and women’s self-reported motivations, barriers, attitudes, and beliefs toward daily walking. It further assesses attitudes toward features of a mobile app designed to sync with a wearable step tracker to increase and maintain levels of daily walking among women.

Methods: A web-based anonymous survey was completed by 400 women, aged 21-75 years. The 31-item survey captured women’s perceived barriers and motivators toward daily walking and attitudes toward mobile apps to support and maintain daily walking. For analysis, responses to the survey were grouped into 2 categories of women: ages 21-49 years and ages 50-75 years. Bivariate analyses were conducted through SPSS (IBM Corp) for each of the survey questions using chi-square for dichotomous variables and 1-tailed t tests for scales and continuous variables to identify significant differences between the groups. One-tailed t tests were run for scaled variables to identify significant differences between the 10-year age increments.

Results: Significant barriers to daily walking were observed in the 21-49–year group for personal and work responsibilities, motivational and psychosocial factors, and physical and environmental factors. Motivators to walk daily in the 21-49–year group were significantly higher to reduce stress and anxiety, and motivators to walk daily in the 50-75–year group were significantly higher to help manage or lose weight and to reduce the risk of chronic illness. Women’s walking preferences, beliefs around their walking behaviors, and their perceived importance of the features of a future mobile app for walking designed specifically for women showed significant variation according to age. When asked about the importance of features for a mobile app, women aged 21–49 years indicated a significantly higher number of positive responses for the following features: digital community support, rewards or point system, and seeing a daily or weekly or monthly progress chart.

Conclusions: Our findings indicate that barriers, motivators, and beliefs around daily walking and the importance of preferred features of a mobile app vary according to women’s ages. Messaging and app features should be tailored to different age groups of women. These study results can be viewed as a foundation for future research and development of mobile health interventions to effectively increase daily walking among women of all ages.

(JMIR Form Res 2024;8:e48668) doi:10.2196/48668

KEYWORDS
mHealth; mobile health; mobile app; walking; physical activity; step counts; women’s health; age; wearable activity tracker; chronic disease; mental health; mobile phone; COVID-19
**Introduction**

Physical activity (PA) has established benefits of preventing and treating many adverse health conditions in women, such as heart disease [1], type 2 diabetes [2], osteoporosis [3], and anxiety [4]. Current guidelines recommend that adults engage in at least 150 minutes of weekly moderate-intensity aerobic PA, such as walking, and suggest that health benefits can be achieved in bouts as short as 10 minutes [5]. However, despite the overwhelmingly positive evidence for performing regular PA, older adults continue to be underactive, with older women in particular being the most inactive segment of the population [6]. This gender difference, which is observed across the life span, widens during midlife (ages 40-60 years) [7,8]. Regular exercise declines for many women just when menopause-related physiological changes increase their risk of weight gain and chronic diseases [7,9-11].

The percentage of all adults in the United States, aged 18 years and older, who met the Centers for Disease Control and Prevention’s Physical Activity Guidelines for Americans for aerobic PA in 2020 was 22.7% [12]. Percentages by gender of all age groups who met both aerobic activity and muscle strengthening guidelines in 2020 show males at 28.3% and females at 20.4%. The prevalence of PA among females declines with age from 28.7% for ages 18-34 years to 22.7% for ages 35-49 years [12]. Prevalence continues to further decline to 17.6% for ages 50-64 years ending with 10.8% for those aged 65 years and older. The decline in males moves from 41.3% (ages 18-34 years) to 15.3% (65 years and older) [12]. Men are more likely than women to meet both PA guidelines across all age groups [12].

Walking is one of the most effective interventions for reducing rates of chronic disease as well as one of the best, low-cost, easily implemented, widely accessible, moderate-intensity forms of PA [10,13-16]. Furthermore, the low risk of injury can allow individuals to remain active in older age [10]. The effectiveness of walking programs relies on initiatives aligning with women’s key barriers and motivators [7]. Previous research on walking levels in midlife women shows a lack of time as the primary barrier [7,10,14]. This combined with professional obligations and family care responsibilities relegates walking to a low daily priority for many women [7,17]. Health issues, poor motivation, and absence of social support are other barriers noted [7]. Women are more likely than men to a cite lack of social support and other social influences (eg, embarrassment due to being overweight) as obstacles to PA engagement [18,19]. Furthermore, the effectiveness of the role of social support for PA may be influenced by cultural norms, which should be considered when creating groups and matching walking partners [18]. Environmental factors, such as poor weather, lack of walking paths, and safety concerns [10], can inhibit daily walking as well [7,19].

Motivators of daily walking for midlife women include the associated health benefits, greater well-being, reduced stress, enjoyment, social support, and accountability to others [20]. Walking for transport was an important facilitator for some. Midlife women are motivated by immediate enjoyment versus long-term benefits [7]. The social aspect of making friends was a primary motivation for participating in health walks [7,21]. A study involving low-income urban mothers identified social connection as their most powerful facilitator alongside “me time” and the opportunity to gain a brief respite from their responsibilities [22].

Encouraging inactive populations, particularly women, to increase walking is an important public health consideration and remains a challenge [23]. The first step to increasing walking is to accurately measure it [24]. The use of technology, including wearable fitness trackers and smartphones, shows a great deal of promise for measuring and encouraging walking among women [25]. Results of a 2015 feasibility study by Arigo [17] showed that a large proportion of midlife women had purchased or intended to purchase a wearable tracking device for personal use after returning the program device used in the study (16/20, 80%). This continued interest in tracking highlights the potential for longer-term behavior change, particularly with novice users, with commercially available wearable technology [18].

Increasing health issues among youths due to a sedentary lifestyle as well as the growing demand for fitness apps for women are key factors propelling the market growth into a multibillion-dollar business [26]. Fitness apps are designed to motivate and persuade behavior change to help their users achieve health and wellness goals. While the industry of fitness apps for women is rapidly evolving, there is a lack of research on the impact of gender-centered design on users’ adoption, usage, perceptions, retention, and outcomes. This study explores the relationship between age and women’s self-reported motivations, barriers, attitudes, and beliefs toward daily walking. It further assesses attitudes toward features of a mobile app designed to sync with a wearable step tracker to increase and maintain levels of daily walking among women.

**Methods**

**Research Design and Participant Recruitment**

Our study design involved a web-based anonymous survey hosted on SurveyMonkey only in English. The inclusion criteria were to be female as sex assigned at birth, or intersex identifying as female, and to be 21-75 years of age living in the United States. Exclusion criteria were to be male as sex assigned at birth and to be younger than 21 years or older than 75 years of age. Proof of gender and age was not required. Recruitment for the survey was carried out using a URL link or QR code, which was distributed on the following digital platforms: email listservs, Facebook, LinkedIn, and Reddit. The survey remained open for approximately 4 months, from February to June 2022, until 400 participants consented to take it.

**Ethical Considerations**

The “Research on Daily Walking Habits Among Women Ages 21 to 75” was approved by The George Washington University Institutional Review Board on February 3, 2021 (NCR224026). All participants were recruited via digital outreach. Privacy and confidentiality protections included anonymous and deidentified data collection. All questions on the survey were broad enough
to avoid any chance of identification of individuals. To ensure informed consent, immediately upon opening the survey, as the first question, participants were greeted and given information about the research to decide whether to complete it or not. The paragraph concluded with the question, “Would you like to continue?” Answering yes opened the survey. Participants could skip questions or opt out of the survey at any time. There was no time limit to complete it. Completion rate by the 400 respondents was 100%. Average completion time was 5 minutes. There was no compensation, and no physical risks were associated with the survey.

**Data Collection**

The survey was comprised of 31 questions (Multimedia Appendix 1). The first question was focused on consent to participate, and the next 5 questions were designed to capture sociodemographic information including sex assigned at birth, age, current relationship status, race, Hispanic, Latinx, or Spanish origin, and employment or student status. Questions 7 to 13 covered dog ownership, employment status, chronic disease diagnosis, advice from a doctor on walking, and the use of mobile apps and devices to track steps. Question 14 asked participants if their walking rates changed since the start of the COVID-19 pandemic. Question 15 asked approximately how many steps a participant walks each day. Questions 16 to 21 were designed to glean information on possible barriers and motivators for daily walking. Questions 22 to 30 were aimed at capturing data on the acceptability of features of a mobile app designed for women to increase daily walking. The last question, question 31, asked participants if they would be willing to use a walking app designed for women and to wear an activity-tracking device, such as a Fitbit, to increase daily walking.

**Data Analysis**

When the survey results were downloaded, a codebook was created and uploaded to SPSS software (version 28.0.0.0-142; IBM Corp). We analyzed our independent variable, ages of women, and two dependent variables: (1) barriers and motivators toward daily walking and (2) attitudes toward features of a mobile app to increase walking. Data analysis included descriptive statistics to describe respondents’ demographics (eg, age, relationship status, employment status, and diagnosis of chronic illness). All data were analyzed with stratification by age. The first type of stratification was based on dividing the population into 2 groups of ages: 21-49 years (21- to 49-year group) and 50-75 years (50- to 75-year group), respectively. The second type of stratification was based on dividing the population into groups with 10-year age increments. Bivariate analyses were conducted through SPSS for each survey question using chi-square for dichotomous variables and 1-tailed t tests for scales and continuous variables to identify significant differences between the 2 groups. One-tailed t tests were run for scaled variables to identify significant differences divided by decades.

**Results**

**Descriptive Statistics of Demographics**

Sociodemographic questions captured in the survey and displayed in Table 1 included information on sex assigned at birth, age, relationship status, race and ethnicity, employment or student status, chronic disease status, doctor’s advice regarding walking, and having a dog that you walk daily. Results showed that all 400 respondents self-identified as female for sex at birth. Most women were in their 20s (n=123, 29.6%), followed by 30s (n=101, 25.4%), 50s (n=78, 19.3%), then 40s (n=52, 13%), 60s (n=32, 8%), and 70s (n=14, 3.5%).

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Full sample (N=400), n (%)</th>
<th>Respondents aged 21-49 years (n=276, 69%), n (%)</th>
<th>Respondents aged 50-75 years (n=124, 31%), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relationship status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>142 (35.5)</td>
<td>123 (44.6)</td>
<td>19 (15.3)</td>
</tr>
<tr>
<td>Married</td>
<td>196 (49)</td>
<td>104 (37.7)</td>
<td>93 (75.2)</td>
</tr>
<tr>
<td>Living with partner</td>
<td>46 (11.5)</td>
<td>43 (15.6)</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Once married</td>
<td>16 (4)</td>
<td>6 (2.2)</td>
<td>10 (8.1)</td>
</tr>
<tr>
<td>Walks a dog</td>
<td>111 (27.7)</td>
<td>64 (23.3)</td>
<td>46 (37.4)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>51 (12.8)</td>
<td>16 (5.8)</td>
<td>35 (28.2)</td>
</tr>
<tr>
<td>Part-time employment</td>
<td>42 (10.5)</td>
<td>16 (5.8)</td>
<td>26 (21)</td>
</tr>
<tr>
<td>Full-time employment</td>
<td>185 (46.3)</td>
<td>128 (46.3)</td>
<td>60 (48.4)</td>
</tr>
<tr>
<td><strong>Student status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time student</td>
<td>80 (20)</td>
<td>79 (28.6)</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Part-time student</td>
<td>42 (10.5)</td>
<td>40 (14.5)</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Diagnosed with a chronic illness</td>
<td>93 (23.3)</td>
<td>56 (20.3)</td>
<td>37 (29.8)</td>
</tr>
</tbody>
</table>
In the 21- to 49-year group, most women were single, while in the 50- to 75-year group, most were married. The majority of the 400 women were White (n=297, 74.3%), followed by Black (n=54, 13.5%), and small percentages were Asian, American Indian or Alaska Native, and others. Only 7.5% (n=30) of the women identified as Hispanic, Latinx, or Spanish origin. In both groups, the majority of women were full-time employed. The 50- to 75-year group had significantly more part-time employment and significantly less students. Women who self-reported chronic disease were 20% (55/276) in the 21- to 49-year group and 29% (36/124) in the 50- to 75-year group. More women owned a dog that they walked daily in the 50- to 75-year group. Almost no women were told by their doctor not to walk 2.3% (n=9), and 39.9% (n=159) of all women were advised by their doctor to increase their daily walking.

Women’s Current Walking Behaviors

Results of current walking behaviors show that approximately the same percentage of women in both groups use a step-tracking device (n=144, 52% in the 21- to 49-year group vs n=57, 45.9% in the 50- to 75-year group) and a mobile app to track steps (n=90, 32.7% and n=39, 31.7%). The number of approximate daily steps self-reported by women showed that the 21- to 49-year group had significantly higher percentages for less than 2000 steps (n=28, 10.1% vs n=4, 3.2%; \( P = .03 \)). Other differences in step counts were not as significant between the 2 groups: 2000 to 3000 steps (n=42, 15.2% vs n=16, 13%), 3001 to 5000 steps (n=72, 26% vs n=29, 23.2%), and 5001 to 8000 steps (n=80, 28.9% vs n=28, 22.7%). The 50- to 75-year group had significantly higher percentages in the 8001 steps or more (n=83, 30% vs n=22, 17.4%; \( P = .001 \)) category.

This survey was conducted from February to June 2022 during the COVID-19 pandemic. Women were asked one question on the survey about COVID-19 to give some context to possible changes in their daily walking habits during the time of the pandemic. Their responses showed that a significantly higher percentage of women in the 21- to 49-year group reportedly reduced their daily walking than the 50- to 75-year group (n=77, 27.9% vs n=15, 12.1%; \( P = .001 \)), while a higher percentage of women in the 50- to 75-year group maintained their amount of daily walking (n=40, 32.2% vs n=55, 19.9%; \( P = .001 \)) or increased it (n=67, 54% vs n=142, 51.4%) during the time period of the survey. Specific groups of women, delineated by age decade, who increased daily step counts during COVID-19 showed that women in their 70s (n=13) self-reported the highest change in walking, and those in their 30s (n=96) self-reported the lowest. Figure 1 illustrates these findings.

Figure 1. Survey results of step tracking with a wearable device and mobile app for walking.

Women’s Barriers to Walk Daily

Regarding barriers to daily walking displayed in Table 2, the 21- to 49-year group had significantly higher percentages than the 50- to 75-year group for the following reasons: personal and work responsibilities (including lack of time, lack of child or older adult care, and work or other scheduling barriers), motivational and psychosocial factors (including lack of motivation to exercise in general, lack of motivation to engage in self-care, lack of support from partner or family, lack of support from work environment, not being able to find a walking group or community to support walking, not enjoying walking or would rather do another form of exercise), and physical or environmental factors (including lack of safe spaces to walk, lack of proper clothing, fear of falling or getting injured, doctor advised not to walk, and difficult or uncomfortable to walk). The most cited barriers with significant differences between the 2 groups, with the 21- to 49-year group higher in all categories, were lack of time (n=172, 62.3% vs n=43, 34.6%; \( P < .001 \)), lack of motivation to exercise (n=119, 43.1% vs n=38, 30.6%;...
work or other scheduling (n=115, 41.7% vs n=28, 22.6%; P<.001), lack of safe spaces to walk (n=57, 20.7% vs n=7, 5.7%; P<.001), and lack of motivation to self-care (n=49, 17.8% vs n=12, 9.7%; P=.03). In Table 2, the category “Most cited barriers” illustrates the top 5 barriers to walking by women’s age groups.

Table 2. Survey results of women’s barriers to walking daily according to age.

<table>
<thead>
<tr>
<th>Domains</th>
<th>Full sample (N=400), n (%)</th>
<th>Respondents aged 21-49 years (n=276), n (%)</th>
<th>Respondents aged 50+ years (n=124), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal and work responsibilities&lt;sup&gt;a&lt;/sup&gt;</td>
<td>256 (64)</td>
<td>205 (74.3)</td>
<td>51 (41.1)</td>
</tr>
<tr>
<td>Motivational and psychological factors&lt;sup&gt;b&lt;/sup&gt;</td>
<td>190 (47.5)</td>
<td>145 (52.5)</td>
<td>45 (36.3)</td>
</tr>
<tr>
<td>Physical and environmental factors&lt;sup&gt;b&lt;/sup&gt;</td>
<td>86 (21.5)</td>
<td>69 (25)</td>
<td>17 (13.7)</td>
</tr>
<tr>
<td>None identified&lt;sup&gt;c&lt;/sup&gt;</td>
<td>89 (22.3)</td>
<td>43 (15.6)</td>
<td>46 (37.1)</td>
</tr>
<tr>
<td>Most cited barriers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of time&lt;sup&gt;a&lt;/sup&gt;</td>
<td>215 (53.8)</td>
<td>172 (62.3)</td>
<td>43 (34.7)</td>
</tr>
<tr>
<td>Lack of motivation to exercise&lt;sup&gt;b&lt;/sup&gt;</td>
<td>157 (39.3)</td>
<td>119 (43.1)</td>
<td>38 (30.7)</td>
</tr>
<tr>
<td>Work or scheduling barriers&lt;sup&gt;a&lt;/sup&gt;</td>
<td>143 (35.8)</td>
<td>115 (41.7)</td>
<td>28 (22.6)</td>
</tr>
<tr>
<td>Lack of safe spaces to walk&lt;sup&gt;a&lt;/sup&gt;</td>
<td>64 (16)</td>
<td>57 (20.7)</td>
<td>7 (5.7)</td>
</tr>
<tr>
<td>Lack of motivation to engage in self-care&lt;sup&gt;b&lt;/sup&gt;</td>
<td>61 (15.3)</td>
<td>49 (17.8)</td>
<td>12 (9.7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>P<.001.  
<sup>b</sup>P<.05.

Motivators for Women to Walk Daily

Motivators to walking daily displayed in Table 3 included the following options: physical factors (feels good physically, gives me more energy, reduces my risk for chronic disease or improves my overall health, burns calories or helps with weight management, and builds strength) and mental or emotional factors (reduces depression, reduces anxiety, reduces stress levels, boosts self-esteem, and allows me to work through issues). Both groups had approximately the same percentage of responses in the categories of feels good physically, gives me more energy, and reduces depression. The 21- to 49-year group had significantly higher percentages than the 50- to 75-year group for the motivator category reduces anxiety (n=181, 65.6% vs n=62, 50%; P=.003). While the 50- to 75-year group had significantly higher percentages for walking to burn calories or help with weight management (n=100, 81% vs n=186, 67.4%; P=.004) and reduce risk of chronic disease or improve overall health (n=95, 76.6% vs n=168, 61%; P=.002). The top 5 motivators to walking by women’s age groups were feels good physically, reduces stress levels, helps with weight management, gives me more energy, and reduces risk for chronic illness or improves overall health as illustrated in the category “Most cited motivators” of Table 3.

Table 3. Survey results of motivators for women to walk daily by age groups.

<table>
<thead>
<tr>
<th>Domains</th>
<th>Full sample (N=400), n (%)</th>
<th>Respondents aged 21-49 years (n=276), n (%)</th>
<th>Respondents aged 50-75 years (n=124), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall physical well-being</td>
<td>387 (96.8)</td>
<td>266 (96.4)</td>
<td>121 (97.6)</td>
</tr>
<tr>
<td>Overall mental well-being&lt;sup&gt;a&lt;/sup&gt;</td>
<td>344 (86)</td>
<td>244 (88.4)</td>
<td>100 (80.7)</td>
</tr>
<tr>
<td>Most cited motivators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feels good physically</td>
<td>336 (84)</td>
<td>235 (85.1)</td>
<td>101 (81.5)</td>
</tr>
<tr>
<td>Reduces stress levels</td>
<td>298 (74.5)</td>
<td>213 (77.2)</td>
<td>85 (68.6)</td>
</tr>
<tr>
<td>Helps with weight management&lt;sup&gt;b&lt;/sup&gt;</td>
<td>286 (71.5)</td>
<td>186 (67.4)</td>
<td>100 (80.7)</td>
</tr>
<tr>
<td>Gives me more energy</td>
<td>283 (70.8)</td>
<td>194 (70.3)</td>
<td>89 (71.8)</td>
</tr>
<tr>
<td>Reduces risk for chronic illness or improves overall health&lt;sup&gt;b&lt;/sup&gt;</td>
<td>263 (65.8)</td>
<td>168 (60.9)</td>
<td>95 (76.6)</td>
</tr>
<tr>
<td>Reduces anxiety&lt;sup&gt;b&lt;/sup&gt;</td>
<td>243 (60.8)</td>
<td>181 (65.6)</td>
<td>62 (50)</td>
</tr>
<tr>
<td>Reduces depression</td>
<td>229 (57.3)</td>
<td>164 (59.4)</td>
<td>65 (52.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>P<.05.  
<sup>b</sup>P<.01.
Daily Current Walking Routines and Preferences

Results on current and preferred daily walking routines showed that the 21- to 49-year group had higher percentages than the 50- to 75-year group in the categories of taking child or children on walks (n=39, 14.1% vs n=3, 2.4%; \( P<.001 \)), walking for work or use of public transit (n=92, 33.3% vs n=14, 11.3%; \( P<.001 \)), and walking on campus (n=84, 30.4% vs n=7, 5.7%; \( P<.001 \)). The 50- to 75-year group scored highest in the categories of walking the dog (n=48, 39% vs n=69, 25%; \( P=.005 \)) and none of the above, referring to all of the options as motivators to walk (n=59, 48% vs n=35, 28.2%; \( P<.001 \)). Findings in Table 4 show that more women in the 21- to 49-year group prefer to walk alone than in the 50- to 75-year group (n=193, 70% vs n=62, 50%; \( P<.001 \)), while more women in the 50- to 75-year group prefer to walk with a friend (n=66, 53.2% vs n=105, 38%; \( P<.001 \)), or walk with their partner (n=44, 35.1% vs n=76, 27.4%), or with their dog (n=45, 36.2% vs n=72, 26.1%; \( P=.04 \)).

Table 4. Survey results of daily current walking routines and preferences by age groups.

<table>
<thead>
<tr>
<th>Domains</th>
<th>Full sample (N=400), n (%)</th>
<th>Respondents aged 21-49 years (n=276), n (%)</th>
<th>Respondents aged 50-75 years (n=124), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question: Where does walking fit into your daily routine?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walking the dog( ^{a} )</td>
<td>117 (29.3)</td>
<td>69 (25)</td>
<td>48 (38.7)</td>
</tr>
<tr>
<td>Take child or children on walks( ^{b} )</td>
<td>42 (10.5)</td>
<td>39 (14.1)</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>For work or to use public transit( ^{b} )</td>
<td>106 (26.5)</td>
<td>92 (33.3)</td>
<td>14 (11.3)</td>
</tr>
<tr>
<td>Part of my job</td>
<td>36 (9)</td>
<td>26 (9.4)</td>
<td>10 (8)</td>
</tr>
<tr>
<td>Walk on campus( ^{b} )</td>
<td>91 (22.8)</td>
<td>84 (30.4)</td>
<td>7 (5.7)</td>
</tr>
<tr>
<td>None of the above( ^{b} )</td>
<td>137 (34.3)</td>
<td>78 (28.3)</td>
<td>59 (47.6)</td>
</tr>
<tr>
<td>Preferred walking company</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone( ^{b} )</td>
<td>254 (63.5)</td>
<td>267 (69.6)</td>
<td>62 (50)</td>
</tr>
<tr>
<td>With their partner</td>
<td>131 (32.8)</td>
<td>97 (35.1)</td>
<td>34 (27.4)</td>
</tr>
<tr>
<td>With a friend( ^{a} )</td>
<td>170 (42.5)</td>
<td>104 (37.7)</td>
<td>66 (53.2)</td>
</tr>
<tr>
<td>In a social group</td>
<td>37 (9.3)</td>
<td>28 (10.1)</td>
<td>9 (7.3)</td>
</tr>
<tr>
<td>With their dog( ^{c} )</td>
<td>117 (29.3)</td>
<td>72 (26.1)</td>
<td>45 (36.3)</td>
</tr>
</tbody>
</table>

\( ^{a} P<.01 \).
\( ^{b} P<.001 \).
\( ^{c} P<.05 \).

Women’s Beliefs Around Increasing Walking Behavior

The results of survey questions about beliefs around walking behaviors showed one significant value in the 50- to 75-year group associated with the survey statement: “I would walk more often if people around me walked more often” (scale: 1=strongly disagree to 5=strongly agree). Overall, respondents scored 3.67, while respondents aged 21-49 years scored 3.78, and respondents aged 50-75 years scored 3.41 (\( P<.001 \)). Other survey statements using the same response scale did not show significance: “I would walk more if I had an app with reminders,” all respondents scored 2.83, ages 21-49 years scored 2.84, and ages 50-75 years scored 2.80; and “I believe I can walk up to 8000 steps on most days with proper support,” all respondents scored 4.04, ages 21-49 years scored 4.00, and ages 50-75 years scored 4.12 (Figure 2).
Women’s Attitudes Toward the Importance of Mobile App Features

Results for survey questions concerning the importance of features for a mobile app showed a significantly higher number of positive responses in the 21- to 49-year group for the following features (scale: 1=low, 2=medium, and 3=high): digital community support ($P=.03$), rewards or point system ($P<.001$), and seeing a daily or weekly or monthly chart of progress ($P<.001$). No significant differences were found for tracking of steps, reminders to walk daily, reminders to wear a step-tracking device, daily in-app motivational messaging, and daily in-app educational messaging.

In Table 5, a deeper analysis of the importance of features on a mobile app was examined by age categories delineated by age decades. The results showed that the youngest age group of women, aged 20-29 years, placed the highest importance on the following features: reminders to walk daily, reminders to wear a step-tracking device, digital community for support ($P<.05$), daily in-app motivational messaging, daily in-app educational messaging, in-app reward or point system ($P<.001$), and daily or weekly or monthly progress reports ($P<.001$). The 50- to 75-year group only placed higher importance on step tracking.

Figures 3-5 map out mobile app features grouped into different categories: the importance of step tracking and reminders, importance of community support and daily messaging, importance of in-app rewards, progress charts (daily, monthly, and weekly), and willingness to use an app designed for women and to wear a step-tracking device. Our results show that approximately the same percentage of women in both groups use a step-tracking device (n=64, 52% in the younger group and n=127, 46% in the older group), and the same applies to using a mobile app to track daily steps (n=91, 33% and n=40, 32%). Responses to the survey question, “I would be willing to use a walking app designed for women and to wear an activity-tracking device to increase daily walking,” showed overwhelmingly positive responses among all women: yes (n=240, 60%), no (n=55, 13.8%), and not sure (n=105, 26.3%).
Table 5. Survey results of the importance of women’s perceptions of mobile app features by age decades.

<table>
<thead>
<tr>
<th>Importance levels&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Full sample</th>
<th>20-29 years</th>
<th>30-39 years</th>
<th>40-49 years</th>
<th>50-59 years</th>
<th>60-69 years</th>
<th>70+ years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking steps</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.8</td>
<td>2.1</td>
<td>1.9</td>
<td>2.1</td>
</tr>
<tr>
<td>Reminders to walk daily</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Reminder to wear a step-tracking device</td>
<td>1.7</td>
<td>1.8</td>
<td>1.6</td>
<td>1.5</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Build digital community for support&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.7</td>
<td>1.8</td>
<td>1.6</td>
<td>1.5</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Daily motivational messaging</td>
<td>1.5</td>
<td>1.6</td>
<td>1.4</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Daily educational messaging</td>
<td>1.5</td>
<td>1.6</td>
<td>1.4</td>
<td>1.5</td>
<td>1.4</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Reward or point system&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.6</td>
<td>1.9</td>
<td>1.6</td>
<td>1.4</td>
<td>1.5</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Seeing a daily or weekly or monthly chart of progress&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.3</td>
<td>2.5</td>
<td>2.4</td>
<td>2.0</td>
<td>2.2</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Willingness to use a walking app designed for women and to wear a step-tracking device</td>
<td>2.4</td>
<td>2.4</td>
<td>2.3</td>
<td>2.5</td>
<td>2.2</td>
<td>2.4</td>
<td>2.3</td>
</tr>
</tbody>
</table>

<sup>a</sup>1=low, 2=medium, and 3=high.

<sup>b</sup><i>P</i><.05.

<sup>c</sup><i>P</i><.001.

Figure 3. Survey results of the importance of step tracking and reminders for women by age decades.

![Importance of step tracking and reminders](image-url)
Discussion

Principal Findings
This study explores the relationship between age and women’s self-reported motivations, barriers, attitudes, and beliefs toward daily walking. It further assesses attitudes toward features of a mobile app designed to sync with a wearable step tracker to increase and maintain levels of daily walking among women. Findings indicate that barriers, motivators, and beliefs around daily walking and the importance of preferred features of a mobile app vary according to women’s ages. Therefore, messaging and app features should be tailored to different age groups of women.

Explanations for Women’s Walking Behaviors
Not surprisingly, women in the 21- to 49-year group, comprised primarily of students, women of childbearing age, young mothers, and women with careers, tended to walk more with children, walk to work or use public transit, or walk on campus.
The majority prefer to walk alone, which may be attributed to the need for alone time in their hectic lives, similar to the results of Jones et al’s [21] study in which women wanted “me time” and a respite from their responsibilities. Women in the 50- to 75-year group also preferred to walk alone, but they were more accepting of walking with a friend. One reason might be that more women in the older group have leisure time to walk with friends. In this study, older women were more inclined to walk more if people around them walked more often, which is an important data point to be considered in any walking intervention for older women. These age-focused findings demonstrate the need for more research on gender-tailored fitness apps that satisfy the specific needs of women’s lifestyles at different stages. Catering to women engaged in childcare, for instance, requires communication reminding them to make time for themselves without feeling guilty, while older women might respond well to messaging about walking with a friend to help keep them motivated. There is a dearth of research that systematically examines the acceptability, user experiences, and outcomes of tailored messaging in walking apps for women and the potential impacts they deliver.

Barriers to Women Walking

The main barriers to walking in the 21- to 49-year group included lack of time, lack of motivation, and work scheduling. This is understandable as many women might be advancing their careers, while also juggling families and other pressures. For the 50- to 75-year group, these same barriers existed but were not as significant and became less significant as age increased. Previous studies have shown lack of time as a primary barrier to PA, as well as professional obligations, family care responsibilities, health issues, poor motivation, and absent social support, which are consistent with our results [7,19]. Addressing these barriers has generated meaningful insights in previous research but has not yet led to a large increase in PA that is sustained over time [25].

This study did not find environmental and safety concerns as common barriers, but these may be related to our sample demographic of mostly White women in either their 20s or 50s. Other studies have shown that in the presence of individual characteristics, such as low self-efficacy and functional limitations, the effects of a prohibitive neighborhood environment on walking behavior in older women were further magnified [6]. More work is desperately needed to address the barriers of environmental concerns, particularly among older women in underserved or unsafe neighborhoods.

Motivators for Women to Walk

Both age groups of women in this study were motivated to walk because it makes them feel good physically, gives them more energy, and reduces depression, suggesting the powerful physical and mental health benefits of walking. The younger group was more focused on reducing anxiety, while the older group noted walking for burning calories, weight management, reducing the risk of chronic disease, and improving overall health. However, given our concurrent findings of relatively high rates of self-reported chronic disease in the 21- to 49-year group, messaging on chronic disease prevention and management in a walking app should be equally targeted to both groups but in different contexts with tailored nuances.

Menopause increases the risk of weight gain and chronic diseases, happening simultaneously as women’s PA levels tend to decline [27]. Research by Sydora et al [10] found that women experiencing menopause were not averse to regular exercise, especially those seeking to avoid the perceived increased health risks of hormonal therapy, and that walking is the preferred type of exercise among menopausal women. In this study, walking to reduce stress and anxiety was less important for the older group, which supports the results of the Hedgeman et al [28] study on perceived stress across midlife, which found self-reported stress decreased for most women as they transitioned across midlife.

The social aspect of walking, whether to make friends or for social support, has been well received in previous studies, particularly in midlife women [7]. Interestingly, our results showed that social- and family-based walking was not as popular as walking with a friend or walking alone in both age groups of women. A study by Cho et al [29] offers insights into how walking with a partner might motivate walking due to social support; however, it often results in reduced speed that may unintentionally reduce health benefits, a trade-off that needs more research. A feature to build a digital community of support was more favored in the younger group. This is admittedly contradictory, as our data for the 21- to 49-year group showed a preference to walk alone. Furthermore, while the younger group may not want to walk in person with others or they cannot fit it into their schedules, they placed importance on a digital community of support through a mobile app. A study by Hollander et al [20] found that a mobile app for walking (not limited to women) created a digital connection among walking group members, and participants felt that walking improved their mental health, helped to relieve stress, and made them feel more connected with friends or family members. More research should be aimed at determining the effects of sociodemographic variables of women (eg, income, location, age, race, and ethnicity) and their preferences for a digital or in-person community to support walking.

Acceptance of Technology

The use of smartphones, mobile apps for fitness, and wearable devices to track steps, such as Fitbit, are accepted by women of all ages, and adoption of these devices and exercise apps increases every year [30]. The Pew Research Center concluded that about 1 in 5 Americans use a smartwatch or fitness tracker, with more women than men wearing one (25% vs 18%), and among age groups, more adults aged 18-49 years (25%) wearing one than people aged 50 years and older (17%) [30,31]. In Tong et al’s [32] study, the use of mobile apps and fitness trackers during the COVID-19 pandemic was associated with higher levels of PA in a sample of educated and likely health-conscious individuals among males and females [32]. Our survey responses showed that while an equal percentage of both groups of women wore fitness trackers in general, during COVID-19, higher percentages of women in the younger group reduced their daily walking, as opposed to the older group who maintained or increased steps. More research is needed to ensure walking is
Successful interventions to increase women’s daily walking should ideally combine a wearable tracking device with support from mobile app features, such as goal setting, self-monitoring, positive feedback, and social support [9,24,25]. In 2022, findings from a systematic literature review on assessing the acceptability and effectiveness of mobile-based PA interventions for midlife women during menopause concluded that mobile apps and wearable activity trackers showed a small to moderate increase in moderate to vigorous PA among midlife women [27]. The most acceptable features of mobile apps were manual goal setting and step tracking plus the attractiveness and comfort of wearable activity trackers [27].

In this study, younger women placed importance on app features for community support, rewards or a point system, and progress charts, and they are more inclined to use technology for lifestyle interventions, from gaining and redeeming points for coffee and grocery purchases to using apps to track sleep and menstrual cycles. Women in the 21- to 49-year group placed high importance on reminders to walk daily, reminders to wear a step-tracking device, motivational messaging, and educational messaging. The 50- to 75-year group placed the highest importance on step tracking, though they did not place significant importance on progress charts. This may be a technology-based generational difference as more younger women engage with progress charts for academics, digital banking, and other platforms and therefore are more inclined to accept them and to possess the digital literacy to use them with confidence. Daily motivational and educational messaging was more accepted in the younger group perhaps because mini modules of communication, such as text messages and direct messaging on social media platforms, are widely used.

Insufficient exercise among women of all ages is a global public health issue [33]. Walking should be encouraged as much as possible using evidence-based tools to obtain sustainable results. Integrating fitness into health care on a large scale continues to be a challenge, but this is changing as the US health care system is moving more toward a value-based model of care with a focus on prevention and population health versus one that does not normally inhibit walking, such as diabetes.

Understanding how and why women are motivated to walk will lead to increasingly effective interventions to manage physical and mental health issues at all ages. As the US population aged 65 years and older is projected to nearly double over the next 3 decades, from 48 million to 88 million by 2050 [37], promoting PA among older adults is an important public health, clinical, and economic issue deserving greater attention [27,31,38]. Ideally, future mHealth walking interventions that are uniquely designed for women will combine a wearable tracking device and mobile app with evidence-based behavioral approaches to promote daily walking [39]. This study is among the few focused on women’s walking habits from ages 21 to 75 years old, providing insights into walking behaviors and the motivations and barriers behind them. Our results provide a foundation on which to guide future research and development in this space.

Strength and Limitations

The major strength of this study was the number of respondents (N=400), all of whom completed the survey 100%. This study had several limitations. First, the survey included self-selected participants who might already be walking enough or would be willing to increase their daily steps. Second, survey question 15 asked approximately how many steps a participant walks each day. Limitations to answering this question include the inherent difficulty to recall and estimate accurately the number of steps one walks daily without using a tracker, which may lead to a lack of reliable data. Inflating daily step counts may also reflect an element of social desirability. Third, the uneven number of participants in each group causes concern as well as the lack of diversity in the sample. Fourth, we did not specifically define chronic disease in the survey. Reporting might be different among women with a chronic disease that significantly reduces mobility, such as rheumatoid arthritis, versus one that does not normally inhibit walking, such as diabetes.

Conclusions

Insufficient PA is a leading risk factor for noncommunicable diseases and can also negatively affect mental health and quality of life for women of all ages. Findings indicate that barriers, motivators, and beliefs around daily walking and the importance of preferred features of a mobile app vary according to women’s ages. Messaging and app features should be tailored to different age groups of women. These study results can be viewed as a foundation for future research and development of mHealth interventions to effectively increase daily walking among women of all ages.
Acknowledgments

The research team would like to thank the Milken Institute School of Public Health, The George Washington University for supporting the institutional review board for this study. This research was completely voluntary; no funding was used for any part of this research. The authors would like to thank anyone who participated in their anonymous survey. No artificial intelligence was used in any portion of this paper.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

CJ conceptualized and designed the study and survey, managed recruitment, collected data, and drafted the paper. SC interpreted and analyzed the data, created the tables and figures, and made contributions to the overall paper. AV assisted with the study and survey design and made significant editorial contributions to the paper. MN helped interpret the data and critically reviewed the paper. All authors read and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Web-based survey instrument with 31 questions designed to measure women’s attitudes, barriers, and motivators toward daily walking and the features of a mobile app to increase daily walking.

References


25. Conger SA, Bassett D, Toth L. A boom in fitness trackers isn't leading to a boom in physical activity—men, women, kids and adults in developed countries are all moving less. ProQuest. URL: https://www.proquest.com/docview/2659367163/abstract/B0DA512D27B44EACPO/1 [accessed 2022-06-12]


Abbreviations

mHealth: mobile health
PA: physical activity
Brief Intervention as a Method to Reduce Z-Hypnotic Use by Older Adults: Feasibility Case Series

Maria Torheim Bjelkarøy1,2, BSc, MChiro; Tone Breines Simonsen2, MSc; Tahreem Ghazal Siddiqui1,2, PhD; Sigrid Halset3, MD; Socheat Cheng1,2, PhD; Ramune Grambaite2,4, PhD; Jūratė Šaltytė Benth1,2, Prof Dr; Jennifer Gerwing2, PhD; Espen Saxhaug Kristoffersen5,6, PhD; Christofer Lundqvist1,2,5, Prof Dr

1Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Campus Ahus, Lørenskog, Norway
2Health Services Research Unit Helsetjenesteforskning, Akershus University Hospital, Lørenskog, Norway
3Department of Geriatrics, Akershus University Hospital, Lørenskog, Norway
4Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway
5Department of Neurology, Akershus University Hospital, Lørenskog, Norway
6Department of General Practice, Institute of Health and Society, University of Oslo, Oslo, Norway

Corresponding Author:
Maria Torheim Bjelkarøy, BSc, MChiro
Institute of Clinical Medicine
Faculty of Medicine
University of Oslo, Campus Ahus
Sykehusveien 25
Lørenskog, 1478
Norway
Phone: 47 67960000
Email: matobj@ahus.no

Abstract

Background: Z-hypnotics or z-drugs are commonly prescribed for insomnia and sleep difficulties in older adults. These drugs are associated with adverse events and dependence and are not recommended for long-term use. Despite evidence of older adults being more sensitive to a wide array of adverse events and clinical guidelines advocating limiting use, inappropriate use in this population is still prevalent. Previous intervention studies have focused mainly on prescriber information. Simple, individually focused intervention designs are less studied. Brief intervention (BI) is a simple, easily transferable method mainly used to treat patients at risk of alcohol overuse.

Objective: Our objective was to design and test the feasibility and acceptability of a BI intervention adapted to address individual, inappropriate use of z-hypnotics among older adults. This preparatory study aimed to optimize the intervention in advance of a quantitative randomized controlled trial investigating the treatment effect in a larger population.

Methods: This feasibility case series was conducted at Akershus University Hospital, Norway, in autumn 2021. We included 5 adults aged ≥65 years with long-term (≥4 weeks) use of z-hypnotics and 2 intervening physicians. Additionally, 2 study investigators contributed with process evaluation notes. The BI consists of information on the risk of inappropriate use and individualized advice on how to reduce use. The focus of the intervention is behavioral and aims, in cooperation with the patient and based on shared decision-making, to change patient behavior regarding sleep medication rather than physician-based detoxification and termination of z-hypnotic prescriptions. Qualitative and descriptive quantitative data were collected from intervening physicians, study investigators, and participants at baseline, immediately after the intervention, and at the 6-week follow-up.

Results: Data were obtained from 2 physicians, 2 study investigators, and 5 participants (4 women) with a median age of 84 years. The average time spent on the BI consultation was 15 minutes. All 5 participants completed the intervention without problems. The participants and 2 intervening physicians reported the intervention as acceptable and were satisfied with the delivery of the intervention. After the intervention, 2 participants stopped their use of z-hypnotics completely and participated in the follow-up interview. Study investigators identified logistical challenges regarding location and time requirements. Identified as aspects that may improve the intervention and reduce dropouts included revising the intervention content, focusing on rebound insomnia, adding an information leaflet, and supporting the patient in the period between the intervention and follow-up.

https://formative.jmir.org/2024/1/e51862
notion that the intervention should best be located and conducted by the patient’s own general practitioner was supported by the participants.

**Conclusions:** We identified important aspects to improve the designed intervention and found that the BI is feasible and acceptable for incorporation into a larger randomized trial investigating the treatment effect of BI for reducing z-hypnotic use by older adults.

**Trial Registration:** ClinicalTrials.gov NCT03162081; http://tinyurl.com/rmzx6brn

**KEYWORDS**

prescription medication misuse; older adults; brief intervention; z-drugs; benzodiazepine-related drugs; BZD-related drugs; z-hypnotic; intervention; feasibility; case series; insomnia; sleep; substance overuse; older adult; treatment; reduction; benzodiazepine; hypnotics

**Introduction**

Sleep difficulties and symptoms of insomnia are common, experienced by up to 50% of older adults [1]. Prevalence is higher in the older population than the younger population, and not staying asleep is the most common complaint, followed by initiating sleep and nonrestorative sleep [1-4]. Sleep difficulties in older adults are often associated with comorbid conditions including obstructive sleep apnea, cardiovascular disease, restless legs, nocturia, depression, and neurological conditions, as well as side effects from medication use [3,5].

Z-hypnotics or z-drugs (zolpidem, zopiclone, and zaleplon) are commonly used in the treatment of insomnia and sleep difficulties in older adults. These drugs have been suggested to be milder and safer options than benzodiazepines with a shorter half-life and selective gamma-aminobutyric acid (GABA)–binding qualities [6]. They are now the dominant prescribed sleeping medication in many countries [7,8]. Even so, these drugs are associated with adverse events and dependence [9,10]. A meta-analysis concluded that 13 patients need to be treated for 1 patient to experience improved sleep, while the number needed to harm (any adverse event) is 6 patients [11]. Z-hypnotics are not recommended for more than 2 weeks to 4 weeks of continuous use [12,13], as such use has been associated with reduced sleep quality, reduced cognitive function, and reduced psychomotor skills, including increased risk of falls [11,14]. In our work investigating hospitalized older Norwegian adults, we found a substantial proportion using z-hypnotics on a regular basis with a further considerable share having addictive behavior related to their use [15]. Reducing the use of these drugs by older adults is associated with improved muscle strength, better sleep quality, improved quality of life, and reduced daytime fatigue [16,17]. Despite the risks associated with z-hypnotics, reducing their use seems difficult for patients, with initial rebound insomnia as a common withdrawal symptom [9].

Evidence supports brief interventions (BIs) for reducing z-hypnotic and benzodiazepine use [18]. A BI is a specific structured intervention method that aims to facilitate and encourage a behavioral change in the patient. We developed a BI based on the framework suggested by Babor and Higgins-Biddle [19]. This framework has previously been used as an early intervention for patients at risk of substance abuse and is effective in treating patients at risk of alcohol overuse. Our research team has demonstrated that the BI method is effective for reducing pain medication use in patients with medication-overuse headache [20,21]. Based on this, we suggest that an adapted BI may also be beneficial for reducing the use of z-hypnotics. The BI scheme is a short one-time intervention based on individual behavioral adaptation and shared decision-making. It offers possibilities for the intervening physician to provide individualized support for the participant. The focus is to reduce the use of potentially addictive medications, in this case z-hypnotics, and not as an intervention to specifically treat insomnia. Cognitive behavioral therapy for insomnia is the recommended first-line treatment for insomnia [22] and includes a range of different components including sleep hygiene. With the risk of experiencing rebound insomnia as a side effect of reducing the use of z-hypnotics, the individualized BI scheme is open to provide advice on sleep hygiene, although this is not the main focus of the BI itself. Our aim was to investigate the logistics, feasibility, and acceptability of using the BI with older adults who have inappropriate use of z-hypnotics, in order to optimize the intervention itself in preparation for a later full-scale randomized controlled trial (RCT).

**Methods**

This feasibility study investigated the logistics, feasibility, and acceptability of our BI design for reducing the use of z-hypnotics by older adults. The study was conducted at Akershus University Hospital, Norway, during September 2021 to November 2021.

**Main Outcomes**

The main outcomes for this study were the feasibility and acceptability of the intervention. Acceptability was tested based on indicated parameters [23-26] with an emphasis on the following: logistics including the costs and practicalities, burden and the perceived effort for participation, affective attitude and self-efficacy toward the intervention, and intervention coherence and perceived effectiveness. This was investigated using the following research questions:

- How do the logistics work out from the participating patient point of view?
- How do the logistics and administration work out from an organizing point of view?
- Are the instruments for data collection acceptable?
• Is the BI framework acceptable to perform from a physician point of view?
• Is the BI understandable and acceptable for the patients?
• How do the patients experience the attempt to reduce use of z-hypnotics?

The collected data consisted of qualitative measures and descriptive data collected at baseline and a 6-week follow-up.

Participants
Adults aged 65 years and older with long-term (≥4 weeks) use of z-hypnotics were invited to participate. Prescription of z-hypnotics in Norway is exclusively approved for the purpose of sleep difficulties, with or without fulfilling the diagnostic criteria for insomnia [27,28]. Participants were recruited from eligible z-hypnotic users in an in-hospital observational study of 246 older adults originally conducted in 2017 and 2018 [15] as well as directly through the geriatric department at the hospital. The participation flowchart is presented in Figure 1.

We included 5 participants aged ≥65 years with previously reported inappropriate use of z-hypnotics. Inappropriate use was defined as using z-hypnotics for ≥5 days per week for ≥4 weeks based on clinical guidelines [12,13]. A score ≥5 on the 15-point Severity of Dependence Scale (SDS) indicated a risk of dependence [29]. Participants were not excluded if there was a discrepancy between previously reported inappropriate use and their current self-reported use during the BI intervention. As part of the individualized intervention, self-reported use as well as the SDS score were incorporated in the discussion as basis for the individual plan (see the description of the BI in Figure 2). Exclusion criteria consisted of having a Mini-Mental State Examination (MMSE) score ≤21, a serious visual or hearing impairment, insufficient Norwegian language skills, and the following pre-existing diagnoses: moderate to severe depression, stroke, dementia, or psychiatric disorders.

Figure 1. Study population flowchart. BI: brief intervention; SDS: Severity of Dependence Scale.
Figure 2. Brief intervention (BI) procedure. SDS: Severity of Dependence Scale.

<table>
<thead>
<tr>
<th>Invitation to participate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS score</td>
</tr>
<tr>
<td>- A score of ≥5 indicates risk of inappropriate use of z-hypnotics.</td>
</tr>
<tr>
<td>- With a score &lt;5, the patient may be offered a shortened version of the BI consisting of general information.</td>
</tr>
<tr>
<td>A full individualized BI was indicated at an SDS score above the cutoff</td>
</tr>
<tr>
<td>- Information regarding limiting use beyond 2-4 weeks</td>
</tr>
<tr>
<td>- Relate to length of patient’s own use</td>
</tr>
<tr>
<td>Information regarding use of sleeping medication</td>
</tr>
<tr>
<td>- Information on who is at risk of inappropriate use: older and especially 75-84 years, women more than men, living alone, pain intensity</td>
</tr>
<tr>
<td>- Information regarding combination with other centrally acting medications</td>
</tr>
<tr>
<td>Information regarding possible consequences of inappropriate use</td>
</tr>
<tr>
<td>- Associated with risk of addiction to sleeping medications</td>
</tr>
<tr>
<td>- May have poorer quality of life</td>
</tr>
<tr>
<td>- Associated with reduced cognitive function, which can lead to difficulties with activities of daily living</td>
</tr>
<tr>
<td>- Long-term use may lead to tolerance, reduced effect, and need of increased dosage for effect</td>
</tr>
<tr>
<td>Advantages of reducing medication</td>
</tr>
<tr>
<td>- Avoid reduced cognitive function due to medication use both in long and short-term use</td>
</tr>
<tr>
<td>- Avoid side effects such as constipation, drowsiness, confusion, loss of coordination, and risk of falls</td>
</tr>
<tr>
<td>- Lower risk of addiction and inappropriate use of sleeping medications</td>
</tr>
<tr>
<td>- Higher quality of life</td>
</tr>
<tr>
<td>Challenges</td>
</tr>
<tr>
<td>- Explanation about rebound insomnia</td>
</tr>
<tr>
<td>- Difficulties may last 1-2 weeks</td>
</tr>
<tr>
<td>Invitation to reduce medication use and formulate individual treatment plan including support</td>
</tr>
<tr>
<td>- Individual plan: reduce or stop medication</td>
</tr>
<tr>
<td>- Evaluate sleep at follow-up</td>
</tr>
<tr>
<td>- Support from doctor in case of difficulties</td>
</tr>
<tr>
<td>- Write down goals together with patient</td>
</tr>
</tbody>
</table>

Study Logistics
The intervention consisted of a baseline consultation and follow-up 6 weeks later. The baseline data collection was conducted in a hospital setting either in a consultation room or at the bedside by the study investigators (MTB and TBS). Subsequently, the BI consultation was conducted separately by 1 of 2 participating physicians (CL and SH). After the intervention session, both physicians and participating patients completed a short, written questionnaire with open-text responses regarding their experiences with the BI. The follow-up 6 weeks later involved a home visit by the study investigators (MTB and TBS) and consisted of questionnaires and a qualitative semistructured interview. This study was conducted during the COVID-19 pandemic, and care was taken with regards to safety measures and infection control.

Brief Intervention
The developed BI scheme was based on the BI framework as described by Babor and Higgins-Biddle [19]. Based on studies previously conducted by our research group, we adapted the BI scheme for older adults using z-hypnotics [30,31].

Application of the BI was conducted in 2 steps. The first step consisted of providing training, information, and communication advice for the intervening physicians. The second step consisted of the physicians performing the BI scheme with the participants included in the study.

The BI conversation with the participant consisted of the following (full procedure outlined in Figure 2): identifying the risk of the participant’s dependency on z-hypnotics using the SDS questionnaire (Table 1), informing participants of their risk of inappropriate use and dependence, providing structured information using fact sheets about difficulties associated with
reducing medication and possible withdrawal symptoms such as rebound insomnia [32,33], and adjusting individualized information to the participant’s own experience and inviting the participants to make a decision toward reducing or stopping use and to make a plan on how to proceed. The plan could include strategies on how to handle withdrawal symptoms including rebound insomnia and strategies for contact and support from a physician if needed.

Table 1. Questions in and scoring of the Severity of Dependence Scale (SDS).

<table>
<thead>
<tr>
<th>Question number</th>
<th>Question content</th>
<th>Answer options and corresponding scoresa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Do you think your use of sleeping pills is out of control?</td>
<td>Never/almost never=0, sometimes=1, often=2, always/nearly always=3</td>
</tr>
<tr>
<td>2</td>
<td>Does the prospect of missing a dose make you anxious or worried?</td>
<td>Never/almost never=0, sometimes=1, often=2, always/nearly always=3</td>
</tr>
<tr>
<td>3</td>
<td>Do you worry about your use of sleeping pills?</td>
<td>Never/almost never=0, sometimes=1, often=2, always/nearly always=3</td>
</tr>
<tr>
<td>4</td>
<td>Do you wish you could stop?</td>
<td>Never/almost never=0, sometimes=1, often=2, always/nearly always=3</td>
</tr>
<tr>
<td>5</td>
<td>How difficult do you find it to stop or go without your sleeping pills?</td>
<td>Not difficult=0, quite difficult=1, very difficult=2, impossible=3</td>
</tr>
</tbody>
</table>

Each is scored on a 4-point scale (0-3), and the total maximum score is 15 points. In older adults using z-hypnotics, the cutoff score was ≥5 [26].

Data Collection and Instruments

The main focus was to test the experience with, acceptability of, and logistics of the entire BI including the incorporation of quantitative assessment instruments that would be needed in a full-scale RCT. In this study, however, the demographic information and instruments were only used as descriptive information. The data collected included demographic information; the MMSE [34]; health-related quality of life measured with the EuroQol Group’s EQ-5D-3L [35]; the 6-item De Jong Gierveld Loneliness Scale [36]; the Hospital Anxiety and Depression Scale [37]; visual analogue scale (VAS) [38] scores for intensity of pain, anxiety, and depression; questions on experiences with pain; questions on experiences with sleep difficulties; the Bergen Insomnia Scale (BIS) [39]; the PROMIS-57 Profile v2.1 for sleep quality [40]; the clock test [41]; Cognistat [42]; the SDS [29]; the Cumulative Illness Rating Score-Geriatrics [43]; the Barthel Index for Activity of Daily Living [44]; the single leg balance test [45]; a record of medications and dosage; the degree of expectations and beliefs in the intervention and outcome measured using a VAS at baseline; and a written plan with individual goals for reducing z-hypnotics. A questionnaire on the experience delivering the BI was collected from physicians after each session. Process evaluation notes were obtained from study investigators.

At the 6-week follow-up, we also collected the experiences with the intervention and outcome using a VAS and a medication diary (z-hypnotics) for the past 6 weeks. We also conducted a semi-structured qualitative interview (interview guide in Multimedia Appendix 1) evaluating the participants’ experiences with the BI conversation and their experiences during the 6-week follow-up period. The qualitative interview was tape-recorded and subsequently transcribed verbatim and analyzed.

Analysis

The qualitative data were analyzed by investigators MTB and TBS using the text condensation method [46], and findings are reported as quotations. ELAN computer software (version 6.3; Max Planck Institute for Psycholinguistics) was used to transcribe the interviews. SPSS for Windows (version 26.0; IBM Corp) was used to record quantitative data. Quantitative data are reported descriptively.

Ethical Considerations

This study was approved by the Regional Committee for Medical and Health Research Ethics (REK) (2016/2289) and the Akershus University Hospital data protection officer (PVO; 17-054). All participation was by written informed consent. Collected data were analyzed and stored de-identified as required by the REK and PVO. Data were stored on a protected server approved by the Akershus University Hospital PVO. Participants received no financial nor other compensation for their participation.

Patient and Public Involvement

The Health Services Research Unit User Advisory Board at Akershus University Hospital reviewed and provided advice about the study. The board includes both patient representatives and representatives of the health services as well as other public representatives.

Results

Participants

We recruited 5 older adults (4 women) with a median age of 84 years and ≥4 weeks of z-hypnotic use. We recruited 3 of the 5 participants from eligible participants in our previous hospital-based observational study [15,29,47-50], and 2 participants were recruited directly through the geriatric department. The study population flowchart is presented in Figure 1.

Of the 5 participants, 4 reported using z-hypnotics ≥6 days per week, and 1 participant reported using z-hypnotics 1 day per week at baseline. The median number of days per week using z-hypnotics was 7 days. The median SDS score was 5. The median BIS score was 27 (max score 42), and 3 participants reported that they believed that they would have greater sleep disturbance. The main reasons for sleep disturbance included extensive daytime napping, nocturia, pain, and lying in bed thinking. Demographic and clinical characteristics of the
participants are presented in Table 2. At baseline, 3 participants reported that their current experience with not regularly taking their z-hypnotics resulted in greater sleep disturbance, and 3 patients reported that they believed that they would have greater sleep disturbance if they did not take their sleeping medication (Table 2).

Table 2. Demographic data for the study population and beliefs about sleeping medication at baseline.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participant 1</th>
<th>Participant 2</th>
<th>Participant 3</th>
<th>Participant 4</th>
<th>Participant 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td>Age (years)</td>
<td>81</td>
<td>70</td>
<td>84</td>
<td>87</td>
<td>89</td>
</tr>
<tr>
<td>MMSE score</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>BIS score</td>
<td>22</td>
<td>4</td>
<td>27</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>Insomnia occurrence of insomnia</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>SDS score</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Use of medication (days per week)</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Medication</td>
<td>Zopiclone</td>
<td>Zopiclone</td>
<td>Zolpidem</td>
<td>Zolpidem</td>
<td>Zopiclone</td>
</tr>
<tr>
<td>EQ-5D VAS score</td>
<td>45</td>
<td>80</td>
<td>70</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Cognistat score</td>
<td>70</td>
<td>67</td>
<td>65</td>
<td>66</td>
<td>61</td>
</tr>
<tr>
<td>Main reason for sleep disturbance</td>
<td>Daytime napping 4-5 times/day</td>
<td>Lay in bed thinking</td>
<td>Nocturia 3 times/night</td>
<td>Pain, unsettled, and thinking</td>
<td>Sleep itself and thinking</td>
</tr>
<tr>
<td>CIRS-G score</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td><strong>Current beliefs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep without meds &quot;How have you slept the nights you have not taken sleeping pills?&quot;</td>
<td>Unsure if there is a difference</td>
<td>Worse</td>
<td>Worse</td>
<td>Worse</td>
<td>Do not know, always take it</td>
</tr>
<tr>
<td>Belief about sleep without meds &quot;How do you think you will sleep if you do not take your sleeping pills?&quot;</td>
<td>As good</td>
<td>Worse</td>
<td>Worse</td>
<td>Worse</td>
<td>Not sure</td>
</tr>
</tbody>
</table>

aMMSE: Mini-Mental State Examination.  
bBIS: Bergen Insomnia Scale (range 0-42).  
cOccurrence of insomnia according to the Bergen Insomnia Scale.  
dSDS: Severity of Dependence Scale.  
eVAS: visual analogue scale.  
fCIRS-G: Cumulative Illness Rating Score-Geriatrics.  
g"How have you slept the nights you have not taken sleeping pills?"  
h"How do you think you will sleep if you do not take your sleeping pills?"

**Intervention Logistics, Physicians’ Experience, and Feasibility**

The median duration of the BI consultation was 15 minutes. The 2 intervening physicians reported that the BI framework for a potentially sensitive subject made the doctor-patient consultation easier to conduct (Textbox 1, quotes 1 and 2). They also reported that performing the BI in a consultation room was preferred over a bedside setting. They further reported that all 5 participants were open and positive toward the BI conversation (Textbox 1, quotes 2 and 3), as demonstrated by the participants listening actively and asking questions. The tools for patient communication that were received during training for the BI were beneficial during the consultation. The most beneficial tools for communication included asking open questions, letting the participants talk about their own experiences, having the participant repeat information, and writing down information together with the participant. The study investigators identified that the number of instruments used and the travel to participants were time consuming and advised making adjustments toward limiting time consumption.
Textbox 1. Individual quotations from physicians immediately after the brief intervention (BI) consultation and individual quotations from the participating patients at the 6-week follow-up.

### Quotations from the intervening physicians immediately after the BI consultation:

1. “Overall positive atmosphere. Patient understood the content and stated that effect on cognitive function made an impression. Patient stated little belief in health advantages of changing medication use and stated that, at her age, it did not matter.”

2. “The BI was overall received with a positive attitude. Patient stated that she found quitting difficult as she was living alone and was afraid of not falling asleep.”

3. “Patient was open-minded, interested, and motivated to try to change medication use. Was aware of possibilities of medication misuse in general.”

### Quotations from the participants at the 6-week follow-up:

4. “I stopped the sleeping pills that same night and have not touched them since. [...] I thought, that pill I will manage without, and that has gone very well. And after that, I have realized that I do not need to take them. [...] It does not take me any longer to fall asleep, and when I sleep, I sleep well. [...] It has been a relief not having to remember to take that pill.”

5. “Trying to stop the sleeping medication has been very hard. [...] I have not slept more than half of what I should have. Normally, I am up for the toilet 1 to 2 times a night, but throughout this period, I have been up 3 to 5 times each night, and that leads to poor sleeping. [...] So, therefore, I had decided to tell you that I cannot have it like this in the future. I am so old, maybe I will live 2 more years. It does not matter. I rather have good nights than to become a hundred years.”

6. “It [the BI] was a mild form of advice. I believe in that type of approach. Doctors have tried to scare me about things before, for example, when they wanted me to quit smoking.”

7. “I experience that what they said at the hospital [the BI] and what my general practitioner has said earlier; that information corresponds. That is good for me to hear.”

### Participants’ Experiences and Feasibility at Baseline

All 5 participants attending the BI consultation were positive overall toward participating and interested in the BI conversation. Expectations and beliefs measured by the VAS immediately after the intervention found median expectations of reducing medication use of 35 (min;max 3;100), median beliefs about improved health with reducing z-hypnotic use of 33 (min;max 2;100), median beliefs about the importance of reducing the medication of 27 (min;max 2;100), and beliefs in one’s own ability to reduce z-hypnotic use of 42 (min;max 1;100). Expectations and beliefs are presented in Table 3.

The participants’ intentions and goals on how to proceed after the BI were recorded, and 2 participants decided to quit z-hypnotics completely and immediately. Of the remaining 3 participants, 1 wanted to reduce dosage but continue regular use, 1 wanted to discuss a change to the type of z-hypnotic used with her general practitioner (GP), and 1 aimed to continue use as before and reported that this was about once a month.

The greatest challenge for reducing or stopping z-hypnotic use was reported by 3 participants as a worry that they would not sleep. One stated that she was not sure if changing medication mattered, as she was old and was soon going to die. One participant reported a need for support from the physician to stay motivated to stop z-hypnotic use, and 4 participants did not set up a direct plan for what they would do if they encountered difficulties.
Table 3. Expectations and motivation at baseline versus experiences at follow-up.

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Participant 1</th>
<th>Participant 2</th>
<th>Participant 3</th>
<th>Participant 4</th>
<th>Participant 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td><strong>Baseline expectations and motivation (VAS(^a) score)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectations for reduction(^b)</td>
<td>100</td>
<td>20</td>
<td>3</td>
<td>54</td>
<td>35</td>
</tr>
<tr>
<td>Belief in health improvement(^c)</td>
<td>100</td>
<td>33</td>
<td>2</td>
<td>84</td>
<td>27</td>
</tr>
<tr>
<td>Importance to reduce meds(^d)</td>
<td>100</td>
<td>27</td>
<td>2</td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td>Belief in ability to reduce meds(^e)</td>
<td>100</td>
<td>53</td>
<td>1</td>
<td>42</td>
<td>10</td>
</tr>
<tr>
<td><strong>Experience at the follow-up (VAS score)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied with the experience(^f)</td>
<td>100</td>
<td>—(^g)</td>
<td>—</td>
<td>15</td>
<td>—</td>
</tr>
<tr>
<td>Health improvement(^h)</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>7</td>
<td>—</td>
</tr>
<tr>
<td>Experienced importance(^i)</td>
<td>83</td>
<td>—</td>
<td>—</td>
<td>93</td>
<td>—</td>
</tr>
<tr>
<td>Ability to adjust meds(^j)</td>
<td>100</td>
<td>—</td>
<td>—</td>
<td>89</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)VAS: visual analogue scale.

\(^b\)“What are your expectations for trying to reduce the use of sleeping pills?” (VAS: 0=no expectations, 100=great expectations).

\(^c\)”How much do you believe you will experience a health improvement if you reduce the use of sleeping pills?” (VAS: 0=no belief, 100=great belief).

\(^d\)”How important is it for you to reduce or stop using the sleeping pill after talking to the doctor?” (VAS: 0=not important, 100=very important).

\(^e\)”How sure are you that you can reduce or stop using sleeping pills if you decide to do so?” (VAS: 0=not sure, 100=very sure).

\(^f\)”How satisfied are you with the experience of trying to reduce the use of sleeping pills?” (VAS: 0=not satisfied, 100=very satisfied).

\(^g\)Not answered by this participant.

\(^h\)”Did you experience a health improvement by reducing the use of sleeping pills?” (VAS: 0=no improvement, 100=good improvement).

\(^i\)”How important was it for you to reduce or stop using the sleeping pills after talking to the doctor?” (VAS: 0=not important, 100=very important).

\(^j\)”To what extent were you able to adjust the consumption of sleeping pills in relation to what you decided after consultation with the doctor?” (VAS: 0=no degree, 100=very great degree).

Logistics and Participants’ Experiences With Reducing Z-Hypnotics

At the 6-week follow-up, 2 participants completed the questionnaire and qualitative interview. Both managed to completely stop their use of z-hypnotics. Although 1 participant reported that stopping was easy and did not notice any difference (Textbox 1, quote 4), the other participant reported that trying to reduce medication had been very hard and aimed to start the medication again in the hope of getting more sleep (Textbox 1, quote 5). Of the 3 participants who decided not to participate in the follow-up interview, 1 stated a preference to engage in conversation about these types of medications with her GP, 1 reported having too much going on at home with disease in the family, and 1 participant was lost to follow-up. Experiences at the 6-week follow-up are presented in Table 3.

Participant Feedback During Interviews at the 6-Week Follow-Up

Regarding the practicalities of participating, the 2 participants stated that it had been very easy, with no burden regarding cost or time requirements. They added that there would have been difficulties with participating at the follow-up if it had not been arranged as a home visit, as travelling to the hospital was difficult. Both stated that the BI conversation itself had been a good experience and that the information conveyed had been understandable (Textbox 1, quotes 6 and 7). They had both been aware of the negative side effects of z-hypnotics before the BI conversation and stated that they had not learned anything new in the BI conversation. At the follow-up, they could not recall any specific piece of information that had stood out regarding risks for long-term use, possible adverse events, or advantage for reducing use.

Discussion

Principal Findings

In this study investigating the feasibility of a BI for older adults with inappropriate use of z-hypnotics, we identified important aspects to improve before we proceed with the intervention. The focus for this study was to test the method, logistics, and intervention coherence, identifying strengths and weaknesses to improve the protocol for a larger study investigating the treatment effect. We tested the logistics and found that it is advantageous for the intervention to be delivered in primary care as opposed to secondary care; we need to reduce the number of data collection instruments; the designed BI communication framework was acceptable overall for both physicians and participants; and the participants found it challenging to participate further and reduce their medication, which in turn calls for further support during the intervention period.

Previous research has concluded that long-term use of z-hypnotics by older adults is to be avoided [12,13] and that the
treatment effect of z-hypnotics is questionable [8]. It is suggested that attention should be directed at educational interventions [8], and such interventions have been found beneficial [51-53]. The BI method has been proven effective at handling substance overuse and dependence [19,20]. Adjusting the BI to benefit older adults with inappropriate use of z-hypnotics is probably appropriate and valuable for both individual patients and health care professionals and systems.

Evaluation of Study Logistics

Female gender and being ≥75 years of age are associated with a greater risk of inappropriate use [15], hence we believe we tested our intervention in a relevant population. Recruiting and intervention were linked to a previous observational study [15] and conducted in a hospital setting, which we decided was the most cost-effective and practical solution to test both the study logistics and intervention under the current circumstances. The hospital setting is a good opportunity for revising patients’ medications according to START/STOP criteria [54]. Conducting the intervention in a specialist environment at the secondary health care level was theorized to carry some impact for adherence. We did, however, experience that this type of intervention may be better suited to the primary care setting where it can occur as a conversation between patients and their GPs. Prescriptions for z-hypnotics, although often initiated during a hospital stay, are commonly continued by the GP. Training GPs to perform the BI may increase awareness of prescription habits regarding z-hypnotics. One study participant reported her preference for discussing issues regarding her use of z-hypnotics directly with her GP as the reason for withdrawing from the follow-up interview. Although the participants reported no burden and low perceived effort with participating in the study, they noted not having to travel to the hospital was important for participation. GPs’ proximity to their patients’ locale and their everyday involvement in their patients’ health issues may provide a more effective and suitable setting for this intervention.

The average duration of the BI was 15 minutes, which is longer than anticipated. During the design, we aimed for a 10-minute intervention, as we have previously experienced this duration to be acceptable [20,30]. We believe that limiting time spent on the intervention could increase feasibility in a busy daily practice for both patients and GPs, and it may be advantageous to revise the BI scheme accordingly. However, the increase in expected duration could be manageable for GPs during a conversation with older patients.

Both at baseline and follow-up, the participants completed a series of questionnaires, patient-reported outcomes forms, and tests. As these were in addition to the BI itself, participating in the study was both time-consuming and demanding. In a future study investigating the treatment effect, limiting data collection would be advantageous for both reducing load and possibly increasing adherence at follow-up.

Evaluation of Patient Intervention Coherence

Core to this study was investigating the participants’ own understanding of the intervention and their thoughts and beliefs regarding it. To get an understanding of how the intervention would affect the participants, we also investigated their thoughts and beliefs regarding sleep disturbance and z-hypnotics.

In our study, reported reasons for not sleeping included extensive daytime napping (poor sleep hygiene), nocturia, and pain. Once awake, lying in bed thinking was reported as the most common reason for not sleeping. When adjusting for comorbidities, pain, depression, and medication use, which are all factors associated with sleep disturbance, the prevalence of insomnia, as defined by the International Classification of Sleep Disorders (3rd edition) [27,28,55] criteria, decreases in older adults [3,56]. This calls into question the appropriateness of pharmacological treatment for sleep disturbance in older adults and underlines the importance of identifying all issues regarding sleep disturbance during diagnosis and treatment. Reasons for sleep disturbance may have an impact on the measured effect of an intervention study, indicating a benefit for adjusting for this in the study design.

When reporting on their pre-intervention experience with the use of z-hypnotics, the participants stated that, during nights when they had not taken their z-hypnotics, they would have a worse night. Further, they reported anticipating having a worse night if they did not take their z-hypnotics. Rebound insomnia is a known symptom of z-hypnotic withdrawal [32]. It is most prevalent during the first days of withdrawal, and the patient may interpret this as confirmation that they cannot sleep without their medication. In addition, research has shown that there is a large placebo effect with z-hypnotics. The difference in sleep latency between active treatment and placebo is only 22 minutes of objectively measured effect in favor of active treatment [8], demonstrating a very low clinical gain of active treatment compared with placebo and underlining anticipation as a considerable portion of the experienced effect. Further, only 50% of patients are aware that long-term use of z-hypnotics can reduce sleep quality, and up to 84% of patients are aware of dependence, interaction with alcohol use, and dizziness. Regardless of the awareness of side effects, only 26% of patients report an interest in reducing their use of z-hypnotics [57]. Not being aware of possible side effects and placebo effects in addition to experiencing rebound insomnia with withdrawal are factors that will affect intervention compliance and patients’ interest in making changes to their z-hypnotics.

At the follow-up interviews, the participants reported that they could not recall information shared during the BI consultation regarding use of z-hypnotics, risk of adverse events, and benefits from reducing use. They stated that they did not learn anything new about z-hypnotics during the consultation but that the BI consultation had confirmed some previous knowledge. This underlines the importance of investigating the patients’ knowledge about the substance in question. Awareness about patients’ knowledge is important on an individual level. It emphasizes the importance of revising the content and delivery of the information about z-hypnotics presented in the BI. It also prompts exploring more aspects of the BI technique such as providing a short information leaflet [19] that summarizes and repeats the core information given at the consultation. In this study, we chose not to use such additions, as we wanted to explore the feasibility of using just a single consultation.
Another aspect of the BI technique we chose not to pursue in this study was to contact participants to offer support during the follow-up period. We regarded 6 weeks to be a relatively short follow-up period and therefore did not actively seek contact. As part of the BI, the participants were invited to make a plan about what to do in case they had challenges reducing their medication. They were given a phone number to contact if they had questions, needed support, or encountered difficulties. One participant reported needing support from the physicians during the follow-up, but 4 did not make a plan for what to do if they met with challenges. Prompting patients to create a plan about what to do if they meet with difficulties and contacting the patients during the weeks after the intervention may improve participation and adherence.

Strengths and Weaknesses
This feasibility case series was performed as part of our development of a complex intervention [58], as suggested by the UK Medical Research Council, in which a central component is conducting a feasibility study [58,60]. It provided useful information regarding the feasibility, acceptability, and logistics. Naturally, such a small feasibility series does not contribute quantitative data nor solve the central question about whether the BI is an effective intervention. It was a valuable opportunity for an in-depth investigation of the BI scheme with older adults using z-hypnotics. It also provided information about older adults’ thoughts and beliefs about sleep disturbance and their medication use. The main weakness was the small sample as well as the attrition of participants: 3 of 5 participants declined the follow-up interview. We found, however, that both the participants who completed and those who withdrew provided valuable information to further design and optimize the BI scheme for older adults using z-hypnotics.

Conclusion
This study assessing the feasibility of the BI design for older adults with inappropriate use of z-hypnotics identified some important aspects with regards to improving the design before proceeding to a larger study investigating the treatment effect. Conducting the BI in primary care, limiting the duration of the BI while emphasizing core information, providing patients with an information leaflet, and contacting patients for support may improve the effect of the intervention.

Acknowledgments
We are grateful for the support during data collection from the geriatric department at Akershus University Hospital. We also recognize the extraordinary commitment of the patients who participated in this study.

This work was funded by ELIB (Stiftelsen Et Liv I Bevegelse) and the Health Services Research Unit of the Akershus University Hospital. CL also received funding from the South Eastern Norway Regional health authority. No grant numbers were provided by the funders. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability
The raw data are closed for public access due to data protection regulations by the Akershus University Hospital data protection officer and the Regional Committees for Medical and Health Research Ethics. However, summarized data may be available upon reasonable request to the corresponding author MTB.

Conflicts of Interest
CL has participated on an advisory board and received payment for lectures arranged by AbbVie Pharma AS, Novartis AS, Lundbeck AS, Teva AS, and Roche AS, Norway. He has also received research sponsorship from AbbVie Pharma and Lundbeck AS. The other authors declare no conflict of interest.

Multimedia Appendix 1
Interview guide.

References


https://formative.jmir.org/2024/1/e51862


Abbreviations

BI: brief intervention
BIS: Bergen Insomnia Scale
GABA: gamma-aminobutyric acid
GP: general practitioner
MMSE: Mini-Mental State Examination
PVO: data protection officer
RCT: randomized controlled trial
REK: Regional Committee for Medical and Health Research Ethics
SDS: Severity of Dependence Scale
VAS: visual analogue scale
Investigating the Impact of Prompt Engineering on the Performance of Large Language Models for Standardizing Obstetric Diagnosis Text: Comparative Study

Lei Wang¹, MS; Wenshuai Bi¹, MS; Suling Zhao¹, MS; Yinyao Ma², MS, MD; Longting Lv¹, MS; Chenwei Meng¹, MS; Jingru Fu¹, MS; Hanlin Lv¹, PhD, MD

¹BGI Research, Shenzhen, China
²The People's Hospital of Guangxi Zhuang Autonomous Region, Guangxi, China

Corresponding Author:
Hanlin Lv, PhD, MD
BGI Research
Building 11, Beishan Industrial Zone
Yantian District
Shenzhen, 518083
China
Phone: 86 18707190886
Email: lvhanlin@genomics.cn

Abstract

Background: The accumulation of vast electronic medical records (EMRs) through medical informatization creates significant research value, particularly in obstetrics. Diagnostic standardization across different health care institutions and regions is vital for medical data analysis. Large language models (LLMs) have been extensively used for various medical tasks. Prompt engineering is key to use LLMs effectively.

Objective: This study aims to evaluate and compare the performance of LLMs with various prompt engineering techniques on the task of standardizing obstetric diagnostic terminology using real-world obstetric data.

Methods: The paper describes a 4-step approach used for mapping diagnoses in electronic medical records to the International Classification of Diseases, 10th revision, observation domain. First, similarity measures were used for mapping the diagnoses. Second, candidate mapping terms were collected based on similarity scores above a threshold, to be used as the training data set. For generating optimal mapping terms, we used two LLMs (ChatGLM2 and Qwen-14B-Chat [QWEN]) for zero-shot learning in step 3. Finally, a performance comparison was conducted by using 3 pretrained bidirectional encoder representations from transformers (BERTs), including BERT, whole word masking BERT, and momentum contrastive learning with BERT (MC-BERT), for unsupervised optimal mapping term generation in the fourth step.

Results: LLMs and BERT demonstrated comparable performance at their respective optimal levels. LLMs showed clear advantages in terms of performance and efficiency in unsupervised settings. Interestingly, the performance of the LLMs varied significantly across different prompt engineering setups. For instance, when applying the self-consistency approach in QWEN, the $F_1$-score improved by 5%, with precision increasing by 7.9%, outperforming the zero-shot method. Likewise, ChatGLM2 delivered similar rates of accurately generated responses. During the analysis, the BERT series served as a comparative model with comparable results. Among the 3 models, MC-BERT demonstrated the highest level of performance. However, the differences among the versions of BERT in this study were relatively insignificant.

Conclusions: After applying LLMs to standardize diagnoses and designing 4 different prompts, we compared the results to those generated by the BERT model. Our findings indicate that QWEN prompts largely outperformed the other prompts, with precision comparable to that of the BERT model. These results demonstrate the potential of unsupervised approaches in improving the efficiency of aligning diagnostic terms in daily research and uncovering hidden information values in patient data.

(JMIR Form Res 2024;8:e53216) doi:10.2196/53216

KEYWORDS
obstetric data; similarity embedding; term standardization; large language models; LLMs
Introduction

The advancement of medical informatization has resulted in the accumulation of vast amounts of electronic medical records (EMRs) in hospitals, giving rise to medical big data [1]. These data hold significant research value. Using obstetrics as an example, the implementation of China’s “three-child” policy in 2021 has led to an increasing proportion of women with advanced maternal age and multiparity. Studies indicate that as maternal age and parity increase, the occurrence of pregnancy complications and adverse pregnancy outcomes also tends to rise, posing new challenges for obstetrics across health care institutions at all levels [2]. Extracting valuable information from obstetric EMRs could significantly benefit clinical research aimed at improving pregnancy success rates.

However, due to varying writing habits among doctors, diagnostic descriptions in medical records lack standardization, which hinders the analysis and use of medical data. Consequently, mapping clinical diagnostic descriptions to a standard terminology database is vital for medical data analysis. This process enables the standardization of medical terms across different health care institutions and regions, preventing misunderstandings and confusion caused by varying terminologies. It positively impacts health care quality, reduces medical costs, enhances doctor-patient relationships, and promotes the development of medical science.

The emergence of large language models (LLMs), represented by ChatGPT, has caused a surge in interest in their application across various fields of research. In the medical domain, LLMs have been extensively used for tasks such as intelligent medical history collection and preliminary diagnosis, personalized treatment and drug recommendations, medical record documentation and report generation, literature retrieval and analysis, and medical education and training [3-6]. Kanjee et al [7] assessed ChatGPT’s ability to accurately diagnose challenging medical cases and suggested that generative artificial intelligence (AI) models hold promise as potential aids to human diagnostic cognition. Research by Agbavor and Liang [8] demonstrated that GPT-3-generated text embeddings can reliably distinguish Alzheimer disease patients from healthy controls and infer cognitive test scores of patients, potentially enhancing early dementia diagnosis. Palanca et al [9] explored ChatGPT’s potential applications in psychological counseling, emotional support, and mental illness screening while discussing related challenges and future research directions.

LLMs have also played a crucial role in medical research. Clinical research often involves large amounts of unlabeled natural language data, and LLMs’ zero-shot learning ability allows them to effectively process such data. Agrawal et al [10] showed that ChatGPT excels in extracting zero-shot and few-shot information from clinical texts. Hu et al [11] revealed ChatGPT’s potential in zero-shot clinical entity recognition tasks. Furthermore, Lamichhane's [12] 3 text-based experiments on mental health classification demonstrated ChatGPT’s potential in zero-shot text classification tasks.

LLMs are chatbot technologies based on natural language processing and deep learning; they learn language patterns and knowledge from a large amount of text data to realize natural conversations with humans. The key to effectively using LLMs is to set an optimal prompt [13].

In few-shot learning, designing appropriate prompts can help LLMs learn better from a small number of training samples and improve performance [13]. Even in zero-shot learning scenarios, appropriate prompts can guide LLMs to use contextual information to output correct results [14]. Prompt engineering has been widely used in various fields of natural language processing, such as question answering, text generation, and sentiment classification, as well as other tasks. By carefully designing prompts, LLMs can better understand the task requirements and context and generate more accurate and useful outputs [13,15,16]. In addition, prompt engineering is an efficient method that does not rely on large-scale computing resources. It can narrow the gap between the pretraining and fine-tuning stages, improve the model’s learning ability and generalization ability on a small amount of data, and fully exploit the model’s potential performance [15].

Chain-of-thought (CoT) prompts were proposed by Wei et al [17], who experimented with the effect of CoT prompts on multiple tasks, including mathematical problems, logical reasoning, reading comprehension, and common sense reasoning; they compared it with other prompt engineering techniques and pointed out that CoT prompts could significantly improve the model’s performance on these tasks and even allow the model to show complex reasoning abilities, such as induction, deduction, and analogy. The basic idea of CoT prompts is that, when giving a question or task, instead of directly asking the model to give an answer or result, the user asks the model to give a CoT, that is, a series of intermediate reasoning steps in which each step is a complete sentence, and the last step is the answer or result. The advantage of this is that it can make the model better understand the meaning and goal of the question or task, avoid irrelevant or wrong outputs, and also make it easier for human users to check and evaluate the model’s output.

The goal of self-consistency prompts is to improve the quality and consistency of the generated results by requiring the model to make consistency judgments on the previously generated text [18]. When using self-consistency prompts, the user first provides an initial text as a prompt and then lets the model continue to generate the subsequent text. Next, the user replaces the “greedy decoding” in the CoT prompt with sampling from the language model’s decoder to generate a set of diverse reasoning paths; finally, the user marginalizes the reasoning paths and aggregates them by selecting the most consistent answer in the final answer. This can force the model to maintain self-consistency when generating text, avoiding contradictions and incoherence.

This paper delves into the potential of LLMs for zero-shot or unsupervised learning in the domain of standardizing diagnostic terminology in obstetrics. By leveraging a composite approach that merges different prompt engineering techniques with LLMs, our goal is to identify the most fitting pipeline for unsupervised scenarios.
As most of the LLMs used in the Chinese domain use the Chinese version of the International Classification of Diseases, 10th revision (ICD-10-CN), as their core training corpus [19], in order to compare the performance of LLMs and supervised learning algorithms horizontally on a baseline, we used standard diagnostic terminology in the ICD-10-CN as the alignment target throughout this study.

**Methods**

**Task Overview**

The approach can be divided into 4 steps: (1) mapping the diagnosis in EMRs to the observation domain of the ICD-10-CN via embedded similarity; (2) collecting the candidate mapping terms with similarity above the threshold as the training data set; (3) using 2 LLMs, ChatGLM2 [20] and Qwen-14B-Chat (QWEN) [21], with zero-shot learning to generate the optimal mapping terms; and (4) using 3 pretrained bidirectional encoder representations from transformers (BERTs), BERT [22], whole word masking BERT (BERT-WWM) [23], and momentum contrastive learning with BERT (MC-BERT) [24], for unsupervised generation of the optimal mapping terms for performance comparison. The entire workflow is illustrated in Figure 1.

![Figure 1](image-url)

**Data Preparation**

In this study, the raw data were collected from the obstetric EMR data of the People’s Hospital of Guangxi Zhuang Autonomous Region from April 2014 to April 2022; these data contained only diagnostic reports. Sample data are shown in Textbox 1.

**Textbox 1.** A sample data of diagnoses for ID 720444 is listed below with a translated version. All data processed in this research were in Chinese.

<table>
<thead>
<tr>
<th>Discharge diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 头位顺产</td>
</tr>
<tr>
<td>2. 单胎活产</td>
</tr>
<tr>
<td>3. 孕1产1妊娠39+4周</td>
</tr>
<tr>
<td>4. 羊水偏少</td>
</tr>
</tbody>
</table>

**Translations**

1. Vertex delivery
2. Singleton live birth
3. Pregnancy: G1P1, 39+4 weeks
4. Oligohydramnios

The raw data set underwent data preprocessing by removing punctuation marks and meaningless special symbols to avoid potential interference with subsequent word segmentation operations.

We implemented LLMs in an intranet security environment. Both ChatGLM2 and QWEN were used exclusively on physically isolated graphical processing units, with access facilitated via OpenAI format and FastAPI (built on PyTorch 2.0). Temperature settings for the LLMs were configured at 0, with max_token parameters tailored on a task-by-task basis.

The standard vocabulary referred to in the following text consists of the diagnostic categories belonging to the observation domain of ICD-10-CN.

https://formative.jmir.org/2024/1/e53216
Embedding Learning

We used the conditional random fields (CRF) model [24] to segment the text and obtained original-diagnosis raw data aligned with standard vocabulary terms. The principle of CRF is to treat word segmentation as a character position classification problem. Character position information is often defined as follows: $B$ represents the beginning of a word, $M$ denotes the middle of a word, $E$ signifies the end of a word, and $S$ indicates a single-character word. Feature functions are constructed to describe the relationship between each character and label and the transition between adjacent labels. Using training data, we learn the weights of feature functions to maximize conditional probability. The Viterbi algorithm predicts new input sequences and finds the most probable label sequence; according to the label sequence, we construct word segmentation results from characters between $B$ and $E$ and single characters $S$. As shown in Textbox 2, we conducted CRF word segmentation on diagnoses in EMRs.

Textbox 2. Sample of word segmentation with the conditional random fields model. The data below represent a preliminary diagnosis of placenta abruption.

<table>
<thead>
<tr>
<th>Original word: 初步诊断为胎盘早剥</th>
<th>After CRF annotation: 初/B 步/M 诊/M 断/E 为/S 胎/B 盘/M 早/M 剥/E</th>
</tr>
</thead>
</table>
| According to the label, the word segmentation result is as follows: 初步诊断为胎盘早剥. |}

To calculate the similarity between diagnoses in the raw data set and terms in the standard vocabulary, we used the BERT-medicine model to transform diagnoses and terms into embeddings for storage. The BERT-medicine model is specifically designed to improve the model’s understanding of medical terms and symptoms by introducing a medical domain–specific vocabulary list, lexicon, and pretraining tasks. The main structure of the BERT-medical model is the BERT, and the main inputs are the raw word vectors of each word or phrase in the text. In this study, we used diagnoses in the raw data set as the input text sequences. The BERT model extracted the contextual information of the text through a self-attention mechanism and learned the bidirectional linguistic representations, so as to obtain a semantic representation of each word in its context. The final output embedding vector is represented by the sum of character embedding, partition embedding, and position embedding, which constitute the input sequence.

Similarity Computation

The feature embedding of the diagnosis is denoted by $\text{[diagnosis embedding]}$, the feature embedding of the standard terms is denoted by $\text{[term embedding]}$, and their similarity is calculated using the cosine similarity with the following formula:

$$
\text{similarity} = \cos\left(\theta, \frac{\text{[diagnosis embedding]} \cdot \text{[term embedding]}}{\|\text{[diagnosis embedding]}\| \cdot \|\text{[term embedding]}\|}\right)
$$

The proposed approach is evaluated through the following steps: Standard terms with a similarity score higher than 0.9 are considered candidates for diagnosis keywords and are then verified by medical experts. The normalized precision and recall are calculated, and the precision-recall curve is obtained. Since a particular diagnosis might have multiple similar standard terms, we aimed to identify as many similar terms as possible, and we thus expected high recall and precision. To obtain candidate terms, we collected the original diagnosis and the 10 most similar standard terms having a similarity score greater than or equal to 0.855.

Optimal Term Selection

To comprehensively evaluate the performance of LLMs in the standardization of obstetric diagnostic terminology, we used 4 different prompts, with the prompt design ranging from simple to complex. This started with the prompt trained on zero samples (the zero-shot learning prompt); next were the prompt trained on a small number of samples, the in-context learning prompt, the CoT prompt, and finally the self-consistency prompt. The specific flow chart of LLM training is shown in Figure 2.
The zero-shot learning prompt was meant to guide the LLMs’ output by directly telling them the purpose of this study. The target task of this study was to find the standard-term expression for the diagnosis, that is, to let the LLMs determine the word with the highest similarity. Therefore, we directly told the LLMs to find the most similar word to the input word among the candidate words for the standard term. The LLMs determined the similarity between words based on their own learned knowledge, and then output the word with the highest similarity to the input word as the output result.

The purpose of in-context learning prompts is to give context hints and let LLMs learn by analogy from few shots to output results that more closely meet the requirements [25]. Its input is in the form of {question, answer}, that is, in the input, the question and result are given to the LLM as a template, and it answers the same type of questions in a specific way according to the specific answer.

The input form of CoT prompts is similar to in-context learning prompts, that is, {question, answer}, with the difference that the answer contains the intermediate steps of thinking. In order to reduce human costs, we used LLMs to generate CoT prompts, and then encapsulated them into the prompt inputs.

The key method of self-consistency prompts in this study was to input the CoT prompts from the previous section multiple times, obtain multiple results, randomly sample a group of output results, and use the majority voting method to decide the final result. Next, we will demonstrate the experimental process with different prompts through specific examples, shown in **Figure 3**.
Figure 3. Detailed illustration of the technical intricacies underlying this study. The process of mapping nonstandardized local diagnostic text to standardized International Classification of Diseases, 10th revision, Chinese version (ICD-10-CN) terms involves preliminary similarity-based selection through the vector database, followed by optimal solution selection performed by large language models (LLMs) based on semantic comprehension.

Evaluation

The evaluation metrics in this study to assess the model’s performance were precision, recall, and $F_1$-score [26]. We classified words that matched the original word and the standard word as positives, and those that did not match as negatives. There were 4 possible classification outcomes: true positive, in which the model correctly identified a positive as positive; false negative, in which the model mistakenly classified a positive as negative; true negative, in which the model correctly identified a negative as negative; and false positive, in which the model mistakenly classified a negative as positive. Using these classification outcomes, we could calculate precision, recall, and $F_1$-score to evaluate the model’s performance in standardizing diagnoses.

Precision, recall, and $F_1$-score (the reconciled mean of precision and recall) were defined as follows:

Ethical Considerations

The study was approved by the People’s Hospital of the Guangxi Zhuang Autonomous Region in China (KT-KJT-2021-67), and all pregnancy data were deidentified and anonymized.

Results

Overview

For similarity computation, according to experimental tests, an average precision of 0.88 met the requirement for high precision and recall. The corresponding threshold value at this point was 0.855. Therefore, the threshold value for calculations of similarity was determined to be 0.855, which was used to filter out standard terms that were not similar enough to the diagnosis.

After collecting the candidate data set, we used 2 LLMs and 4 techniques for prompt engineering. Subsequently, we mapped the LLM outputs to the most suitable candidate terms from the ICD-10-CN standard vocabulary, enabling us to calculate precision, recall, and $F_1$-score. In order to undertake entity normalization, we selected the classic BERT series, comprising BERT, MC-BERT, and BERT-WWM, as our comparison models. We then compared their performance with the results obtained using the LLMs with 4 different prompts. The outcomes of this comparison are presented in Table 1.
Table 1. Metric performance comparison across large language model and bidirectional encoder representations from transformers (BERT) series.

<table>
<thead>
<tr>
<th>Model and prompt engineering approach</th>
<th>Precision, %</th>
<th>Recall, %</th>
<th>$F_1$-score, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT(^a)</td>
<td>91.93</td>
<td>91.95</td>
<td>91.94</td>
</tr>
<tr>
<td>Momentum contrastive learning with BERT (MC-BERT)(^a)</td>
<td>92.34</td>
<td>92.37</td>
<td>92.35</td>
</tr>
<tr>
<td>BERT-whole word masking(^a)</td>
<td>92.13</td>
<td>92.17</td>
<td>92.15</td>
</tr>
<tr>
<td>ChatGLM2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero shot</td>
<td>75.02</td>
<td>89.90</td>
<td>81.79</td>
</tr>
<tr>
<td>BERT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In context</td>
<td>85.13</td>
<td>86.60</td>
<td>85.85</td>
</tr>
<tr>
<td>Chain of thought</td>
<td>86.52</td>
<td>88.93</td>
<td>82.51</td>
</tr>
<tr>
<td>Self consistency</td>
<td>88.53</td>
<td>90.11</td>
<td>89.31</td>
</tr>
<tr>
<td>Qwen-14B-Chat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero shot</td>
<td>84.01</td>
<td>86.72</td>
<td>85.53</td>
</tr>
<tr>
<td>In context</td>
<td>88.25</td>
<td>91.18</td>
<td>89.69</td>
</tr>
<tr>
<td>Chain of thought</td>
<td>89.92</td>
<td>91.30</td>
<td>90.60</td>
</tr>
<tr>
<td>Self consistency</td>
<td>90.91</td>
<td>92.13</td>
<td>91.51</td>
</tr>
</tbody>
</table>

\(^a\)Prompt engineering not applicable to these models.

It is evident from the table that the LLMs and BERT displayed comparable performance at their optimal levels, indicating that the LLMs provided a performance and time advantage under unsupervised conditions. Furthermore, the LLMs exhibited varied performance under different prompt engineering setups. Taking QWEN as an example, the implementation of the self-consistency approach improved the $F_1$-score by 5% and precision by 7.9% compared to the zero-shot method. Similarly, the same proportion of correctly generated responses was observed in ChatGLM2’s performance, with a range from 9.19% to 18.02%. Thus, QWEN achieved better performance than ChatGLM2 in all 4 prompt engineering approaches.

The BERT series were additional comparison models and exhibited more comparable results in this task. Among the 3 models shown in Table 2, MC-BERT delivered the best performance. However, in this study, the disparity between the 3 versions of BERT was relatively small.

Table 2. Cluster results of standardized terms. Original words in Chinese translated to English via ChatGPT.

<table>
<thead>
<tr>
<th>ID</th>
<th>Word 0</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Word 3</th>
<th>Word 4</th>
<th>Word 5</th>
<th>Word 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>α-Thalassemia</td>
<td>β-Thalassemia</td>
<td>δ-β-Thalassemia</td>
<td>Intermediate thalassemia</td>
<td>Major thalassemia</td>
<td>Combined thalassemia</td>
<td>Thalassemia</td>
</tr>
<tr>
<td>23</td>
<td>Acute mixed-type fetal distress</td>
<td>Acute fetal distress</td>
<td>Acute fetal heart-type fetal distress</td>
<td>Acute anamniotic fluid-type fetal distress</td>
<td>Chronic fetal distress</td>
<td>Chronic fetal - heart type fetal distress</td>
<td>Chronic anamniotic fluid-type fetal distress</td>
</tr>
<tr>
<td>55</td>
<td>Fetal cardiac malformations</td>
<td>Fetal limb malformations</td>
<td>Fetus with multiple malformations</td>
<td>Fetal ear malformations</td>
<td>Fetal malformations</td>
<td>Fetal structural anomalies</td>
<td>Fetal kidney malformations</td>
</tr>
<tr>
<td>73</td>
<td>Uterine interstitial leiomyoma</td>
<td>Uterine suberosal leiomyoma</td>
<td>Uterine intramural leiomyoma</td>
<td>Uterine submucosal leiomyoma</td>
<td>Uterine mucosal leiomyoma</td>
<td>Uterine leiomyoma</td>
<td>Uterine multiple leiomyoma</td>
</tr>
<tr>
<td>76</td>
<td>Intrahepatic bile duct stones</td>
<td>Hepatobiliary stones</td>
<td>Biliary stones</td>
<td><em>a</em></td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>25</td>
<td>Severe pulmonary arterial hypertension</td>
<td>Mild pulmonary arterial hypertension</td>
<td>Moderate pulmonary arterial hypertension</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>20</td>
<td>Central pelvic stenosis</td>
<td>Pelvic stenosis</td>
<td>Pelvic outlet stenosis</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>36</td>
<td>Acute bronchitis</td>
<td>Acute tracheitis</td>
<td>Chronic bronchitis</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>91</td>
<td>Pregnancy-related reproductive tract infection</td>
<td>Pregnancy-related urinary tract infection</td>
<td>Pregnancy-related urethral infection</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)Not applicable.
Additional Research
In this study, we used the Louvain algorithm to mine terms from the standard data set output by the LLMs and obtained 1100 relatively common diagnostic terms. In the medical field, different medical institutions and professionals may use different terms to describe the same or similar clinical diagnoses, which can cause difficulties and misunderstandings in data exchange, statistics, and analysis. Therefore, standardizing clinical diagnostic terms is an important task. The standardized terms can be used to unify treatment plans and disease statistics, as well as to build clinical diagnostic knowledge bases. The data in our study were clustered into 107 clusters, and each cluster was analyzed separately, resulting in a diagnostic clustering table. Part of the results of the clustering table are shown in Table 2.

Discussion

Principal Results
This paper proposes an effective unsupervised standardization method for obstetric diagnosis. Through a multi-metrics comparison of different LLMs under various prompt engineering strategies, we found that unsupervised LLMs coupled with effective prompt engineering can achieve performance comparable to supervised learning.

A comparison of different prompt engineering strategies showed that although the models’ baseline performance under zero-shot settings varied, they generally showed significant improvement after incorporating strategies such as CoT, which also highlights the importance of effective prompts for LLMs.

The goal of our alignment in this study is the ICD-10-CN terminology, which belongs to the core vocabulary of the Chinese medical field. LLMs trained on Chinese language data usually include it as part of the training corpus [19], and the performance of the baseline model allows prompt engineering to further improve the alignment performance.

Comparison With Prior Work
Compared to previous research that primarily relied on BERT-based methods to map diagnostic descriptions from EMRs to standard terminologies, this study explores a novel approach based on LLMs. Among BERT models, we identified MC-BERT as the top performer, achieving an $F_1$-score of 0.9235.

Beyond the conventional BERT methods, we examined 4 mainstream prompt strategies and found that the self-consistency method outperformed the others, achieving an $F_1$-score of 0.9233. This level of performance matches that of supervised learning, opening up new possibilities for terminology mapping research in the medical domain.

Limitations
As all data were sourced from real-world patient information, and even though we anonymized the data through multiple strategies and only used a portion of the diagnostic text information without any personal identifying information, there is still a risk associated with uploading patient data to an open network. Additionally, as our research objective was to align and standardize Chinese text based on Chinese target terminologies, the choice of LLMs used in this study was limited. The development of LLMs in the Chinese domain is advancing rapidly, and there are many newly released versions that we have yet to explore.

Moreover, our alignment target was for scientific exploration. In future studies, we will attempt to train target vocabulary that is more suited to the scientific research context into the model through methods such as global optimization and exploring semantic alignment scenarios.

Conclusions
This paper investigates the capability of LLMs in standardizing clinical medical terms. By using LLMs to standardize diagnostic terms extracted from real-world obstetric EMRs and designing 4 different prompts for LLMs, we were able to compare their output results with those of the BERT model. Our findings demonstrate that QWEN mostly achieved the best performance and had precision on par with the BERT model, which illustrates that an unsupervised approach improved the efficiency of aligning diagnostic terms in daily research and to uncover the hidden value of patient data information.

Acknowledgments
This study was supported by Guangxi Key Research and Development Program (AB22035056). We thank the China National GeneBank for technical support.

Data Availability
The data sets generated during and/or analyzed during this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


Abbreviations

AI: artificial intelligence
BERT: bidirectional encoder representations from transformers
BERT-WMM: whole word masking bidirectional encoder representations from transformers
CoT: chain-of-thought
CRF: conditional random fields
EMR: electronic medical record
ICD-10-CN: Chinese version of the International Classification of Diseases, 10th revision
LLM: large language model
MC-BERT: momentum contrastive learning with bidirectional encoder representations from transformers
QWEN: Qwen-14B-Chat

Edited by A Mavragani; submitted 29.09.23; peer-reviewed by M Harris, H Chen; comments to author 20.12.23; revised version received 25.12.23; accepted 11.01.24; published 08.02.24.

Please cite as:
Wang L, Bi W, Zhao S, Ma Y, Lv L, Meng C, Fu J, Lv H
Investigating the Impact of Prompt Engineering on the Performance of Large Language Models for Standardizing Obstetric Diagnosis
Text: Comparative Study
JMIR Form Res 2024;8:e53216
URL: https://formative.jmir.org/2024/1/e53216
doi: 10.2196/53216
PMID: 38329787

©Lei Wang, Wenshuai Bi, Suling Zhao, Yinyao Ma, Longting Lv, Chenwei Meng, Jingru Fu, Hanlin Lv. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Efficacy of Digital Outreach Strategies for Collecting Smoking Data: Pragmatic Randomized Trial

Lauren E Kearney1*, MD; Emily Jansen2*, MPH; Hasmeena Kathuria1, MD; Katrina Steiling1, MD, MSc; Kayla C Jones3, MA; Allan Walkey1,3, MD, MSc; Nicholas Cordella2, MD, MSc

1The Pulmonary Center, Boston University, Boston, MA, United States
2Department of Quality and Patient Safety, Boston Medical Center, Boston, MA, United States
3The Evan’s Center for Implementation & Improvement Sciences, Boston University, Boston, MA, United States

*these authors contributed equally

Corresponding Author:
Lauren E Kearney, MD
The Pulmonary Center
Boston University
72 East Concord Street
Boston, MA, 02118
United States
Phone: 1 9788070286
Email: lekearn@bu.edu

Abstract

Background: Tobacco smoking is an important risk factor for disease, but inaccurate smoking history data in the electronic medical record (EMR) limits the reach of lung cancer screening (LCS) and tobacco cessation interventions. Patient-generated health data is a novel approach to documenting smoking history; however, the comparative effectiveness of different approaches is unclear.

Objective: We designed a quality improvement intervention to evaluate the effectiveness of portal questionnaires compared to SMS text message–based surveys, to compare message frames, and to evaluate the completeness of patient-generated smoking histories.

Methods: We randomly assigned patients aged between 50 and 80 years with a history of tobacco use who identified English as a preferred language and have never undergone LCS to receive an EMR portal questionnaire or a text survey. The portal questionnaire used a “helpfulness” message, while the text survey tested frame types informed by behavior economics (“gain,” “loss,” and “helpfulness”) and nudge messaging. The primary outcome was the response rate for each modality and framing type. Completeness and consistency with documented structured smoking data were also evaluated.

Results: Participants were more likely to respond to the text survey (191/1000, 19.1%) compared to the portal questionnaire (35/504, 6.9%). Across all text survey rounds, patients were less responsive to the “helpfulness” frame compared with the “gain” frame (odds ratio [OR] 0.29, 95% CI 0.09-0.91; P<.05) and “loss” frame (OR 0.32, 95% CI 11.8-99.4; P<.05). Compared to the structured data in the EMR, the patient-generated data were significantly more likely to be complete enough to determine LCS eligibility both compared to the portal questionnaire (OR 34.2, 95% CI 3.8-11.1; P<.05) and to the text survey (OR 6.8, 95% CI 3.8-11.1; P<.05).

Conclusions: We found that an approach using patient-generated data is a feasible way to engage patients and collect complete smoking histories. Patients are likely to respond to a text survey using “gain” or “loss” framing to report detailed smoking histories. Optimizing an SMS text message approach to collect medical information has implications for preventative and follow-up clinical care beyond smoking histories, LCS, and smoking cessation therapy.

(JMIR Form Res 2024;8:e50465) doi:10.2196/50465

KEYWORDS
electronic health records; EHR; informatics; learning health system; lung cancer screening; smoking history
Introduction

Tobacco use is an important risk factor for multiple diseases, including lung cancer, and is one of the leading contributors to preventable death in the United States [1]. The collection of nuanced, complete, and accurate tobacco use histories has significant implications for clinical care. For example, determination of lung cancer screening (LCS) eligibility (eligibility criteria: adults aged between 50 and 80 years with a 20 pack-year smoking history and who are either currently smoking or have quit within the past 15 years) [2] requires full documentation of pack-years (calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked) [3]. Clinicians across specialties discuss smoking histories with patients and record them in structured (eg, within a dedicated place in social history) and unstructured data fields (eg, within the clinical note) in the electronic medical record (EMR). Despite this theoretical wealth of longitudinal smoking information, past research illustrates that smoking history data documented in the health record are usually inaccurate, internally inconsistent, incomplete, or outdated [3-8]. Further, while unstructured data may be more accurate, it is limited in its ability to be extracted quickly and easily for clinical care [9,10]. Interventions aimed at improving the documentation and use of patient tobacco histories may have significant implications for interventions that seek to accurately identify patients eligible for LCS, cardiovascular risk reduction, and smoking cessation interventions [4,11,12].

A learning health system (LHS) systematically integrates clinical care, informatics, and research and engages patients to provide opportunities to implement new knowledge rapidly and iteratively [13]. Patient-generated health data (PGHD) is one promising modality to engage patients and build components of a LHS [14]. One challenge, however, is determining the best approach to engaging patients for self-reported data and scaling interventions for use across the health care system. Implementation of initiatives using patient surveys to generate health data are most effective when systematically designed and studied to determine effectiveness and scalability [15,16]. For example, which technology to use and which message framing to use are important considerations to optimize PGHD. Framed messages describe a choice in terms of how participation may provide gain or loss to the individual or helpfulness to the clinician.

We used a PGHD approach to address the issue of poorly documented smoking history, a previously highlighted barrier to uptake of LCS at our institution [7]. Given the clinical relevance of smoking histories to LCS and smoking cessation counseling, the primary objective of this study was to evaluate the impact of patient-generated methods to improve smoking history documentation. To this end, we designed a quality improvement intervention to evaluate three questions about patient-generated smoking history data: (1) “What is the effectiveness of portal questionnaires versus SMS text message–based surveys?” (2) “What is the most effective message framing accompanying the survey link?” and (3), “What is the optimal approach to following up on uncompleted surveys to increase response rates?”

Methods

Setting and Cohort

We conducted this trial at a large academic safety net hospital in the northeast United States [17]. The institutional review board determined this project qualified for an exemption determination as quality improvement research. We carried out our pragmatic trial from October 2022 to January 2023 in the general internal medicine practice, the largest adult primary care clinic at our hospital. Our hospital uses EPIC (Verona), referred to as “EMR” throughout this manuscript.

Participants

A quality analyst generated random patient lists from the EMR for portal questionnaires and text survey cohorts. Our inclusion criteria were a history of tobacco use (current or former), being aged between 50 and 80 years, and having English as a preferred language. We based the patient’s smoking status on their recorded substance use history within the structed social history section of the medical record. In addition to smoking status, this included entry fields for cigarette smoking start date, quit date, cigarette packs per day, cigarette use years, and a pack-year calculated field, which multiplied the packs per day and years. Not all fields were complete for every patient. Because this trial was designed to increase uptake of LCS by gathering an accurate smoking history, we excluded patients who had LCS or an existing LCS order pending since presumably a more accurate smoking status already existed. Additionally, the portal cohort had to have an active portal account, whereas the SMS text message survey cohort needed to have a recorded mobile or home number documented. Finally, we excluded patients from the text survey cohort if they received the portal questionnaire message, so that each cohort was mutually exclusive.

The portal questionnaire cohort consisted of 500 patients, and the text survey cohort consisted of 1000 patients from general internal medicine clinics. The sample size was determined through a judgment sampling approach [18]. Time and resource limitations (our text survey contract was capped at 1000 patients) played a role in the determination of the judgment sample size. The chosen sample size sought to balance meaningful insights and adherence to practical constraints.

Smoking History Query Interventions

We evaluated 2 modalities: an electronic health record portal questionnaire using EPIC’s MyChart (Verona) and a text survey using Patient Navigation Manager CareTour (Philips Healthcare), a texting platform. The EPIC MyChart questionnaire is referred to as the “portal questionnaire,” and the Philips Healthcare texting platform is referred to as the “text survey” throughout this manuscript. For both modalities, questions pertaining to obtaining an accurate smoking history were designed based on a review of the literature and in consultation with pulmonary and critical care specialists with expertise in tobacco dependence treatment and written in plain language to promote readability and interpretability (Table S1 and Figures S1 and S2 in Multimedia Appendix 1 provide information on the survey questions and user interface). We
also used an intentional phased approach for both surveys to assess technical issues, identify remediable issues, and scale more widely. Given that this study used patient-generated smoking history queries, which required participant comprehension and engagement, blinding participants to the study’s purpose was deemed impractical.

**Message Framing**

We reviewed the behavioral economic theory literature to inform our approach and to use message framing that we hoped would best engage patients [15,16,19,20]. We chose to evaluate 3 message framings: “gain,” “loss,” and “helpfulness” (Table 1). Gain- or loss-framed messages have been shown to be effective in smoking cessation, cancer prevention, and vaccination work [21,22]. We also included a helpfulness message, which has been explored within the web-based industry and marketing research but has been underexplored in health care settings [23,24]. We tested these different message frames in the text survey. The EPIC MyChart questionnaire portal system is configured to send users a general email that reads, “You have a portal questionnaire message.” This is a global configuration setting that cannot be modified on a per-project basis. Thus, we did not test different message frames in the portal questionnaire.

**Survey Modalities**

**Electronic Health Record Portal Questionnaire**

The portal questionnaire to assess smoking status consisted of up to 6 questions, which were a combination of multiple-choice or open-ended questions with answers restricted to numeric-only values (Table S1 and Figure S2 in Multimedia Appendix 1). The portal questionnaire design permitted the conditional display of questions tailored to their smoking status. For example, only people who formerly smoked saw the question, “How old were you when you stopped smoking?” Once the message was accessed, a “helpfulness” frame (Table 1) was shown in the portal message’s subject and body (Multimedia Appendix 1 provides full details of surveys).

**Table 1.** Comparison of features tested in portal questionnaire and text survey.

<table>
<thead>
<tr>
<th>Survey modality</th>
<th>Message framing options</th>
<th>Conditional display of messages</th>
<th>Nudge message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portal questionnaire (EPIC’s MyChart; Verona)</td>
<td>• “Helpfulness” frame for all patients: Help your [hospital name] healthcare team give you the best care possible by answering a few questions about your health. (Msg/data rates may apply. Reply STOP to stop msgs)</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
| Text survey (Patient Navigation Manager CareTour texting platform; Philips Healthcare) | • “Gain”: Please answer a few questions for your [hospital name] healthcare team to help you get screenings to keep you healthy. (Msg/data rates may apply. Reply STOP to stop msgs)  
• “Loss”: Please answer a few questions for your [hospital name] healthcare team so you don't miss out on any screenings. (Msg/data rates may apply. Reply STOP to stop msgs)  
• “Helpfulness”: Help your [hospital name] healthcare team give you the best care possible by answering a few questions about your health. (Msg/data rates may apply. Reply STOP to stop msgs) | No                              | Yes            |

We deployed the portal questionnaire in 3 steps to support iterative optimization (Figure 1). The portal notified patients that there was a new questionnaire in their portal through a general email or a pop-up on their smartphone app. Patients were required to have an active portal account and had to log into the account to complete the questionnaire. For both round 1 and round 2, we held our date and time fields constant by sending the message on Tuesdays at 10:20 AM Eastern Daylight Time (EDT).
**Figure 1.** Portal questionnaire and text survey phased rollout and formative modifications. A total of 16 participants were sent the survey both in the pilot and in the subsequent rounds of the portal questionnaire rollout (5 in round 1 and 11 in round 2). These participants were only included for analysis based on their initial participation in the pilot and were excluded from analysis in round 1 and round 2.

### Portal Questionnaire

- **Pilot**
  - N=20
  - Set required fields
  - Refined patient logic to exclude all patient with LCS ordered whether it was completed or not

- **Round 1**
  - N=45
  - Assessed for problems with initial survey responses
  - Proceeded to scale

- **Round 2**
  - N=439
  - Identified opportunity to refine question about current smoking status to include only one selection (current, former, never) and applied it to the text survey

- N=504

### Text Survey

- **Round 1**
  - Index + Nudge
  - N=50
  - Set all questions as required
  - Observed unstructured free text and that we could not set fields to numeric only. Added capitalized words to emphasize question (i.e., STARTED and STOPPED)

- **Round 2**
  - Index + Nudge
  - N=950

- N=1000

---

**SMS Text Message Survey**

The text survey to assess smoking status consisted of 4 questions, which were either multiple-choice or open-ended (Table S1 and Figure S1 in Multimedia Appendix 1). This software would not allow for the restriction of numeric-only values. Conditional display of questions was not possible with the text survey platform, so all patients saw all questions. As a result, we added the leading text “If you stopped” to the question “How old were you when you stopped smoking?” We also evaluated response rates by message framing (Figure 1).

As in the portal questionnaire, we deployed the text survey using a 2-round phased approach to support iterative optimization (Figure 1). We randomly assigned patients to receive 1 of 3
messages: “gain,” “loss,” or “helpfulness” (Table 1). Patients were randomized in Excel (Microsoft Corporation) using the RAND function to assign each participant a random number and the RANK and ROUNDUP functions to evenly distribute participants in each of the 3 message groups. We sent an initial SMS text message on Tuesdays at 10:20 AM EDT, consistent with the portal questionnaire. Participants could either respond, not respond, or unsubscribe. We then sent a nudge or second message to all nonresponders who had not unsubscribed, such that 50% of the nonresponders received the same message as the index message, while 50% received a different message, split evenly among the other 2 framing options (Figure S3 in Multimedia Appendix 1 depicts the message trial schema). We sent the nudge message 2 days after the initial message on Thursday at 10:20 AM EDT.

Statistical Approach

All statistical analyses were performed using R (version 4.1.0; The R Project for Statistical Computing). For analyses including 2 variables (portal vs text survey and same vs different second push messages), Fisher exact test was used. For comparison of framing, we performed a random effects logistic regression model for outcome of response and exposure of survey with random intercept for patients’ to account for repeated measures.

The primary outcome measure was the proportion of surveyed patients who responded based on survey modality (portal vs text survey). As a secondary outcome, we assessed response rates for the text survey based on framing and repeated pushes (first or second). Odds ratios (ORs) were calculated for each of these comparisons.

As an additional secondary outcome, we also compared the data obtained from survey responses to those already existing in the smoking history captured in EPIC. Using Fisher exact test, we compared the number of patients with complete smoking histories, defined as adequate information to determine LCS eligibility (pack-years and time since quitting). We also analyzed the concordance between EPIC data and the data gathered from completed surveys for LCS eligibility, smoking status, and pack-years reported. We used the Cohen κ coefficient to compare LCS eligibility and smoking status. We used a 2-way random effects intraclass correlation coefficient to compare agreement in reported pack-years between the EMR and completed surveys [25]. A level of significance of α=.05 was used.

Results

Survey Response Rates

Overall, the characteristics of responders and nonresponders for both survey modalities were similar, except that responders to the text survey were more likely to identify as White and to have stopped smoking compared with text nonresponders (Table 2).

Table 2. Characteristics of responders and nonresponders to the portal questionnaire and text survey.

<table>
<thead>
<tr>
<th>Race or Ethnicity, n (%)</th>
<th>Portal questionnaire responders (n=35)</th>
<th>Portal questionnaire nonresponders (n=469)</th>
<th>Text survey responders (n=191)</th>
<th>Text survey nonresponders (n=809)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>59.11 (8.6)</td>
<td>59.04 (6.9)</td>
<td>59.31 (6.66)</td>
<td>60.23 (7.77)</td>
</tr>
<tr>
<td>Race or Ethnicity, n (%)</td>
<td>1 (2.9)</td>
<td>7 (1.5)</td>
<td>2 (1.0)</td>
<td>8 (1)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (2.9)</td>
<td>7 (1.5)</td>
<td>2 (1.0)</td>
<td>8 (1)</td>
</tr>
<tr>
<td>Black</td>
<td>17 (48.6)</td>
<td>259 (55.2)</td>
<td>99 (51.8)</td>
<td>530 (65.5)</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>4 (11.4)</td>
<td>50 (10.7)</td>
<td>25 (13.1)</td>
<td>61 (7.5)</td>
</tr>
<tr>
<td>White</td>
<td>12 (34.3)</td>
<td>127 (27.1)</td>
<td>56 (29.3)</td>
<td>177 (21.9)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (2.9)</td>
<td>6 (1.3)</td>
<td>1 (0.5)</td>
<td>11 (1.4)</td>
</tr>
<tr>
<td>Declined</td>
<td>0 (0)</td>
<td>20 (4.3)</td>
<td>8 (4.2)</td>
<td>22 (2.7)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>14 (40)</td>
<td>223 (47.5)</td>
<td>103 (53.9)</td>
<td>471 (58.2)</td>
</tr>
<tr>
<td>Individuals who reported current tobacco use (as recorded in EPIC), n (%)</td>
<td>13 (37.1)</td>
<td>189 (40.2)</td>
<td>69 (36.1)</td>
<td>397 (49.1)</td>
</tr>
</tbody>
</table>

The response rate for the portal questionnaire was 6.9% (35/504) and the response rate for the SMS text message–based survey was 19.1% (191/1000) with an OR of 3.18 (95% CI 2.16-4.79; P<.05) (Figure 2). Across all survey rounds, patients were less responsive to the “helpfulness” message compared with the “gain” message (OR 0.29, 95% CI 0.08-0.99; P<.05) and compared with the “loss” message (OR 0.32, 95% CI 0.09-0.91; P<.05); however, there was no difference in responses between the “gain” and “loss” messages (OR 0.89, 95% CI 0.31-2.55, P=.82) (Figure 2). There was no significant difference in response rates to different message frames when comparing responses from only the first push survey round or responses from only the second push survey round. In reference to the first push message frame, there was also no difference in response rate if the same message framing or a different message framing was used in the second push (OR 1.3, 95% CI 0.74-2.17; P=.44). The overall unsubscribe rate for the text survey was 5.5% (55/1000). There was no significant difference in unsubscribe rates depending on the message framing.
Completed Smoking History Data Analysis
Both the portal questionnaire and text survey were significantly more likely to obtain complete data compared to the available data in the medical chart (Table 3).
Of the responses that did not provide completed data, patients reported “do not remember” (n=0 in the portal and n=3 in the text), submitted nonnumeric answers such as “only tried briefly” (n=0 in the portal and n=7 in the text), or submitted answers that were extreme outliers (n=4 in the portal and n=5 in the text), leading to exclusion. Furthermore, 6 individuals who reported current smoking and responded to the text survey reported gaps in their smoking history. Since we could not discern whether these “gaps” represented periods where patients had stopped smoking versus incomplete data, we instead used current age to calculate total pack-years.

The portal questionnaire generated 24 newly complete smoking histories to determine LCS eligibility for those whose EPIC data were not complete. The number of patients who had complete data for both the existing information in EPIC and the new data obtained from the portal questionnaire was low (7 out of 504). This small sample size limited our statistical analysis of the concordance for LCS eligibility and current smoking status between the medical chart data and the data gathered in the portal questionnaire. However, our raw data demonstrates relative agreement, with 6 of 7 responses in agreement for LCS eligibility (eligible or ineligible) and 7 of 7 responses in agreement for smoking status (current or former). The average number of pack-years recorded in EPIC for this group was 10.82, and the average number of pack-years recorded by the portal questionnaire was 9.27, with a good correlation between the 2 sets of data (intraclass correlation [ICC] 0.81, 95% CI 0.28-0.96). The text survey generated newly complete data to determine LCS eligibility for 89 patients, whose EPIC data were incomplete. Complete data to determine LCS eligibility in both the medical chart and the text survey were available for 87 patients. However, there was poor agreement between the existing data in the medical chart and those collected by the text survey. The 2 sets of data were discordant in identifying whether patients were eligible for LCS (Cohen κ 0.32, 95% CI 0.029-0.62) and in identifying current smoking status (Cohen κ 0.008, 95% CI –0.0077 to 0.024). The average number of pack-years for this group recorded in the medical chart was 14.83, and the average number of pack-years recorded by the text survey was 9.81, with a poor correlation between the 2 data sets (ICC 0.27, 95% CI 0.04-0.48).

**Discussion**

Improving our health systems’ ability to capture accurate and complete smoking histories could have significant implications for the delivery of care, specifically for LCS and smoking cessation counseling. We found that a PGHD approach using patient-generated survey data is a feasible way to engage patients and collect smoking histories. Our trial provides a model for robust, pragmatic evaluation of digital interventions for quality improvement. We were able to test 2 types of survey delivery methods and 3 different survey message framings, all with significant equipoise in the literature. Overall, the portal questionnaire was less effective in generating responses compared to the text survey, and the “helpfulness” framing was less effective in generating responses compared to the “gain” and “loss” framings. A major finding was that both the text survey and the portal questionnaire generated more complete smoking histories to determine LCS eligibility when compared to the existing information available in the EMR.

Previous studies report a wide range of response rates to web-based surveys, with multiple factors contributing to decisions to respond, such as type of information collected, framing, number of reminders, and patients’ health care use [5,26-30]. Our SMS text message survey generated a response rate of 19.1% (191/1000), which is within the range reported in previous studies evaluating web-based surveys and significantly higher than that generated by the portal questionnaire [5,26-30]. The lower response rate to the portal questionnaire may have been influenced by the inability to test framing or deliver nudges messages. However, if we maintain the assumption of an equivalent increase in response rate due to framing (1.8%) and separately due to nudging (6.3%), then we can infer a response rate of 15% (compared to the actual response rate of 35/504, 6.9%). This demonstrates that while framing and a lack of nudges likely had a significant impact, other factors also contributed. One possible explanation is that while the text survey used an interruptive design of direct messaging that could be accessed immediately, the configuration of the portal messaging required access to an app or email and a separate login into the portal, making it less accessible.

Our findings demonstrated improved engagement with “gain” and “loss” framing as opposed to “helpfulness” framing. It is well documented that “gain” and “loss” framing improves patient engagement, attitudes, and motivation [31-36], which, based on our data, likely extends to patient engagement to report smoking history data. However, direct comparisons to “helpfulness” messaging are limited. In fact, the use of “helpfulness” messaging is better documented in nonmedical survey methodology [24]. One potential explanation for the superior performance of “gain” and “loss” framing compared with “helpfulness” framing is that the messages used for the “gain” and “loss” framing center on the implications of responding for the patient, while the “helpfulness” framing centers on the implications of responding for the health care provider. User-centered design has been shown to improve patient engagement [37-39]. Further evaluation of the reasons underlying differential engagement based on message framing, for example, with qualitative analysis, is needed in future studies.

<table>
<thead>
<tr>
<th></th>
<th>EPIC data, n/N (%)</th>
<th>Survey data, n/N (%)</th>
<th>ORa (CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portal questionnaire</td>
<td>93/504 (18%)</td>
<td>31/35 (88%)</td>
<td>34.2 (11.8-99.4)</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Text survey</td>
<td>431/1000 (43.1%)</td>
<td>176/191 (89%)</td>
<td>6.8 (3.8-11.1)</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

aOR: odds ratio.

Table 3. Comparison of smoking history completeness to determine lung cancer screening (LCS) eligibility.

https://formative.jmir.org/2024/1/e50465

XSL FO RenderX
As in this study, where the PGHD generated more complete smoking histories compared to the EMR, other studies have highlighted how dedicated structured data fields are missing key information needed to assess eligibility, such as packs per day, pack-years, and years since quitting [3-6,8,40]. In our health system, some individuals, such as medical assistants and tobacco treatment specialists, may update smoking history in structured fields, but many clinicians are more likely to document smoking history in unstructured notes. As a result, the structured data fields may be more vulnerable to becoming outdated over time and more likely to underreport smoking histories [4]. Survey instruments, in contrast, can be designed to force complete data entry and can be clarified with the patient in the context of a shared decision-making conversation. This could further alleviate completeness issues.

While the data generated from the surveys were more complete than the EMR, it is difficult to ascertain which history is most accurate. Previous research has found substantial agreement between self-reported smoking history comparing a baseline survey and a 1-month follow-up, with a higher likelihood of inconsistent reporting from individuals currently smoking as compared with individuals who have stopped smoking [41]. This reproducibility could be considered a proxy for accuracy. Other studies have treated the history obtained from shared decision-making as the source of truth. Modin et al [6] found a high degree of underreporting in the health record compared to the history obtained in a shared decision-making discussion. It is difficult to know whether the robustness of a nuanced history elicited by a trained clinician could produce a more true result or if other factors, such as time from the last history ascertainment, may influence accuracy. Which of the tobacco use histories is most accurate is particularly salient for the text survey, which demonstrated significant disagreement with the data in the EMR. While the portal questionnaire data suggested closer agreement with the EMR, this is likely an effect of the very small sample size of patients who had complete portal questionnaire data and EMR data. Future studies might use larger sample sizes and repeated measures to better ascertain the connection between completeness and accuracy.

The feasibility of web-based surveys to engage patients and obtain medical data has important implications for a LHS beyond tobacco use history. Consistent, easily accessible structured medical data can be used to target interventions in preventive and follow-up care, for example, the use of web-based symptom checkers to remotely triage patient concerns [42,43]. While innovations such as artificial intelligence and large language models are being proposed as a way to better use unstructured data, these technologies are not yet commercially available or integrated into the EMR and may be costly [44]. Using existing infrastructure, such as portal questionnaires and text surveys, is a low-cost, readily available way to asynchronously engage with patients and gather more structured data for clinical use.

This study was strengthened by testing multiple modalities and multiple message frames to determine the most effective method for engaging patients to self-report tobacco use history. However, this study has some limitations. The use of specific web-based technologies to obtain smoking history data and only English-language data may limit generalizability. Judgment sampling may introduce selection bias, which larger sample sizes would mitigate. However, the cohort characteristics would suggest a diverse cohort that was balanced across intervention groups. Furthermore, obtaining information on smoking history, regardless of modality, can be impacted by social desirability bias, recall bias, and recency bias, all of which may have contributed to the data obtained in this trial [4,45]. We were unable to test message frames in the MyChart Questionnaire portal system due to the unmodifiable global configuration of the initial email message. Also, the CareTour texting platform is designed for appointment reminder functionality. As we were using it for an alternate purpose, we were limited by the lack of conditional display and the inability to restrict data entry on field types, leading to the collection of unusable data. While this modality was able to capture dynamic histories of starting and stopping smoking, it was unclear how to easily translate this into a functionally detailed history. In fact, it remains unclear whether open-ended, “yes” or “no” questions would be most effective in capturing enough of a detailed smoking history to identify patients for interventions such as LCS [46]. Identifying the optimal questions and technology to navigate these limitations should be prioritized in future projects.

This study demonstrates that patients are likely to engage with a text-based survey using “gain” or “loss” framing to report detailed and complete smoking histories. Optimizing web-based surveys to collect tobacco use history and other medical data directly from patients is an appealing approach to improving health care delivery, especially if fully integrated into the EMR, as this could allow health care providers to proactively engage patients in LCS shared decision-making or smoking cessation counseling. Future work should focus on the validation of patient-generated history and the patient experience with receiving and completing a self-reported smoking history survey to allow for further optimization and implementation.

Acknowledgments
The authors would like to thank Corey Dolan, Philips Implementation Manager, for help developing and deploying the text-based survey, David Meter, EpicCare Ambulatory Analyst, for help developing and deploying the portal questionnaire, and all participants who engaged with our surveys.

Conflicts of Interest
None declared.
Multimedia Appendix 1
Images outlining the survey questions as they appeared to participants and trial schema.

[PDF File (Adobe PDF File), 580 KB - formative_v8i1e50465_app1.pdf]

References


Abbreviations

EHR: electronic health record
EMR: electronic medical record
LCS: lung cancer screening
LHS: learning health system
PGHD: patient-generated health data

©Lauren E Kearney, Emily Jansen, Hasmeena Kathuria, Katrina Steiling, Kayla C Jones, Allan Walkey, Nicholas Cordella. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
mHealth App Usability Questionnaire for Stand-Alone mHealth Apps Used by Health Care Providers: Canadian French Translation, Cross-Cultural Adaptation, and Validation (Part 1)

Abstract

Background: An increasing number of health care professionals are using mobile apps. The mHealth App Usability Questionnaire (MAUQ) was designed to evaluate the usability of mobile health apps by patients and providers. However, this questionnaire is not available in French.

Objective: This study aims to translate (from English to Canadian French), cross-culturally adapt, and initiate the validation of the original version of MAUQ for stand-alone mobile health apps used by French-speaking health care providers.

Methods: A cross-cultural research study using a well-established method was conducted to translate MAUQ to Canadian French by certified translators and subsequently review it with a translation committee. It was then back translated to English. The back translations were compared with the original by the members of the committee to reach consensus regarding the prefinal version. A pilot test of the prefinal version was conducted with a sample of 49 potential users and 10 experts for content validation.

Results: The statements are considered clear, with interrater agreement of 99.14% among potential users and 90% among experts. Of 21 statements, 5 (24%) did not exceed the 80% interrater agreement of the experts regarding clarity. Following the revisions, interrater agreement exceeded 80%. The content validity index of the items varied from 0.90 to 1, and the overall content validity index was 0.981. Individual Fleiss multirater κ of each item was between 0.89 and 1, showing excellent agreement and increasing confidence in the questionnaire’s content validity.

Conclusions: This process of translation and cultural adaptation produced a new version of MAUQ that was validated for later use among the Canadian French–speaking population. An upcoming separate study will investigate the psychometric properties of the adapted questionnaire.

(JMIR Form Res 2024;8:e50839) doi:10.2196/50839

KEYWORDS
cross-cultural adaptation; French language; mHealth App Usability Questionnaire; MAUQ; mobile health; mHealth; mobile app; questionnaire translation; usability; validation; health care providers; French translation
Introduction

Background

Mobile health (mHealth) is increasingly used in health care systems, and this fast-growing technology has immeasurable potential to improve the quality and accessibility of health care worldwide [1]. There is no exact figure for the number of mHealth apps available worldwide, as this number is constantly changing owing to the launch of new apps and the removal of existing ones. Globally, the number of mHealth apps for patients and health care providers exceeded 350,000 in 2021 [2]. For care providers, mHealth apps are a fast and effective way to improve communication between patients and interdisciplinary teams. They also enable more accurate data collection at patients’ bedside, facilitate documentation, and increase the availability of care for people living in rural or remote areas [1]. The effectiveness and efficacy of mHealth apps must be guaranteed to optimize the use of this limitless resource, improve user experience, and benefit from the subsequent reduction in health care system costs [3].

Ensuring the usability of mHealth apps is an important step in their development and evaluation. However, literature indicates a lack of evidence regarding the quality of mHealth apps and no legal framework at the national policy level [4,5]. In mHealth, usability refers to the ease and efficiency with which users will use a tool to satisfactorily accomplish a specific task [6]. This includes aspects such as ease of use, operability, clarity of instructions, risk of errors and possibility of correcting them, and user-friendliness of the interface [7].

Currently, several questionnaires are available for evaluating mobile apps. The Mobile Application Rating Scale (MARS) and the user version of MARS for evaluating the quality of mHealth apps in the broadest sense are among the most widely used measures [8,9]. A systematic review, including 87 studies published between 2000 and 2018 [10], highlighted that the usability scales used to evaluate mHealth apps were all initially created to obtain the perspective of developers and researchers. Questionnaires for assessing the usability of mHealth apps by different users have since been created but not yet validated. One of these questionnaires is the multidimensional App Quality Assessment Tool for Health-Related Apps that can be used by experts and users to quickly determine the quality of health-related and mental health-related apps [11]. The mHealth App Usability Questionnaire (MAUQ) is the only questionnaire specifically validated for stand-alone mHealth apps used by health care providers [12]. It is originally available in English, and MAUQ for stand-alone mHealth apps for patients was translated to Malay and validated by a Malaysian research team [13].

The Need for a Canadian French Questionnaire

It is well known that cultural differences can influence how participants respond to questions associated to the measurement tools owing to dissimilarities in language and social and professional norms [14]. Therefore, cultural bias can creep into study results and influence their interpretation [15]. Ensuring the translation, cross-cultural adaptation, and validation of measurement tools beforehand is a recognized process for minimizing this bias and ensuring the validity of study results [16,17].

Within 321 million speakers, French is the fifth most spoken language in the world [18]. As a member state of Francophonie, Canada has a vast territory that is rich in linguistic diversity. Spanning 5514 km between the Pacific and Atlantic oceans, Canada has 2 official languages: English and French. The proportion of Canadians with French as their mother tongue is 20.9% [19], and the number of French-speaking researchers is 63,455 [20]. Although both languages are spoken across the country, French remains as the majority language in the province of Quebec, which accounts for 85.5% of Canadians with French as their mother tongue [19]. Francophone Canadian researchers are also interested in the contribution of mobile technologies to health but have access to very few reliable and valid instruments in French. Clearly, the lack of valid measurement tools in French affects the ability to study this population [20]. This puts francophone health care providers at a disadvantage, as they are often left out of studies available exclusively to anglophone participants. Thus, their experiences are less represented in literature [21].

Currently, there are only few measurement tools available in French such as MARS [22] or Unified Theory of Acceptance and Use of Technology 2 [23,24]. So far, there is no French version of MAUQ. Consequently, there is a necessity for a measurement tool that is translated, cross-culturally adapted, and validated for use with Canadian French health care providers.

This study was the first of 2 phases of a methodological study. The aims of the first phase were the Canadian French translation and cross-cultural adaptation of MAUQ and the initiation of its validation to allow Canadian French health care providers to eventually evaluate the usability of mHealth apps.

Methods

This paper has described the Canadian French translation, cross-cultural adaptation, and validation of the original version of MAUQ. The second step in the assessment of the psychometric properties of the translated version will be described in a later publication.

Instrument

The original version of MAUQ was developed to quantitatively measure the usability of mHealth apps by patients and health care providers regarding ease of use, interface design, user satisfaction, and usefulness, before their launch to the general public [12]. Originally in English, MAUQ was created and validated by health informatics professor Leming Zhou and his colleagues at the University of Pittsburgh [12]. The authors point out that there are no licensing fees for using the questionnaire, and it is not necessary to request permission before using it. The questionnaire is freely accessible on the website [25] and is available in 4 versions, according to app type (interactive or stand-alone) and target population (patients or health care providers). This study was conducted using MAUQ for stand-alone mHealth apps used by health care providers.

https://formative.jmir.org/2024/1/e50839

JMIR Form Res 2024 | vol. 8 | e50839 | p.2466
(page number not for citation purposes)
MAUQ for stand-alone mHealth apps used by health care providers consists of a short guideline for completing the questionnaire, followed by 18 statements and an open question for comments. The statements address 3 domains: ease of use (questions 1-5), interface and satisfaction (questions 6-12), and usefulness of the mobile app (questions 13-18). The statements were developed based on a systematic literature review of 312 unique questionnaire statements from 38 questionnaires. People completing MAUQ are asked to rate their level of agreement on a Likert scale ranging from 1 (disagree) to 7 (agree). App usability is determined by the total average of all scored items for each participant: the higher the average, the better the app’s usability. It is also possible to evaluate the responses to each item to assess a specific component of usability and compare the averages.

The validity study conducted by MAUQ authors used only the 2 patient versions [12]. The authors report that the differences between the patient and health care provider versions are negligible. Initially conducted with 128 participants from the University of Pittsburgh’s academic community, the validity study of MAUQ designed for stand-alone mHealth apps demonstrated strong internal reliability, with an overall Cronbach α value of .914 for the entire questionnaire and .847, .908, and .717 for ease of use, interface and satisfaction, and usefulness, respectively [12].

**Translation, Adaptation, and Validation Processes**

The accepted method of instrument translation and cultural adaptation suggested by Sousa and Rojjanasrirat [17] was retained for this study (Table 1). This 7-step sequential method incorporates the recommendations of the most established methodological approaches in a clear and detailed guideline. Moreover, it aims to provide a symmetrical translation, which is the most recommended because it remains true to the intended meaning and linguistic expression in equal measure between the 2 languages (that of the source instrument and the target instrument) [17,26,27]. Ultimately, the objective of this method was to achieve equivalence between the original and translated versions of the questionnaire. The cross-cultural equivalence is broken down by Flaherty et al [28] into 5 mutually exclusive equivalences of semantic, technical, conceptual, content, and criterion origin (defined in Textbox 1).

<table>
<thead>
<tr>
<th>Steps</th>
<th>Equivalences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Completed in full in this study</strong></td>
<td></td>
</tr>
<tr>
<td>Obtaining authorization to translate the tool; independent, double</td>
<td></td>
</tr>
<tr>
<td>translation from English to French</td>
<td>✓</td>
</tr>
<tr>
<td>Comparison of the 2 translated versions, discussion, and consensus</td>
<td>✓</td>
</tr>
<tr>
<td>regarding a preliminary French version</td>
<td></td>
</tr>
<tr>
<td>Independent double back translation from French to English</td>
<td>✓</td>
</tr>
<tr>
<td>Comparison of the 2 back-translated versions with the original; dis-</td>
<td>✓</td>
</tr>
<tr>
<td>cussion and consensus regarding the prefinal French version</td>
<td>✓</td>
</tr>
<tr>
<td>Pilot study to test the prefinal French version</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Separate studies (upcoming)</strong></td>
<td></td>
</tr>
<tr>
<td>Preliminary psychometric test with a bilingual sample (French-English)</td>
<td>✓</td>
</tr>
<tr>
<td>Complete psychometric test</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Table 1. The 7-step guideline for translation, cross-cultural adaptation, and validation according to Sousa and Rojjanasrirat [17] and the respective equivalences achieved.
The theoretical constructs evaluated and the concepts used are the same in both cultures.

• The content of each questionnaire statement is relevant to each culture.

Semantic
• The meaning of each statement is the same in each culture after translation.

Technical
• The data collection method (in this case, a questionnaire) is comparable in each culture in terms of the data it reports.

Criterion
• The interpretation of the measurement of each variable and of the results is the same when compared between the 2 cultures.

Conceptual
• The theoretical constructs evaluated and the concepts used are the same in both cultures.

Step 1
Step 1 consisted of the independent and anonymous forward translation of the original MAUQ from English to French by 2 professional translators with French as their mother tongue. One of the 2 translators was familiar with digital health terminology, and the other was familiar with the cultural and linguistic nuances of French.

Step 2
In step 2, a comparison of the 2 translated versions with the original was performed by a team comprising the 2 translators, a nurse (JG; member of the research team), and a third-party translator to assess the degree of equivalence of the translation. Ambiguities and differences between words, phrases, grammar, and meanings were discussed in a virtual meeting to reach consensus regarding the first version of the translated MAUQ.

Step 3
Step 3 involved the independent and anonymous back translation of the translated version to English by 2 other certified translators with English as their mother tongue and no previous knowledge about the original MAUQ. They had to consider the French version as the original.

Step 4
In step 4, a committee (n=6) compared these 2 back-translated versions with the original MAUQ version to assess the degree of equivalence of the back translations. This committee, which included all 4 bilingual and bicultural translators who worked in steps 1 and 3, a nurse (JG; member of the research team), and a health care provider (an experienced acute care nurse), was formed to discuss ambiguities and differences between words, phrases, grammar, and meanings. Consensus for each of the statements was established during a virtual meeting, ensuring consistency and clarity of formulation according to the Canadian French language and culture. The prefinal version of the translated and adapted questionnaire was consolidated and named MAUQ en français (MAUQ-FR).

At each of these first 4 steps, 1 of the 2 certified translators was familiar with digital health terminology, ensuring that the constructs of the tool were understood. All the involved individuals were bilingual experts.

Step 5
Step 5 consisted of pilot-testing MAUQ-FR with target users and a panel of unilingual experts. For the target population, Sousa and Rojjanasrirat [17] define participants as people whose language is the target language of the instrument and who should be recruited from the target population in which the instrument will be used. A group of 49 registered nurses with French as their mother tongue completed a 5-minute SurveyMonkey (Symphony Technology Group) questionnaire asking them to rate the clarity of the instructions and each translated MAUQ statement dichotomously (clear or unclear) [17,29]. If they selected unclear, a textbox appeared, so that they could indicate how to rewrite the statement to make it clear. Recruitment with voluntary sampling was conducted among graduate nurses from a Quebec university.

The same approach was used with an expert panel, in addition to rating the relevance of each statement regarding their experience with Canadian health care. To achieve this, a Likert scale ranging from 1 (not relevant) to 4 (very relevant) was used to avoid a neutral position [30,31]. Sousa and Rojjanasrirat [17] indicate that the panel should consist of experts “who are knowledgeable about the content areas of the construct of the instrument and the target population in which the instrument will be used and whose mother language is the target language of the instrument.” A search was conducted across Canada to find experts who are using mobile technology at work and with French as their mother tongue. Following the target number of 6 to 10 experts [30,32], the 10 people who assessed content validity were 2 (20%) professors in nursing, 1 (10%) person in public health who works on the evaluation of information and communication technologies and its specific terminology, 3 (30%) doctoral candidates and professors in nursing, 1 (10%) physician and clinical professor in medicine, 1 (10%) mobile app developer, and 2 (20%) health-related practitioners who use mHealth (1 nurse manager and 1 medical specialist). Experts were recruited from the Canadian provinces of Manitoba, Ontario, Quebec, and New Brunswick through networking...
As recommended by Sousa and Rojjanasrirat [17], this first study only covered steps 1 to 5. Steps 6 and 7 involving the evaluation of the psychometric properties (Cronbach $\alpha$) and the measurement of the internal consistency reliability (Lin concordance correlation coefficient) of MAUQ-FR with a bilingual, French-English sample (target $n=90$) and the target population (target $n=180$) will be conducted in 2 subsequent studies.

### Analyses for the Validation of the Instrument

The quantitative data obtained during the pilot test were extracted directly from the SurveyMonkey website and analyzed using descriptive statistics presented as frequencies and percentages, including intrarater agreement. The minimum intrarater agreement was set at 80% [17]. The research team revised and reevaluated the statements rated as unclear by at least 20% (2/10) of the sample, in addition to considering all feedback obtained from unclear responses to improve MAUQ-FR.

Data collected from the expert panel made it possible to assess content validity with the content validity index (CVI): CVI at item level (I-CVI) and CVI at scale level (S-CVI). Relevance scores were previously dichotomized: scores of 1 and 2 were coded as 0 (not relevant) and scores of 3 and 4 were coded as 1 (relevant) [30]. With 10 experts, the minimum thresholds to reach were at least 0.79 for I-CVI [30] and at least 0.80 for the averaging calculation at S-CVI [32,33]. Considered as the average of the proportion of items deemed relevant across the various judges, S-CVI was calculated by adding I-CVIs and dividing by the number of items [33].

Members of the research team considered and discussed the statements with a relevance score of 1 (not relevant) or 2 (unable to assess relevance). Items that failed to meet the previously indicated I-CVI thresholds were revised and reevaluated by the expert panel. New validity indices were then calculated until acceptable I-CVIs were reached. The modified $\kappa$ coefficient of agreement (Fleiss multirater $\kappa$) was also calculated to determine intrarater agreement among experts [34,35]. A $\kappa$ of 0.60 is considered as the minimum acceptable coefficient to determine good agreement, whereas a value $\geq 0.75$ is considered as excellent [34,36]. All statistical analyses were performed using Microsoft Excel.

### Ethical Considerations

After submission to the research ethics board at University of Ottawa, an approval from the research ethics board will be required only for subsequent stages (psychometric testing), since this study is regarded as a quality improvement study. All participants received the information about the objectives of the study, procedures involved, and confidentiality of the data. Informed consent was obtained from all participants. In accordance with the chosen methodology, the completed questionnaires were entirely anonymous and did not collect sociodemographic data. Authorization to translate MAUQ was obtained in advance from the authors.

### Results

#### Steps 1 to 4: Translation

Steps 1 to 4 helped to achieve conceptual, semantic, and content equivalence. The translated version includes the 3 domains of the original version of MAUQ, which have been similarly broken down into 18 statements. In more detail, the step-2 consensus phase made it possible to work on semantic equivalence, ensuring that there was no change in the meaning of the words used in the original questionnaire. The committee met virtually for 1 hour. As there was hesitation in choosing the right terms, the translators were encouraged to indicate all possible options for certain words to clarify their connotations and jointly make the best decision (eg, user-friendliness vs usability).

The step-3, independent, double back translation clarified the words and sentences used in the translation to determine the accuracy of the translation by identifying the differences between the 2 English versions (semantic equivalence). In a 2-hour virtual meeting, the step-4 committee discussion validated each statement and established conceptual, semantic, and content equivalences. Professor Zhou, author of the original questionnaire [12], was contacted to clarify the intended meaning of the term “social settings” in the ninth statement. Then, 4 statements were modified between the step-2 and step-4 consensuses (Table 2). Following these modifications, the committee unanimously reached consensus that the words and concepts used complied with the language and each cultural perspective.

### Step 5: Pilot Test (Target Population)

For face validation, the French-speaking registered nurses ($n=49$) considered the statements to be clear, with intrarater agreement of 99.14% (Table 3). In total, 5 comments were collected and considered to improve the questionnaire. This pilot test provided additional support for conceptual equivalence.
(clarity) and content equivalence (relevance) in the Canadian French cultural context.

**Table 3.** Interrater agreement on statement clarity among the target population and the expert panel during the pilot test.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Target population (n=49), n (%)</th>
<th>Round-1 experts (n=10), n (%)</th>
<th>Round-2 experts (n=9), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>45 (92)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Directives</td>
<td>49 (100)</td>
<td>10 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 1</td>
<td>48 (98)</td>
<td>10 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 2</td>
<td>49 (100)</td>
<td>10 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 3</td>
<td>46 (94)</td>
<td>8 (80)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Item 4</td>
<td>49 (100)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 5</td>
<td>49 (100)</td>
<td>10 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 6</td>
<td>49 (100)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 7</td>
<td>49 (100)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 8</td>
<td>49 (100)</td>
<td>8 (80)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Item 9</td>
<td>49 (100)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 10</td>
<td>49 (100)</td>
<td>10 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 11</td>
<td>49 (100)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 12</td>
<td>49 (100)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 13</td>
<td>49 (100)</td>
<td>10 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 14</td>
<td>49 (100)</td>
<td>8 (80)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Item 15</td>
<td>49 (100)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 16</td>
<td>48 (98)</td>
<td>9 (90)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 17</td>
<td>49 (100)</td>
<td>10 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td>Item 18</td>
<td>49 (100)</td>
<td>8 (80)</td>
<td>8 (89)</td>
</tr>
<tr>
<td>Conclusion</td>
<td>49 (100)</td>
<td>6 (60)</td>
<td>9 (100)</td>
</tr>
</tbody>
</table>

\( ^a\) Interrater agreement within the target population=99.14.

\( ^b\) Interrater agreement among round-1 experts=90.

\( ^c\) Interrater agreement among round-2 experts=93.

\( ^d\) N/A: not applicable.

**Step 5: Expert Panel**

The experts (n=10) considered the statements to be clear, with interrater agreement of 90% (Table 3). The 5 statements that did not exceed 80% interrater agreement were revised by the research team and reevaluated by the expert panel (9/10, 90%) to achieve content-related validity. Interrater agreement for the modified statements was 93%.

I-CVI for each statement ranged from 0.90 to 1, and S-CVI was 0.981 (Table 4). Individual Fleiss multirater \( \kappa \) for each item ranged from 0.89 to 1, increasing confidence in the questionnaire’s content validity [35]. There were 32 comments, which improved the accuracy of the statements.

Step 5 helped to reinforce the conceptual, semantic, and content equivalence and prepare a translated and adapted version. The sample sizes required were achieved and even exceeded for the pilot test with the target population.

In short, all the items and the title, instructions, and conclusion met the thresholds for psychometric testing. The example in Table 5 illustrates the entire process of steps 1 to 5.
Table 4. Content validity index (CVI) of item relevancy and Fleiss \(\kappa\) agreement by the expert panel during the pilot test.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Rating of 1 or 2 (n=10), n (%)</th>
<th>Rating of 3 or 4 (n=10), n (%)</th>
<th>Item-level CVI(^a)</th>
<th>Fleiss (\kappa)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Directives</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 2</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 3</td>
<td>1 (10)</td>
<td>9 (90)</td>
<td>0.90</td>
<td>0.89</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 4</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 5</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 6</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 7</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 8</td>
<td>1 (10)</td>
<td>9 (90)</td>
<td>0.90</td>
<td>0.89</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 9</td>
<td>1 (10)</td>
<td>9 (90)</td>
<td>0.90</td>
<td>0.89</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 10</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 11</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 12</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 13</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 14</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 15</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 16</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 17</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Item 18</td>
<td>0 (0)</td>
<td>10 (100)</td>
<td>1.00</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
<tr>
<td>Conclusion</td>
<td>1 (10)</td>
<td>9 (90)</td>
<td>0.90</td>
<td>0.89</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

\(^a\)Overall scale CVI=0.981.

Table 5. Translation process: example (item 16).

<table>
<thead>
<tr>
<th>Original English version</th>
<th>Forward translation (English to French)</th>
<th>Consensus</th>
<th>Back translation (French to English)</th>
<th>Consensus</th>
<th>Pilot test</th>
</tr>
</thead>
<tbody>
<tr>
<td>This app has all the functions and capabilities I expected it to have.</td>
<td>• Translator 1: L’application possède toutes les fonctions et capacités auxquelles je m’attendais.</td>
<td>L’application comportait toutes les fonctions et capacités auxquelles je m’attendais.</td>
<td>• Translator 3: The app had all the features and functions I was expecting.</td>
<td>L’application comportait toutes les fonctions et capacités auxquelles je m’attendais.</td>
<td>L’application comportait toutes les fonctionnalités auxquelles je m’attendais.</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

This study made it possible to translate, cross-culturally adapt, and initiate the validation of the original, English MAUQ in Canadian French. Study results indicate that MAUQ-FR has high content validity. CVI is high for all individual items (>0.90) and for the overall scale (0.981), exceeding the minimum thresholds of 0.79 and 0.80, respectively [30,32]. These results are comparable with I-CVIs of the version translated to Malay, which varied between 0.9 and 1, and the overall S-CVI of 0.983 [13]. It is important to distinguish between item-level (I-CVI) and scale-level (S-CVI) content validity as it helps to identify specific elements of the scale that do not effectively measure the desired construct. The \(\kappa\) statistic showed excellent interexpert agreement. These results suggest that MAUQ-FR has been accurately translated and adapted for future francophone users in Canada.

A renowned, systematic method was used to ensure linguistic and cultural equivalence [28,29,37]. The guideline by Sousa and Rojjanasrirat [17] for achieving the objectives provided a clear and precise approach. Translation and cross-cultural adaptation studies must follow a rigorous process, as instruments simply translated from one language to another may lose their validity and no longer measure what they intended to measure, in addition to jeopardizing safety and research ethics [38]. As
Sperber [39] points out in his methodological paper, the translation, cross-cultural adaptation, and validation processes are often treated as afterthoughts in research protocols. Forward translation by uncertified translators is also a commonly used methodological approach [17]. Nevertheless, it is not only essential to translate words in the literary sense but they must, most importantly, also be closely related to the context [40]. This premise is especially critical here, considering that idiomatic expressions vary in each French-speaking region of the vast Canadian territory. Being aware of possible variations involving colloquialisms and jargon, the certified translators, consensus committee members, and experts involved in this study sought to use the most common and neutral vocabulary possible, while paying close attention to cultural nuances (hence the importance of involving translators who come from both cultures or who are bicultural). Beck et al [41] used the same qualitative approach for their cross-cultural study, in which the authors highlighted the need to go beyond the search for equivalence in the denotative meaning of words. Rather, there is a great need to grasp their meaning and connotation within the cultural context they are used. All things considered, the approach was a success, and the results were validated by the group of experts from different French-speaking regions of Canada.

Despite the methodological process, the adaptation of measuring instruments between 2 cultures rarely results in perfect transposition [28]. Some equivalences are more strongly achieved than others. In this study, combining the expertise of translators, researchers, IT specialists, and health care providers favorably contributed to the thorough evaluation of semantic and conceptual equivalences [42]. Choosing qualified and certified translators and having a second independent team for back translation enabled the development of a high-quality instrument by minimizing idiosyncratic bias. Moreover, back translation has long been recognized as a key method for achieving semantic equivalence by ensuring that the translation matches the characteristics of the original instrument [26]. It also allowed the research team to verify the quality of the translation by comparing the 2 English versions of MAUQ (the original and retranslated versions). Few errors were found, attesting to the quality of the previously completed translation and consensus work.

The sequential form of the study allowed for the progressive improvement of MAUQ-FR by identifying ambiguities or terminological imprecision that had not been raised by the translation team. For example, the translation of the item, “The navigation was consistent when moving between screens,” did not exceed the 80% threshold of agreement between the experts, highlighting a lack of clarity in the translated version. Corrections were made by the research team based, among other things, on the feedback received. 1-CVI of the revised version of this item finally reached 100%, ensuring conceptual equivalence. An essential element in the process was the outstanding collaboration with the principal author of the original questionnaire, which enabled fluid communication and clarification of the original meaning of certain items. Finally, the equivalences were deemed to have been satisfactorily achieved, making it possible to proceed to the evaluation of the psychometric properties of MAUQ-FR.

Once the cross-cultural validation process is completed, it will be possible to use MAUQ-FR in a comparable way in different cultures while ensuring data comparability. This will ultimately make it possible to distinguish significant differences between cultures. MAUQ-FR will enable even unilingual anglophone researchers to collect data from francophone Canadians, a population that is currently understudied [20]. Given that French is the world’s fifth most spoken language [18], MAUQ’s French translation can help to create opportunities for other cultural adaptations.

Limitations
This study has its limitations. First, the sociodemographic data of the participants were not collected, as they are not required by the chosen method [17]. However, this prevents certain factors from being considered during the validation process, such as professional experience, age, sex, and gender.

Another limitation is that the pilot test in the target population was conducted exclusively by nurses, whereas the questionnaire could be used by other health care providers. This excluded other potential participants, such as physicians, physiotherapists, respiratory therapists, and other health care providers. In addition, the target population sample was drawn from a university in Quebec (Canada), the province with the largest number of French speakers in the country [19]. These 2 constraints make it impossible to generalize the results to all French-Canadian health care providers. The same applies to the experts surveyed. Although they come from different Canadian provinces, it would be essential to eventually include participants from other French-speaking minority regions such as the Yukon Territory and British Columbia [19]. In addition, the recruitment of experts through networking may have induced a selection bias within the panel. The participants selected were nonetheless representative of the majority of mHealth app users and able to provide a reliable evaluation of the questionnaire.

Although the pilot test allowed for the assessment of conceptual equivalence, question comprehension, and content validity, it does not guarantee construct validity, internal consistency reliability, or fidelity [17,29]. Additional studies must be conducted with full psychometric testing of a large sample of health care providers to establish Cronbach α, internal consistency reliability (Lin concordance correlation coefficient), stability reliability (test-retest), homogeneity, construct-related validity with scale and item analysis, Pearson correlations, and exploratory and confirmatory factor analysis.

Conclusions
In summary, this study is based on the domains of equivalence by Flaherty et al [28] and was conducted in accordance with the methodology by Sousa and Rojjanasrirat [17] to achieve the translation from English to Canadian French, cross-cultural adaptation, and initiation of the validation of MAUQ. Initial tests performed with MAUQ-FR show excellent validity. As part of a doctoral research project, this adaptation was necessary to meet the specific circumstances of the population to be studied, and to ensure the methodological rigor of future studies.
Finally, this study was the first phase of a methodological study and will enable the continuation of work with the psychometric evaluation of MAUQ-FR. The data collected will be shared with the authors of the original MAUQ to undertake further analyses and improve the use of the questionnaire.

Acknowledgments

The authors would like to thank Dr Leming Zhou for granting the permission to translate the mHealth App Usability Questionnaire from English to French. The authors would also like to thank all the participants and experts involved in this study. Finally, the authors would like to thank Lucie Charbonneau, research officer at Université du Québec, for her feedback following the expert panel’s modifications. This study is a part of a PhD project funded by the Canadian Institutes of Health Research Doctoral Research Award (20211FBD-476880-67262), Fonds de recherche du Québec Santé (2022-2023-BF2-319284), and University of Ottawa.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. Recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: [https://www.who.int/publications/i/item/9789241550505] [accessed 2023-06-05]


25. PITT usability questionnaire. Health and Rehabilitation Informatics Lab. URL: https://ux.hari.pitt.edu/v2/portal/#/ [accessed 2022-10-18]


Abbreviations

- **CVI:** content validity index
- **I-CVI:** content validity index at item level
- **MARS:** Mobile Application Rating Scale
- **MAUQ:** mHealth App Usability Questionnaire
- **MAUQ-FR:** mHealth App Usability Questionnaire en français
- **mHealth:** mobile health
- **S-CVI:** content validity index at scale level

---

Edited by A Mavragani; submitted 14.07.23; peer-reviewed by P Nohl-Deryk, N Musrafa; comments to author 11.01.24; revised version received 17.01.24; accepted 19.01.24; published 13.02.24.

Please cite as:
Gagnon J, Probst S, Chartrand J, Lalonde M
mHealth App Usability Questionnaire for Stand-Alone mHealth Apps Used by Health Care Providers: Canadian French Translation, Cross-Cultural Adaptation, and Validation (Part 1)
JMIR Form Res 2024;8:e50839
URL: https://formative.jmir.org/2024/1/e50839
doi: 10.2196/50839
PMID: 38349710

©Julie Gagnon, Sebastian Probst, Julie Chartrand, Michelle Lalonde. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Predictive Criterion Validity of the Parsley Symptom Index Against the Patient-Reported Outcomes Measurement Information System-10 in a Chronic Disease Cohort: Retrospective Cohort Study

Hants Williams¹*, RN, PhD; Sarah Steinberg²*, MD, PhD; Kendall Leon³, PhD; Ryan Vingum¹, MA; Mengyao Hu¹, PhD; Robin Berzin², MD; Heather Hagg², PhD; Patrick Hanaway⁴, MD

¹Applied Health Informatics, School of Health Professions, Stony Brook University, Stony Brook, NY, United States
²Parsley Health, New York, NY, United States
³Untold Content, Cincinnati, OH, United States
⁴Family to Family, Weaverville, NC, United States
* these authors contributed equally

Corresponding Author:
Hants Williams, RN, PhD
Applied Health Informatics
School of Health Professions
Stony Brook University
101 Nicolls Road
Stony Brook, NY, 11794
United States
Phone: 1 631 444 2252
Email: hantsawilliams@gmail.com

Abstract

Background: Approximately 60% of US adults live with chronic disease, imposing a significant burden on patients and the health care system. With the rise of telehealth, patient-reported outcomes measures (PROMs) have emerged as pivotal tools for managing chronic disease. While numerous PROMs exist, few have been designed explicitly for telehealth settings. The Parsley Symptom Index (PSI) is an electronic patient-reported outcome measure (ePROM) developed specifically for telehealth environments.

Objective: Our aim is to determine whether the PSI predicts changes in the established Patient-Reported Outcomes Measurement Information System-10 (PROMIS-10) Global Health, a 10-question short form.

Methods: We conducted a retrospective cohort study using data from 367 unique patients, amassing 1170 observations between August 30, 2017, and January 30, 2023. Patients completed the PSI and the PROMIS-10 multiple times throughout the study period. Using univariate regression models, we assess the predictive criterion validity of the PSI against PROMIS-10 scores.

Results: This study revealed significant relationships between the PSI and PROMIS-10 physical and mental health scores through comprehensive univariate analyses, thus establishing support for the criterion validity of the PSI. These analyses highlighted the PSI’s potential as an insightful tool for understanding and predicting both mental and physical health dimensions.

Conclusions: Our findings emphasize the importance of the PSI in capturing the nuanced interactions between symptomatology and health outcomes. These insights reinforce the value of the PSI in clinical contexts and support its potential as a versatile tool in both research and practice.

(JMIR Form Res 2024;8:e53316) doi:10.2196/53316

KEYWORDS
chronic disease; eHealth; ePROM; mHealth; Parsley Symptom Index; patient-reported outcome measure; PROM; PSI; telehealth; telemedicine; validation; web-based
Introduction

Approximately 60% of US adults live with one or more chronic diseases [1]. Due to the growing population of adults aged 65 years or older and increased risk factors, chronic disease is expected to impact over 221 million people in the United States by 2050 [2]. Developing tools and strategies to promote health is increasingly important as a way to alleviate the enormous burden that chronic disease places on patients, providers, and health care systems [3]. Compared with people without chronic disease, people with chronic conditions have higher health care costs and require more time to manage their care than primary care providers have available [4]. Many new care models, such as the Chronic Care Model, have been implemented [5-7] to overcome these challenges; however, more work is needed to increase access to effective chronic disease care that reduces resource constraints and improves patients’ health.

Emerging telehealth tools, which have become increasingly popular and widespread since the COVID-19 pandemic [8], are proving capable of creating meaningful changes in chronic disease management [9]. Care provided through telehealth has been shown to alleviate many of the burdens of chronic diseases, such as lowering health care costs, reducing missed appointments, and increasing access to timely care [10]. Telehealth tools have also been shown to encourage collaborative disease management, incentivizing patients to participate in their care [11,12].

Patient-reported outcomes measures (PROMs) are patient-oriented, self-reporting tools that can be implemented in a range of settings to improve care processes and track outcomes [13]. PROMs have been found to help facilitate patient-clinician communication [14] and save valuable time and resources for both patients and providers [15], making them a crucial tool in chronic disease management.

Several powerful PROMS exist to capture patients’ perceptions of their health and well-being, such as the Patient Reported Outcomes Measurement Information System (PROMIS) [16,17], the 36-Item Short Form Health Survey [18,19], and the Medical Symptom Toxicity Questionnaire [20]. However, few validated PROMS were designed for telehealth settings first, as opposed to paper and pen PROMs retrofitted for a telehealth environment.

In response to a need for a validated, digital-first electronic patient-reported outcome measure (ePROM), Parsley Health—a subscription-based, holistic medical practice—designed the Parsley Symptom Index (PSI) [21]. To our knowledge, the PSI is the only multi-item ePROM (ie, the Patient-Reported Outcomes Measurement Information System–10 [PROMIS-10] Global Health, a 10-question short form) [16,17]. We aimed to ascertain whether the PSI could predict alterations in the widely accepted PROMIS-10 tool.

Methods

Ethical Considerations

This study used patient-reported survey data that were recorded so that participants were unidentifiable to the researchers. The institutional review board at Stony Brook University considered this study exempt (IRB2020-00429) from the Code of Federal Regulations Title 45 requirements.

Study Design

This retrospective cohort study took place at the “Family to Family” medical clinic in the Southeast region of the United States between August 30, 2017, and January 30, 2023, among a sample of 367 participants with a range of chronic diseases. Additionally, for the purpose of PSI to PROMIS T-score calibration, an independent data set consisting of 122,591 assessments from 29,353 customers of Parsley Health was used to establish the PSI T-score conversion table detailed in Multimedia Appendix 1.

The PSI

The PSI is a 45-item ePROM, similar to a Review of Symptoms, focusing on bodily domains and the most commonly reported symptoms associated with chronic diseases for each domain. The PSI assesses a patient’s perception of symptom burden. The PSI was developed using the Federal Drug Agency’s guidance for PROM development [24]. Items are grouped into 9 systems and ranked on a scale from 0 (asymptomatic) to 10 (extremely symptomatic). A total score is calculated with the following 4 cutoff ranges: “well” (0-24), “symptomatic” (25-43), “very symptomatic” (44-71), and “sick” (≥71). The PSI has shown clinical validity for use in clinical practice [22].

The PROMIS-10

The PROMIS-10 is a single, generalizable, and validated PROM that can be used for various diseases and conditions. It is a shortened version of PROMIS that was developed to minimize respondent burden. This version is a 10-item, patient-reported questionnaire that was created as a general health assessment tool. Nine out of 10 questions on the PROMIS-10 are answered...
using a 5-point Likert scale, with the tenth question answered using a numeric rating scale. Results can be tracked in three different ways: (1) answers to each of the 10 questions can be evaluated separately, (2) answers can be grouped together to provide a global summary score, or (3) answers can be split into 2 groups to provide a global physical health score and a global mental health score.

We compared the PSI to the PROMIS-10, as it is similar to other general health short-form surveys and is widely adopted due to its ease of use. The PROMIS-10 has been shown to be valid and reliable in clinical settings for patients from the general population [16] and those living with chronic diseases [25,26]. Similar to its more extensive counterpart, the PROMIS-10 has undergone rigorous testing and validation across diverse age groups, including younger and older adults [27,28], and has proven to be reliable across a variety of clinical populations [29-31].

Study Setting and Population

Family to Family is a hybrid (remote and in-person) functional and holistic medicine clinic for adults and children located in the Asheville, North Carolina, metropolitan area. The average patient age was 53.7 years old, and patients predominantly identified as female (73%). While race data were not available, the 2 clinicians at this practice report that their patients are predominantly White.

Procedure

Patients and their caregivers were prompted to complete both the PSI and the PROMIS-10 through a password-protected electronic medical record web-based portal before each clinical visit. The PSI was added as a PROM to complete along with the PROMIS-10 because the clinicians believed it provided different insight as a Review of Symptoms to capture a more comprehensive view of patients’ symptomatology and progress over time.

Patients were required to complete the PSI and PROMIS before their first clinical visit. If both ePROMs were not completed before a patient’s first clinical visit at Family to Family, the visit was postponed or rescheduled. For all subsequent visits, completing the ePROM was optional but encouraged. Participants were not compensated for completing the ePROMs. When preparing for the patient’s visit, Family to Family clinicians could view responses to both ePROMs in a patient’s electronic health record and use these responses to guide a clinical encounter. Clinicians were able to ask targeted questions about a patient’s symptoms and identify triggers that might contribute to the symptoms.

Data Analysis Software

The data analyses were conducted using Python (version 3.10; Python Software Foundation) [32]. Statistical Methods

We conducted an analysis to explore the relationship between the PSI and the PROMIS-10. Initially, the raw scores of the PSI and the PROMIS-10 were transformed into T-scores [33,34]. The approach for PSI T-score conversion is detailed in Multimedia Appendix 1. Following T-score conversions, the underlying distribution characteristics of PSI T-scores, PROMIS physical T-scores, and PROMIS mental T-scores were evaluated for normality and distribution. D’Agostino and Pearson normality test were applied to each set of scores to assess the normality. Measures of skewness and kurtosis were calculated for each set of scores to provide insights into the distribution’s symmetry. Histograms with overlaid box plots were created for each set of scores to visually inspect their distributions.

Univariate Regressions

We performed 2 univariate regression models to assess the predictive criterion validity of the PSI T-scores on the PROMIS physical and mental T-scores. Due to the observed nonnormal distribution of the PSI T-scores ($\chi^2_{1166}=183.324; P<.001$), generalized linear models with a Gaussian family and identity link function were chosen as the appropriate modeling approach. This choice accommodates the nonnormal distribution of the PSI T-scores by allowing for a linear relationship between the predictors and response without assuming that the residuals are normally distributed. The flexibility in the Gaussian family made it suitable for modeling the specific distributional properties of the PSI T-scores.

Our first univariate model examined the relationship between PSI T-scores (an independent variable) and PROMIS mental T-scores (a dependent variable), aiming to understand how the PSI is predictive of mental health as quantified by the PROMIS scale. The second univariate model focused on the relationship between PSI T-scores (an independent variable) and PROMIS physical T-scores (a dependent variable), aiming to understand how the PSI is predictive of physical health as quantified by the PROMIS scale. This approach provides insights into the effects of PSI on mental and physical health that are robust to distributional assumptions. Coefficients, SEs, and significance levels were reported to highlight the specific relationships.

Tables

Pivot tables were used to summarize the mean (SD) of the PSI T-scores, PROMIS physical T-scores, and PROMIS mental T-scores. The data were stratified by time order, reflecting different periods of assessment. Multiple pivot tables were generated to encapsulate the mean (SD) for each measurement, organized by the time period.

Results

Overview

In our data set, we analyzed a total of 1170 observations from 367 unique patients recorded between August 30, 2017, and January 30, 2023, from Family to Family. On average, participants completed the PSI 3.2 times and the PROMIS 3.4 times during the study period. Adhering to the guidelines for good reporting practices, the CHERIES (Checklist for Reporting Results of Internet E-Surveys) [35] is included in Multimedia Appendix 2. Detailed patient demographics and general descriptions of the sample are delineated in Table 1. On average, participants reported experiencing 8 distinct symptoms or conditions. Excluding nutrient deficiencies, the most commonly reported diseases and health problems, as classified...
by *International Statistical Classification of Diseases, Tenth Revision (ICD-10)* codes, were other fatigue (82/367, 22.3%), anxiety disorder (81/367, 22.1%), hypothyroidism (81/367, 22.1%), and chronic fatigue (65/367, 17.7%), as outlined in Table 2.

The mean value for the PSI T-score is 40.808 (SD 7.00), with a minimum and maximum range of values between 25 and 56, indicating a broad spectrum of reported symptom states within the sample. Since the expected average for a typical population is 50, this lower mean suggests that the sample population exhibits a higher level of symptoms or less optimal health than the general population. The mean value for the PROMIS physical score is 47.952 (SD 8.114), which is slightly below the expected average of 50. This result also implies that the physical health of the sample population is somewhat below average. The minimum and maximum values for the physical T-scores range from 19.9 to 67.7, indicating a broad spectrum of physical health states within the sample. The mean value for the PROMIS mental score is 46.638 (SD 8.551), which is also below the expected average of 50. This suggests that the mental health of the sample population is also somewhat lower compared to the general population. The PROMIS mental ranges from a minimum of 21.2 to a maximum of 67.6, further indicating variation in mental health states within the sample. Additional descriptives for PSI and PROMIS T-scores across time are provided in Table 3.

**Table 1. Patient descriptives.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Total, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>53.7 (17.1)</td>
<td>367</td>
</tr>
<tr>
<td>Annual income (US $), median (IQR)</td>
<td>70,344 (54,597-72,575)</td>
<td>360</td>
</tr>
<tr>
<td>Annual income (US $), mean (SD)</td>
<td>98,849 (40,259)</td>
<td>361</td>
</tr>
<tr>
<td>Number of symptoms and conditions, mean (SD)</td>
<td>8.1 (5.4)</td>
<td>367</td>
</tr>
<tr>
<td>Number of PROMIS\textsuperscript{a} surveys completed, mean (SD)</td>
<td>3.4 (3.1)</td>
<td>367</td>
</tr>
<tr>
<td>Number of PSI\textsuperscript{b} surveys completed, mean (SD)</td>
<td>3.2 (3.1)</td>
<td>367</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td>367</td>
</tr>
<tr>
<td>Male</td>
<td>99 (27)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>268 (73)</td>
<td></td>
</tr>
<tr>
<td><strong>Insurance, n (%)</strong></td>
<td></td>
<td>332</td>
</tr>
<tr>
<td>Yes</td>
<td>325 (97.9)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>7 (2.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Relationship to the insured\textsuperscript{c}, n (%)</strong></td>
<td></td>
<td>332</td>
</tr>
<tr>
<td>Self</td>
<td>302 (91)</td>
<td></td>
</tr>
<tr>
<td>Spouse</td>
<td>26 (7.8)</td>
<td></td>
</tr>
<tr>
<td>Dependent</td>
<td>4 (1.2)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}PROMIS: Patient-Reported Outcomes Measurement Information System.

\textsuperscript{b}PSI: Parsley Symptom Index.

\textsuperscript{c}Relationship to the insured refers to the participant’s status as the primary beneficiary of the insurance policy. “Self” indicates the participant holds the policy in their own name. “Spouse” denotes the participant is covered under a policy held by their married partner. “Dependent” means the participant is covered under a policy due to their status as a dependent, typically a family member without independent coverage.

---

https://formative.jmir.org/2024/1/e53316

JMIR Form Res 2024 | vol. 8 | e53316 | p.2479

(page number not for citation purposes)
<table>
<thead>
<tr>
<th>Name</th>
<th>ICD&lt;sup&gt;a&lt;/sup&gt; code type</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deficiency of multiple nutrient elements</td>
<td>E</td>
<td>139 (37.9)</td>
</tr>
<tr>
<td>Vitamin D deficiency, unspecified</td>
<td>E</td>
<td>93 (25.3)</td>
</tr>
<tr>
<td>Other fatigue</td>
<td>R</td>
<td>82 (22.3)</td>
</tr>
<tr>
<td>Anxiety disorder, unspecified</td>
<td>F</td>
<td>81 (22.1)</td>
</tr>
<tr>
<td>Hypothyroidism, unspecified</td>
<td>E</td>
<td>81 (22.1)</td>
</tr>
<tr>
<td>Chronic fatigue, unspecified</td>
<td>R</td>
<td>65 (17.7)</td>
</tr>
<tr>
<td>Essential fatty acid deficiency</td>
<td>E</td>
<td>64 (17.4)</td>
</tr>
<tr>
<td>Pure hypercholesterolemia, unspecified</td>
<td>E</td>
<td>62 (16.9)</td>
</tr>
<tr>
<td>Other abnormal glucose</td>
<td>R</td>
<td>52 (14.2)</td>
</tr>
<tr>
<td>Irritable bowel syndrome with diarrhea</td>
<td>K</td>
<td>51 (13.9)</td>
</tr>
<tr>
<td>Autoimmune thyroiditis</td>
<td>E</td>
<td>50 (13.6)</td>
</tr>
<tr>
<td>Mixed irritable bowel syndrome</td>
<td>K</td>
<td>50 (13.6)</td>
</tr>
<tr>
<td>Abnormal level of hormones in specimens from other organ or tissue</td>
<td>R</td>
<td>47 (12.8)</td>
</tr>
<tr>
<td>Other disorders involving the immune mechanism, Not elsewhere classified.</td>
<td>D</td>
<td>45 (12.3)</td>
</tr>
<tr>
<td>Gastroesophageal reflux disease without esophagitis</td>
<td>K</td>
<td>41 (11.2)</td>
</tr>
<tr>
<td>Essential (primary) hypertension</td>
<td>I</td>
<td>39 (10.6)</td>
</tr>
<tr>
<td>Major depressive disorder, recurrent, unspecified</td>
<td>F</td>
<td>39 (9.3)</td>
</tr>
<tr>
<td>Disorder involving the immune mechanism, unspecified</td>
<td>D</td>
<td>36 (9.8)</td>
</tr>
<tr>
<td>Irritable bowel syndrome with constipation</td>
<td>K</td>
<td>34 (9.3)</td>
</tr>
<tr>
<td>Impaired glucose tolerance (oral)</td>
<td>R</td>
<td>32 (8.7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>ICD: International Classification of Diseases.
Table 3. Descriptive statistics by time order.

<table>
<thead>
<tr>
<th>Time point</th>
<th>Sample size, n</th>
<th>PSI T-score, mean (SD)</th>
<th>PROMIS physical T-score, mean (SD)</th>
<th>PROMIS mental T-score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>362c</td>
<td>39.1 (6.7)</td>
<td>46.8 (8.3)</td>
<td>45.7 (8.6)</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>41.4 (7.0)</td>
<td>48.7 (8.1)</td>
<td>46.8 (8.7)</td>
</tr>
<tr>
<td>3</td>
<td>144</td>
<td>40.9 (7.2)</td>
<td>47.8 (8.4)</td>
<td>46.4 (8.5)</td>
</tr>
<tr>
<td>4</td>
<td>109</td>
<td>41.3 (6.7)</td>
<td>48.1 (8.2)</td>
<td>46.6 (9.0)</td>
</tr>
<tr>
<td>5</td>
<td>92</td>
<td>41.2 (6.7)</td>
<td>48.7 (7.2)</td>
<td>47.8 (8.3)</td>
</tr>
<tr>
<td>6</td>
<td>65</td>
<td>41.0 (6.5)</td>
<td>48.4 (6.7)</td>
<td>46.6 (8.0)</td>
</tr>
<tr>
<td>7</td>
<td>48</td>
<td>41.2 (6.0)</td>
<td>47.3 (7.5)</td>
<td>47.8 (8.1)</td>
</tr>
<tr>
<td>8</td>
<td>35</td>
<td>43.0 (7.4)</td>
<td>47.9 (8.7)</td>
<td>46.8 (7.5)</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>41.7 (7.2)</td>
<td>48.2 (8.5)</td>
<td>46.5 (7.8)</td>
</tr>
<tr>
<td>10</td>
<td>18</td>
<td>42.6 (7.1)</td>
<td>49.2 (7.7)</td>
<td>47.8 (8.1)</td>
</tr>
<tr>
<td>11</td>
<td>14</td>
<td>42.9 (7.9)</td>
<td>48.7 (7.8)</td>
<td>47.4 (9.3)</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>43.2 (9.5)</td>
<td>47.7 (6.5)</td>
<td>48.3 (9.4)</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>47.8 (8.4)</td>
<td>54.8 (7.1)</td>
<td>53.5 (9.3)</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>45.8 (6.9)</td>
<td>50.5 (11.8)</td>
<td>49.6 (6.2)</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>48.9 (6.0)</td>
<td>57.9 (3.2)</td>
<td>53.6 (6.9)</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>51.5 (3.5)</td>
<td>59.8 (3.0)</td>
<td>55.4 (10.0)</td>
</tr>
</tbody>
</table>

aPSI: Parsley Symptom Index.

bPROMIS: Patient-Reported Outcomes Measurement Information System.

1Of the 367 unique participants, 5 were excluded from time point 1 due to corrupted data files for the PSI or PROMIS assessments. This resulted in a sample size of 362 for the initial time point.

Distributions And Normality

The distribution characteristics of the PSI and PROMIS T-scores were assessed for normality. The PSI T-scores were found to follow a non-normal distribution ($\chi^2_{1169}$ = 183.324; $P < .001$). The skewness of 0.577 in the PSI T-scores indicates a distribution with a longer right tail and a concentration of scores on the left, reflecting a higher frequency of lower scores and thus a less healthy population. The negative kurtosis of –0.858 signifies a platykurtic kurtosis and that extreme outliers (very high or low) are less frequent in this data set than they would be in a normally distributed data set. In contrast, the PROMIS physical and mental T-scores were found to follow a normal distribution ($P = .10$ and $P = .46$), with a minor skewness of –0.145 and a kurtosis of 0.100 for the physical, while an almost perfect skewness of –0.015 and a slight platykurtic kurtosis of –0.173 for the mental.

Univariate Regressions

Model 1: PSI T-Scores Predicting PROMIS Mental T-Scores

The first generalized linear regression model was fit using the Gaussian family with an identity link function, revealing a significant positive association between the PSI T-scores and PROMIS mental T-scores (Table 4). Specifically, a 1-unit increase in the PSI T-score corresponded to a 0.627-unit increase in the PROMIS mental T-score (95% CI 0.567-0.687; $z = 20.462$; $P < .001$). The intercept was estimated at 21.0487 (95% CI 18.562-23.536). The model’s pseudo $R^2$ value (CS) was 0.3008, indicating that it explained approximately 30.08% of the variability in the PROMIS mental T-scores. The deviance statistic, which measures the goodness of fit, was 62,925, and the Pearson chi-square value was approximately 62,900, further supporting the model’s fit to the data.

Model 2: PSI T-Scores Predicting PROMIS Physical T-Scores

The second generalized linear regression model was fit using the Gaussian family with an identity link function, revealing a significant positive association between the PSI T-scores and PROMIS physical T-scores (Table 4). Specifically, a 1-unit increase in the PSI T-score corresponded to a 0.6479 unit increase in the PROMIS physical T-score (95% CI 0.593-0.703; $z = 23.064$; $P < .001$). The intercept was estimated at 21.0487 (95% CI 18.562-23.536). The model’s pseudo $R^2$ value (CS) was 0.3008, indicating that it explained approximately 30.08% of the variability in the PROMIS mental T-scores. The deviance statistic, which measures the goodness of fit, was 62,925, and the Pearson chi-square value was approximately 62,900, further supporting the model’s fit to the data.
Table 4. Summary of generalized linear model regression models predicting Patient-Reported Outcomes Measurement Information System (PROMIS) mental and physical T-scores from Parsley Symptom Index (PSI) T-scores.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dependent variable</th>
<th>Coefficient</th>
<th>95% CI</th>
<th>P value</th>
<th>Pseudo $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>PROMIS mental</td>
<td>0.6270</td>
<td>0.567-0.687</td>
<td>&lt;.001</td>
<td>0.3008</td>
</tr>
<tr>
<td>Model 2</td>
<td>PROMIS physical</td>
<td>0.6479</td>
<td>0.593-0.703</td>
<td>&lt;.001</td>
<td>0.3653</td>
</tr>
</tbody>
</table>

Discussion
Overview
Previous studies found the PSI to be a valid tool that can be deployed, completed, and helpful to both patients and clinicians [21,22,36]. This study examined differences in use between the PSI as compared to the PROMIS-10 short form when used in clinical settings with patients with chronic disease. The PSI and PROMIS-10 were chosen for this retrospective study because the clinicians believed it was useful for patients to complete both forms as they provide slightly different insights into how patients perceive their health status. While the differences between the PSI and PROMIS-10 reveal ways that each has its place in clinical practice, they overlap enough to demonstrate a moderate correlation to support validation of the PSI when compared to the PROMIS-10.

The overall mean (SD) statistics for the PSI and PROMIS-10 paint a picture of a sample population that is generally less healthy than the average population, both in terms of physical and mental well-being, as well as symptomatology. The lower PSI score, in particular, stands out, indicating a higher level of symptoms. These findings set the stage for further analysis to understand the underlying factors, relationships, and potential interventions that may be relevant to this specific population.

Univariate Regression Analysis: Generalized Linear Model
The univariate generalized linear regression analyses conducted in this study used the Gaussian family with an identity link function to explore the relationship between the PSI T-scores and both mental and physical health as measured by the PROMIS scales. The consistency in the direction and significance of the relationships between the PSI across both mental and physical health domains defined by the PROMIS, as revealed in the univariate analyses, lends credibility to the models and provides a robust foundation for further exploration. The positive associations demonstrate the criterion validity of the PSI, illustrating its potential to predict changes in both mental and physical health as measured by the PROMIS scales. The substantial explanatory power of the models, as evidenced by the pseudo $R^2$ values and supported by the deviance statistics and Pearson chi-square values, adds to the robustness of the findings and their potential implications for the PSI as a validated health assessment. This validation underscores the use of the PSI and its capability of offering insights into overall well-being. Future research may benefit from examining these relationships in different populations or contexts, potentially extending the applicability of the PSI.

Limitations
This study bears several notable limitations. First, our data emanate from a single clinic where a significant majority of Family to Family participants identified as female (73%), with an average age of 53 years. Although race and ethnicity data were not available, the clinic reported that the patient population was predominantly White. Such skewness constrains the ecological validity of our findings.

Our sample size was not large enough to support a robust longitudinal analysis, thereby limiting the depth of insights we could derive. It is imperative for future validation studies to explore the PSI’s use within a more diverse demographic profile and over a more extended time frame. Such studies would not only deepen the understanding of symptom trajectories but also facilitate the evaluation of patient outcomes across a wider demographic landscape.

In terms of the PSI questionnaire itself, the Family to Family participants engaged with an earlier iteration. Based on patient feedback, Parsley Health implemented minor revisions to this version to enhance readability. Consequently, we made retrospective adjustments to ensure alignment with the updated version of the PSI, which encompassed an additional item but was more concise in terms of completion time since responses were no longer categorized as “resolved” or “ongoing.”

Additionally, the nature of this being a retrospective cohort study meant that the PSI and PROMIS-10 items were not presented to participants in a randomized manner, which could have potentially mitigated response biases. We advocate for the implementation of randomization, or A/B testing, in subsequent studies.

Conclusions
Although we know that telehealth tools can be used to deliver effective care to patients with chronic conditions, few—if any—tools exist that are designed as digital-first ePROMS. This predictive criterion study compared the PSI—a digital-first ePROM—to the PROMIS-10, a traditional PROM, in a functional medicine clinic for patients with a range of chronic conditions. This study revealed significant relationships between the PSI and PROMIS physical and mental health scores through comprehensive univariate analyses, thus establishing support for the criterion validity of the PSI. These analyses highlighted the PSI’s potential as an insightful tool for understanding and predicting both mental and physical health dimensions.

Overall, the findings of this study emphasize the importance of the PSI as a versatile clinical instrument. Future research is warranted to further dissect these relationships and enhance our understanding of the PSI’s applicability in various health contexts.
Acknowledgments

This study was fully supported by Parsley Health. The funder had the following involvement with the study: study design, research, and preparation of the manuscript.

Authors' Contributions

HW, SS, RB, and HH contributed to the conception of the study design, manuscript preparation, and data collection. KL and RV contributed to the manuscript preparation. All authors read and approved the final version of the manuscript. This statement confirms that this manuscript has been submitted solely to this journal and is not published, in press, or submitted elsewhere.

Conflicts of Interest

All authors are either employees or consultants to Parsley Health at the time of analysis. All authors declare no other competing interests.

Multimedia Appendix 1

Parsley Symptom Index (PSI) reverse coding guide.
[DOCX File, 42 KB - formative_v8i1e53316_app1.docx]

Multimedia Appendix 2

CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist.
[PDF File (Adobe PDF File), 66 KB - formative_v8i1e53316_app2.pdf]

References


35. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]


Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
ePROM: electronic patient-reported outcome measure
ICD-10: International Statistical Classification of Diseases, Tenth Revision
PROM: patient-reported outcome measure
PROMIS: Patient-Reported Outcomes Measurement Information System
PSI: Parsley Symptom Index
SRH: single-rated health

Please cite as:
Predictive Criterion Validity of the Parsley Symptom Index Against the Patient-Reported Outcomes Measurement Information System-10 in a Chronic Disease Cohort: Retrospective Cohort Study
JMIR Form Res 2024;8:e53316
URL: https://formative.jmir.org/2024/1/e53316
doi:10.2196/53316
PMID:38363587

©Hants Williams, Sarah Steinberg, Kendall Leon, Ryan Vingum, Mengyao Hu, Robin Berzin, Heather Hagg, Patrick Hanaway. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
**Original Paper**

**Daily Activity Lifelogs of People With Heart Failure: Observational Study**

Teketo Kassaw Tegegne¹, BSc, MPH, PhD; Ly-Duyen Tran², MSc; Rebecca Nourse¹, MSc; Cathal Gurrin², PhD; Ralph Maddison¹, PhD

¹Deakin University, Burwood, Australia
²Dublin City University, Dublin, Ireland

**Corresponding Author:**
Teketo Kassaw Tegegne, BSc, MPH, PhD
Deakin University
221 Burwood Highway
Burwood, 3125
Australia
Phone: 61 0406722673
Email: teketo.tegegne@deakin.edu.au

**Abstract**

**Background:** Globally, heart failure (HF) affects more than 64 million people, and attempts to reduce its social and economic burden are a public health priority. Interventions to support people with HF to self-manage have been shown to reduce hospitalizations, improve quality of life, and reduce mortality rates. Understanding how people self-manage is imperative to improve future interventions; however, most approaches to date, have used self-report methods to achieve this. Wearable cameras provide a unique tool to understand the lived experiences of people with HF and the daily activities they undertake, which could lead to more effective interventions. However, their potential for understanding chronic conditions such as HF is unclear.

**Objective:** This study aimed to determine the potential utility of wearable cameras to better understand the activities of daily living in people living with HF.

**Methods:** The “Seeing is Believing (SIB)” study involved 30 patients with HF who wore wearable cameras for a maximum of 30 days. We used the E-Myscéal web-based lifelog retrieval system to process and analyze the wearable camera image data set. Search terms for 7 daily activities (physical activity, gardening, shopping, screen time, drinking, eating, and medication intake) were developed and used for image retrieval. Sensitivity analysis was conducted to compare the number of images retrieved using different search terms. Temporal patterns in daily activities were examined, and differences before and after hospitalization were assessed.

**Results:** E-Myscéal exhibited sensitivity to specific search terms, leading to significant variations in the number of images retrieved for each activity. The highest number of images returned were related to eating and drinking, with fewer images for physical activity, screen time, and taking medication. The majority of captured activities occurred before midday. Notably, temporal differences in daily activity patterns were observed for participants hospitalized during this study. The number of medication images increased after hospital discharge, while screen time images decreased.

**Conclusions:** Wearable cameras offer valuable insights into daily activities and self-management in people living with HF. E-Myscéal efficiently retrieves relevant images, but search term sensitivity underscores the need for careful selection.

(JMIR Form Res 2024;8:e51248) doi:10.2196/51248

**KEYWORDS**

heart failure; self-management; lifelogs; daily activity; wearable camera; E-Myscéal; activities of daily living; ADL; intervention; self-report method; wearable; chronic condition

**Introduction**

Heart failure (HF) refers to when the heart does not work as well as it should in pumping blood and oxygen around the body [1]. HF is one of the most common chronic conditions, affecting 64 million people worldwide [2]. Despite advancements in medical treatment, people diagnosed with HF remain at an increased risk of hospitalization, hospital readmission, and
premature death [3]. People with HF are encouraged to engage in a range of daily self-care activities to help manage their condition and prevent deterioration of their condition, collectively referred to as self-management. Those strategies typically include taking prescribed medication, making lifestyle changes, self-weighing, and monitoring signs and symptoms, and responding accordingly [4]. Previous meta-analyses of interventions to support self-management have reported significant benefits for people living with HF, including reduced hospitalizations [5-7], improved quality of life, and reduced all-cause mortality [5,7]. Moreover, self-management empowers people living with HF to take an active role in their care and facilitate effective management of their condition [4].

Understanding what behaviors people with HF engage in, and when, is important to understand how to support people with HF to better manage their condition. However, previous research has often relied on self-report methods to achieve this; self-report is subject to recall and social desirability biases, and memory impairment [8-11]. Digital technologies, such as wearable cameras, can provide valuable insights into health management behaviors, including contextual information such as location, time of day, and setting. Further, wearable cameras also offer a unique tool for researchers and health care providers to gain a better understanding of people’s lived experiences, which could lead to more tailored and effective interventions and treatments [12].

However, using wearable cameras generates large data sets, requiring the retrieval, processing, and analysis of relevant images for practical utility. This involves assigning semantic contexts like visual descriptions, time, and location [13,14]. Various computer vision models are employed, such as object detection, activity recognition, and optical character recognition, in addition to embedding models [13,15-17]. Retrieval systems incorporate techniques such as query enhancement, visual similarity search, and temporal search [13,16]. Previous studies [18-20], including our own [21], have manually reviewed camera images, which is time-consuming. In this study, we used the E-Myscéal system [22] for efficient retrieval and review of relevant images depicting the daily activities of people living with HF.

Daily activities significantly impact the health and well-being of people with HF, and a balanced approach to activities such as physical activity, screen time, eating, drinking, and medication intake is crucial for effective symptom management and improved quality of life [23-25]. Thus, the overall aim of this study was to determine the utility of wearable cameras to better understand activities of daily living (ADLs) in people living with HF. This study focused on 7 daily activities: physical activity, gardening, shopping, screen time, drinking, eating, and medication intake behaviors. This study also aimed to evaluate the sensitivity of an image-processing software tool for identifying these activities.

Methods

Study Design

The Seeing is Believing (SIB) study was a large prospective observational pilot study that evaluated the feasibility and acceptability of using wearable cameras and point-of-care testing for self-care in patients with HF [21,26]. As a pilot study, no formal power calculations were performed to determine the optimum sample size. It involved 30 patients with HF in Melbourne, Victoria, Australia, who wore wearable cameras for a month and conducted regular self-assessments. However, in this study, we analyzed the wearable camera image data.

Study Population, Recruitment, and Setting

Participants, aged 18 years or older, were recruited from a single-center HF outpatient clinic in Melbourne if they had a documented HF diagnosis, had previous HF hospitalization, were on maximum tolerated medication, and were able to read and understand English. Exclusion criteria included severe HF symptoms (New York Heart Association class IV), advanced malignancy, cognitive impairment, and end-of-life care. Recruitment occurred during outpatient HF clinic visits by a cardiologist or researcher, followed by screening for eligibility by a Deakin University researcher. Eligible individuals underwent a baseline assessment and provided written consent.

Study Procedures

Participants were asked to attach the “Narrative Clip” wearable camera to their shirt or blouse during waking hours for a maximum of 30 days. The camera captured images every 30 seconds, resulting in a data set of approximately 2.2 million images. Baseline assessment included medical history, demographics, and physical measurements. A research assistant collected point-of-care test data, hospitalization information, and other variables twice a week during the 30-day study [26]. At the end of this study, participants underwent a brief interview about their camera usage experience. In the original SIB study, 10 individuals were hospitalized [26].

Wearable Camera Image-Processing and Retrieval

The wearable camera image data were processed and retrieved using the E-Myscéal web-based lifelog retrieval system. The E-Myscéal system uses deep learning algorithms to create embeddings (vector representations) of various lifelog data types such as images, text, and audio, enabling intuitive and web-based cross-media querying [22]. E-Myscéal uses the CLIP model to retrieve images similar in content to descriptive textual search terms, allowing users to search through lifelog data with great flexibility [27]. E-Myscéal was ranked as the top retrieval system at the Lifelog Search Challenge (LSC’22 challenge), outperforming others in terms of finding the highest number of relevant items in the shortest time across various retrieval tasks [28].

With E-Myscéal, users can enter any search terms, leading to a list of related images along with date or time and location metadata [22]. For instance, when “eating” was used as a search term, E-Myscéal retrieved all food-related camera images, enabling researchers to assess the frequency of the wearer’s
Before applying the E-Mycéal system to our wearable camera data, we evaluated its capability to detect ADLs using a publicly available data set [29]. The results indicated a high precision in detecting ADLs such as physical activity, screen time, and shopping, with precision ranging from 0.8 to 1.0 [28]. Given these results, we decided to use the system with our data set. Detailed information on how E-Mycéal determined image counts can be found in Multimedia Appendix 1.

Figure 1. E-Mycéal user interface and event view window. Cam: camera.

Search Terms
Wearable camera images from the SIB study were uploaded to the E-Mycéal system [22]. Based on expert knowledge we developed search terms for each of the 7 daily activities (physical activity, gardening, shopping, screen time, drinking, eating, and taking medication), which were entered into E-Mycéal. Different search terms were used for each activity to identify the most appropriate ones for use. For example, for physical activity, search terms included “running,” “walking,” “yoga,” or “exercise,” which are common terms used to describe physical activity. Similarly, for gardening, search terms included “horticulture,” “cultivation,” “planting,” or “watering,” which are commonly used terms in gardening. This study assessed the relevance of search terms for each activity by analyzing the number and relevance of images returned.

Data Analysis
We performed a sensitivity analysis using the Wilcoxon test to examine significant differences in the number of images retrieved for both searches (with and without outliers). Following the sensitivity analysis, descriptive statistics were performed for ADLs using the most sensitive search terms. The results are presented in tables and figure. Furthermore, we used the Wilcoxon test to determine if there were significant temporal differences in daily activity patterns (ie, before and after midday). Further, we examined the ADLs of the 10 participants who were hospitalized during the 30-day study period. We compared their ADLs before they were admitted to the hospital and after they were discharged and returned home.

Ethical Considerations
Ethical approval was granted by the Deakin University Human Research Ethics Committee (HREC/16/MH/55) and the Western Health Human Research Ethics Committee (2016.071). Participants were provided with study information through an information sheet and verbal explanation, with the option to withdraw at any time while retaining data collected to that point. All participants had an opportunity to ask questions about this study before they gave written informed consent. Participants could delete images from their wearable camera during this study. Upon completion, participants received an Aus $40 (US $28) voucher.

Results
Characteristics of Study Participants
In total, 30 adults (18 men) with HF agreed to participate and wore the wearable camera for up to 30 days. The median age...
of participants was 84 years, with a range of 47-96 years. Out of 30 participants, 20 were in New York Heart Association class III, and 10 were in class II. Among the 30 participants, 18 were diagnosed with HF within 5 years, and 10 were readmitted to the hospital due to HF exacerbation. For additional sample details, please refer to our previous publications [21,26]. No serious adverse events were reported with the use of the wearable camera.

**E-Myscéal Sensitivity to Search Terms**

For each search term, the E-Myscéal system yielded varying numbers of images for each daily activity (Table 1). For example, the difference in the number of images for each daily activity ranged from 362 for screen time to 7015 for gardening. The Wilcoxon test confirmed a significant difference in the number of images retrieved using these search terms (Table 1). Based on the sensitivity analysis, the final search terms selected for each daily activity were as follows: physical activity, gardening, retail therapy, screen viewing, fluid intake, food intake, and taking medication.

### Table 1. Sensitivity of the E-Myscéal system to search terms related to daily activities in a wearable camera pilot study on people with HF

<table>
<thead>
<tr>
<th>Daily activities and search terms</th>
<th>Images retrieved, n</th>
<th>Wilcoxon test P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical activity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td>12,575</td>
<td>&gt; .05</td>
</tr>
<tr>
<td>Exercise</td>
<td>11,083</td>
<td></td>
</tr>
<tr>
<td><strong>Screen time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screen viewing</td>
<td>16,968</td>
<td>&gt; .05</td>
</tr>
<tr>
<td>Screen time</td>
<td>16,606</td>
<td></td>
</tr>
<tr>
<td><strong>Taking medication</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taking medication</td>
<td>21,750</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Medication intake</td>
<td>16,985</td>
<td></td>
</tr>
<tr>
<td><strong>Shopping</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail therapy</td>
<td>23,318</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Shopping</td>
<td>21,221</td>
<td></td>
</tr>
<tr>
<td><strong>Drinking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluid intake</td>
<td>23,773</td>
<td>&gt; .05</td>
</tr>
<tr>
<td>Hydrating</td>
<td>19,020</td>
<td></td>
</tr>
<tr>
<td><strong>Gardening</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gardening</td>
<td>25,585</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Horticulture</td>
<td>18,570</td>
<td></td>
</tr>
<tr>
<td><strong>Eating</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food intake</td>
<td>26,812</td>
<td>&gt; .05</td>
</tr>
<tr>
<td>Food consumption</td>
<td>24,672</td>
<td></td>
</tr>
</tbody>
</table>

*aHF: heart failure.*

**Activity Identification**

For each of the 7 daily activities, the number of images returned ranged from 12,575 to 26,812. Eating had the highest number of images returned at 26,812, followed by gardening with 25,583 images, drinking with 23,773 images, and shopping with 23,318 images (Table 2).
Table 2. Daily activities of people living with HF\(^a\) in a wearable camera pilot study (N=30).

<table>
<thead>
<tr>
<th>Daily activity</th>
<th>Images retrieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical activity</td>
<td>12,575</td>
</tr>
<tr>
<td>Screen time</td>
<td>16,968</td>
</tr>
<tr>
<td>Taking medication</td>
<td>21,750</td>
</tr>
<tr>
<td>Shopping</td>
<td>23,318</td>
</tr>
<tr>
<td>Drinking</td>
<td>23,773</td>
</tr>
<tr>
<td>Gardening</td>
<td>25,583</td>
</tr>
<tr>
<td>Eating</td>
<td>26,812</td>
</tr>
</tbody>
</table>

\(^a\)HF: heart failure.

Activity Patterns During the Day

Most of the activities were observed before midday (Table 3). For example, 17,980 (67.1\%) of images returned using search terms related to “eating” activity were captured before midday. Similarly, 17,065 (66.7\%) gardening images, 15,245 (65.4\%) shopping, and 15,219 (64\%) drinking images were captured before midday. Additionally, there were statistically significant differences between daily activity patterns pre- and postmidday.

Table 3. Daily activities of people living with HF\(^a\) based on time of day in a wearable camera pilot study (N=30).

<table>
<thead>
<tr>
<th>Daily activities</th>
<th>Before midday, n (%)</th>
<th>After midday, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical activity</td>
<td>7572 (60.2)</td>
<td>5003 (39.8)</td>
</tr>
<tr>
<td>Screen time</td>
<td>10,383 (61.2)</td>
<td>6585 (38.8)</td>
</tr>
<tr>
<td>Taking medication</td>
<td>13,299 (61.1)</td>
<td>8451 (38.9)</td>
</tr>
<tr>
<td>Shopping</td>
<td>15,245 (65.4)</td>
<td>8073 (34.6)</td>
</tr>
<tr>
<td>Drinking</td>
<td>15,219 (64)</td>
<td>8554 (36)</td>
</tr>
<tr>
<td>Gardening</td>
<td>17,065 (66.7)</td>
<td>8518 (33.3)</td>
</tr>
<tr>
<td>Eating</td>
<td>17,980 (67.1)</td>
<td>8832 (32.9)</td>
</tr>
</tbody>
</table>

\(^a\)HF: heart failure.

Activity Patterns Before and After Hospitalization

For the 10 participants that were hospitalized, there were marked differences in the percentage of ADLs images captured by a wearable camera pre- and posthospitalization, except for physical activity (Table 4). To illustrate, a total of 10,144 medication images were identified using E-Myscéal. Of those, 3827 (37.7\%) were observed before hospital admission, while 6317 (62.3\%) were observed posthospital discharge. In contrast, there was a decrease in screen time images after hospital discharge 946 (32.1\%), compared to before admission 2003 (67.9\%).

Table 4. Activities of daily living in patients with HF\(^a\) before and after hospitalization in a wearable camera pilot study (N=10).

<table>
<thead>
<tr>
<th>Daily activities</th>
<th>Before hospital admission, n (%)</th>
<th>After hospital discharge, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical activity</td>
<td>2709 (50.5)</td>
<td>2659 (49.5)</td>
</tr>
<tr>
<td>Screen time</td>
<td>2003 (67.9)</td>
<td>946 (32.1)</td>
</tr>
<tr>
<td>Taking medication</td>
<td>3827 (37.7)</td>
<td>6317 (62.3)</td>
</tr>
<tr>
<td>Shopping</td>
<td>4161 (42.5)</td>
<td>5633 (57.5)</td>
</tr>
<tr>
<td>Drinking</td>
<td>3958 (35.4)</td>
<td>7211 (64.6)</td>
</tr>
<tr>
<td>Gardening</td>
<td>3142 (23.8)</td>
<td>10,061 (76.2)</td>
</tr>
<tr>
<td>Eating</td>
<td>5516 (40.7)</td>
<td>8038 (59.3)</td>
</tr>
</tbody>
</table>

\(^a\)HF: heart failure.
Discussion

Principal Findings

This study aimed to determine the potential utility of wearable cameras to better understand ADLs in people living with HF. Overall, we showed that wearable cameras can be used to capture specific daily activities, which could be used to help identify areas where interventions may be targeted to improve people’s overall health and well-being. The E-Myscéal search engine was critical to the potential utility of this technology by rapidly retrieving images of relevance. Beyond, these general observations, specific issues are discussed below.

Typically, the use of wearable cameras results in large data sets, which means we have to retrieve, process, and analyze images of relevance for these data to be useful. The most common approach to organizing, retrieving, and analyzing data from wearable cameras involves assigning semantic contexts to images, like visual descriptions, time, and location [13,14]. Various computer vision models are employed to extract visual information from the images, including object detection, activity recognition, optical character recognition [13,15], and embedding models [16,17]. A typical retrieval system would also incorporate different techniques, namely, query enhancement [13], visual similarity search [16], and temporal search [16]. We previously reviewed the SIB image data manually to determine the feasibility and acceptability of wearable cameras to assess self-care in people living with HF [21]. Other studies have also manually reviewed camera images [18-20], which is time-consuming and laborious. In this study, we used the E-Myscéal system [22] to rapidly retrieve and review images of relevance. The E-Myscéal system eliminated the need for manual image review, enabled cross-media querying, supported temporal queries, allowed data filtering and clustering, and provided fast retrieval of the most relevant images.

E-Myscéal was the top-ranked retrieval system at the LSC’22 challenge [28], achieving the highest number of relevant items in the shortest time, and represents the current state-of-the-art in lifelog inquiry and retrieval systems. E-Myscéal exhibits a heightened sensitivity toward certain search terms. For instance, the search term “taking medication” generated considerably higher scores compared to “medication intake,” leading to a significant difference in the number of images retrieved. Hence, the user needs to consider what search terms will work well for any category or daily activities. The flexibility of the E-Myscéal system allowed a wide range of search terms to be tested and evaluated for any category, which the user needs to consider in future research.

Using the E-Myscéal system, the most frequently observed activities were eating, drinking, and screen time. These findings suggest that people living with HF are sedentary and spend considerable time in the presence of some form of screen (eg, television and computer). However, it was not possible to determine from the images whether the person was watching the television. As highlighted above, the Narrative Clip used in this study captured images every 30 seconds and captured the presence of objects in front of the person (eg, television, book, or dinner plate), as well as the context (eg, sitting at a table, being inside or outside, or seated in a car); however, unless the camera captured the person lifting a cup to their mouth it was difficult to determine whether they actually performed the behavior. Thus, while the technology offers a potential for capturing potential activities, a process for confirming what the person was doing is needed. In previous studies [18-20], wearable cameras have been used to support traditional self-report methods or to provide a primary record of dietary intake, however, no studies have used wearable cameras as the primary method of data collection; future research is needed to address this.

If the recorded images do represent a fair reflection of what people with HF were doing throughout the day, then they highlight some interesting patterns. For example, excessive screen or sedentary time might highlight people’s preferences (eg, sitting and watching television) [30-32] or could indicate fatigue, a common symptom of HF. Further, the high number of recorded images of eating and drinking could give some indication of what people with HF were eating (eg, foods high in salt) [33], and the frequency of drinking fluids, which may be an issue for people with HF who are on fluid restriction [4,33]. Despite the importance of taking prescribed medications and engaging in physical activities such as gardening and shopping, the number of images recorded by wearable cameras for these activities was low. The lower number of medication-taking images may have resulted from people not wearing their cameras first thing in the morning or later in the evening when medications were taken. However, it is more likely that the sensitivity of E-Myscéal for detecting medication taking was lower than for other ADLs used in this study. In terms of physical activities, it is plausible that participants removed the camera when shopping or doing physical activity. However, this is unlikely. In the original SIB study, adherence to wearing the camera was high [26]. Moreover, previous research has shown that people wear these types of wearable cameras when undertaking physical activity and they are useful for providing contextual information on activity and sedentary behavior [34,35].

We also found that most images of ADLs were observed before midday. Further, I possible reason is that participants did not wear the camera in the afternoon—they may have removed the camera or forgotten [36] to put it back on, which would result in a lower number of images captured during this time. Another possibility is that the camera battery might not have been fully charged during the previous night, which resulted in the device turning off [37] in the afternoon and fewer images being captured. Additionally, it is possible that participants were spending their afternoons sleeping by taking naps or resting [38-40], which would result in fewer activities [41] being captured during this time. Lastly, issues with the wearable camera time set-up, which may have led to inaccurate time stamps on the images captured, could also have contributed to the observed differences in activity patterns.

Furthermore, there were observed changes in the participants’ daily activities both before hospital admission and after discharge, which could be attributed to the time participants wore the device or differences in performed ADLs. If the latter,
these findings suggest that participants in specific ADLs may have been influenced by their health condition or recovery process. For instance, if they were less active before hospitalization or had mobility limitations, it could result in fewer captured images during that period. The increase in medication images after discharge suggests a heightened focus on medication management and adherence during recovery. Conversely, the decrease in screen time images implies a shift in attention or engagement with electronic devices, possibly due to increased social interaction or involvement in other activities. These findings provide insights into potential changes in participants’ daily routines and behaviors, but further analysis, interpretation, and context are needed to fully understand the reasons behind these variations and their impact on overall well-being.

The temporal differences in the number of images recorded might have implications for managing HF in older adults, as health care professionals could use this information to optimize future self-care strategies such as designing customized interventions to promote medication taking, physical activity, minimizing sedentary behavior, and promoting a healthy diet [4,33]. The use of wearable cameras to augment self-care interventions was highlighted in a previous scoping review [42]. In that review, the authors suggested that people with a new diagnosis of HF could wear a camera for several days. On their return to an outpatient clinic, a nurse specialist or other health professional could review images alongside the individual to identify specific activities and use that as an opportunity to question them about self-care practices and offer tailored suggestions for improvement. The E-Myscéal platform would permit such rapid review.

A strength of this study was the use of wearable cameras to record first-person perspectives of real-life experiences and associated contexts for people living with HF. This approach provides a rich source of data to better understand people’s lived experiences and context for self-management and could be used to enhance patient outcomes by enabling health care providers to access more personalized and precise information about a person’s condition, which could ultimately lead to better care and treatment. This study is also the first to use the E-Myscéal system, a flexible and efficient solution for processing large volumes of images in a short time, addressing image preprocessing and classification challenges. However, a limitation of this approach is the sensitivity of E-Myscéal to specific search terms, which may influence the number and types of images retrieved, which could affect interpretation. It is important to consider these factors when interpreting this study’s findings and drawing conclusions about the daily activity patterns of individuals living with HF. Future research is needed to investigate the reasons behind these differences and develop strategies to improve the accuracy and reliability of wearable camera data collection.

Conclusions
Wearable cameras are a valuable tool for understanding daily activities and self-management in people living with HF. E-Myscéal efficiently retrieves images, emphasizing the need for careful search term selection. These findings suggest a potential for tailored HF interventions based on temporal activity patterns, despite challenges in confirming specific behaviors from images. Further research is needed to address observed activity variations and enhance data accuracy.

Acknowledgments
The National Heart Foundation of Australia (Vanguard Grant #101348) provided funding for this study. However, the funding agency did not play a part in the design, execution, analysis, or interpretation of this study. The authors would like to express their gratitude to the recruitment staff at Western Health, as well as all the participants who took part in this study.

Data Availability
The data sets generated and analyzed during this study are not publicly available as the camera images include identifiable information. However, other data are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The method by which E-Myscéal determines the number of images for a given query.

References

https://formative.jmir.org/2024/1/e51248


Abbreviations

ADL: activity of daily living  
HF: heart failure  
LSC’22: Lifelog Search Challenge  
SIB: Seeing is Believing
Designing and Validating a Novel Method for Assessing Delay Discounting Associated With Health Behaviors: Ecological Momentary Assessment Study

Amanda Luken1, MHS; Jill A Rabinowitz1, PhD; Jonathan L Wells2, MHS; David W Sosnowski1, PhD; Justin C Strickland1, PhD; Johannes Thrul1,4,5, PhD; Gregory D Kirk6,7, MPH, MD, PhD; Brion S Maher1, PhD

1Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States
2Department of Epidemiology, School of Population Health, Virginia Commonwealth University, Richmond, VA, United States
3Department of Psychiatry and Behavioral Sciences, School of Medicine, Johns Hopkins University, Baltimore, MD, United States
4Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, United States
5Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia
6Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States
7School of Medicine, Johns Hopkins University, Baltimore, MD, United States

Abstract

Background: Delay discounting quantifies an individual’s preference for smaller, short-term rewards over larger, long-term rewards and represents a transdiagnostic factor associated with numerous adverse health outcomes. Rather than a fixed trait, delay discounting may vary over time and place, influenced by individual and contextual factors. Continuous, real-time measurement could inform adaptive interventions for various health conditions.

Objective: The goals of this paper are 2-fold. First, we present and validate a novel, short, ecological momentary assessment (EMA)–based delay discounting scale we developed. Second, we assess this tool’s ability to reproduce known associations between delay discounting and health behaviors (ie, substance use and craving) using a convenience-based sample.

Methods: Participants (N=97) were adults (age range 18-71 years), recruited on social media. In phase 1, data were collected on participant sociodemographic characteristics, and delay discounting was evaluated via the traditional Monetary Choice Questionnaire (MCQ) and our novel method (ie, 7-item time-selection and 7-item monetary-selection scales). During phase 2 (approximately 6 months later), participants completed the MCQ, our novel delay discounting measures, and health outcomes questions. The correlations between our method and the traditional MCQ within and across phases were examined. For scale reduction, a random number of items were iteratively selected, and the correlation between the full and random scales was assessed. We then examined the association between our time- and monetary-selection scales assessed during phase 2 and the percentage of assessments that participants endorsed using or craving alcohol, tobacco, or cannabis.

Results: In total, 6 of the 7 individual time-selection items were highly correlated with the full scale (r>0.89). Both time-selection (r=0.71; P<.001) and monetary-selection (r=0.66; P<.001) delay discounting rates had high test-retest reliability across phases 1 and 2. Phase 1 MCQ delay discounting function highly correlated with phase 1 (r=0.76; P<.001) and phase 2 (r=0.45; P<.001) time-selection delay discounting scales. One or more randomly chosen time-selection items were highly correlated with the full scale (r>0.94). Greater delay discounting measured via the time-selection measure (adjusted mean difference=5.89, 95% CI 1.99-9.79), but not the monetary-selection scale (adjusted mean difference=–0.62, 95% CI –3.57 to 2.32), was associated with more past-hour tobacco use endorsement in follow-up surveys.

Corresponding Author: Brion S Maher, PhD
Department of Mental Health
Bloomberg School of Public Health
Johns Hopkins University
624 N Broadway
Baltimore, MD, 212055
United States
Phone: 1 4432878287
Email: brion@jhu.edu
Conclusions: This study evaluated a novel EMA-based scale’s ability to validly and reliably assess delay discounting. By measuring delay discounting with fewer items and in situ via EMA in natural environments, researchers may be better able to identify individuals at risk for poor health outcomes.

**KEYWORDS**
delay discounting; measurement; Monetary Choice Questionnaire; ecological momentary assessment; substance use; substance abuse; questionnaire; validity; validation; measurement; monetary; reward; rewards; survey; mobile phone

**Introduction**

**Background**

Delay discounting quantifies a subject’s relative preference for more smaller, immediate rewards over larger, delayed rewards. Stated another way, delay discounting can be defined as the perceived value of a reward based on the temporal delay in the receipt of the reward. Greater delay discounting (ie, preference for smaller, short-term rewards over larger, long-term rewards) increases the risk for several health conditions, including obesity [3], gambling disorder [4], anxiety, depression [5], substance use disorder [6], and poor substance use disorder treatment response [7-10]. Although some research describes delay discounting as an immutable trait [11], findings from laboratory and clinical studies suggest delay discounting varies by place, time, and other contextual factors [12,13]. Delay discounting’s malleability may represent a promising avenue for interventions or treatment to reduce the risk for myriad negative health outcomes. In this study, we present a novel tool to rapidly assess delay discounting in remote studies and natural environments. Below, we describe traditional methods to measure delay discounting, the potential for ecological momentary assessment (EMA) to improve delay discounting measurement, and this study’s goals.

**Traditional Measurement of Delay Discounting**

Delay discounting rates quantify consequence devaluation (eg, a monetary reward’s devaluation) by delay [14]. One method to calculate delay discounting rates considers the devaluation process as a hyperbolic decay [15], as shown in equation 1.

\[
V = A / (1 + kD) \tag{1}
\]

\(V\) represents the indifference point or a small immediate reward’s value, \(A\) represents the long-term reward value, \(k\) represents the delay discounting rate, and \(D\) represents the long-term reward’s delay (ie, long-term reward’s waiting time). Short-term reward preference corresponds to a greater delay discounting rate (ie, future rewards’ steeper devaluation by delay), whereas long-term reward preference corresponds to a shallower delay discounting rate.

Researchers often measure delay discounting with the Monetary Choice Questionnaire (MCQ) [16-19]. To obtain the delay discounting rate, the MCQ asks participants to choose between hypothetical smaller, sooner monetary rewards and various larger, later rewards varied by magnitude and delay [16,18]. Researchers often use the 27-item MCQ due to its intuitive administration and straightforward calculations [16]. There are numerous MCQ variations, including a 21-item version [17] and a 9-item version [19].

Although many studies use the MCQ, a continuous indifference point measure may better capture delay discounting rates due to reduced ceiling effects and faster data saturation. The MCQ has been criticized as constraining delay discounting rates, resulting in ceiling effects in populations with high delay discounting rates [18]. For example, although the MCQ was originally tested with people who use heroin, the MCQ’s current versions limit comparisons across or within populations because people who use drugs can reach maximum delay discounting rates (ie, ceiling effects) [18]. Directly measuring indifference points along a continuous measure (eg, how long are you willing to wait to receive US $100 instead of US $30 today?) may prevent ceiling effects. Some studies have implemented a continuous measure for the indifference point [20], but few have used EMA-based continuous measures. A continuous indifference point measure increases the set of possible discounting rate estimates [20] and may capture the delay discounting rate with fewer items than the MCQ. Repeatedly administering the MCQ may fatigue participants [21], but a continuous indifference point measure may allow researchers to quickly reach data saturation (eg, obtaining consistent delay discounting rates over fewer repeated measures) [20].

**EMA of Delay Discounting**

EMA-based delay discounting research can measure behavior in context, abbreviate delay discounting scales, and minimize bias compared to more traditional data collection methods. First, EMA methods collect repeated measures on participants’ behaviors in situ [22] throughout the day, capturing ephemeral behaviors and moods traditional methods may miss or mismeasure [23]. Indeed, EMA may identify place and time cues that increase the risk for greater delay discounting (ie, preference for sooner, smaller rewards) and increased maladaptive behavior risk [24]. For example, one EMA study identified a relationship between delay discounting and acute substance use withdrawal [25]. The study found constant MCQ scores for the first few hours after substance use, but MCQ scores peaked 4-6 hours after alcohol and cannabis use and after 2 hours for stimulant use [25]. Second, repeated delay discounting assessments may shorten scales with good reliability. Fewer questions varied in presentation may reduce participant burden and fatigue [21] and straightlining (ie, giving the same answer to all questions) [26]. Other researchers have shortened existing delay discounting assessments to reduce participant burden [18], but EMA-based scales could further abbreviate scales with good reliability. Finally, EMA may have less bias than traditional data collection.
Less latency between the event occurrence and survey completion relative to traditional methods may reduce recall bias and measurement error [27]. Capturing delay discounting in situ may also decrease social desirability bias associated with stigmatized behaviors (eg, substance use) compared to face-to-face interviews [27].

The goals of this paper are 2-fold. First, we present and validate a novel, short, EMA-based delay discounting scale we developed. Second, we assess this tool’s ability to reproduce known associations between delay discounting and health behaviors (ie, substance use and craving) using a convenience-based sample.

Methods

Participants

We posted study advertisements on Facebook and Instagram via a Facebook profile. Phase 1 study advertisements appeared on mobile and desktop newsfeeds from December 2020 to February 2021. Study advertisements directed participants to “MetricWire” (MetricWire Inc), a digital research platform. Participants were instructed to download MetricWire’s iPhone- and Android-compatible smartphone app. We administered all surveys via MetricWire, a widely used research app designed for in-the-moment, contextual data collection. Study staff verified participants’ email addresses and phone numbers. Then, interested participants created a password-protected account and answered the screener in the app. Throughout the study, participants could anonymously contact study staff with questions or concerns through the app’s instant messaging (IM) system. Researchers emailed and invited phase 1 participants to join phase 2 in July 2021.

US residents aged 18 years or older with smartphones were eligible to participate. Eligible participants read study and consent materials through the MetricWire app. To ensure participants understood the study objectives, participants had 3 attempts to correctly answer 3 multiple-choice questions regarding the study’s purpose, length, and potential risks. Study incentives and anonymous recruitment risked individuals feigning their country of residence and reregistering under fake accounts. As a result, MetricWire removed participants with IP addresses already registered or abroad.

Ethical Considerations

This study received institutional review board approval from Johns Hopkins Bloomberg School of Public Health (IRB00011160). All participants provided informed consent and were informed they would receive US $10 for the completion of phase 1 and a maximum of US $75 for the completion of phase 2. The data analyzed were anonymous and deidentified.

Sample Size

For a diverse sample, we sought to equally recruit participants from 6 categories based on age (ie, 18-30, 31-50, and >50 years) and race (eg, White and non-White). In recruitment, we considered adults identifying as 2 or more races as non-White adults. We as aimed to recruit 33 adults for each quota, or 198 participants in total. These 198 participants would yield approximately 2821 observations (198 participants × [1 baseline survey + (6 days × 3 follow-up surveys)], assuming 75% compliance).

Study Design

We used a 2-phase EMA study via MetricWire’s smartphone app. In phase 1, participants completed the 8-item MCQ, a sociodemographic questionnaire, and our 2 novel 7-item continuous delay discounting surveys. Researchers invited participants who completed phase 1 to participate in phase 2 a few months later. Phase 2 comprised a survey on the first day (baseline) and 3 daily follow-up surveys on their smartphone for 6 consecutive days, amounting to 19 total surveys. Participants received the 3 follow-up surveys on their smartphones at random times between their self-reported wake and sleep times. In phase 2, the baseline survey included the original MCQ, health behavior assessments (eg, substance use and craving), and our 2 novel EMA-based delay discounting measures. At each daily follow-up, participants completed our EMA-based delay discounting tool and a health outcomes questionnaire (eg, past-hour substance use and craving). The phase 2 baseline survey duration approximated 20 minutes, and each follow-up survey approximated 5 minutes.

Participants received compensation based upon adherence. Phase 1 completion renumerated participants with US $10 credit for Tango, a third-party gift card provider. In phase 2, participants received US $20 Tango credit for baseline assessment completion, US $5 credit each day if they completed at least 1 follow-up survey, and US $25 bonus credit for completing at least 75% (14/19) of surveys upon the study’s end. Participants received links to their accrued credit through the app’s IM system. To improve adherence, text in the consent form encouraged participants to enable MetricWire notifications on their smartphones and to request technical support through the app’s IM system. Halfway through phase 2, study staff messaged participants their adherence rate and a reminder about the US $25 bonus credit.

Measures

Time- and Monetary-Selection Measures

We presented participants with 2 continuous delay discounting measures. For the first type, participants chose how long they would wait to receive the long-term reward rather than the short-term reward (eg, \( D \) from equation 1), hereafter referred to as time-selection items. In the second type, participants chose how much money they would need to receive today rather than the long-term reward (eg, \( V \) from equation 1), hereafter referred to as monetary-selection items. Textbox 1 outlines the time-selection and monetary-selection items.
Textbox 1. Modified ecological momentary assessment (EMA)–based time-selection and monetary-selection delay discounting items for a 2-phase EMA study assessing the validity of a novel delay discounting measure. Items were randomized during the study.

<table>
<thead>
<tr>
<th>Time-selection items (participants could choose between 0 and 52 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How long would you be willing to wait to get US $100 instead of US $50 today?</td>
</tr>
<tr>
<td>• How long would you be willing to wait to get US $100 instead of US $70 today?</td>
</tr>
<tr>
<td>• How long would you be willing to wait to get US $100 instead of US $10 today?</td>
</tr>
<tr>
<td>• How long would you be willing to wait to get US $100 instead of US $80 today?</td>
</tr>
<tr>
<td>• How long would you be willing to wait to get US $100 instead of US $40 today?</td>
</tr>
<tr>
<td>• How long would you be willing to wait to get US $100 instead of US $30 today?</td>
</tr>
<tr>
<td>• How long would you be willing to wait to get US $100 instead of US $99 today?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monetary-selection items (participants could choose between US $1 and US $99)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How much money would you take today instead of US $100 in a year?</td>
</tr>
<tr>
<td>• How much money would you take today instead of US $100 in 1 month?</td>
</tr>
<tr>
<td>• How much money today would you take today instead of US $100 in 6 months?</td>
</tr>
<tr>
<td>• How much money would you take today instead of US $100 in 3 months?</td>
</tr>
<tr>
<td>• How much money would you take today instead of US $100 in 2 weeks?</td>
</tr>
<tr>
<td>• How much money would you take today instead of US $100 in 1 week?</td>
</tr>
<tr>
<td>• How much money would you take today instead of US $100 tomorrow?</td>
</tr>
</tbody>
</table>

Figure 1 displays the time- and monetary-selection items’ in-app appearance. Participants used a sliding scale to select their wait time or monetary reward. Our novel measures comprised 7 time-selection and 7 monetary-selection items. Participants answered time- and monetary-selection items at phase 1, phase 2 baseline, and phase 2 follow-up. From the measured indifference point and long-term reward delay, the delay discounting rate ($k$) was directly calculated via equation 1, since we knew $V$ (the immediate reward’s value), $A$ (the delayed reward’s value), and $D$ (the time delay). The geometric mean across all items was then calculated to determine the average delay discounting rate for the time- and monetary-selection items.

Figure 1. Time-selection item (left) and a monetary-selection item (right) examples for a 2-phase ecological momentary assessment study assessing the validity of a novel delay discounting measure.
Original MCQ

The original MCQ was administered in both phase 1 and 2 baseline surveys to validate our time- and monetary-selection measures. Participants answered 7 “Would you rather receive US $50 today or US $100 in 1 year?” variants. The long-term reward consistently displayed US $100, but the time delay and short-term reward amount varied. The time delay ranged from 2 months to a year, while the short-term reward amount ranged from US $10 to US $99. See Multimedia Appendix 1 for all administered items.

Substance Use

To reproduce established associations between delay discounting and a health condition, we asked participants if they used or craved substances within the last hour during phase 2. Surveyed substances included (1) alcohol; (2) tobacco, cigarettes, cigarillos, cigars, vaping, and nicotine; (3) cannabis, marijuana, pot, grass, and hash; (4) cocaine, coke, and crack; (5) prescription stimulants (eg, Ritalin and Concerta); (6) methamphetamines (eg, speed, crystal meth, and ice); (7) inhalants (eg, nitrous oxide and glue); (8) sedatives of sleeping pills (eg, Valium); (9) hallucinogens (eg, lysergic acid diethylamide and acid); (10) street opioids (eg, heroin); (11) prescription opioids (eg, fentanyl); and (12) others. We created individual-level variables reflecting the total percentage of phase 2 surveys endorsing using or craving alcohol, cannabis, or tobacco. We did not examine other use or craving of any other substance with very low base rates.

Statistical Analysis

Overview

Prior to conducting analyses, we examined the time- and monetary-selection items’ distribution. Participants identified as outliers or who provided invalid responses (ie, always selecting maximum or minimum values) were dropped. Additionally, we examined the delay discounting distributions for violations of normality and the need for log transformations.

The study objectives included identifying noninformative items in our time- and monetary-selection scales, examining scale stability across phase 1 and 2 baselines, validating time- and monetary-selection items with the phase 1 and 2 baseline MCQ, and determining the minimally sufficient set of time- and monetary-selection items required to capture delay discounting. All analyses were performed in R (version 4.0.4; R Foundation for Statistical Computing).

Objective 1: Identify Noninformative Items Within New Measures

We examined Pearson correlations between the phase 1 geometric average delay discounting rate of the 7-item time-selection scale and each time-selection item’s delay discounting rate. This was repeated for the monetary-selection items in phase 1. Then, we repeated the item-scale correlations separately for time- and monetary-selection items in phase 2. Items uncorrelated with the full scale were then dropped from analyses.

Objective 2: Examine New Measures’ Test-Retest Reliability

For the second objective, we compared the delay discounting function from phase 1 baseline to phase 2 baseline via Pearson correlations.

Objective 3: Validate New Measures

For the third goal, we compared time- and monetary-selection scales to the study’s gold standard—the traditional MCQ—via Pearson correlations. The correlation between the delay discounting rate derived from phase 1’s traditional MCQ and the delay discounting rate derived from both the time- and monetary-selection tools at phase 1 and phase 2 was examined. The correlation between the delay discounting rate derived from phase 2’s traditional MCQ and the time- and monetary-selection tools at phase 2 was additionally assessed.

Objective 4: Shorten New Scale

For the fourth goal, we iteratively examined the Pearson correlation between the geometric average delay discounting rate of a randomly chosen set of the informative items and the full 7-item scale using phase 1 data.

Objective 5: Assess the Association Between New Delay Discounting Measures and Substance Use

Finally, we tested the association between delay discounting and substance use to assess our time- and monetary-selection measures’ predictive use. Linear regression analyses were conducted to examine associations between the predictors (ie, phase 2 baseline delay discounting rates calculated from our time- and monetary-selection measures) with 6 outcomes (ie, percentage of assessments participants reported using or craving alcohol, cannabis, or tobacco). Participant’s age, sex, and completed number of surveys were adjusted. We mean-centered continuous predictor variables and conducted analyses with completed surveys only.

Results

Overview

In phase 1, a total of 186 participants were recruited, of whom 111 agreed to participate in phase 2. To identify potential outliers, the delay discounting rate’s SD was calculated at each phase separately for the 7 time- and monetary-selection responses. The 4 resulting plots were then visually inspected. From the phase 1 monetary-selection geometric average delay discounting rate distribution, 6 data points were identified as outliers that were 3 SDs from the standardized mean. For phase 1 time-selection delay discounting items, individuals who consistently selected only the minimum time delay (1 day, n=2) or the maximum time delay (52 days, n=1) were removed. We excluded 8 total participants, noting them as outliers, from our phase 1 analytical sample. For phase 2 time-selection delay discounting items, 3 individuals were identified as outliers that were 3 SDs above the standardized mean—one individual consistently selected the minimum time delay (1 day) and 2 individuals consistently selected the maximum time delay (52 days). For phase 2 monetary-selection delay discounting items, 5 individuals were excluded as outliers that were 3 SDs above
the standardized mean. The phase 2 raw sample included 9 outliers in total. The 8 participants from phase 1 and 9 participants from phase 2 were excluded from the analyses. Our analytic sample included individuals with valid phase 1 and phase 2 data, resulting in a sample of 97 participants (Figure 2).

Figure 2. Analytical sample flowchart for a 2-phase ecological momentary assessment (EMA) study assessing the validity of a novel delay discounting measure. Participants were recruited from Facebook and Instagram to participate in a mobile-based EMA study.

There were some differences in demographic characteristics between those who completed phase 1 only versus those in our analytic sample who completed phases 1 and 2. In particular, there was a greater proportion of individuals who identified as African American or Black who completed phase 1 only (33/89, 37%) compared to those who completed phases 1 and 2 (9/97, 9%; \( \chi^2 = 21.4; P < .001 \)). A greater proportion of individuals identified as White (55/97, 57%), Asian (25/97, 26%), or 2 or more races (33/97, 34%) who completed phases 1 and 2 relative to phase 1 only (White: 39/89, 44%; Asian: 12/86, 14%; 2 or more races: 8/97, 8%). There were no differences in terms of the completion of the study based on participant sex, ethnicity, or income.

Among participants who completed phase 1 and phase 2 (ie, our analytic sample; N=97), most participants identified as female (n=72, 74%) and as White (n=56, 58%). Approximately 80% (n=79) of participants were 50 years and younger of age, and 70% (n=68) reported an annual income below US $75,000 (Table 1). On average, participants completed 16 (SD 4) of 19 surveys. On average, 4% (SD 9%; n=0.8) of a given participant’s surveys endorsed alcohol use, 6% (SD 21%; n=1.22) endorsed tobacco use, 7% (SD 22%; n=0.004) endorsed craving tobacco, and 8% (SD 18%; n=1.60) endorsed craving alcohol.
Table 1. Characteristics of US adults participating in a novel 2-phase ecological momentary assessment–based study assessing the validity of a novel delay discounting tool, 2020-2021 (N=97).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>25 (26)</td>
</tr>
<tr>
<td>Female</td>
<td>72 (74)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>24 (25)</td>
</tr>
<tr>
<td>Black</td>
<td>9 (9)</td>
</tr>
<tr>
<td>White</td>
<td>56 (58)</td>
</tr>
<tr>
<td>2 or more races</td>
<td>8 (8)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>28 (29)</td>
</tr>
<tr>
<td>26-35</td>
<td>25 (26)</td>
</tr>
<tr>
<td>36-50</td>
<td>26 (27)</td>
</tr>
<tr>
<td>51-70</td>
<td>16 (17)</td>
</tr>
<tr>
<td>&gt;71</td>
<td>2 (2)</td>
</tr>
<tr>
<td><strong>Income (US $), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;$30,000</td>
<td>30 (31)</td>
</tr>
<tr>
<td>$30,000-$49,999</td>
<td>19 (20)</td>
</tr>
<tr>
<td>$50,000-$74,999</td>
<td>19 (20)</td>
</tr>
<tr>
<td>$75,000-$100,000</td>
<td>19 (20)</td>
</tr>
<tr>
<td>&gt;$100,000</td>
<td>10 (10)</td>
</tr>
<tr>
<td><strong>Number of surveys completed per person</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>16 (4)</td>
</tr>
<tr>
<td>Range</td>
<td>1-19</td>
</tr>
<tr>
<td><strong>Percentage of surveys reflecting substance use or craving</strong></td>
<td></td>
</tr>
<tr>
<td>Alcohol use</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Range</td>
<td>0-53</td>
</tr>
<tr>
<td>Cannabis use</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>4 (14)</td>
</tr>
<tr>
<td>Range</td>
<td>0-73</td>
</tr>
<tr>
<td>Tobacco use</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>6 (21)</td>
</tr>
<tr>
<td>Range</td>
<td>0-100</td>
</tr>
<tr>
<td>Alcohol craving</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>8 (18)</td>
</tr>
<tr>
<td>Range</td>
<td>0-100</td>
</tr>
<tr>
<td>Cannabis craving</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>8 (20)</td>
</tr>
<tr>
<td>Range</td>
<td>0-100</td>
</tr>
<tr>
<td>Tobacco craving</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>7 (22)</td>
</tr>
</tbody>
</table>
The delay discounting rate distribution generated from the time- and monetary-selection items and the original MCQ, which manifested a positively skewed distribution, was inspected. Thus, the delay discounting rates were log-transformed to obtain normally distributed variables for analyses examining the relationship between delay discounting and other measured characteristics. Below, we present findings from the study’s 5 objectives.

**Objective 1: Identify Noninformative Items Within New Measures**

The Pearson correlations between each item’s delay discounting rate and the 7-item scale’s geometric average delay discounting rate were examined. In phase 1, we found high correlations ($r$ range=0.89-0.96) for 6 of the 7 time-selection items (Multimedia Appendix 2). The poorly performing item ($r=0.38$ in phase 1) asked participants to choose between US $99 today or US $100 in the future. The small difference in the immediate versus delayed reward may have yielded poor discrimination between individuals with high versus low discounting. In subsequent analyses, this item was dropped. We reproduced the finding in phase 2 (Multimedia Appendix 3)—6 of the items strongly correlated with the scale ($r$ range=0.88-0.94), and phase 1’s poor-performing item continued to perform poorly in phase 2 ($r=0.35$ in phase 2). In phase 1, a relatively low correlation was found ($r$ range=0.45-0.85) between individual money-selection items and the 7-item scale’s geometric average delay discounting rate (Multimedia Appendix 4). Moreover, the item-to-scale correlations were not consistent between phase 1 and phase 2, and only 3 items were consistently correlated ($r$>0.75) across phases 1 and 2 (Multimedia Appendix 5).

**Objective 2: Examine New Measures’ Test-Retest Reliability**

Delay discounting rates from phase 1 to phase 2 for both the novel time- and monetary-selection measures were compared. We found moderate to high correlations between phase 1 and 2 time-selection ($r=0.66; P<.001$) and money-selection delay discounting ($r=0.41; P<.001$), indicating good and acceptable test-retest reliability, respectively (Table 2).

### Table 2. Correlations of log-transformed delay discounting variables across 2 phases of a mobile-based, ecological momentary assessment study assessing the validity of a novel delay discounting tool between 2020 and 2021 (N=97).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Phase 1 MCQ</th>
<th>Phase 1 time-selection delay discounting variable</th>
<th>Phase 1 monetary-selection delay discounting variable</th>
<th>Phase 2 MCQ</th>
<th>Phase 2 time-selection delay discounting variable</th>
<th>Phase 2 monetary-selection delay discounting variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>1</td>
<td>0.76</td>
<td>0.57</td>
<td>0.51</td>
<td>0.45</td>
<td>0.22</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.03</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Phase 1 time-selection delay discounting variable</td>
<td>$r$</td>
<td>0.76</td>
<td>1</td>
<td>0.66</td>
<td>0.65</td>
<td>0.71</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Phase 1 monetary-selection delay discounting variable</td>
<td>$r$</td>
<td>0.57</td>
<td>0.66</td>
<td>1</td>
<td>0.42</td>
<td>0.49</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Phase 2 MCQ</td>
<td>$r$</td>
<td>0.51</td>
<td>0.65</td>
<td>0.42</td>
<td>1</td>
<td>0.72</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Phase 2 time-selection delay discounting variable</td>
<td>$r$</td>
<td>0.45</td>
<td>0.71</td>
<td>0.49</td>
<td>0.72</td>
<td>1</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Phase 2 monetary-selection delay discounting variable</td>
<td>$r$</td>
<td>0.22</td>
<td>0.38</td>
<td>0.42</td>
<td>0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.03</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aMCQ: Monetary Choice Questionnaire.

bNot available.
Objective 3: Validate New Measures

The correlations between the traditional MCQ and the novel time- and monetary-selection approaches were examined (Table 2). Excellent and acceptable correlations were observed between the phase 1 traditional MCQ-derived delay discounting rate with the phase 1 time-selection delay discounting rate ($r=0.76; P<.001$; Table 2 and Figure 3) and the phase 2 time-selection delay discounting rate ($r=0.45; P<.001$), respectively. Moderate to small positive correlations were noted between the phase 1 MCQ-derived delay discounting rate with phase 1 monetary-selection delay discounting rate ($r=0.57; P<.001$; Figure 4) and phase 2 money-selection delay discounting rate ($r=0.22; P=.03$). The phase 2 traditional MCQ delay discounting rate was highly positively associated with the phase 2 time-selection ($r=0.72; P<.001$) and moderately positively associated with the monetary-selection delay discounting rate ($r=0.50; P<.001$).

Figure 3. Correlations between phase 1 MCQ log(delay discounting rate) and phase 1 six-item time-selection log(delay discounting rate). Data are from a mobile-based ecological momentary assessment study assessing the validity of a novel delay discounting tool across 2 phases between 2020 and 2021. Participants were US adults recruited from social media. An example of a time-selection item is “How long would you be willing to wait to get US $100 instead of US $50 today?” MCQ: Monetary Choice Questionnaire.
Figure 4. Correlations between phase 1 MCQ log(delay discounting rate) and phase 1 seven-item monetary-selection log(delay discounting rate). Data are from a mobile-based ecological momentary assessment study assessing the validity of a novel delay discounting tool across 2 phases between 2020 and 2021. Participants were US adults recruited from social media. An example of a monetary-selection item is "How much money would you take today instead of US $100 in a year?" MCQ: Monetary Choice Questionnaire.

Objective 4: Shorten New Scale
We also examined if an abbreviated scale could approximate the full 6-item scale of the time-selection approach. To construct the abbreviated scale, we randomly selected N (between 1 and 5) items across 100 iterations at each N and calculated the correlation between the N item’s geometric average delay discounting rate and the full 7-item time-selection scale’s geometric average delay discounting rate. A high correlation was found between randomly selected 1 ($r=0.94$) or 2 ($r=0.97$) item scales and the full 6-item scale, indicating a relatively small number of items can approximate the full scale (Figure 5). The analyses were repeated with the monetary-selection approach, and a relatively weak correlation was found between randomly selected 1 ($r=0.74$) or 2 ($r=0.86$) item scales and the full 7-item scale (Figure 6).
Objective 5: Assess the Association Between New Delay Discounting Measures and Substance Use

Participants with greater phase 2 baseline delay discounting rates from the MCQ and our time-selection measure had a higher proportion of surveys endorsing tobacco craving and use. There were trend-level significant associations between the original MCQ delay discounting rate (adjusted mean difference=3.46, 95% CI –0.63 to 7.55; \(P=0.10\)) and the time-selection delay discounting rate with tobacco craving (adjusted mean difference=3.92, 95% CI –0.12 to 7.95; \(P=0.06\); Table 3). Significant associations were observed between the original
MCQ delay discounting rate (adjusted mean difference=4.32, 95% CI 0.29–8.34; \( P = .04 \)) and time-selection delay discounting rate (adjusted mean difference=5.89, 95% CI 1.99–9.79; \( P = .003 \)) with the percentage of surveys endorsing tobacco use. No significant associations were observed between the phase 2 baseline monetary-selection delay discounting rate and substance use or craving (all \( P > .05 \)). There were no significant differences between alcohol or cannabis use or craving and the delay discounting rate derived from both the MCQ and our novel measures (all \( P > .05 \)).

Table 3. Associations among phase 2 log-transformed delay discounting measures with the percentage of surveys endorsing alcohol, cannabis, or tobacco use or craving (N=97).\(^a\)

<table>
<thead>
<tr>
<th>Measures and variable</th>
<th>Adjusted mean difference (95% CI)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase 2 baseline MCQ(^c)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of alcohol use</td>
<td>1.00 (–0.75 to 2.75)</td>
<td>.26</td>
</tr>
<tr>
<td>Percentage of cannabis use</td>
<td>0.20 (–2.80 to 3.20)</td>
<td>.89</td>
</tr>
<tr>
<td>Percentage of tobacco use</td>
<td>4.32 (0.29 to 8.34)</td>
<td>.04</td>
</tr>
<tr>
<td>Percentage of alcohol craving</td>
<td>2.73 (–0.84 to 6.31)</td>
<td>.13</td>
</tr>
<tr>
<td>Percentage of cannabis craving</td>
<td>1.02 (–3.16 to 5.20)</td>
<td>.63</td>
</tr>
<tr>
<td>Percentage of tobacco craving</td>
<td>3.46 (–0.63 to 7.55)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Phase 2 baseline time-selection delay discounting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of alcohol use</td>
<td>0.64 (–1.10 to 2.38)</td>
<td>.47</td>
</tr>
<tr>
<td>Percentage of cannabis use</td>
<td>–1.07 (–4.04 to 1.89)</td>
<td>.47</td>
</tr>
<tr>
<td>Percentage of tobacco use</td>
<td>5.89 (1.99 to 9.79)</td>
<td>.003</td>
</tr>
<tr>
<td>Percentage of alcohol craving</td>
<td>2.65 (–0.89 to 6.20)</td>
<td>.14</td>
</tr>
<tr>
<td>Percentage of cannabis craving</td>
<td>–0.36 (–4.51 to 3.78)</td>
<td>.86</td>
</tr>
<tr>
<td>Percentage of tobacco craving</td>
<td>3.92 (–0.12 to 7.95)</td>
<td>.06</td>
</tr>
<tr>
<td><strong>Phase 2 baseline monetary-selection delay discounting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of alcohol use</td>
<td>0.97 (–0.27 to 2.22)</td>
<td>.13</td>
</tr>
<tr>
<td>Percentage of cannabis use</td>
<td>–1.17 (–3.30 to 0.97)</td>
<td>.28</td>
</tr>
<tr>
<td>Percentage of tobacco use</td>
<td>–0.62 (–3.57 to 2.32)</td>
<td>.68</td>
</tr>
<tr>
<td>Percentage of alcohol craving</td>
<td>1.78 (–0.79 to 4.34)</td>
<td>.17</td>
</tr>
<tr>
<td>Percentage of cannabis craving</td>
<td>–1.29 (–4.27 to 1.69)</td>
<td>.39</td>
</tr>
<tr>
<td>Percentage of tobacco craving</td>
<td>–1.07 (–4.03 to 1.90)</td>
<td>.48</td>
</tr>
</tbody>
</table>

\(^a\)Data are from a mobile-based ecological momentary assessment study assessing the validity of a novel delay discounting tool between 2020 and 2021. Participants were US adults recruited from social media. An example of a time-selection item is “How long would you be willing to wait to get US $100 instead of US $50 today?” An example of a monetary-selection item is “How much money would you take today instead of US $100 in a year?”

\(^b\)Linear regression models adjusted for number of completed surveys and participant sex and age. Delay discounting rates are log-transformed.

\(^c\)MCQ: Monetary Choice Questionnaire.

**Discussion**

**Principal Findings**

In this study, a novel EMA-based tool was developed to overcome the limitations of existing delay discounting measures. In contrast to traditional delay discounting measures captured at a single time point (eg, the MCQ), our EMA-based delay discounting measure used a continuous indifference point to minimize ceiling effects, abbreviate delay discounting scales, capture state-like fluctuations in delay discounting, and reduce measurement error. We showed that a time-selection, EMA-based method could accurately assess delay discounting. In addition, the novel tool successfully reproduced some associations between delay discounting and substance use behaviors observed in the literature, specifically the established association between delay discounting and tobacco use.

To develop the EMA-based scale, we analyzed the scales’ reliability across 2 phases and its validity in comparison to the MCQ. In total, 6 of the 7 time-selection items correlated with the full scale. We dropped the uninformative, 7th time-selection item—“How long would you be willing to wait to get US $100 instead of US $99 today?” The US $1 difference may have failed to provide sufficient response variability, because even short delays (eg, 1 week) would result in devaluing US $100 over US $99 [28]. Overall, the time-selection EMA measure reliably and validly measured delay discounting. However, the monetary-selection items did not consistently contribute to delay discounting measurement. There are a
number of reasons to explain this lack of performance of this measure. First, we believe participants simply may have struggled to decide which amount they would take today over the longer-term reward. Second, the monetary-selection measure comprised fuzzy units (eg, cost of one’s patience for a specific period), and studies have shown fuzzy units increase short-term reward preferences compared to discrete units (eg, how long one would wait) [29,30]. Participants may not have as much behavioral experience to quantify the abstract monetary cost to wait for an already specified duration. Participants, however, may rely on personal experience and past behaviors to assess their capacity for patience, as conveyed in the time-selection items. Third, framing time as a date (as in the time-selection items) or as days (as in the monetary-selection items) may have influenced participants’ delay discounting rates [30]. Finally, monetary-selection items may have needed a larger, long-term reward to capture variability in long-term reward devaluation. When asked to wait a year to receive US $100, participants choose a short-term reward as close as possible to the long-term reward, but if given a larger, long-term reward, participants may choose a smaller, short-term reward relative to the long-term reward [31].

After assessing the measures’ validity and reliability, we determined that randomly selecting 1 to 2 time-selection items could sufficiently capture delay discounting. Similarly, another study also captured delayed discounting with 2 items with no sensitivity loss [32]. An abbreviated EMA-based scale can minimize participant fatigue and sustain participant attention during data collection.

Additionally, our tool successfully reproduced some established associations between delay discounting and substance use or craving. Based on the time-selection items and the original MCQ, tobacco use and cravings often increased with greater delay discounting. Other studies have similarly evidenced a preference for smaller, immediate rewards over larger, long-term rewards (ie, greater delay discounting) when involving nicotine use [33,34]. On the other hand, we found no relationship between delay discounting and alcohol or cannabis use or craving. Other studies have found positive, modest associations between delay discounting and alcohol and marijuana use and craving in other samples [35-37]. Similarly, a recent meta-analysis showed a small association between delay discounting and cannabis use [38].

**Limitations**

Selection and response biases limit the generalizability of our study’s findings for a number of reasons. First, our study sample differs from our target population with regard to age, race, and sex. We created age and race recruitment quotas to avoid immigrative selection bias, but we ultimately did not meet the quota for older adults identifying as non-White, and we unintentionally oversampled female participants. Other studies similarly had greater ease recruiting White older adults over non-White older adults on social media [39,40]. This study, however, successfully recruited participants across income levels, which may improve generalizability to more socioeconomically diverse populations. Our sample also exceeded other measurement development studies’ sample sizes [41]. Second, few participants endorsed substance use. Our study advertisements did not target individuals who use substances, which likely contributed to low substance use endorsement in our sample and limited ability to identify associations between delay discounting and alcohol and cannabis use and craving. Our advertisement strategy, however, demonstrated our measures’ use for the general population and minimized participant self-selection by substance use frequency. Moreover, we did not incentivize substance use reporting, which may have failed to counter social desirability bias and underreporting, but our strategy improved our confidence in minimal false reports [42]. Finally, our analyses examining the association between delay discounting and substance use and craving did not account for temporality, limiting our ability to draw causal inferences regarding the directionality of effects. Future studies should examine the tool’s use in more specific health populations, such as individuals who use substances to excess or engage in other addictive behaviors, and examine the causal relationship between the new tool and substance use.

**Conclusions**

We found that 1 or 2 randomly selected items from our novel EMA-based time-selection measure can sufficiently assess delay discounting. Our abbreviated EMA-based scale may overcome data collection barriers related to participants’ attentional capacity and measurement barriers due to ceiling effects [18]. Beyond aiding delay discounting–specific research, our transdiagnostic tool may help with intervention assessment and rapid detection of individuals at risk for specific health conditions. In terms of intervention assessment, delay discounting may serve as an outcome for measuring the effectiveness and efficacy of behavior change interventions. For example, an intervention targeting binge eating may seek to assess the intervention’s momentary effect on how an individual values the immediate reward of binge eating over the long-term health benefits of abstaining from binge eating. In this case, delay discounting may serve as a rapid proxy to assess the intervention’s effect on reward valuation. Additionally, our tool may also help detect individuals at higher risk in a high-risk population associated with greater impulsivity. For example, one study suggested delay discounting tools may be adapted to discern HIV risk among individuals with and without cocaine use disorder [43]. Future studies should test the time-selection measures’ predictive use in clinical, high-risk populations.

**Acknowledgments**

This work was supported by R01 DA047064 (to GDK and BSM). AL was also supported by a NIDA T32 DA007292.
Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
GDK and BSM conceptualized the study. GDK, JAR, JT, and BSM designed the time- and monetary-selection items. AL, JAR, and BSM recruited participants and assisted participants with compensation and study questions. JAR, JLW, DWS, and BSM conducted and aided in the interpretation of statistical analyses. AL drafted the initial paper, which all authors reviewed and edited.

Conflicts of Interest
JCS has received research-related funding from Canopy Growth Corporation and DynamiCare Health Inc in the past 3 years.

Multimedia Appendix 1
Monetary Choice Questionnaire–administered items. [DOCX File, 15 KB - formative_v81ie48954_app1.docx ]

Multimedia Appendix 2
Phase 1 time-selection per-item log(delay discounting rate) correlation with full 7-item log(delay discounting rate). [PNG File, 10 KB - formative_v81ie48954_app2.png ]

Multimedia Appendix 3
Phase 2 time-selection per-item log(delay discounting rate) correlation with full 7-item log(delay discounting rate). [PNG File, 10 KB - formative_v81ie48954_app3.png ]

Multimedia Appendix 4
Phase 1 monetary-selection per-item log(delay discounting rate) correlation with full 7-item log(delay discounting rate). [PNG File, 10 KB - formative_v81ie48954_app4.png ]

Multimedia Appendix 5
Phase 2 monetary-selection per-item log(delay discounting rate) correlation with full 7-item log(delay discounting rate). [PNG File, 10 KB - formative_v81ie48954_app5.png ]

References


Abbreviations

- EMA: ecological momentary assessment
- IM: instant messaging
- MCQ: Monetary Choice Questionnaire

©Amanda Luken, Jill A Rabinowitz, Jonathan L Wells, David W Sosnowski, Justin C Strickland, Johannes Thrul, Gregory D Kirk, Brion S Maher. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e48954

JMIR Form Res 2024 | vol. 8 | e48954 | p.2511 (page number not for citation purposes)

Edited by A Mavragani; submitted 12.05.23; peer-reviewed by H Guyer, L Narbey; comments to author 05.12.23; revised version received 15.12.23; accepted 02.01.24; published 27.02.24.
The Influence of Human Connections and Collaboration on Research Grant Success at Various Career Stages: Regression Analysis

Akiko Hashiguchi1, PhD; Makoto Asashima2, PhD; Satoru Takahashi1, MD, PhD

1Institute of Medicine, University of Tsukuba, Tsukuba, Japan
2Advanced Comprehensive Research Organization (ACRO), Teikyo University, Itabashi, Japan

Corresponding Author:
Akiko Hashiguchi, PhD
Institute of Medicine
University of Tsukuba
1-1-1 Tennodai
Tsukuba, 305-8575
Japan
Phone: 81 298537276
Email: hashiguchi.akiko.ge@u.tsukuba.ac.jp

Related Article:
This is a corrected version. See correction statement: https://formative.jmir.org/2024/1/e58733

Abstract

Background: Documenting the grant acquisition characteristics of a highly selective group of researchers could provide insights into the research and faculty development of talented individuals, and the insights gained to foster such researchers will help university management strengthen their research capacity.

Objective: This study examines the role of human connections in the success of biomedical researchers in Japanese universities.

Methods: This study used grant data from the Grants-in-Aid for Scientific Research (GIA) program, the largest competitive research funding program in Japan, to collect information on projects and their implementation systems obtained throughout the participants’ careers. Grant success was measured by the number and amounts of the awards obtained while participants occupied the role of principal investigator. Human connections were quantified by the number of projects in which the participants took part as members and were classified by their relationship with the project leader. Data were matched with information on career history, publication performance, and experience of the participants with government-funded programs apart from GIA and were analyzed using univariate and multivariate regression analyses.

Results: Early-career interpersonal relationships, as measured using the h-index value of the researchers who provided the participants with their initial experience as project members, had a positive effect on grant success. The experience of contributing to prestigious research programs led by top researchers dramatically increased the cumulative amount of GIA awards received by the participants over time. Univariate logistic regression analyses revealed that more interactions with upper-level researchers resulted in fewer acquisitions of large programs (odds ratio [OR] 0.67, 95% CI 0.50-0.89). Collaboration with peers increased the success rate of ≥2 research grants in large programs in situations in which both the participant and project leader were professors (OR 1.16, 95% CI 1.06-1.26). Tracking the process of research development, we found that collaboration during the periods of 10 to 14 years and 15 to 19 years after completing a doctorate degree determined the size of the project that the participant would obtain—interactions with peer researchers and subordinates during the 10- to 14-year postdegree period had positive effects on ≥2 large-program acquisitions (OR 1.51, 95% CI 1.09-2.09 and OR 1.31, 95% CI 1.10-1.57, respectively), whereas interactions with subordinates during the 15- to 19-year postdegree period also had positive effects (OR 1.25, 95% CI 1.06-1.47). Furthermore, relationships that remained narrowly focused resulted in limited grant success for small programs.

Conclusions: Human networking is important for improving an individual’s ability to obtain external funding. The results emphasize the importance of having a high-h-indexed collaborator to obtain quality information early in one’s career; working
with diverse, nonsupervisory personnel at the midcareer stage; and engaging in synergistic collaborations upon establishing a research area in which one can take more initiatives.

(JMIR Form Res 2024;8:e49905) doi:10.2196/49905

KEYWORDS

biomedical researchers; grant success; human connection; peer researchers; synergistic collaborations; research development

Introduction

Background

The university sector has played a major role in the production of scientific knowledge by facilitating research based on the unique conception of researchers and has supported important societal benefits by helping companies source innovation [1]. As the world becomes increasingly unstable, the university sector will be required to create more knowledge and supply more human resources, leading to breakthrough innovation. However, as knowledge accumulates and science becomes more specialized, fewer remarkable discoveries are made [2]. In addition, conducting research that answers simple but major questions is becoming difficult under the recent system of university performance evaluation, which emphasizes short-term outcomes against a backdrop of the demand for accountability in the allocation of public funds [3,4]. This emphasis on short-term results is significant from the perspective of university management in countries where public funds are the primary source of university revenue. For example, Japanese universities need to equip faculty members with strong research skills to receive more government resource allocation, especially after 2019, when performance-based allocations were introduced [5].

High competitiveness in research is measured by the number and impact of publications and the acquisition of external funding; the latter is particularly important from the perspective of university management. Importantly, in the long run, the ability to draw a vision will influence the future direction of the research field and attract young researchers. This innovative power—that is, creating a common understanding that a research field that did not previously exist will exist in the future—is expected to materialize as large research projects supported by significant funding. Therefore, it is important to understand how personnel are capable of obtaining vast amounts of external funding. By examining the existing funding framework in Japan, this study observed that the larger the funding program, the more the disciplines are grouped to form review committees. Therefore, project leaders must convince reviewers from diverse fields of the significance of a project [6]. Project leaders of a group of researchers spanning multiple disciplines are required to be capable of discussing the worldview that the research can present.

Factors such as the exchange of ideas and expansion of knowledge are important for fostering visionary researchers. Accordingly, the effectiveness of international mobility and academic industry collaboration has been evaluated from the perspectives of higher education, research, and innovation policy [7-9]. In addition, the literature has emphasized the importance of “productive interactions” between researchers and social actors in studies that assess the social impact of scientific research [10,11]. These examples demonstrate that learning from others regarding perspectives beyond the scope of one’s research can lead to innovative ideas. However, few studies have focused on the impact of such “productive exchanges” on the human resource development of individual researchers.

Chan et al [12] analyzed the coauthorship patterns of Nobel laureates and found that encounters of heterogeneous ideas from different researchers, which occur early in the collaboration life cycle, generate the most innovative ideas that emerge from the collaborative relationship. This cross-fertilization of ideas is considered effective not only in the research development process of Nobel laureates but also in the process of cultivating creativity among researchers in general. Studies on creativity have shown that acting as a broker of valuable knowledge (i.e., having a high level of “betweenness centrality” in a network) can facilitate information flow so as to generate new ideas [13].

A study of young researchers in biomedical sciences showed that effective collaboration with “nonsupervisor” peers is important for learning [14]. In the academic community, where the culture of apprenticeship is strong, handing over a research theme between a supervisor and an apprentice and taking over a network of people working on such a theme are effective methods for increasing one’s visibility in the field. In this context, how researchers relate not only to their immediate supervisors but also to their colleagues and researchers in other institutions is key to improving their competitiveness in grant acquisition, as previously demonstrated using selective groups of researchers with high betweenness centrality [15].

Objectives

This study explored the objective metrics of success in grant acquisition using information from Grants-in-Aid for Scientific Research (GIA) projects and focusing on interpersonal relationships. By doing so, we aimed to identify the requirements for developing researchers capable of winning external funding for university management and who, ultimately, can influence research trends through innovative ideas and high-level capabilities in project realization.

Methods

Overview

In Japan, GIA provides fundamental financial assistance for academic research activities that cover all fields with the objective of promoting scholarship and advancing creative research [6]. It is the largest competitive research funding program in Japan, which began in the 1950s. Even after the introduction of government funding programs for various purposes, it continues to account for most funds that support “bottom-up” researcher-led projects [16,17]. GIA includes basic

https://formative.jmir.org/2024/1/e49905
programs (classified according to the amount awarded), innovation programs that take on new challenges, and specially promoted research programs that support outstanding original research that pioneers new scholarship. Other programs are also available to young researchers. Screening of GIA projects through peer review, although the breadth of the reviewers’ fields varies according to the program size [5]. Therefore, GIA award acquisition performance is considered an indicator that quantifies the academic creativity of researchers who have reached a certain level regardless of external conditions such as top-down policy requirements. This study focused on the implementation structure and member composition of projects supported by GIA.

Data Sources
The grant acquisition history of the participants, that is, the research projects in which they were involved, was obtained from the GIA database, wherein participants’ names can be used as search terms to obtain information on their positive GIA awards (National Institute of Informatics, Japan) [18]. Although the GIA system has undergone several modifications and programs have been established or abolished over time, we focused on programs that most researchers currently in professorial positions are familiar with. Information on rejected projects was not included in this study.

Data on degree types were obtained from the doctoral dissertation database [19]. These data were introduced as an alternate indicator for understanding the effect of age as the exact age of each researcher is not publicly available. This indicator discriminates whether a degree was obtained by either enrolling in a graduate school and submitting a dissertation or submitting a dissertation and passing an examination even if the person was not enrolled in a graduate school (doctorate obtained by thesis). The latter is a unique system in Japan, and the degrees of this type are primarily awarded to people who conducted their research as employees of a company.

The productivity, contributions, and research impact of each researcher, as measured by the citation counts, were calculated from the publication history data obtained from the SciVal database under the license of the University of Tsukuba using the Scopus IDs as the search key (Elsevier) [20]. Factors that could not be determined through the aforementioned indicators were examined using a database that covered most government funding programs by entering the participants’ names (BIOIMPACT, Japan) [21].

Ethical Considerations
This study did not require ethics approval as it did not collect data from human or animal participants. The names of the professors, external funding obtained, and publication records are publicly available. The 2 databases of grants were constructed to ensure transparency in public research funding allocation and allow users to search for the latest research information in Japan—one records the GIA program of the Ministry of Education, Culture, Sports, Science, and Technology of Japan [22] and the other records other government funding programs of the Cabinet Office; Ministry of Internal Affairs and Communications; Ministry of Health, Labour, and Welfare; Ministry of Agriculture, Forestry, and Fisheries; Ministry of Economy, Trade, and Industry; Ministry of Land, Infrastructure, Transport, and Tourism; Ministry of the Environment; and Ministry of Defense [23]. The analysis plan was not preregistered as this was a secondary analysis of open data extracted from public databases.

Researcher Characteristics
This study considered the following characteristics in terms of the researchers’ competitiveness: the sum of the maximum allocated amount of projects obtained as principal investigators, total allocation earned as a principal investigator during the first 10 years after obtaining a PhD, and the number of projects obtained as principal investigators in large programs, such as Grant-in-Aid for Specially Promoted Research and Grant-in-Aid for Scientific Research (S), or small programs, such as Grant-in-Aid for Scientific Research (C). In terms of their personal attributes, we considered the number of years since obtaining a doctoral degree, whether the doctorate was obtained by thesis, sex, experience in nonuniversity institutions, the rankings of the university in which the researcher obtained their doctorate, and where the researcher is currently affiliated. To measure each researcher’s productivity, we considered the number of papers published, first-authored papers, and last-authored papers.

To clarify the effect of human connections, we included the following attributes: the number of co-researchers connected with through GIA projects throughout their career, the betweenness centrality score based on people-to-people connections in GIA projects, the h-index of the researcher with whom the participant researcher first became a member of GIA projects, the number of government-funded programs participated in as a project member, and the total h-indexes of the researchers with whom the participant researcher interacted through other government-funded programs. To measure the researchers’ interpersonal relationships, we considered the number of researchers who designated the participants as their GIA project members (upper level, peer, or subordinate) throughout their careers. We also considered the number of projects a researcher participated in during each period after obtaining a doctorate (10-14 y and 15-19 y after the doctorate), categorized by the relationship of the project leader to the participant (upper level, peer, or subordinate).

Samples and Analysis Method

Analysis 1
This study focused on a single area of biomedical science to obtain detailed microstructural data on the relationship between research implementation and subsequent improvements in research conception skills. Biochemistry was selected as the target field as it is the basis of today’s medicine given its focus on molecular mechanisms and as the 54 professors in this field included 5 (9%) awardees of Grant-in-Aid for Specially Promoted Research or Grant-in-Aid for Scientific Research (C) and 16 (30%) awardees of Grant-in-Aid for Scientific Research (A). These numbers exceed those in the field of hematology, with 3% (1/35) of awardees of Grant-in-Aid for Specially Promoted Research or Grant-in-Aid for Scientific Research (S) and 9%...
(3/35) of awardees for Grant-in-Aid for Scientific Research (A), and dermatology, with 5% (2/39) of awardees for Grant-in-Aid for Specially Promoted Research or Grant-in-Aid for Scientific Research (S) and 18% (7/39) of awardees for Grant-in-Aid for Scientific Research (A). This indicates that biochemistry is a suitable field for analyzing the competitiveness of researchers in Japan.

The names of the researchers who held professorships in biochemistry at the medical schools of Japanese universities between 2020 and 2022 were extracted from the websites of each university. In total, the study identified 54 researchers, including 2 (4%) female researchers, from 11 schools belonging to the top university group and other national universities. Academic and professional histories were obtained from their websites. In terms of education, 72% (39/54) graduated from schools of medicine, 87% (47/54) obtained PhDs in medicine, and 69% (37/54) obtained their bachelor’s and doctorate degrees in the same university. The mean number of years since receiving the degree was 29 (SD 6). We extracted 1473 GIA projects, of which 803 (54.51%) were implemented by the 54 participants as principal investigators (Table 1). For analysis 1, we summed the maximum allocated amount that each researcher won as the principal investigator and used it as the dependent variable. The independent variables were calculated using data obtained from each source. We controlled for sex (female) and experience at nonuniversity institutions (nonuniversity institutions). In this analysis, nonuniversity institutions refers to corporations and their associations. A lagged variable was designated for the total allocation of projects won by the participants as principal investigators for 10 years after each researcher obtained a PhD. Finally, researchers with missing data were excluded, and 52 were included in the analysis. Multiple regression analysis was performed using the XLSTAT statistical software (Addinsoft).

### Table 1. Data on the researcher population used in the study and their characteristics.

<table>
<thead>
<tr>
<th>Analysis 1</th>
<th>Analysis 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants, N_1</td>
<td>52</td>
</tr>
<tr>
<td>Female participants, N_1, n (%)</td>
<td>2 (3.8)</td>
</tr>
<tr>
<td>Number of projects obtained by the participants^a, N_2</td>
<td>1473</td>
</tr>
<tr>
<td>Number of projects obtained by the participants as PI^b,c, N_2, n (%)</td>
<td>803 (54.51)</td>
</tr>
<tr>
<td>Number of projects participated in as project member^a,b, N_2, n (%)</td>
<td>710 (48.20)</td>
</tr>
</tbody>
</table>

^aThe count includes duplicates because of, for example, changes in principal investigator. N_1 is the number related to participants and N_2 is the number related to acquired projects.

^bThe analysis includes Grant-in-Aid for Scientific Research (S), (A), (B), or (C); Grant-in-Aid for Specially Promoted Research; and Grant-in-Aid for Young Scientist (S) and (A) for analysis 2.

^cPI: principal investigator.

### Analysis 2

We obtained people-to-people connections up to the third level by counting 54 participants based on the team member list of projects throughout their careers. The researchers extracted in the second tier of connections, that is, those who had primary relationships with the 54 participants, totaled 982 and conducted 11,437 projects, of which 4381 (38.31%) were implemented by the 982 participants as principal investigators (Table 1). The types of colleagues were analyzed using the data of the research members for each project. We categorized the researchers who designated the participants as their project members based on whether they were upper-level or peer researchers of the participants. If they were peer researchers, we further categorized the person according to position rank. Connections with the participant were then classified according to their relationship. Analysis 2 focused not only on the total amount allocated for each project but also on the type of programs won. The number of large or small projects won as principal investigators was used as the dependent variable. Furthermore, we categorized the projects in which each participant took part as a project member according to the period in which each of them obtained a PhD to obtain information on when interactions are important over time.

Principal component analysis and multiple regression analysis were conducted using the XLSTAT statistical software.

### Results

#### Analysis 1

Multimedia Appendix 1 presents the descriptive statistics of the variables used in the analysis. When we regressed the total highest amount allocated for each project of the 52 selected researchers on their h-index, which represents the overall competitiveness of a given researcher with respect to publications, we found a positive correlation (Figure 1A). This result suggests that the variable defined previously is consistent with a researcher’s perceived competitiveness. The relationship between the dependent variable and the number of projects in which each researcher participated as a project member is shown in Figure 1B. We analyzed the determinants that contributed to success in obtaining GIA awards as principal investigators and focused on interpersonal relationships in the projects in which they participated as project members (models 1-4 in Table 2).
Figure 1. (A) Regression of the $h$-index value against the amount allocated to each participant as principal investigator (PI; includes Grants-in-Aid for Scientific Research, Grants-in-Aid for Specially Promoted Research, and Grants-in-Aid for Young Scientists). (B) Regression of the amount allocated to each participant as PI against the number of projects in which they participated as project members (N=52).
control variables, the coefficients of not having experience at to be highly influential throughout the analysis. Regarding the University rank at various stages in the researchers' career. to mitigate endogeneity, and model 3 includes variables for the number of papers and last-authored papers published. Model To examine the effects of publication performance, we included connections expressed through the using the value of betweenness centrality and quality of connections was examined using networking level as measured (A). Grant-in-Aid for Scientific Research (S), (A), (B), or (C); Grant-in-Aid for Specially Promoted Research; and Grant-in-Aid for Young Scientists (S) or (A). —0.49f (0.18) —0.48f (0.19) —0.32f (0.14)
Betweenness centralityf, unstandardized coefficient (SE) 0.28 (0.24) 0.38 (0.24) 0.32 (0.24) —
h-index of the researcher with whom the participant first became a project member, unstandardized coefficient (SE) 0.19f (0.09) 0.18f (0.09) 0.13 (0.10) 0.16 (0.08)
Number of papers, unstandardized coefficient (SE) −0.16 (0.24) −0.15 (0.24) −0.22 (0.25) —
Number of last-authored papers, unstandardized coefficient (SE) 0.67b (0.24) 0.55f (0.24) 0.65f (0.24) 0.65b (0.13)
F test (df) 12.68b (8, 43) 12.30b (9, 42) 10.53b (12, 39) 16.88b (6, 45)
Adjusted R2 0.647 0.666 0.652 0.651
AICb 954,178 952,072 954,918 951,888

aThe dependent variable was the sum of the highest amount allocated for each project number of projects obtained as principal investigator. This included Grant-in-Aid for Scientific Research (S), (A), (B), or (C); Grant-in-Aid for Specially Promoted Research; and Grant-in-Aid for Young Scientists (S) or (A).
bP<.001.
cThese variables were quantified based on The Times Higher Education World University Rankings 2023 and have 7 levels of classification, with a higher ranking having a smaller numerical value, that is, 1 for the top 100 ranking, 2 for 201 to 600, a value of 3 for 601 to 1000, a value of 4 for 1001 to 1200, a value of 5 for 1201 to 1500, a value of 6 for ≤1501, and 7 for out of ranking.
dVariables were not included in the model.
eThe total allocation earned as principal investigator during the first 10 years after obtaining a PhD.
fP<.05.
gThe Cytoscape software (version 3.4.0; Cytoscape Team) was used to calculate the value using people-to-people connections up to the third level starting from the participant based on the team member lists of all GIA projects throughout their career.
hP<.01.

Model 1 is the base model that includes the researchers’ attributes, human connections in the history of GIA grant success, and research performance. The effect of human connections was examined using networking level as measured using the value of betweenness centrality and quality of connections expressed through the h-index of the coresearcher. To examine the effects of publication performance, we included the number of papers and last-authored papers published. Model 2 controls for early career success in obtaining GIA awards to mitigate endogeneity, and model 3 includes variables for university rank at various stages in the researchers’ career. Model 4 is a simplified model with variables that were found to be highly influential throughout the analysis. Regarding the control variables, the coefficients of not having experience at nonuniversity institutions were significantly negative (P<.001). Difference based on sex exhibited no significant effect, but the number of female researchers in the analysis was small; thus, this result is not definitive. Considering research performance, the impact of last authorship was large (models 1-4).

Regarding human connections, researchers with fewer collaborative partners tended to win more grants; however, this trend was inconsistent with the correlation coefficients shown in Multimedia Appendix 1. Betweenness centrality, which indicates the visibility of each researcher within the collaborative network, did not exert a significant effect on lifetime grant acquisition (models 1-4). The h-index value of the researcher who first invited the participant to become a project member had a significant positive impact on the participant’s grant
success (model 1). The effect of these 3 variables remained significant even after incorporating variables related to grant success in the early stages of the researchers’ career (model 2). The correlation between university rank and grant success shows that researchers affiliated with higher-ranked universities tend to be more successful (Multimedia Appendix 1); however, when considering the regression models, the rankings of the universities where a researcher obtained their PhD degree and where they were currently working did not, by themselves, affect grant success. When university ranks were incorporated into the regression model, the effect of the $h$-index value of the researcher who first invited the participant to become a project member was not significant (model 3). Overall, the result of the $h$-index value indicates the importance of having a good research guide as the first step but also suggests that a researcher’s grant success can be affected by the new connections established during career development. Finally, model 4, which retains this variable together with last authorship and number of collaborative partners, has the smallest Akaike information criterion value and was considered the best-fit model.

Initially, we assumed that the researchers who first invited the participants to become a project member were laboratory heads and direct superiors, but this was not always the case (Table 3). Only 29.6% (32/108) of the researchers were direct superiors, 5.5% (6/108) were subgroup heads other than professors, and 7.4% (8/108) were peers or subordinates in their own institution. Notably, 8.3% (9/108) were external upper-level professionals (ie, from other institutions; Table 3). The correlation matrix in Multimedia Appendix 1 shows that being a project member of a government-funded program leads to a significant positive correlation with grant success. Therefore, as an in-depth examination of the effect of human resources, we tested whether project experience in a prestigious government-funded program, often led by top researchers, would increase the likelihood of GIA success for the participants using the Japan Science and Technology Agency Strategic Basic Research Program (SBRP) as a case study [17].

Table 3. Relationship with the researcher with whom the participant researcher first became a project membera.

<table>
<thead>
<tr>
<th></th>
<th>Same institution, n (%)</th>
<th>Different institution, n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper level (n=47)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor</td>
<td>38 (35.1)</td>
<td>9 (8.3)</td>
<td>47 (43.5)</td>
</tr>
<tr>
<td>Other than professor</td>
<td>32 (29.6)</td>
<td>8 (7.4)</td>
<td>40 (37.0)</td>
</tr>
<tr>
<td>Peer (n=7)</td>
<td>6 (5.6)</td>
<td>1 (0.9)</td>
<td>7 (6.4)</td>
</tr>
<tr>
<td>Subordinate (n=6)</td>
<td>5 (4.6)</td>
<td>2 (1.9)</td>
<td>7 (6.4)</td>
</tr>
<tr>
<td>Unidentified (n=1)</td>
<td>0 (0)</td>
<td>1 (0.9)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Total (N=108)</td>
<td>84 (77.8)</td>
<td>24 (22.2)</td>
<td>108 (100)</td>
</tr>
</tbody>
</table>

aAll the projects won in the year in which each participant obtained projects for the first time were counted.

Figure 2 plots the cumulative average GIA award values obtained as principal investigators over time; the 2 groups are established based on the presence or absence of SBRP experience among the 52 participants. Researchers who had been project members of the Core Research for Evolutional Science and Technology subprogram were selected as the treatment group from a total sample of 52, whereas those who had been PIs of this program were excluded. Researchers with similar $h$-index values and career lengths as those in the treatment group were then selected as the comparison group. The mean $h$-index value was 63.7 (SD 40.4) for the treatment group and 61.3 (SD 20.9) for the control group, and the mean career length was 33.4 (SD 5.2) and 25.1 (SD 5.4), respectively. The timeline was anchored on 0 for the start year of the projects in which the treatment researchers participated. For the control group, we fixed the 13th year since obtaining a doctoral degree to 0, which is the average career length at the time participants in the treatment group joined the SBRP projects (n=7 for each group).

The 2 groups begin at the leftmost point with low cumulative grant amounts. In the period leading up to the zero point, we observe no evident difference between the 2 groups, with similar funding gains. However, after participating in the SBRP, the gains of the treatment groups increased dramatically (Figure 2). By the end of the observation period, the treatment group received a higher award amount than the control group.
Analysis 2

To conduct a broader analysis of interpersonal relationships, we included 982 researchers, including 55 (5.6%) female researchers, identified as having collaborated with the researchers in the previous section (Table 1). The relationship between the number of years since obtaining a PhD and the amount earned as a principal investigator peaked at approximately 40 years after receiving the degree (Figure 3A). Using principal component analysis to evaluate each researcher's grant success in terms of research impact, namely, the number of projects obtained, project acquisition rate for large programs, and project acquisition rate for small programs, we divided the researchers into 3 groups (Figure 3B). The term “%Small” indicates the percentage of projects from the Grant-in-Aid for Scientific Research (C) (either number of projects or amount awarded), which represents the smallest category, in the total number of projects obtained as PI. Similarly, the term “%large” indicates the Grant-in-Aid for Specially Promoted Research, which are considered large categories.

We conducted univariate logistic regression analyses to identify the factors that produced researchers with high grant success records in both large- and small-program categories (Table 4). Career length had a positive effect for large programs (odds ratio [OR] 1.07, 95% CI 1.04-1.10 for stratum 1) and a negative effect for small programs (OR 0.95, 95% CI 0.94-0.97 for stratum 1). Conversely, the variable “doctorate obtained by thesis,” which was introduced to observe the effect of age, had no significant effect. However, among the 30.8% (184/597) of researchers who received their PhD later in life, there were prominent researchers who had served as university presidents or on government committees. Earning a degree earlier or later in life did not uniformly affect a researcher’s competitiveness, and individual differences are likely to have a greater impact. Regarding differences based on sex, male researchers exhibited a negative effect on project acquisition of small programs (OR 0.29, 95% CI 0.12-0.70 for stratum 4). This indicates that male researchers are likely to move from small to larger programs.
Figure 3. (A) Regression of the amount allocated to each participant as principal investigator (PI) against years after doctorate; includes Grants-in-Aid for Scientific Research, Grants-in-Aid for Specially Promoted Research, and Grants-in-Aid for Young Scientists. (B) Principal component analysis for the researchers’ grant success performance. Researchers were divided into 3 groups according to their performance (N=982). The term %Large indicates % large grants and the term %Small indicates % small grants.

Upon examining the impact of interaction with upper-level and peer researchers on grant success as principal investigators, we observed the following results (Table 4). More interactions with upper-level researchers resulted in fewer acquisitions of large programs (OR 0.67, 95% CI 0.50-0.89 for stratum 1) and more acquisitions of small programs (OR 1.21, 95% CI 1.07-1.36 for stratum 1) compared with the reference stratum. The differences in the ORs among the strata with different numbers of projects awarded indicated that the stronger the relationship with upper-level researchers, the higher the success rate in the smaller programs (OR 1.43, 95% CI 1.27-1.60 for stratum 4; Table 4).

After examining the impact of interaction with peer researchers separately at each stage of their careers, such as professor, associate professor, and assistant professor, we found that, in large programs, professor-professor interaction had a significant impact on the success rate of ≥2 research grants (OR 1.16, 95% CI 1.06-1.26 for stratum 2), which is not the case for the success of only 1 project (Table 4). More professor-professor interactions led to fewer acquisitions of ≥2 projects in the small category (OR 0.85, 95% CI 0.77-0.93 for stratum 2; Table 4). Interaction with peer researchers at the associate professor level and below had no significant effect on either large or small programs.
Table 4. Factors that affect grant success in terms of the relationship with the coresearchers (2-tailed $\chi^2$ test)\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Variable and stratum\textsuperscript{b}</th>
<th>Odds ratio (95% CI)</th>
<th>Significance of the model, $P$ value\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of large programs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of years since doctoral degree (n=982)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>1.07\textsuperscript{d} (1.04-1.10)</td>
<td>&lt;.001\textsuperscript{d}</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.07\textsuperscript{d} (1.03-1.11)</td>
<td></td>
</tr>
<tr>
<td>Doctorate obtained by thesis (n=597)\textsuperscript{e}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=16)</td>
<td>1.35 (0.43-4.23)</td>
<td>.88</td>
</tr>
<tr>
<td>2 (n=13)</td>
<td>1.01 (0.31-3.32)</td>
<td></td>
</tr>
<tr>
<td>Female (no; n=982)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>0.77 (0.23-2.58)</td>
<td>.82</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.60 (0.21-12.01)</td>
<td></td>
</tr>
<tr>
<td>Number of upper-level researchers who designated the participants as their project members (n=982)\textsuperscript{f}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>0.67\textsuperscript{d} (0.50-0.89)</td>
<td>.003\textsuperscript{d}</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>0.72\textsuperscript{b} (0.53-0.99)</td>
<td></td>
</tr>
<tr>
<td>Number of peer researchers who designated the participants as their project members (professor; n=982)\textsuperscript{f}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>1.08 (0.98-1.19)</td>
<td>.002\textsuperscript{d}</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.16\textsuperscript{d} (1.06-1.26)</td>
<td></td>
</tr>
<tr>
<td>Number of peer researchers who designated the participants as their project members (associate professor; n=982)\textsuperscript{f}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>—</td>
<td>.83</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>0.69 (0.22-2.22)</td>
<td></td>
</tr>
<tr>
<td>Number of peer researchers who designated the participants as their project members (assistant professor; n=982)\textsuperscript{f}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>—</td>
<td>.80</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>0.56 (0.10-3.12)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of small programs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of years since doctoral degree (n=982)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=219)</td>
<td>0.95\textsuperscript{d} (0.94-0.97)</td>
<td>&lt;.001\textsuperscript{d}</td>
</tr>
<tr>
<td>2 (n=190)</td>
<td>0.95\textsuperscript{d} (0.93-0.97)</td>
<td></td>
</tr>
<tr>
<td>3 (n=140)</td>
<td>0.96\textsuperscript{d} (0.94-0.98)</td>
<td></td>
</tr>
<tr>
<td>4 (n=174)</td>
<td>0.96\textsuperscript{d} (0.94-0.98)</td>
<td></td>
</tr>
<tr>
<td>Doctorate obtained by thesis (n=597)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=133)</td>
<td>1.27 (0.77-2.10)</td>
<td>.18</td>
</tr>
<tr>
<td>2 (n=118)</td>
<td>1.37 (0.81-2.31)</td>
<td></td>
</tr>
<tr>
<td>3 (n=76)</td>
<td>1.64 (0.88-3.06)</td>
<td></td>
</tr>
<tr>
<td>4 (n=113)</td>
<td>0.83 (0.50-1.37)</td>
<td></td>
</tr>
<tr>
<td>Female (no; n=982)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=219)</td>
<td>0.74 (0.28-1.96)</td>
<td>.05\textsuperscript{b}</td>
</tr>
<tr>
<td>2 (n=190)</td>
<td>0.47 (0.19-1.18)</td>
<td></td>
</tr>
<tr>
<td>Variable and stratum&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Odds ratio (95% CI)</td>
<td>Significance of the model, &lt;i&gt;P&lt;/i&gt; value&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>3 (n=140)</td>
<td>0.46 (0.17-1.23)</td>
<td></td>
</tr>
<tr>
<td>4 (n=174)</td>
<td>0.29&lt;sup&gt;d&lt;/sup&gt; (0.12-0.70)</td>
<td>&lt;.001&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

**Number of upper-level researchers who designated the participants as their project members (n=982)**

| 1 (n=219) | 1.21<sup>e</sup> (1.07-1.36) | <.001<sup>d</sup> |
| 2 (n=190) | 1.30<sup>d</sup> (1.16-1.47) |
| 3 (n=140) | 1.42<sup>d</sup> (1.26-1.60) |
| 4 (n=174) | 1.43<sup>d</sup> (1.27-1.60) |

**Number of peer researchers who designated the participants as their project members (professor; n=982)**

| 1 (n=219) | 0.94 (0.88-1.01) | .002<sup>f</sup> |
| 2 (n=190) | 0.85<sup>d</sup> (0.77-0.93) |
| 3 (n=140) | 0.85<sup>f</sup> (0.76-0.95) |
| 4 (n=174) | 0.91<sup>b</sup> (0.84-0.99) |

**Number of peer researchers who designated the participants as their project members (associate professor; n=982)**

| 1 (n=219) | 1.75 (0.46-2.00) |
| 2 (n=190) | 1.01 (0.67-1.51) |
| 3 (n=140) | 1.25 (0.85-1.84) |
| 4 (n=174) | 0.93 (0.60-1.44) |

**Number of peer researchers who designated the participants as their project members (assistant professor; n=982)**

| 1 (n=219) | 1.33 (0.78-2.28) |
| 2 (n=190) | 1.06 (0.57-1.97) |
| 3 (n=140) | 1.22 (0.65-2.27) |
| 4 (n=174) | 1.51 (0.88-2.57) |

<sup>a</sup>The number of projects obtained by the participants as principal investigators was used as the dependent variable. The term large indicates projects from the Grant-in-Aid for Scientific Research (S) and Grant-in-Aid for Specially Promoted Research, which are considered large categories. Similarly, the term small indicates the Grant-in-Aid for Scientific Research category (C), which is considered the smallest category.<br> <sup>b</sup>Number of projects obtained. Cases with no corresponding acquisitions were designated as stratum 0 and used as a reference. Cases with ≥2 corresponding acquisitions were defined as stratum 2 for large categories, and cases with ≥4 acquisitions were defined as stratum 4 for the small category.<br> <sup>c</sup>Wald test (<i>P</i> < .05).<br> <sup>d</sup><i>P</i> < .001.<br> <sup>e</sup>Doctoral degree conferred by submitting a doctoral thesis and passing its examination. We used this as a control variable and conducted an analysis by assigning a value of 0 for a regular degree and 1 for a doctorate obtained by thesis.<br> <sup>f</sup>Interaction among researchers in the same position was expressed as the number of researchers who made the participant a project member.<br> <sup>g</sup><i>P</i> < .01.<br> <sup>h</sup><i>P</i> < .05.<br> <sup>i</sup>No results were output by XLSTAT software for this stratum.

**Figure 4** shows the number of projects in which the researcher participated as a coinvestigator based on the years after obtaining a PhD. The researchers were divided into 3 groups based on their performance level, as shown in **Figure 3B**. The groups with higher research achievements had a greater number of projects during the 10- to 14-year postdegree period. Group 3, the group with the highest research achievements, had the largest number of projects during the 15- to 19-year postdegree period (**Figure 4**). **Table 5** focuses on these years as key periods and presents the average scores of several indicators of grant success and interpersonal relationships. Although the most active researchers (group 3) obtained large-program grants and interacted with more researchers during the indicated periods, the less active researchers (group 1), who primarily won small-program grants, interacted with fewer researchers during this period, many of whom were upper-level researchers (**Table 5**).
Figure 4. The average number of projects in which the researcher participated as a coinvestigator by number of years after obtaining a PhD and by research performance.

Table 5. Grant success and interaction with colleagues (2-tailed).

<table>
<thead>
<tr>
<th>Group (performance)</th>
<th>All (n=982), mean (SE)</th>
<th>Group 1 (low; n=436), mean (SD)</th>
<th>Group 2 (moderate; n=476), mean (SD)</th>
<th>Group 3 (high; n=70), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of years since doctoral degree</td>
<td>31.19 (10.92)</td>
<td>26.32 (10.24)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>34.57 (9.95)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>38.51 (8.23)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of projects obtained as a PI&lt;sup&gt;d&lt;/sup&gt;</td>
<td>4.11 (2.55)</td>
<td>2.86 (1.71)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.79 (2.53)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>7.31 (2.59)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Amount awarded as a PI (million yen, JP ¥&lt;sup&gt;e&lt;/sup&gt;)</td>
<td>59.73 (109.49)</td>
<td>11.10 (7.71)</td>
<td>59.31 (44.57)</td>
<td>365.45 (215.95)</td>
</tr>
<tr>
<td>Percentage of small grants in total number of acquisitions (projects)</td>
<td>0.55 (0.43)</td>
<td>0.96 (0.14)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.25 (0.28)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.06 (0.10)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Percentage of large grants in total number of acquisitions (projects)</td>
<td>0.02 (0.08)</td>
<td>0 (0)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.24 (0.17)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of projects participated in 10-14 y after doctorate</td>
<td>1.16 (1.90)</td>
<td>0.85 (1.38)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.40 (2.17)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.46 (2.04)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of upper-level researchers who designated the participants as their project members (10-14 y)</td>
<td>0.38 (0.81)</td>
<td>0.40 (0.78)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.39 (0.86)&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>0.19 (0.60)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of projects participated in 15-19 y after doctorate</td>
<td>0.26 (1.29)</td>
<td>0.20 (0.84)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.24 (0.83)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.67 (3.67)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of upper-level researchers who designated the participants as their project members (15-19 y)</td>
<td>0.19 (0.56)</td>
<td>0.27 (0.64)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.14 (0.52)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.03 (0.17)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>P<.05 against groups with b and c.
<sup>b</sup>P<.05 against groups with a and c.
<sup>c</sup>P<.05 against groups with a and b.
<sup>d</sup>PI: principal investigator.
<sup>e</sup>A currency exchange rate of JP ¥1=US $0.0067 is applicable.

Table 6 presents the results of the univariate logistic regression analyses on the impact of the frequency and quality of connections using the number of project acquisitions in large and small programs as dependent variables. When the frequency of interactions during the periods was categorized by the job relationships between the partner and the participant, the results showed that interactions with peer researchers and subordinates during the 10- to 14-year postdegree period had positive effects on ≥2 large-program acquisitions (OR 1.51, 95% CI 1.09-2.09 and OR 1.31, 95% CI 1.10-1.57, respectively; Table 6). Interactions with subordinates during the 15- to 19-year postdegree period also had positive effects (OR 1.25, 95% CI 1.25-1.07). In contrast, for the small programs, interaction with upper-level researchers was important both in the 10- to 14-year and the 15 to 19-year postdegree periods, with significant positive effects (Table 6). Notably, these results show that the frequency and quality of human interaction had opposite effects on acquisitions of large and small programs.
Table 6. Factors that affect grant success based on the timing of experience as a project member (2-tailed $\chi^2$ test).a

<table>
<thead>
<tr>
<th>Variable and stratum$^b$</th>
<th>Odds ratio (95% CI)</th>
<th>Significance of the model, $P$ value$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of large programs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of projects participated in 10-14 y after doctorate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of projects by upper-level researchers (n=968)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=41)</td>
<td>0.62 (0.34-1.13)</td>
<td>.11</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>0.58 (0.28-1.24)</td>
<td></td>
</tr>
<tr>
<td>Number of projects by peer researchers (n=968)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=41)</td>
<td>1.12 (0.75-1.66)</td>
<td>.04d</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.51d (1.09-2.09)</td>
<td></td>
</tr>
<tr>
<td>Number of projects by subordinates (n=968)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=41)</td>
<td>1.15 (0.94-1.42)</td>
<td>.007e</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.31e (1.10-1.57)</td>
<td></td>
</tr>
<tr>
<td>Number of projects participated in 15-19 y after doctorate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of projects by upper-level researchers (n=922)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>0.32 (0.09-1.18)</td>
<td>.23</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.4 (1.00-1.96)</td>
<td>.06</td>
</tr>
<tr>
<td>Number of projects by peer researchers (n=922)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>1.4 (1.00-1.96)</td>
<td>.02d</td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.36 (0.9-2.05)</td>
<td></td>
</tr>
<tr>
<td>Number of projects by subordinates (n=922)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=42)</td>
<td>1.12 (0.94-1.33)</td>
<td></td>
</tr>
<tr>
<td>2 (n=28)</td>
<td>1.25e (1.06-1.47)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of small programs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of projects participated in 10-14 y after doctorate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of projects by upper-level researchers (n=968)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=210)</td>
<td>1.20 (0.88-1.62)</td>
<td>&lt;.001f</td>
</tr>
<tr>
<td>2 (n=186)</td>
<td>1.48f (1.11-1.97)</td>
<td></td>
</tr>
<tr>
<td>3 (n=140)</td>
<td>1.88f (1.42-2.49)</td>
<td></td>
</tr>
<tr>
<td>4 (n=174)</td>
<td>1.87f (1.42-2.45)</td>
<td></td>
</tr>
<tr>
<td>Number of projects by peer researchers (n=968)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=210)</td>
<td>0.87 (0.69-1.10)</td>
<td>.12</td>
</tr>
<tr>
<td>2 (n=186)</td>
<td>0.69f (0.52-0.93)</td>
<td></td>
</tr>
<tr>
<td>3 (n=140)</td>
<td>0.81 (0.61-1.08)</td>
<td></td>
</tr>
<tr>
<td>4 (n=174)</td>
<td>0.80 (0.61-1.05)</td>
<td></td>
</tr>
<tr>
<td>Number of projects by subordinates (n=968)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (n=210)</td>
<td>1.02 (0.89-1.17)</td>
<td>.43</td>
</tr>
<tr>
<td>2 (n=186)</td>
<td>0.9 (0.75-1.07)</td>
<td></td>
</tr>
<tr>
<td>3 (n=140)</td>
<td>0.88 (0.72-1.08)</td>
<td></td>
</tr>
<tr>
<td>4 (n=174)</td>
<td>0.92 (0.78-1.10)</td>
<td></td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

This study aimed to identify factors that influence researchers’ potential to obtain external research funding by surveying their stage of career development and the types of people they interacted with using the GIA project implementation structure. Early-career interpersonal relationships, as measured using the h-index value of the researcher who provided the participants with their initial experience as project members, had a positive effect on grant success (Table 2). The results revealed the importance of having a good guide. We propose that a good guide can broaden project members’ perspectives by demonstrating the “behind-the-scenes” elements of effective project implementation. A good guide, not necessarily an immediate supervisor, also serves as a channel for informing fellow researchers of the perspectives required to obtain larger funds.

The results based on nonuniversity experiences (Table 2) suggest that creating an attractive proposal based solely on individual research curiosity may be difficult. The breadth of scientific expertise expressed within a research group rarely matches that expressed by an academic committee [24]. Enhancing one’s perspective by participating in large, purpose-driven projects such as those conducted by companies is important. The fact that experiencing the SBRP (a prestigious government program) as a project member facilitated subsequent grant success also confirms this hypothesis (Figure 2). A unique feature of the SBRP is that a star researcher, as the research director, conducts various interventions to modify the proposed research plan [17]. This provides the project members with opportunities to learn not only about the research conception of the principal investigator but also about the overall view of the research field held by the star researcher above the principal investigators and the strategic goals determined by policy objectives.

The effect of human connections varies depending on the career stage at which the research collaboration occurs. We found that the signs of the coefficients for the number of coauthors were inconsistent between Multimedia Appendix 1 and Table 2. This may be due to the fact that Table 2 shows the actual relevance of the indicator to researchers’ competitiveness at the individual level, whereas Table 2 shows the macro trends. Although a larger number of collaborators indicates a larger number of projects and more grant amounts obtained in general (positive coefficient in Multimedia Appendix 1), it also suggests the importance of implementing a small number of elite projects.
with selected collaborators to obtain large, trend-setting projects (negative coefficient in Table 2).

Our results suggest that greater collaboration among professors increases the number of large projects obtained (Table 4). After establishing one’s specialty and becoming a professor, collaborating with researchers in different fields and leveraging synergies to obtain greater funding is easier than when one is young [25]. It is assumed that highly competitive researchers who become professors early in their careers have more opportunities to conduct collaborative research among professors, and this collaboration and friendly competition with peers may stimulate their motivation to generate new ideas worthy of being supported by large programs. Meanwhile, factors that influence midcareer grant success remain largely unexplored despite challenging expectations regarding human resource development at universities and research institutions. This study did not present significant results regarding the midcareer level (Table 4); this is because collaboration during the earlier period may include protected time until each researcher refines their research and reconciles it with that of other researchers, after which truly meaningful collaboration occurs [26,27]. Tracking the process of research development, we found that interaction with others during the periods of 10 to 14 years and 15 to 19 years after obtaining a PhD determines the size of the project that the participant will obtain (Tables 5 and 6).

This study initially attempted to identify the factors that produce researchers with high grant success records, but as grant success depends on various factors, including the assignment of reviewers and other random factors, and given cases in which initial success may have been leveraged in subsequent years [28], it was difficult to obtain clear results when focusing on large programs alone (Tables 5 and 6). Interpreted in conjunction with the results from the project acquisition rate of small programs (Tables 5 and 6), midcareer relationships that remain narrowly focused, such as immediate supervisors, keep participants’ grant success limited to small programs throughout their careers and do not lead to the acquisition of large programs (Table 6). Liu [29] pointed out that the relationship between scholar productivity and tie strength exhibits an inverted U shape using data from tourism scholars. Researchers who devote their efforts to others’ research cannot concentrate on deepening their own studies. Considering the trade-off in collaboration between acquiring ideas and paying for effort instead of undertaking part of the supervisors’ initiatives, getting involved in diverse projects by peers and subordinates is important. In particular, participating in projects by subordinates is an effective way to be exposed to the fresh ideas of a younger person and look over their projects critically as an experienced person. This will ultimately help researchers become established figures who can conduct large-scale research projects as principal investigators.

Limitations
A limitation of this study is that we experienced some difficulties in obtaining clear data and subsequent results on the factors that influence the most prominent figures, such as those who had ≥3 projects in large programs (8/982, 0.8% in analysis 2), because of their rarity and the particular nature of their careers and research histories. These researchers tended to obtain large projects early in their careers instead of obtaining projects gradually increasing in size, which is related to the fact that they returned after international education pursuits or worked at nonuniversity institutions. Although initial success is likely to be influenced by almost uniform factors that apply to all researchers, such as publication performance relative to age, continued success is likely to be heavily influenced by individual enthusiasm and willingness to acquire large projects. Therefore, contextual analysis, such as interview surveys, will be necessary to identify the factors that produce prominent figures with outstanding achievements.

Acknowledgments
The authors would like to thank Exmart Inc for their assistance with data acquisition. This work was supported by a grant from the Watanabe Memorial Foundation for the Advancement of Technology awarded to AH. The authors would like to thank Editage for English-language editing. The authors did not use generative artificial intelligence during the preparation of this work.
Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request. Requests to access the data set should be directed to AH.

Authors’ Contributions
AH contributed to conceptualization, data curation, investigation, visualization, and writing—original draft preparation. MA contributed to writing—review and editing. ST contributed to resources and writing—review and editing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Descriptive statistics and correlation matrix.

References
16. Collecting the data study for evaluating the achievement of the S and T basic plans. National Institute of Science and Technology Policy. URL: https://histep.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item&item_id=4455&item_no=1&page_id=13&block_id=2 [accessed 2024-01-29]
23. Understand the current state of Japanese research!. BioImpact. URL: https://bioimpact.co.jp/media/researcher-jp [accessed 2024-02-02]

Abbreviations

GIA: Grants-in-Aid for Scientific Research
OR: odds ratio
SBRP: Strategic Basic Research Program

©Akiko Hashiguchi, Makoto Asashima, Satoru Takahashi. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Identifying Unmet Needs in Major Depressive Disorder Using a Computer-Assisted Alternative to Conventional Thematic Analysis: Qualitative Interview Study With Psychiatrists

Michelle A Worthington¹,², MA, MPhil; Richard H Christie¹, MD, PhD; Aaron J Masino¹,³, PhD; Sarah M Kark¹, PhD

¹AiCure, New York, NY, United States
²Department of Psychology, Yale University, New Haven, CT, United States
³The School of Computing, Clemson University, Clemson, SC, United States

Corresponding Author:
Sarah M Kark, PhD
AiCure
214 Sullivan Street 6C
New York, NY, 10012
United States
Phone: 1 (800) 570 0448
Email: sarah.kark@aicure.com

Abstract

Background: The development of digital health tools that are clinically relevant requires a deep understanding of the unmet needs of stakeholders, such as clinicians and patients. One way to reveal unforeseen stakeholder needs is through qualitative research, including stakeholder interviews. However, conventional qualitative data analytical approaches are time-consuming and resource-intensive, rendering them untenable in many industry settings where digital tools are conceived of and developed. Thus, a more time-efficient process for identifying clinically relevant target needs for digital tool development is needed.

Objective: The objective of this study was to address the need for an accessible, simple, and time-efficient alternative to conventional thematic analysis of qualitative research data through text analysis of semistructured interview transcripts. In addition, we sought to identify important themes across expert psychiatrist advisor interview transcripts to efficiently reveal areas for the development of digital tools that target unmet clinical needs.

Methods: We conducted 10 (1-hour-long) semistructured interviews with US-based psychiatrists treating major depressive disorder. The interviews were conducted using an interview guide that comprised open-ended questions predesigned to (1) understand the clinicians’ experience of the care management process and (2) understand the clinicians’ perceptions of the patients’ experience of the care management process. We then implemented a hybrid analytical approach that combines computer-assisted text analyses with deductive analyses as an alternative to conventional qualitative thematic analysis to identify word combination frequencies, content categories, and broad themes characterizing unmet needs in the care management process.

Results: Using this hybrid computer-assisted analytical approach, we were able to identify several key areas that are of interest to clinicians in the context of major depressive disorder and would be appropriate targets for digital tool development.

Conclusions: A hybrid approach to qualitative research combining computer-assisted techniques with deductive techniques provides a time-efficient approach to identifying unmet needs, targets, and relevant themes to inform digital tool development. This can increase the likelihood that useful and practical tools are built and implemented to ultimately improve health outcomes for patients.

(JMIR Form Res 2024;8:e48894) doi: 10.2196/48894

KEYWORDS
consumer health informatics; interview; major depressive disorder; medical informatics applications; needs assessment; psychiatry and psychology
Introduction

Digital health tools have the potential to advance health care efficiency, precision medicine, and patient health outcomes. Yet even the most high-performing digital tools using cutting-edge artificial intelligence and machine learning techniques are likely to be shelved if they do not address shared priorities among stakeholders [1] and their associated real-world unmet needs, leaving the promise of artificial intelligence and machine learning in health care underrealized [2]. There are currently 350,000 digital health apps worldwide that aim to address a range of functions (eg, condition management, wellness and prevention, and patient experience) and health conditions (eg, diabetes [3], neurological conditions [4], and psychiatric illnesses [5,6]). However, most of these apps are not regulated or clinically validated, and most are not widely used or integrated into clinical practice [7,8]. Although evidence suggests digital health tools are acceptable to both patients and clinicians, there are diverging needs, priorities, and attitudes among stakeholder groups within the digital tool ecosystem [1].

Gathering insights from key stakeholders is an essential step to ensure the development of digital tools that meet the needs of both patients and clinicians toward the goal of providing high-quality patient-centered care [9,10]. Using upstream stakeholder engagement methods can reveal actionable—but otherwise unforeseen—needs for targeted design and development of clinically impactful patient-facing digital tools with increased potential for widespread adoption [11,12]. Previous work has shown the benefits of qualitative research to assess hypothetical smartphone apps with content and function designs [13,14]. However, that work does not consider the motivation, or genuine need, for the app. Here we focus on the use of upstream qualitative research to first identify areas of real-world unmet need to serve as the foundation for hypothetical smartphone app ideation, design, and prototyping.

Qualitative research often involves designing and conducting a group of individual interviews and applying thematic analysis to transcripts of the resulting data. Conventional thematic analysis typically involves multiple researchers reviewing a subset of interview transcripts to identify themes [15] and developing a hierarchical system (codebook) of themes and subthemes (codes) to apply to sections of text (segments). The results of a conventional thematic analysis are typically presented in a table with broad summary themes, nested subthemes, and illustrative quotes from interview transcripts. The process of constructing the codebook can be time-consuming (eg, 3-5 times the amount of time taken to collect the data needed to review each interview [16]), subjective, and not always replicable [17,18].

While conventional thematical approaches are powerful for extensive investigations into a particular research area, a more time-efficient, scalable, and reproducible method is needed for digital health tool developers working in industry to swiftly identify key areas for the development of clinically meaningful tools. Indeed, in many settings of the digital health ecosystem, it is not feasible for researchers to dedicate substantial time to complete conventional thematic analysis. In some cases, it might be more appropriate to apply an automated text analysis that can be easily implemented by researchers, developers, and clinician scientists [19] to guide the identification of unmet needs and potential solutions. For example, recent work has applied text analysis to large qualitative research data sets to quickly identify common themes based on word frequency analysis [20-22]. Word combination frequency analysis provides a data-driven and repeatable approach to quickly identify frequently mentioned topics across a set of qualitative interview transcripts and potentially reduce the introduction of personal bias. Such approaches enable a faster and more convenient method to analyze a large amount of qualitative text data obtained from interviews with stakeholders [23]. Further, hybrid approaches that combine conventional thematic analysis with a data-driven inductive approach have the potential to leverage the strengths of both methods [24]. Importantly, such computer-assisted approaches to upstream qualitative research can be applied to engage and research any stakeholder group, including patients, to inform the development of clinically meaningful digital tools.

Given the limitations of conventional qualitative analysis—including the substantial time required for theme development—we demonstrate the utility of a hybrid approach leveraging the simplicity and accessibility of text analysis to identify stakeholder themes to support the initial stage of concept development for digital health tools. Specifically, we applied word combination analysis to a set of semistructured interviews with US-based psychiatrists specializing in treating outpatients with major depressive disorder (MDD) to reduce the amount of text for thematic analysis fivefold to facilitate uncovering common themes and unmet needs of clinicians and patients across advisors. We opted to demonstrate the use of simple text analysis over some of the more advanced natural language processing techniques to increase accessibility to those without advanced backgrounds in coding or computational techniques. We propose that this approach could serve as a straightforward and repeatable framework to identify unmet needs before concept development and implementation. We will discuss our findings around unmet needs in treating MDD and how this might affect digital tool development in psychiatry as an end-to-end demonstration of this method.

Methods

Ethical Considerations

The interviewed psychiatrists were originally recruited for market research purposes using the Guidepoint Expert Network (Guidepoint Global). The BRANY Institutional Review Board determined this study was exempt from review under category 4ii in 45 CFR 46.104(d). Participants were compensated for their time and participation. All reported data were stored in password-protected databases accessible only to approved study personnel.

Recruitment and Data Collection

We used purposive criterion sampling to recruit 10 US-based psychiatrists specializing in treating MDD outpatients in a variety of practice settings, including academic medical centers and teaching hospitals, community-based mental health clinics,
and private practices spanning urban and rural settings in the United States (Table 1). With only 2 exceptions, we recruited and interviewed psychiatrists who spent most of their time (≥50%) on direct MDD outpatient patient care and who had 10 or more years of experience postresidency. The interviewed psychiatrists were originally recruited for market research purposes using the Guidepoint Expert Network (Guidepoint Global).

Table 1. Stakeholder characteristics and treatment settings.

<table>
<thead>
<tr>
<th>Stakeholder setting/US region</th>
<th>Clinic experience (years)</th>
<th>Direct patient care (% time)</th>
<th>Outpatient facing (% time)</th>
<th>MDD(^a) patient load (patients/week)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Academic medical center or university teaching hospital</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>25</td>
<td>75</td>
<td>75</td>
<td>120</td>
</tr>
<tr>
<td>West</td>
<td>25</td>
<td>90</td>
<td>80</td>
<td>250</td>
</tr>
<tr>
<td>Southwest</td>
<td>20</td>
<td>30</td>
<td>99</td>
<td>20</td>
</tr>
<tr>
<td><strong>Community hospital</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>25</td>
<td>95</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>Southwest</td>
<td>10</td>
<td>98</td>
<td>60</td>
<td>35</td>
</tr>
<tr>
<td>Southwest</td>
<td>7</td>
<td>100</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td><strong>Group or private practice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>28</td>
<td>95</td>
<td>95</td>
<td>35</td>
</tr>
<tr>
<td>Southwest</td>
<td>13</td>
<td>95</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>Mid-Atlantic</td>
<td>13</td>
<td>90</td>
<td>85</td>
<td>50</td>
</tr>
<tr>
<td>West</td>
<td>12</td>
<td>100</td>
<td>100</td>
<td>30</td>
</tr>
</tbody>
</table>

\(^a\)MDD: major depressive disorder.

Regarding sample size, previous work suggests most qualitative research data sets reach saturation—the point during data collection in which themes begin to repeat, new insights begin to wane, and future data collection becomes redundant—between 9 and 17 interviews [25]. Moreover, our inclusion criteria were designed to recruit a relatively homogenous sample of experienced clinicians specializing in outpatient MDD care, as we did not require any between-subject comparisons.

One researcher conducted individual, 1-hour, single-blinded, semistructured, recorded audio interviews following a predesigned interview guide (Multimedia Appendix 1). The interview guide was designed to include open-ended questions primarily targeted at understanding the clinicians’ experience in care management for MDD as well as the clinicians’ perceptions of patient experience in care management for MDD. We used MAXQDA Analytics Pro 2022 (version 22.1.1; VERBI GmbH) [26] to process and analyze the quality-checked interview transcripts (their official website provides help with using specific features). However, one can use other available software or open-source approaches to execute the simple word combination text analysis steps below.

Data Preparation and Overview

All transcription text segments were auto-coded as either “interviewer” or “advisor,” depending on the speaker. Only “advisor” text was used during the text analysis stage.

The data-driven hybrid approach to identifying relevant themes involved a 4-step process (Figure 1): (1) During the computer-assisted stage, use MAXQDA (or other) software to identify the most frequent n-grams (n-word word combinations, n=2+ words) across the interview data set. (2) Extract the sentences containing each of those n-grams, along with the preceding and succeeding sentences relative to the sentence containing each n-gram. (3) During the “hybrid stage,” work iteratively with at least 1 other researcher to read and reread each computer-extracted text segment in order to inductively identify key content categories and assign a content category label to each text segment until 100% agreement is reached among researchers. Discordant content category assignments among researchers can be resolved during discussion. (4) Finally, in the “deductive stage,” manually examine the full list of key content categories to develop overarching themes and nest the key content categories under the broader themes characterizing priorities in the process of care management in treating MDD.
Data Analysis

Identifying Word Combinations

First, we used the “word combination” feature in MAXQDA to identify the most frequently used phrases in the advisor’s responses. We used MAXQDA, but any software capable of word combination analysis can be used. We used n-word combinations instead of single-word frequencies to increase the specificity of theme identification. Within the search parameters, we required the resulting words to be at least 4 characters long. All words were lemmatized in English, and we applied the MAXQDA English language stop list [27] to remove articles, conjunctions, and other words likely to be redundant in the analysis. We additionally added word combinations that co-occurred with our disease state of interest to the stop list (eg, “major depression” and “mental health”), as these were unlikely to yield meaningful insights. Next, we filtered the resulting word combination frequencies to word combinations that also appeared in a majority (at least 60%) of the interviews. We opted for a majority threshold cut-off based on our sample (word combinations that appeared in at least 60% of interviews) to limit the subsequent research steps to a more focused set of word combinations that were not likely spoken just by random chance (ie, ≤50%). Our goal was to maximize the chance of developing a solution that would be impactful for the majority (>50%) of clinicians. However, depending on the goals, users may elect to use different thresholds. For completion, word combinations that appeared in 50% or more of the interviews are shown in the “Identifying Word Combinations” section, but only those that appeared in 60% or more of the interviews were included in the subsequent analysis.

Broadly, this first step served as a data-driven text analysis to identify commonly discussed themes or frequent expressions when discussing the care management process of the disease state of interest. Identifying key phrases used across stakeholders in a sample can give a sense of widely applicable needs and daily experiences. Identifying n-grams or word combinations is the crucial time-saving inductive step; however, to build meaning around the phrases, we recommend identifying key content areas by conducting follow-up deductive analyses of the context surrounding the word combinations.

Identifying Key Content Categories

Next, we extracted the resulting n-grams and their surrounding context for further analysis of the context in which these word combinations were uttered. Specifically, for each n-gram, we extracted the sentence containing the n-gram along with the 2 sentences before and after the appearance of each n-gram to provide context.

Next, 2 researchers systematically read the sentences surrounding each extracted word combination to understand the context in which the word combination was uttered. The researchers then worked iteratively through discussion and multiple independent readings to generate and assign relevant content categories for each n-gram. For expediency, the 2 researchers reconciled discordant labels through discussion until they reached 100% agreement. The researchers aimed to create content category labels for each word combination with definitions that were broad enough to accommodate multiple text segments but specific enough to distinguish among text segments from a given word combination. These researcher-generated content categories ultimately reflected the context surrounding the n-grams and provided a more in-depth understanding beyond the n-grams alone. Extracting and reviewing text around frequent and common n-grams helped to both focus the reading and deductive analysis of the text and substantially reduce the amount of text needed for review.

Next, we selected representative quotations for each content category to provide canonical examples of each of the chosen themes (Table S1 in Multimedia Appendix 2). This step represents a bridge between the text-analytic approach and the conventional deductive approach, in which researchers manually evaluate all the text to determine the appropriate content category represented in the segment. During this state, researchers review the text segments and apply their subject-matter expertise to sort the resulting text into subthemes.
However, the text analytic step saves time and increases process transparency by examining only the segments in each interview where the n-gram appears. This should ultimately limit the scope of the analysis to pertinent segments constrained by the list of word combinations established in the first step.

Researchers may choose to stop at this level of analysis and proceed with the development of digital tool concepts if unmet needs and potential challenges are adequately identified. The word combinations and content categories may be substantial enough to provide sufficient context to understand how to proceed with digital tool development. Here, we further analyzed the word combinations and content categories in order to identify overarching themes, bringing the results of the hybrid method even closer to those yielded from conventional analysis.

Developing Overarching Themes

Once the content categories were identified as described in the previous step, key overarching themes emerged across the content categories and word combinations. At this stage, we also included word combinations that appeared in 50% of interviews to help guide the identification of more robust overarching themes that accommodated more word combinations. During reading and analysis of the resulting n-grams and their corresponding text segments, themes should start to emerge that may have been deduced using the more conventional analysis approach. In essence, the word combinations and key content categories established in the first and second steps are essentially transposed to develop these overarching summary themes, highlighting areas of need. In this approach, the word combinations are reorganized and presented with word combinations nested within broad summary themes. The goal of this final step is to summarize the findings from the first 2 steps into an alternative table structure that may assist in conceptualizing unmet needs for patient-facing digital tool development. By following the first 2 steps, theme identification in this final step becomes a much more efficient and transparent process as compared to a more conventional approach.

Results

Identifying Word Combinations

All word combinations appearing in 50% or more of the interviews are shown in Table 2. Only bigrams resulted from the analyses. There were no word combinations of ≥3 words. For a full summary of word combinations, content categories, detailed descriptions, and representative quotations, please refer to Table S1 in Multimedia Appendix 2.

Table 2. Frequently used word combinations listed by the percentage of interviews in which they were uttered and the total number of utterances across all interviews. The table elements are listed in descending order by the percentage of interviews in which they were uttered.

<table>
<thead>
<tr>
<th>Word combination</th>
<th>Percentage of the interviews containing the word combination, %</th>
<th>Total utterances of the word combination across all interviews, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side effect</td>
<td>100</td>
<td>49</td>
</tr>
<tr>
<td>Make sure</td>
<td>100</td>
<td>24</td>
</tr>
<tr>
<td>Family history</td>
<td>70</td>
<td>12</td>
</tr>
<tr>
<td>Primary care</td>
<td>60</td>
<td>18</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>60</td>
<td>9</td>
</tr>
<tr>
<td>Energy level</td>
<td>60</td>
<td>7</td>
</tr>
<tr>
<td>Really want</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>Treatment plan</td>
<td>50</td>
<td>9</td>
</tr>
<tr>
<td>Treatment resistant</td>
<td>50</td>
<td>9</td>
</tr>
<tr>
<td>Come back</td>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td>Bipolar depression</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>Facial expression</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>Suicidal thought</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>Family member</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>Patient come</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>Feel comfortable</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>Good thing</td>
<td>50</td>
<td>5</td>
</tr>
</tbody>
</table>

Critically, after extracting the sentences containing the set of word combinations and their preceding and succeeding sentences, we observed that the total advisor text word count to manually review reduced to 21% of the original stakeholder text word count, demonstrating the use of the approach to assist with focusing theme extraction on key areas of text containing n-grams (ie, the highly repeatable text reduction step).

Identifying Key Content Categories

After each researcher completed multiple readings of all word combinations and their surrounding contextual sentences, up to
5 content categories for each word combination were identified, resulting in a total of 14 unique content categories across phrases (Table S1 in Multimedia Appendix 2): administration, care management, common side effects, conceptualization, desired clinical information, differential diagnosis, medical comorbidities, medical history, medication monitoring and management, patient experience, risk assessment, risk for substance use, side effect monitoring, and treatment planning. The researchers aimed for ≤5 content categories for each word combination to balance specificity with generality. However, this target can be changed to accommodate larger or smaller studies that might require a higher or lower threshold to balance these goals. In total, there were 14 unique content categories, as some content category labels emerged under multiple word combinations (eg, care management emerged as a content category relevant to 2-word combinations, “side effects” and “primary care”).

Developing Overarching Themes

In the final step, we identified overarching summary themes by transposing the word combinations and content categories identified in steps 1 and 2. In this analysis, we included word combinations uttered by 50% of advisors, which returned 11 additional word combinations (eg, “bipolar depression” and “facial expression”). The full results expanding on these overarching themes can be found in Table 3, with the word combinations listed as relevant word combinations that support content categories and overarching themes. Through this process, we identified four overarching summary themes: (1) evaluation, (2) medication decisions, (3) tracking symptom progression, and (4) factors contributing to treatment adherence.

### Table 3. Overarching themes identified through transposing word combinations and content categories.

<table>
<thead>
<tr>
<th>Overarching theme (with definition) and relevant key content categories</th>
<th>Relevant word combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evaluation:</strong> Providers expressed day-to-day challenges related to ruling out co-occurring psychiatric disorders (such as bipolar depression) and medical comorbidities and that they would benefit from more data about their patients.</td>
<td></td>
</tr>
<tr>
<td>Differential diagnosis</td>
<td>“bipolar depression”, “make sure”, “substance use/abuse”</td>
</tr>
<tr>
<td>Medical comorbidities</td>
<td>“make sure”, “primary care”</td>
</tr>
<tr>
<td>Risk assessment</td>
<td>“suicidal thought”, “substance use/abuse”</td>
</tr>
<tr>
<td>Conceptualization</td>
<td>“suicidal thought”, “treatment resistant”, “facial expression”, “family history”, “family member”</td>
</tr>
<tr>
<td>Desired clinical information</td>
<td>“substance use/abuse”, “energy level”, “facial expression”</td>
</tr>
<tr>
<td><strong>Medication decisions:</strong> Providers revealed their treatment planning choices are driven by leveraging known medication side effects to counteract patient symptom profiles (typically based on baseline energy levels).</td>
<td></td>
</tr>
<tr>
<td>Medical history</td>
<td>“patient come”, “primary care”</td>
</tr>
<tr>
<td>Medication management and monitoring</td>
<td>“make sure”, “primary care”, “treatment resistant”</td>
</tr>
<tr>
<td>Treatment planning</td>
<td>“treatment plan”, “energy level”</td>
</tr>
<tr>
<td>Patient experience</td>
<td>“side effect”, “feel comfortable”</td>
</tr>
<tr>
<td>Common side effects</td>
<td>“side effect”</td>
</tr>
<tr>
<td><strong>Tracking symptom progression:</strong> Providers were concerned about their lack of ability to track patient safety, triaging to a higher level of care, and co-occurring substance use outside of the clinic.</td>
<td></td>
</tr>
<tr>
<td>Risk assessment</td>
<td>“suicidal thought”</td>
</tr>
<tr>
<td>Risk for substance [ab]use</td>
<td>“substance use/abuse”</td>
</tr>
<tr>
<td>Side effect monitoring</td>
<td>“side effect”, “energy level”</td>
</tr>
<tr>
<td><strong>Factors contributing to treatment adherence:</strong> Providers highlighted the impact of adverse side effects on adherence and the critical role family members play in both patient adherence and providing them with symptom insights.</td>
<td></td>
</tr>
<tr>
<td>Side effects</td>
<td>“side effect”</td>
</tr>
<tr>
<td>Patient experience</td>
<td>“feel comfortable”, “side effect”, “come back”</td>
</tr>
<tr>
<td>Care management</td>
<td>“family member”, “primary care”</td>
</tr>
<tr>
<td>Administrative</td>
<td>“come back”, “make sure”</td>
</tr>
</tbody>
</table>

To create overarching themes, the researchers applied a similar process to the content category labels as they did to the word combinations—identifying a label and definition that would provide an appropriate umbrella term under which to nest multiple of the key content categories. To illustrate how we arrived at these broader themes, we will walk through the development of the “evaluation” theme. Analysis of the retrieved segments in the sentences surrounding the word combinations “bipolar depression,” “make sure,” and “substance [ab]use” converged on a content category best summarized as “differential diagnosis.” Through a similar process, segments surrounding the word combinations “suicidal thought,”
“treatment resistant,” “facial expression,” “family history,” and “family member” converged on a content category best summarized as “conceptualization.” Ultimately, “differential diagnosis” and “conceptualization” fit together with “medical comorbidities,” “risk assessment,” and “desired clinical information” under the broad and overarching theme of “evaluation.”

**Discussion**

**Overview**

To ensure the development of clinically meaningful digital tools, developers must address real-world, unmet stakeholder needs. Understanding these needs at the outset of concept ideation and tool development ensures that developers are solving the most urgent challenges. Qualitative research is a powerful tool to understand the daily experiences of patients and clinicians and to better understand how potential digital tools will provide value and integrate seamlessly into a given environment. Here, we presented a hybrid data-driven approach to facilitate time-efficient discovery of unmet clinician needs to inform the direction of digital tool development in health care settings. With this approach, in the context of care management in MDD, we identified several key areas that are of interest to clinicians and would be appropriate targets for digital tool development.

**Using Key Content Categories to Inform Digital Tool Development**

The key content categories are consistent with challenges identified in the literature, suggesting this approach can highlight real-world problems and point developers to further information in previously unknown areas of research literature. For example, the phrase “side effect” consisted of the following content categories: treatment planning, side effect monitoring, care management, patient experience, and common side effects. Further investigation into the literature encompassing this topic provides additional support for the selection of this area as a target of digital tool development. Burdensome treatment-related side effects are a leading cause of nonadherence and discontinuation of antidepressant medications [28] and have a negative impact on treatment outcomes [29]. Patients frequently experience side effects early in antidepressant treatment [30], but a lack of understanding of side effects [31] and barriers to communicating these side effects to clinicians, especially primary care providers [32], lead to early discontinuation and poor outcomes [29]. Some evidence exists that interventions addressing these early barriers to adherence, which include side effects, could improve adherence, communication between patient and provider, and, ultimately, treatment outcomes [33,34].

In the context of MDD care management, building digital tools that address patient and clinician concerns related to side effects, medication adherence due to side effects, and medication decisions based on side effects might best address one aspect of the current needs of clinicians as expressed through the interviews. The strengths of digital tools that could be leveraged in this context include facilitating side effect monitoring and reporting in between visits using smartphone capabilities, increasing patients’ understanding and expectations around side effects with on-demand psychoeducation through a mobile app or customized website, increasing clinicians’ awareness of the emergence of side effects through smartphone-based remote monitoring, or increasing and enhancing patient-provider communication through digital platforms.

Using a single topic as a guide for concept development may be a reasonable starting point for digital tool development; however, it is important to consider other related content categories that arise from this procedure. Incorporating other themes or content categories may bolster the initial concept and increase the likelihood that patients or clinicians will adopt a tool in clinical practice. In the context of MDD care management, the content category “administration” emerged as an important consideration for clinicians, signaling that workflow and administrative components of care management are also important to keep in mind while developing digital tools to ensure consideration of practical integration into workflows.

**Incorporating New Ideas With Existing Best Practices**

Overall, the topics identified through the approach outlined in this study should line up with current thinking and best practices around digital tool development in health care. This includes, but is not limited to, prioritizing ethical considerations around data sharing and privacy [35,36], considering the legal implications of digital tool implementation [37], and considering the balance between addressing unmet needs and integrating tools into the current care management workflow [38]. While identifying specific themes and content categories for a given disease state will elucidate current unmet needs for clinicians, established guidelines around digital tool development should be included as well.

**Limitations**

The approach outlined in this study has the potential to facilitate digital tool development across numerous clinical environments due to its ease of use and relative efficiency. Nevertheless, the comprehensive nature of conventional qualitative analysis may yield more nuanced findings from stakeholder interviews using a more deductive coding process. Moreover, limiting the amount of text for analytical review to those sentences surrounding frequent key word combinations across advisors is both a strength and a limitation of this method. While this approach increases efficiency, there could be a loss of sensitivity to important learnings outside of the extracted text segments, as well as some important one-off learnings uttered by only 1 advisor that could potentially be excluded from the text extracted for deductive analysis. Although the current sample size falls within a reasonable range to reach theme saturation across interviews [25], a more expanded sample size has the potential to reveal even more insights and accommodate analysis by psychiatry subspecialties (eg, addiction, child, adolescent, and geriatric psychiatrists). Furthermore, given that this method seeks to find consensus among spoken terms describing themes and unmet needs among clinicians and patients, it is possible that diverging opinions among stakeholders might be obscured. Finally, although the word combination analysis is 100% repeatable, further work would be needed to determine
confirmability (ie, results confirmed by an independent set of researchers) [39].

**Future Directions**

Here, we first interviewed clinicians because they are uniquely positioned within the digital health ecosystem to simultaneously identify and communicate patient needs, unmet clinician needs, potential for real-world clinical impact, and influence patient engagement with appropriate apps due to the trust patients place in their clinicians [1,7]. One challenge that blocks patient engagement with relevant digital tools is integration into clinician workflow [1], because clinicians have limited bandwidth to integrate patient-facing digital tools into their workflow. However, to ensure the most clinically relevant tools are created for and with patients, future work is needed to identify overlapping priorities between patients and clinicians. The method outlined here could be applied to patient interviews to reveal common and frequent unmet needs among patients.

The current framework provides a foundation for developers in the technology space to identify concepts that are worthy of further investigation and validation. Researchers following this framework may consider validating findings from this approach through follow-up methods such as: (1) triangulation studies (eg, test for convergence of results from this framework with focus group results or text from other sources, such as peer-reviewed articles about MDD), (2) quantitative survey methods, and (3) member-checking by presenting responses with the results to confirm the interpreted data and identified concept results resonate with their experience (respondent validation) [40] before moving further into conceptualization and prototyping.

**Conclusions**

We presented a hybrid computer-assisted method for identifying unmet needs expressed in semistructured interviews, which provides an efficient and user-friendly approach to this problem. The presented method offers some of the efficiencies of a purely analytic approach (eg, topic modeling), while the incorporation of manual analysis of the surrounding context sentences offers some of the benefits of finding more interpretable and relevant concerns, as in a traditional qualitative analysis. By contrast, in topic modeling alone, researchers often consider the topics (lists of words) outside of their surrounding context when trying to interpret their meaning. Thus, this hybrid approach falls between these 2 approaches and gives digital health researchers a feasible approach for conducting upstream research to inform ideation and the development of high-impact patient-facing digital tools.

**Acknowledgments**

The authors would like to thank the participating psychiatrists for their time and insights.

**Conflicts of Interest**

All authors were employed at and had a financial interest in AiCure, LLC, at the time of the study.

**Multimedia Appendix 1**

Semistructured interview guide.

[DOCX File, 21 KB - formative_v8i1e48894_app1.docx]

**Multimedia Appendix 2**

Word combinations, key content categories, descriptions and representative quotations for word combinations appearing in at least 60% of all interviews.

[DOCX File, 20 KB - formative_v8i1e48894_app2.docx]

**References**


Abbreviations

MDD: major depressive disorder

Edited by A Mavragani; submitted 10.05.23; peer-reviewed by M Gasmi, C Weir; comments to author 28.12.23; revised version received 16.01.24; accepted 17.01.24; published 01.03.24.

Please cite as:

Worthington MA, Christie RH, Masino AJ, Kark SM
Identifying Unmet Needs in Major Depressive Disorder Using a Computer-Assisted Alternative to Conventional Thematic Analysis: Qualitative Interview Study With Psychiatrists
JMIH Form Res 2024;8:e48894
URL: https://formative.jmir.org/2024/1/e48894
doi:10.2196/48894
PMID:38427407
Risk Identification in Perinatal Health Care Settings via Technology-Based Recruitment Methods: Comparative Study

Jessica R Beatty¹, PhD; Logan Zelenak¹, BA; Spencer Gillon¹, BA; Lucy McGoron¹, PhD; Gregory Goyert², MD; Steven J Ondersma³,4, PhD

¹Merrill Palmer Skillman Institute for Child & Family Development, Wayne State University, Detroit, MI, United States
²Maternal Fetal Medicine, Women's Health Services, Henry Ford Health, Detroit, MI, United States
³Charles Stewart Mott Department of Public Health, College of Human Medicine, Michigan State University, Flint, MI, United States
⁴Department of Obstetrics, Gynecology, and Reproductive Biology, College of Human Medicine, Michigan State University, East Lansing, MI, United States

Corresponding Author:
Steven J Ondersma, PhD
Charles Stewart Mott Department of Public Health
College of Human Medicine
Michigan State University
200 East 1st Street
Room 368
Flint, MI, 48502
United States
Phone: 1 313 444 9797
Email: onders12@msu.edu

Abstract

Background: Digital screening and intervention tools have shown promise in the identification and reduction of substance use in health care settings. However, research in this area is impeded by challenges in integrating recruitment efforts into ongoing clinical workflows or staffing multiple study clinics with full-time research assistants, as well as by the underreporting of substance use.

Objective: The aim of the study is to evaluate pragmatic methods for facilitating study recruitment in health care settings by examining recruitment rates and participant characteristics using in-person–based versus flyer approaches.

Methods: This study compared recruitment rates at a Women’s Health clinic in the Midwest under 2 different recruitment strategies: in person versus via a flyer with a QR code. We also examined the disclosure of substance use and risk screener positivity for the 2 strategies. We also obtained information about the current use of technology and willingness to use it for study participation.

Results: A greater percentage of patients recruited in person participated than those recruited via flyers (57/63, 91% vs 64/377, 17%). However, the final number recruited in each group was roughly equal (n=57 vs n=64). Additionally, participants recruited via flyers were more likely to screen positive for alcohol use risk on the Tolerance, Annoyed, Cut Down, Eye-Opener alcohol screen than those recruited at the clinic (24/64, 38% vs 11/57, 19%; χ²=4.9; P=.03). Participants recruited via flyers were also more likely to screen positive for drug use risk on the Wayne Indirect Drug Use Screener than those recruited at the clinic (20/64, 31% vs 9/57, 16%; χ²=4.0; P=.05). Furthermore, of the 121 pregnant women, 117 (96.7%) reported owning a smartphone, 111 (91.7%) had an SMS text message plan on their phone, and 94 (77.7%) reported being willing to receive SMS text messages or participate in a study if sent a link to their phone.

Conclusions: The distribution of flyers with a QR code by medical staff appears to be an efficient and cost-effective method of recruitment that also facilitates disclosure while reducing the impact on clinic workflows. This method of recruitment can be useful for data collection at multiple locations and lead to larger samples across and between health systems. Participant recruitment via technology in perinatal health care appears to facilitate disclosure, particularly when participants can learn about the research and complete screening using their own device at a place and time convenient for them. Pregnant women in an urban Midwestern hospital had access to and were comfortable using technology.
Introduction

Screening, brief intervention, and referral for treatment (SBIRT) approaches proactively address substance use in primary care settings and potentially reach those at risk, regardless of willingness to seek treatment. Large proportions of at-risk groups can be reached with SBIRT, particularly in the perinatal period where most pregnant women seek prenatal care. The consequent need for proactive screening, together with the promising efficacy of brief interventions for alcohol use [1], has led to recommendations that SBIRT be a standard element of prenatal care [2]. However, studies comparing self-report of drug use to objective indicators show that underreporting is common [3-5], especially in settings where disclosure can have heightened negative consequences such as during pregnancy [4,6,7]. The disclosure of substance use during pregnancy can be both socially stigmatizing and increase the woman’s risk for potential legal consequences. Currently, 18 states view substance use during pregnancy as child abuse, and 15 states have laws stating that health care workers are mandated reporters for drug abuse during pregnancy. Laws such as these increase the social stigma and the internal shame and guilt women may feel. This in turn limits the proportion of women who are willing to disclose substance use to their providers, suppressing disclosure and impacting the health of them and their unborn child [8-10]. This underreporting is a substantial obstacle to proactive screening efforts that seek to identify at-risk pregnant and postpartum women, the majority of whom do not seek treatment for substance use [11].

Additionally, the implementation of SBIRT approaches has been challenging. First, there are considerable time, financial, and logistic obstacles to integrating screening and brief intervention programs into ongoing medical practice [12,13]. For example, one estimate suggests that conducting all recommended prevention-related activities would take a primary care physician 4.4 hours per working day [14]. This issue is exacerbated by the fact that such services are only recently and not consistently being reimbursed by third-party payers. Second, many medical professionals express discomfort with the screening and intervention process and report doubts about its effectiveness—when voluntarily participating in a formal demonstration program [13]. This discomfort and skepticism may in part explain findings of very low levels of physician adherence to recommended brief intervention guidelines, even after training [12,14,15]. Training in brief approaches such as motivational interviewing is expensive, time-consuming, and may have modest or transient effects [16]. Technology provides an exciting option. It can be implemented consistently across patients, with minimal staff involvement, and conducted during natural waiting periods, integrating easily within the workflow of the clinic [17-19]. It has also been shown to improve disclosure of substance use in anonymous studies [20]. However, studies involving technology in health care settings often struggle with recruitment, particularly given time constraints on the part of clinic staff who must provide an initial introduction to the study. Typically, clinical trials are addressed via multisite trials using face-to-face recruitment. Despite being a time-tested gold standard, several limitations to this approach exist. First, the combination of a low base rate of substance use during pregnancy with high levels of underreporting makes recruitment lengthy and challenging even across multiple sites. Second, even multisite trials are only able to measure a limited range of participant characteristics specific to only a few geographic locations. Third, well-funded and tightly controlled trials often use methods (eg, a research assistant [RA] or study nurse) that do not readily translate to how the program could be implemented without research funding. Fourth, multisite research can also quickly become impractical if staffing each clinic with a full-time RA is required. There is increasing recognition of the need for highly pragmatic trials that take translational and implementation issues into account [21]. Research is therefore also needed on pragmatic methods for facilitating recruitment in these settings. The provision of flyers describing the study and allowing enrollment on the web is a possible solution, but relative recruitment rate for this approach, as compared to traditional approaches, is not known and is partly dependent on rates of technology ownership.

This study analyzes data exploring how to best leverage technology to identify risk during pregnancy, particularly whether different approaches in recruitment can increase disclosure. The study had 3 goals. The first goal was to obtain current substance use risk levels of women attending their prenatal care intake at a large Midwestern hospital’s outpatient clinic. The second goal was to compare the disclosure of substance use risk under 2 different recruitment strategies, in person versus via flyers, and determine recruitment rates for the 2 approaches. The last goal was to better understand the access and comfort of using smartphones and SMS text messaging for study participation. It was hypothesized that in-person recruitment would have a higher acceptance rate for study enrollment, but that participating in the study on their own device in the privacy of their home would increase disclosure.

Methods

Participants

Participants were 121 pregnant women attending a new pregnancy intake at an outpatient clinic that is part of a large health system in the Midwest. Eligibility criteria included being 18 years or older of age, understanding spoken English, and being pregnant with the intention to carry the pregnancy to term.

Recruitment

Data collection began in September 2018 and concluded in May 2019. An RA was present at the clinic on 2 half days per week;
during this time, willing participants were introduced to the RA by the nurse who was completing the intake with the patient. At all other times, clinic staff gave patients a flyer describing the study and provided a website (via QR code, along with a unique login ID) through which patients could enroll in and complete the study. The flyer was provided by intake nurses at the end of the appointment, and participants used their own device to complete the study.

**Procedures**

Participants who enrolled in the study completed a series of screening questions regarding substance use before and during pregnancy, as well as questions related to demographics, general health, and technology access. Those recruited by the RA were given a tablet to complete the study at the clinic following their intake appointment. Those enrolling in the study via the flyer used their own device and completed screening at a time and place convenient for them. In addition, those who screened positive for any substance were offered within the app to participate in a subsequent extended assessment (duration of 10 to 20 minutes) with separate consent. The app would link participants to the next screening if they agreed to participate.

**Ethical Considerations**

All procedures for the study were approved by both the university (#085518B3A) and health system (#12267) institutional review boards. The app used for data collection read aloud the consent form that explained to participants that the study has 2 parts. Electronic information sheets were used for the study. Participants agreed to participate in the study in the computerized questionnaire by clicking a box and then answering the questions. There were no physical copies of the consent form. Part I included content regarding broad health behaviors such as nutrition and sleep, as well as brief questions on smoking, alcohol, and marijuana use during the month before they became pregnant. For part II, patients who screened positive for smoking, alcohol, or marijuana use in the month before becoming pregnant were invited to complete a 15- to 20-minute survey asking additional questions about risk factors. This assessment included more sensitive information regarding substance use, traumatic experiences, partner violence, depression, and anxiety. Participants who completed part I received a US $10 Target gift card, and those who were eligible and completed part II received an additional US $20 Target gift card. No identifying information was collected until after participants completed the assessment items. Once participants completed the portions of the study they were eligible for, they were linked (within the survey) to a separate survey, where they entered their email or phone number to receive their gift cards and a copy of the consent form. The data participants gave in order to send the gift cards were kept in a separate password-protected spreadsheet from the rest of the data and were destroyed once the study was complete. Participants who completed the survey in the clinic received the consent form and gift card directly from the RA.

**Measures**

All participants were asked to complete 47 items regarding alcohol, marijuana, and tobacco use before pregnancy and during the past month, as well as questions about pregnancy and general health. These items included, but are not limited to, the following:

1. The Tolerance, Annoyed, Cut Down, Eye-Opener (T-ACE) alcohol screen [22] is a 4-item alcohol risk screening questionnaire that asks about the amount of drinks to feel high, if people have annoyed you by criticizing your drinking, if you have ever thought you should cut down, or if you need to have a drink first thing in the morning. Scores of 2 or higher result in a positive screen.

2. The Wayne Indirect Drug Use Screener (WIDUS) [23] is a 6-item screening instrument that identifies risk for drug use in the perinatal period by asking about correlates of drug use without directly asking about use. Scores above 3 are considered positive. Examples of true or false questions include “most of my friends smoke cigarettes” and “I get mad easily and feel the need to blow off steam.”

3. The National Institute on Drug Abuse (NIDA) Quick Screen [24] consists of 4 questions asking respondents to indicate the frequency with which they had 4 or more drinks in a day, use of illegal drugs, use of prescription drugs for nonmedical reasons, or use of tobacco products in the past year. The alcohol and drug use items have been validated as single-item questionnaires [25,26]. These items were adapted to evaluate use in the past month rather than the past year and to include a separate item for cannabis use.

4. Participants were also given 4 technology questions regarding technology access and use (smartphone ownership, having an SMS text messaging plan, willingness to receive SMS text messages, and willingness to participate in research via a link sent to their phone).

**Statistical Analysis**

Chi-square analyses compared differences between in-person– and flyer-based recruitment as well as differences in disclosure on the T-ACE, WIDUS, and each item of NIDA Quick Screen. Chi-square analyses used all available screening information from each participant. However, participants with missing items were dropped from that specific analysis. Two individuals had missing data for the NIDA Quick Screen binge drinking and tobacco questions. One person had missing data for the NIDA Quick Screen prescription drug and illegal drug use questions. There were no missing data for the T-ACE or WIDUS.

**Results**

**Participant Characteristics**

Study participants were primarily Black and African American (92/121, 76%) and had a mean age of 27.7 (SD 4.9) years (Table 1). Approximately half (66/121, 54.5%) of the participants had completed some education beyond high school.
Table 1. Participant race, ethnicity, and important demographic characteristics (N=121).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>27.7 (4.9)</td>
</tr>
<tr>
<td>Race and ethnicity, n (%)</td>
<td></td>
</tr>
<tr>
<td>Arabic</td>
<td>2 (1.7)</td>
</tr>
<tr>
<td>Asian</td>
<td>4 (3.3)</td>
</tr>
<tr>
<td>Black and African American</td>
<td>92 (76)</td>
</tr>
<tr>
<td>Hispanic and Latino</td>
<td>5 (4.1)</td>
</tr>
<tr>
<td>White</td>
<td>10 (8.3)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>4 (3.3)</td>
</tr>
<tr>
<td>Chose not to answer</td>
<td>4 (3.3)</td>
</tr>
<tr>
<td>High school or General Educational Development test or higher, n (%)</td>
<td>66 (54.5)</td>
</tr>
<tr>
<td>Planned pregnancy, n (%)</td>
<td>45 (37.2)</td>
</tr>
<tr>
<td>First pregnancy, n (%)</td>
<td>34 (28.1)</td>
</tr>
<tr>
<td>Legally married, n (%)</td>
<td>31 (25.6)</td>
</tr>
</tbody>
</table>

Risk Screen Positivity

Between-group differences in positivity rates were examined for 2 validated screening tools, the WIDUS and the T-ACE. A total of 20 (31%) out of 64 participants recruited through flyers screened positive for drug use risk on the WIDUS versus 9 (16%) out of 57 participants recruited at the clinic ($\chi^2=4.0; P=.05$). Additionally, a total of 24 (38%) out of 64 women recruited through flyers screened positive for alcohol risk on the T-ACE versus 11 (19%) out of 57 participants recruited at the clinic ($\chi^2=4.9; P=.03$; Table 2).

Table 2. Disclosure rates for substance risk indicators for in-person– and flyer-based recruitment methods.

<table>
<thead>
<tr>
<th>Substance risk indicator</th>
<th>In person (n=57), n (%)</th>
<th>Flyer (n=64), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIDUS&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9 (16)</td>
<td>20 (31)</td>
</tr>
<tr>
<td>T-ACE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>11 (19)</td>
<td>24 (38)</td>
</tr>
<tr>
<td>Past month alcohol binge</td>
<td>2 (4)</td>
<td>3 (5)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Past month tobacco</td>
<td>8 (14)</td>
<td>11 (18)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Past month opioid painkiller use</td>
<td>3 (5)</td>
<td>1 (2)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Past month other drugs</td>
<td>14 (25)</td>
<td>9 (14)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>WIDUS: Wayne Indirect Drug-Use Screener.
<sup>b</sup>T-ACE: Tolerance, Annoyed, Cut Down, Eye-Opener.
<sup>c</sup>n=62.
<sup>d</sup>n=63.

Disclosure of Past Month Substance Use

Chi-square analyses compared the disclosure of substance use (on the NIDA Quick Screen) for participants recruited in person versus via the flyer. Questions included the frequency of binge drinking (4 or more drinks per day), tobacco use, prescription drugs for nonmedical purposes, and illegal drugs in the past month. Each NIDA Quick Screen item was treated as dichotomous reflecting either any or no reported use (Figure 1). There were no significant differences across groups in binge drinking ($\chi^2=0.1; P=.72$), tobacco use ($\chi^2=0.3; P=.58$), prescription drug use for nonmedical reasons ($\chi^2=2.0; P=.26$), or illegal drugs ($\chi^2=2.0; P=.15$).
Recruitment Rates for Each Method of Recruitment

Of the 121 participants recruited overall, 57 were recruited directly by the RA, and 64 responded to the flyer. Nurses handed out 377 flyers resulting in 64 participants, representing 17% of those given the flyer. In contrast, of 109 patients completing a new pregnancy intake when the RA was in the clinic, 63 (57.8%) were introduced to the RA by the intake nurse, and 57 (91%) of the 63 agreed to participate and completed the screener (57/109, 52.3% of all available patients). Notably, flyer recruitment showed a stable increase over the course of the study (Table 2). This increase occurred following the introduction of a new approach, in which the RA and project coordinator began attending monthly staff meetings and updating nurses on the study progress, bringing in snacks, and building upon the relationships with the intake nurses in the clinic. This change in approach started in January, with an increase in nurse engagement and enthusiasm for the study happening in the next few months, resulting in an increase in flyer recruitment because many more were handed out. During the winter months of February and March, there were fewer intakes overall because of the weather. This decrease corresponds with an expected decrease in in-person recruitment during those months.

Technology Accessibility and Willingness to Use for Participation

Of the 57 participants who were recruited in the clinic and used the tablet provided by the RA, 56 (98%) reported owning a smartphone, and 55 (96%) reported having an SMS text message plan on their phone. In total, 44 (77%) of these participants said that they would be willing to receive SMS text messages as part of a research study, and these participants also said that they would be willing to participate in additional surveys or programs if sent a link on their phone.

Discussion

Principal Findings

This study was set up to obtain substance use risk levels for women attending prenatal care at a large Midwestern hospital’s outpatient clinic, compare 2 different recruitment methods to examine which had higher recruitment rates and disclosure rates, and document participants’ access and comfort using smartphones and SMS text messaging for study participation. Recruitment via flyers distributed by health care staff was less efficient than when those same staff introduced patients to an on-site RA (57/109, 52.3% vs 64/377, 17% enrollment). However, participants in the flyer group were more likely to report substance use risk than those in the on-site RA group (20/64, 31% vs 9/57, 16% for the WIDUS and 24/64, 38% vs 11/57, 19% for the T-ACE). Most study participants owned a smartphone (56/57, 98%) and had an SMS text message package on their phone (55/57, 96%). Additionally, of the 121 participants, 94 (77.7%) were willing to receive SMS text messages or a link to further study participation on their devices.

Despite the lower overall enrollment compared to the on-site RA, the flyer approach requires less effort for medical staff and removes the need for a full-time RA at each study clinic. The flyer approach was also associated with greater disclosure on some measures of substance use. As is often the case, maintaining regular communication with clinic staff was particularly important in the flyer-based recruitment approach. These findings suggest that eligibility determination for substance use studies may be more successful and more representative (because of the wider possible reach with the same level of staffing) when using electronic screening with flyers rather than relying on full-time staff in the clinic. Flyer-driven recruitment appears to be a practical approach, given the high levels of access to technology among the pregnant urban participants and their willingness to use their personal...
devices for research. These latter findings are consistent with national survey data suggesting that smartphone ownership rates are high [27], and research suggesting that low-income patients are willing to use their own smartphone to participate in research [28].

Limitations
The sample size, homogeneity of the sample, and preliminary nature of this research all contribute to clear limits in the generalizability of these findings. In addition, our sample size limited the ability to understand what variables may contribute to higher disclosure within the flyer recruitment group (ie, maternal age, parity, past substance use, or socioeconomic status).

Conclusions
Using electronic methods for eligibility determination appears to facilitate disclosure and, thus, recruitment efficiency. Although flyer-based approaches are less efficient than in-person recruitment with an on-site RA, they may also facilitate disclosure and can allow cost-effective recruitment at multiple sites. Even low-income patients in perinatal settings are very likely to own a smartphone and be willing to use their own device to participate in research. This method can allow for larger study samples by decreasing the amount of money needed to support full-time RAs in each recruitment site. Instead, 1 RA could be used across multiple sites, which can free up funds for a larger number of site locations. This can allow for a wider variety of participants across the country and could be more translatable and easier to replicate or continue once funding ends.

Acknowledgments
This research was supported by internal funds from the Office of Vice President of Research at Wayne State University, who were not involved in the review and approval of the paper for publication. The authors wish to gratefully acknowledge the invaluable assistance of the women who participated in this study, Ms Edith Combs, and the medical staff from the Henry Ford Health New Center One Women’s Health clinic.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


Abbreviations

NIDA: National Institute on Drug Abuse
RA: research assistant
SBIRT: screening, brief intervention, and referral for treatment
T-ACE: Tolerance, Annoyed, Cut Down, Eye-Opener
WIDUS: Wayne Indirect Drug Use Screener

https://formative.jmir.org/2024/1/e48823

JMIR Form Res 2024 | vol. 8 | e48823 | p.2545

(page number not for citation purposes)
Machine Learning–Based Approach for Identifying Research Gaps: COVID-19 as a Case Study

Alaa Abd-alrazaq1*, PhD; Abdulqadir J Nashwan2*, MSc; Zubair Shah3, PhD; Ahmad Abujaber4, PhD; Dari Alhuwail5,6, PhD; Jens Schneider3, PhD; Rawan AlSaad1, PhD; Hazrat Ali7, PhD; Waleed Alomoush8, PhD; Arfan Ahmed1, PhD; Sarah Aziz1, MSc

1AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar
2Department of Nursing, Hamad Medical Corporation, Doha, Qatar
3Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
4Nursing Department, Hamad Medical Corporation, Doha, Qatar
5Information Science Department, College of Life Sciences, Kuwait University, Kuwait, Kuwait
6Health Informatics Unit, Dasman Diabetes Institute, Kuwait, Kuwait
7Faculty of Computing and Information Technology, Sohar University, Sohar, Oman
8School of Information Technology, Skyline University College, Sharjah, United Arab Emirates
*these authors contributed equally

Abstract

Background: Research gaps refer to unanswered questions in the existing body of knowledge, either due to a lack of studies or inconclusive results. Research gaps are essential starting points and motivation in scientific research. Traditional methods for identifying research gaps, such as literature reviews and expert opinions, can be time consuming, labor intensive, and prone to bias. They may also fall short when dealing with rapidly evolving or time-sensitive subjects. Thus, innovative scalable approaches are needed to identify research gaps, systematically assess the literature, and prioritize areas for further study in the topic of interest.

Objective: In this paper, we propose a machine learning–based approach for identifying research gaps through the analysis of scientific literature. We used the COVID-19 pandemic as a case study.

Methods: We conducted an analysis to identify research gaps in COVID-19 literature using the COVID-19 Open Research (CORD-19) data set, which comprises 1,121,433 papers related to the COVID-19 pandemic. Our approach is based on the BERTopic topic modeling technique, which leverages transformers and class-based term frequency-inverse document frequency to create dense clusters allowing for easily interpretable topics. Our BERTopic-based approach involves 3 stages: embedding documents, clustering documents (dimension reduction and clustering), and representing topics (generating candidates and maximizing candidate relevance).

Results: After applying the study selection criteria, we included 33,206 abstracts in the analysis of this study. The final list of research gaps identified 21 different areas, which were grouped into 6 principal topics. These topics were: “virus of COVID-19,” “risk factors of COVID-19,” “prevention of COVID-19,” “treatment of COVID-19,” “health care delivery during COVID-19,” “and impact of COVID-19.” The most prominent topic, observed in over half of the analyzed studies, was “the impact of COVID-19.”

Conclusions: The proposed machine learning–based approach has the potential to identify research gaps in scientific literature. This study is not intended to replace individual literature research within a selected topic. Instead, it can serve as a guide to
formulate precise literature search queries in specific areas associated with research questions that previous publications have earmarked for future exploration. Future research should leverage an up-to-date list of studies that are retrieved from the most common databases in the target area. When feasible, full texts or, at minimum, discussion sections should be analyzed rather than limiting their analysis to abstracts. Furthermore, future studies could evaluate more efficient modeling algorithms, especially those combining topic modeling with statistical uncertainty quantification, such as conformal prediction.

(JMIR Form Res 2024;8:e49411) doi:10.2196/49411

KEYWORDS
research gaps; research gap; research topic; research topics; scientific literature; literature review; machine learning; COVID-19; BERTopic; topic clustering; text analysis; BERT; NLP; natural language processing; review methods; review methodology; SARS-CoV-2; coronavirus; COVID

Introduction

Background

Scientific research relies on applying systematic scientific methods and actions to increase knowledge in fields or specific topics [1]. One essential first step in engaging in scientific research is identifying research gaps [2], where insufficient data, knowledge, or understanding limit our ability to draw conclusions in a given field or topic [3]. Research gaps can also be referred to as unanswered questions that have not yet been addressed or are underexplored in the existing body of knowledge, either due to a lack of studies or inconclusive results [4,5]. Research gaps can also serve as a starting point for research as well as motivate further research [6]. Researchers have classified research gaps into seven categories [3,6,7]: (1) evidence gaps—where contradictions in the findings of the previous research exist; (2) knowledge gaps—where knowledge may either not exist in the literature or the results deviated from what was expected; (3) practical-knowledge conflict gaps—where the goal is to discover the reasons and scope of differences between professionals’ behaviors versus their advocated behavior; (4) empirical gaps—where there is a need to empirically evaluate and verify research findings or propositions; (5) methodological gaps—where shortcomings may arise due to having a single methodology influencing the research results; (6) theoretical gaps—related to examining gaps that exist in theories and their models and compare them with prior research; and (7) population gaps—where a population is not adequately represented or underresearched in prior studies.

The sheer volume and accelerated pace of scientific research output present both opportunities and challenges. Identifying potential interventions, best practices, and policy recommendations, all backed up by evidence is made easier by the wealth of information available. However, researchers face challenges with staying up-to-date with the latest findings, identifying redundancies in research, and objectively identifying research gaps that need to be addressed. Therefore, it becomes paramount to accurately identify the research gaps to advance our understanding of the issue or topic; better use the allocation of resources, and better inform evidence-based policy making.

Traditionally, several methods are used to identify research gaps, including literature reviews, systematic reviews, expert opinions, and consensus-building activities (eg, developing guidelines). Yet, such methods require an intensive time commitment, are prone to bias, and can be labor intensive. Additionally, such methods may not be suitable to address issues, where the evidence or research subject is rapidly increasing in volume and pace or is time sensitive (eg, COVID-19). Consequently, there is a need for innovative and scalable approaches to systematically assess existing literature, identify research gaps, and prioritize areas for further study in the topic or field of interest.

Machine learning (ML) techniques have demonstrated great potential applications of scientific insights and discoveries by addressing challenges related to information retrieval, knowledge discovery, and natural language processing [8]. ML is a branch of computational science and a subset of artificial intelligence (AI); it focuses on the development of algorithms that enable machines to learn from [9-12] and make predictions or decisions based on data, without being directly programmed [13]. Broadly, ML algorithms can “learn” through supervised learning, unsupervised learning (eg, reinforcement learning), or a mixture thereof, referred to as semisupervised learning [14,15].

Research Problem and Aim

In the context of identifying research gaps, ML techniques may facilitate the discovery of research gaps by analyzing large volumes of scientific evidence in a systematic, scalable, and efficient manner. To understand the current status quo of scientific evidence available, several studies leveraged ML to perform natural language processing, bibliometric analysis, and text mining, which have yielded promising results in several domains, including health care [9-12], social sciences [16-19], and environmental sciences [20]. However, the application of ML and its potential for identifying research gaps in rapidly evolving fields remains underexplored. To the best of our knowledge, there are no previous studies that leveraged ML-based techniques for identifying research gaps.

This paper aims to propose an ML-based approach to detect research gaps in the literature. In this work, we use the novel COVID-19 pandemic, caused by the SARS-CoV-2 virus, as a case study due to the fast and time-critical pace it evolved, along with the urgent need to comprehend this global pandemic. Since its emergence in late 2019, COVID-19 has had a profound global impact which not only halted many activities, triggered economic fallouts, and caused significant hardships, but it also claimed more than 6.8 million lives. In turn, the scientific community united to address the various challenges presented by the global spread of the virus.

https://formative.jmir.org/2024/1/e49411
Through unprecedented levels of scientific activities, researchers generated a vast amount of research output, with thousands of research papers being published each month, which discuss various aspects of the virus, its transmission, diagnosis, treatment, and prevention strategies [9,21,22]. However, with the rapid pace of information generation and dissemination, there is an increased risk of overlooking research gaps or underexplored areas that may be critical to understanding and mitigating the virus’s impact. Thus, this aims to propose an ML-based approach to detect research gaps in the literature using COVID-19 as a case study.

Methods

Study Data Selection
We used the COVID-19 Open Research (CORD-19) data set, produced by the Allen Institute for AI and made available on the Kaggle platform [23]. It was updated every week until June 2, 2022, to include the most recently published COVID-19 papers. We used the final version, which contains 1,121,433 entries. The search terms used by Allen Institute for AI to retrieve these studies were “Coronavirus” OR “Corona virus” OR “COVID-19” OR “2019-nCoV” OR “SARS-CoV” OR “Severe Acute Respiratory Syndrome” OR “MERS-CoV” OR “Middle East Respiratory Syndrome” [24]. The sources of studies in the CORD-19 data set were PubMed Central, PubMed, the World Health Organization’s (WHO) COVID-19 Database, arXiv, bioRxiv, and medRxiv [24]. The data set includes a CSV file containing meta-information about all the papers, such as article DOI, CORD UID, PMCID, PUBMED ID, title, abstract, journal, authors, and publication date. We removed duplicate papers, entries with empty and non-English abstracts, and papers published before January 1, 2020. We selected abstracts containing the keywords “novel coronavirus,” “coronavirus 2019,” “2019-nCov,” “COVID-19,” “COVID 2019,” “severe acute respiratory syndrome coronavirus 2,” and “SARS-COV-2” to include only COVID-19–related papers.

Data Preprocessing
We cleaned the abstracts and removed nonalphanumeric characters, punctuations, and sectioning keywords “BACKGROUND,” “OBJECTIVE,” “METHOD,” “RESULT,” “CONCLUSION.” We used the Python programming language in a Jupyter Notebook environment and Python libraries such as pandas, NumPy, langdetect, re, string, and TextBlob. We then searched abstracts that mentioned any term related to research gap: “unknown,” “not known,” “little is known,” “unrevealed,” “uncertain,” “undetermined,” “understudied,” “unexplored,” “not fully understood,” “literature gap,” “research gap,” “knowledge gap,” “future studies,” “future research,” “research problem,” “more studies,” “more research,” “further studies,” and “further research.” We decided to use these terms in an abstract search rather than a full-text search for 2 reasons. First, using these terms in a full-text search may lead to the inclusion of a significant number of unrelated studies, increasing the likelihood of inaccurately identifying research gaps that are not pertinent to the subject of interest. Second, the CORD-19 data set that we used in this study does not contain the full text of the studies. With this process, we identified 33,206 abstracts of scholarly papers published after January 1, 2020, related to COVID-19 and containing the gap words. We analyzed these abstracts, and from each abstract, we extracted 3 sentences: 1 sentence that includes the gap word, 1 sentence before, and 1 sentence after the sentence containing the gap word. We used a full stop (".") as a sentence marker. We converted the selected sentences to lowercase text. Next, we used the Python NLTK library to remove the stop words and tokenize the sentences. Finally, we used the clean sentences after the preprocessing steps for clustering.

Analysis

Overview
For clustering the sentences into semantically similar topics, we used the BERTopic algorithm [25]. The BERTopic algorithm is an unsupervised learning algorithm for topic modeling. It uses the Bidirectional Encoder Representations from Transformers (BERT). BERTopic does not require labeled data as it extracts topics from an input text in a supervised way [26]. BERTopic gained popularity due to its potential to capture context-aware information in a given input text and does not rely on a predefined number of topics [26]. Besides topic modeling, BERTopic is also useful in the clustering of documents and text summarization [26]. The BERTopic topic modeling method produces dense clusters by combining class-based term frequency-inverse document frequency (TF-IDF) and transformers (BERT embeddings). In addition, it makes it simple to comprehend and visualize the generated topics. The 3 stages in the BERTopic algorithm are presented in Figure 1 and discussed in the next subsections.
Figure 1. Stages of the BerTopic modeling algorithm. c-TF-IDF: class-based term frequency-inverse document frequency; MMR: maximum marginal relevance; HDBSCAN: hierarchical density-based spatial clustering of applications with noise; UMAP: uniform manifold approximation and projection.

**Embedding Documents**

In this stage, the BERTopic algorithm extracts document embeddings using BERT sentence transformers. Each document embedding is generated using word embeddings, which represent each word in the document in multidimensional space. It ensures that words with related meanings have comparable representations. This way, words are represented by numbers in a vector space, where vectors are defined by TF-IDF weights. TF-IDF describes the importance of a term relative to a document in a corpus. Neural networks are the primary foundation of embedding models. Typically, word embeddings are used to compute document embedding into 2 phases. The word embedding is first applied to every word in the text and then the word embeddings are aggregated by averaging over each dimension to produce document embeddings.

**Clustering Documents**

Documents embeddings produced in the previous step are usually very sparse. Therefore, our first stage of the analysis uses uniform manifold approximation and projection to decrease the dimensionality of the embeddings [27]. Then, we use the hierarchical density-based spatial clustering of applications with a noise approach to cluster reduced embeddings and produce clusters of texts with comparable semantic properties.

**Representing Topics**

In this stage, class-based TF-IDF weights are used to extract topics. These topics are reduced and further improved by finding the coherence of words using maximum marginal relevance. We used BERTopic to cluster the extracted sentences and found 191 clusters. Then, we applied the topics reduction technique of BERTopic and found a total of 50 clusters. We assigned labels to each cluster by checking their representative words. When it was challenging to assign labels to a cluster based on its representative words, we reviewed sentences containing the specified research gap terms, along with the sentence before and after in most studies within that cluster. After that, we merged clusters that had similar labels to identify 21 unique labels (research gaps). Finally, we grouped these clusters into 6 broader categories. Multimedia Appendix 1 shows the code used for the analysis of bibliographic data to identify research gaps.

**Results**

**Search Results**

By June 2, 2022, the CORD-19 data set contained 1,121,433 papers (Figure 2). Of those, we excluded 1,088,227 papers for the following reasons: (1) abstracts were unavailable (n=300,540); (2) the papers were published before January 1, 2020 (n=225,964); (3) the papers were written in a language other than English (n=6499); (4) papers did not contain search terms related to COVID-19 (n=195,498); and (5) papers did not contain search terms related to research gap (n=359,726). Consequently, we included 33,206 papers in the analysis of this study.
Results of Gap Identification

Overview

As mentioned earlier, 191 clusters of the 33,206 papers were generated by our analysis. Then, the number of clusters was reduced to 50 when we applied a topics reduction technique. We deleted 6 clusters as we could not identify the research gap from them. Further, 23 clusters were merged with other clusters as the same research gaps were identified in these clusters. Overall, we identified 21 different research gaps from 4646 papers. As shown in Table 1, these research gaps were grouped into 6 topics, which are discussed in the next subsections.
Table 1. Topics and the corresponding subtopics identified in this study (N=4646).

<table>
<thead>
<tr>
<th>Topics and subtopics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic 1: virus of COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Origin of COVID-19</td>
<td>318 (6.8)</td>
</tr>
<tr>
<td>Emerging variants</td>
<td>202 (4.3)</td>
</tr>
<tr>
<td>Transmission of COVID-19</td>
<td>75 (1.6)</td>
</tr>
<tr>
<td>Role of the immune system</td>
<td>52 (1.1)</td>
</tr>
<tr>
<td><strong>Topic 2: risk factors of COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>77 (1.7)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>65 (1.4)</td>
</tr>
<tr>
<td><strong>Topic 3: prevention of COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Vaccination</td>
<td>447 (9.6)</td>
</tr>
<tr>
<td>Precautionary measures</td>
<td>124 (2.7)</td>
</tr>
<tr>
<td>Wastewater surveillance</td>
<td>60 (1.3)</td>
</tr>
<tr>
<td><strong>Topic 4: treatment of COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Medications of COVID-19</td>
<td>170 (3.7)</td>
</tr>
<tr>
<td>Herbal medicine</td>
<td>66 (1.4)</td>
</tr>
<tr>
<td>Support system</td>
<td>55 (1.2)</td>
</tr>
<tr>
<td><strong>Topic 5: health care delivery during COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Telehealth</td>
<td>305 (6.6)</td>
</tr>
<tr>
<td>Surgeries</td>
<td>158 (3.4)</td>
</tr>
<tr>
<td>Organ transplantation</td>
<td>92 (2)</td>
</tr>
<tr>
<td><strong>Topic 6: impact of COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Health complications</td>
<td>1552 (33.4)</td>
</tr>
<tr>
<td>Special groups</td>
<td>434 (9.3)</td>
</tr>
<tr>
<td>Education</td>
<td>136 (2.9)</td>
</tr>
<tr>
<td>Media and communication</td>
<td>109 (2.3)</td>
</tr>
<tr>
<td>Mortality</td>
<td>87 (1.9)</td>
</tr>
<tr>
<td>Food security</td>
<td>62 (1.3)</td>
</tr>
</tbody>
</table>

**Topic 1: Virus of COVID-19 (n=647, 13.9%)**

**Origin of COVID-19**

There are 318 studies related to the origin of COVID-19. Despite various theories and pieces of evidence, the exact source of SARS-CoV-2 remains unknown according to the majority of studies on this topic.

**Emerging Variants**

According to the 202 studies on this topic, the full extent of the impact of new variants on the course of the pandemic and the effectiveness of current prevention and treatment strategies on the new variants are still not known. Furthermore, it is uncertain how the emergence of new variants may affect the duration of immunity after recovery. In addition, it is not clear how accurate the prediction models are in predicting the emergence of new variants and their impact.

**Transmission of COVID-19**

A total of 75 studies have revealed several unknown aspects of the virus such as the role of asymptomatic carriers in the spread of the virus, the true level of global spread, and the extent to which COVID-19 may be circulating undetected in some areas, and the influence of environmental factors (e.g., temperature and humidity) on the transmission of COVID-19.

**Role of the Immune System**

In 52 studies related to this topic, several research gaps were identified. Specifically, while it is known that the virus elicits an immune response, it is still unclear the specifics of how the immune system responds, the duration of immunity after recovery, the relationship between the severity of illness and the strength of the immune response, and how the immune response may differ between individuals.
Topic 2: Risk Factors of COVID-19 (n=142, 3.1%)

Obesity
From 77 studies related to this topic, we found that the relationship between obesity and COVID-19 is still largely a mystery. While some studies have suggested that obesity may increase the risk of severe COVID-19 outcomes, the exact mechanisms behind this are still unknown. Additionally, it is unclear whether weight loss can reduce the risk of severe illness and death from COVID-19 in individuals with obesity. Furthermore, there is no consensus on the optimal BMI cut-off for identifying individuals with obesity who are at increased risk of severe illness and death from COVID-19.

Ethnicity
A total of 65 studies identified a significant research gap in understanding the relationship between ethnicity and COVID-19. Despite the numerous studies conducted, the reasons for these disparities remain largely unknown.

Topic 3: Prevention of COVID-19 (n=631, 13.6%)

Vaccination
A total of 447 studies have revealed that the long-term efficacy and safety of COVID-19 vaccines remain largely unknown. Although initial clinical trials have demonstrated promising results, it is uncertain how long the protection provided by the vaccines will last and what potential long-term side effects may be. Further, vaccine hesitancy and the factors that contribute to it are not fully understood. Moreover, the safety of COVID-19 vaccines in patients with chronic diseases (eg, chronic hepatitis, diabetes, heart failure, renal failure, and epilepsy) remains unknown.

Precautionary Measures
A total of 124 studies were related to precautionary measures implemented to prevent the spread of the virus. According to these studies, the effectiveness of many precautionary measures such as social distancing, wearing masks, hand hygiene, respiratory etiquette (eg, covering mouth and nose when coughing and sneezing), and ventilation systems is still unknown. Further, the duration of quarantine or isolation for individuals with COVID-19 remains uncertain.

Wastewater Surveillance
In total, 60 studies were related to using wastewater surveillance to detect the presence of the virus in communities. According to these studies, it is unclear how accurately wastewater surveillance can predict or track COVID-19 outbreaks and what the potential false-positive or false-negative rates may be.

Topic 4: Treatment of COVID-19 (n=291, 6.3%)

Medications of COVID-19
The role of medications (eg, remdesivir and corticosteroids) in managing severe cases of COVID-19 was discussed in 170 studies. According to these studies, there are still many unknowns about the medication for COVID-19, including the potential for long-term effects on the heart, lungs, and other organs; the optimal treatment protocol (eg, timing, dose, and duration) for different stages of the disease and age groups; the effectiveness of existing medications against new variants and new medications against multiple variants; the safety and efficacy of a combination of medications; the best combinations and dosages; the safety and efficacy of medication in children; and the effect of the combination of medication and dietary supplements.

Herbal Medicine
Herbal medicine was the main topic in 66 studies. According to these studies, much remains unknown about the effectiveness, safety, standardization, dosage, and potential interactions with other treatment regimens.

Support System
A total of 55 studies were related to support systems that provide several services (eg, social support, mental health, and financial assistance) to individuals affected by COVID-19. According to these studies, the effectiveness and long-term impacts of support systems on patients with COVID-19 remain largely unknown.

Topic 5: Health Care Delivery During COVID-19 (n=555, 11.9%)

Telehealth
The COVID-19 pandemic has seen a dramatic surge in the use of telehealth; yet, a total of 305 studies have revealed that there are still many unknowns about its effectiveness and accessibility. These unknowns have the potential to create new privacy and security risks, as well as to impact health care disparities.

Surgeries
According to the 158 studies discussing this topic, there is still no consensus on planning to maintain surgical care preparedness in ongoing and future pandemics. This includes understanding the risk of transmission, perioperative testing criteria, postoperative outcomes in specific populations, and effective strategies for resource allocation.

Organ Transplantation
A total of 92 studies have revealed that there is still much to be discovered about the best way to manage organ transplantation during the COVID-19 pandemic. Specifically, there is still much to be uncovered about COVID-19 in relation to organ transplant recipients, including identifying risk factors, developing strategies to minimize transmission, understanding outcomes for transplant recipients who contract COVID-19, and determining the best treatment for COVID-19 in organ transplant recipients considering their immunocompromised status.

Topic 6: Impact of COVID-19 (n=2380, 51.2%)

Health Complications
From 1552 studies, we found that the long-term health complications of COVID-19 remain largely a mystery. To be more precise, there is a lack of evidence on the long-term impact of COVID-19 on the respiratory system (eg, asthma and chronic obstructive pulmonary disease), cardiovascular system (eg, hypertension, deep vein thrombosis, and pulmonary embolism), neurological system (eg, confusion, dizziness, and headache), endocrine system (eg, diabetes, hepatic system (eg, liver injury),
Special Groups

According to 434 studies, the impact of COVID-19 on special groups is still largely unknown. Specifically, there are still many unanswered questions related to the impact of COVID-19 on pregnancy. For example, what is its impact on the risk of preterm birth and respiratory distress? What is the long-term effect on fetuses and babies? And how is the virus transmitted from mother to baby during pregnancy, labor, or delivery? Further, there is still much to learn about the impacts of the virus on pediatric populations and patients with cancer in terms of diagnosis, treatment, and complications. The long-term effects of COVID-19 on health care workers remain uncovered.

Education

The COVID-19 pandemic has profoundly impacted education, with schools and universities shutting down or moving to web-based learning to slow the spread of the virus. A total of 136 studies on this subtopic revealed that there is still much unknown about the impact of COVID-19 on education.

Media and Communication

The COVID-19 pandemic has had a profound impact on media and communication, and while much research and analysis have been conducted, there are still many unknowns. A total of 109 studies explored this topic. From these studies, we found that there are still many unanswered questions related to media and communication. For example, Will the increased media consumption habits that have emerged during the pandemic persist in the long term, or will people revert to their prepandemic habits? How will media organizations adapt to the financial challenges posed by the pandemic, and will it lead to long-term changes in the media landscape? What will be the long-term effects of social media on the spread of misinformation and society as a whole? How will the increased reliance on digital communication tools, such as video conferencing and messaging apps, impact our relationships and social dynamics in the long term?

Mortality

A total of 87 studies were related to the mortality of COVID-19. According to these studies, there is a lack of evidence on factors affecting mortality rates over time and on whether new strains of the virus are more deadly or have different mortality rates than the original strain.

Food Security

In total, 62 studies have revealed that the impacts of COVID-19 on food security remain largely unknown. While some studies have suggested that the pandemic may worsen food insecurity in certain populations, the true extent of the problem is still unclear. In addition, the long-term effects of food insecurity during the pandemic on health and well-being are yet to be determined.

Discussion

Principal Findings

This study proposed an ML-based method to identify research gaps in the literature. We used COVID-19 literature as a case study. Our proposed method enabled us to identify research gaps in 21 subtopics that were grouped into 6 topics. The largest topic that was identified in more than half of the analyzed studies is the “impact of COVID-19” (topic 6). This is hardly surprising, since we, as a society, are still struggling to come to terms with the long-lasting impact of the pandemic, which continues to affect many aspects of our lives. Long-term health complications are still poorly understood, given the relatively short time since the initial outbreak, which is less than 4 years. In addition, understanding the precise impact of the disease on body systems and organs proves challenging given that it is a complex disease affecting multiple organs, there is a lack of data about the virus, and its impact may vary among different populations.

The second largest topic identified in this study is the “virus of COVID-19” (topic 1). The lack of knowledge about the COVID-19 virus is attributed to its novelty as a new strain of coronavirus not previously identified in humans. Further, tracing the origins of zoonotic diseases can be difficult as they may pass through multiple animal hosts before reaching humans. Moreover, due to its rapid evolution into new variants, studying each variation of the COVID-19 virus within a short timeframe poses a significant challenge.

Topic 3 (ie, prevention of COVID-19) was the third largest topic in this paper. The lack of knowledge in this area may be attributed to the limited long-term data on vaccine safety and efficacy. Additionally, the emergence of new variants may impact vaccine efficacy, raising concerns about the effectiveness of the current vaccines against these new variants. Further, a significant challenge in assessing the long-term effects of the COVID-19 vaccine is vaccine hesitancy, which is not fully understood and can hinder data gathering on long-term effects.

The COVID-19 pandemic has impacted health care delivery, and this was topic 5 in this study. One of the subtopics in this topic is telehealth, which has seen a significant surge. Telehealth is a relatively new technology that has only recently become widely available. As a result, there may not yet be enough data available to fully evaluate its effectiveness. Telehealth is often used in conjunction with other health care interventions, such as in-person visits or medications, which can make it difficult to separate the impact of telehealth from other factors. Importantly, there may be biases in the types of patients who are most likely to use telehealth. Thus, it is difficult to generalize findings to the broader population. Another subtopic in this topic is organ transplantation. Research gaps related to this subtopic may be attributed to the following reasons: (1) while there have been some studies on the impact of COVID-19 on organ transplant patients, the number of patients in these studies is often relatively small. These limited data can make it difficult to draw definitive conclusions about the impact of COVID-19 on this population; (2) organ transplant patients are a heterogeneous population, and the impact of COVID-19 may vary depending on factors such as age, comorbidities, and the
type of organ transplant; (3) organ transplant patients may be at increased risk of complications from COVID-19 due to their underlying health conditions and the immunosuppressant drugs they take to prevent organ rejection. Therefore, it can be difficult to isolate the impact of COVID-19 from these other factors.

We found that there are research gaps related to the treatment of COVID-19 (topic 4). This may be due to several factors. First, addressing research gaps related to this topic needs many rigorous clinical trials, which usually take several years to complete all 3 phases before the licensing stage. Second, such clinical trials are very expensive, and therefore only researchers with large funds can carry out these trials.

Critical gaps in our understanding of the relationship among obesity, ethnicity, and severe COVID-19 outcomes exist (topic 2). The relationship among obesity, ethnicity, and COVID-19 outcomes is likely influenced by a complex interplay of biological, behavioral, and environmental factors, making it difficult to determine the precise nature and extent of these relationships. Disentangling the effects of these different factors on health outcomes can be challenging, and limitations in the quality and availability of data on obesity and ethnicity in certain populations can make it difficult to accurately measure the relationship between these factors. Moreover, known statistical challenges are related to controlling confounding factors such as age, sex, and comorbidities that affect the nature, behavior, and interpretation of the relationship among these factors.

In February 2020, a 2-day meeting organized by the WHO brought together more than 400 participants worldwide [28]. The objective was to develop a research road map that would facilitate and expedite global research efforts aimed at controlling the transmission of COVID-19 [28]. The research road map identified many knowledge gaps and grouped them into 8 areas. All research gaps identified in this review were mentioned in the road map except for topic 6 (impact of COVID-19). This may be attributed to the fact that this topic was less important at that stage of the pandemic (2 months after the onset of COVID-19).

After 3 months of the WHO research road map, a mixed methods study was conducted on 4087 participants (researchers, policy makers, health care workers, etc) to check which of the early WHO road map priorities are still most pressing and identify any newly emerging priorities that warrant attention [29]. The study revealed that the WHO research road map is still globally applicable. However, it identified a number of new research priorities that align with the evolving nature of the pandemic and provide insights into areas where knowledge gaps exist. One of these new research priorities is the impact of COVID-19, which aligns with topic 6 in our study.

About 10 months after the onset of COVID-19, a team from the WHO Southeast Asia Region conducted a web-based survey of 48 experts to identify COVID-19 research priorities in the Southeast Asia Region [30]. The study identified 27 research priorities, which include all 6 research topics identified in this study.

Our findings are also in agreement with previous work [31] that similarly used the CORD-19 data set to determine research priorities during the COVID-19 pandemic. The earlier study identified 10 hotspots, 4 of which overlap with our identified research gaps: virus of COVID-19, risk factors of COVID-19, prevention of COVID-19, and treatment of COVID-19. Notably, our analysis further includes 2 additional topics, which are health care delivery during COVID-19 and the impact of COVID-19. The earlier study also identified additional areas: nursing and health care, diagnosis and testing, drugs and vaccines, social psychology, infection process, and clinical characteristics. The discrepancy between the 2 sets of findings may be attributed to the period when the studies were conducted. The prior study was conducted during the initial phase of COVID-19-related research (January to September 2020), during which the emphasis was predominantly on diagnosis, testing, infection processes, and clinical characteristics. At that early stage, the short- and long-term impacts of COVID-19 were less clear, as was its effect on health care delivery.

Limitations
This study has several limitations that need to be considered when interpreting the results. First, clusters generated by the BERTopic modeling algorithm were subject to noise. In other words, we noticed that several studies in a cluster are not relevant to that cluster. Therefore, not all studies on a topic reported a research gap related to that topic.

Second, the studies included in our analysis were limited to those published up to June 2022, given the Allen Institute for AI stopped updating the data set, and new research has been published since then. Therefore, it is likely that several research gaps identified in this review have been addressed and new research gaps have emerged.

Third, it is likely that this study missed other important research gaps for several reasons: (1) studies in the CORD-19 data set were retrieved from only 5 databases; therefore, the CORD-19 data set did not include studies from other common databases such as Scopus, Web of Science, Embase, and PsycINFO; (2) our analysis relied on abstracts rather than full texts, in which especially introduction and discussion sections commonly identify research gaps; (3) this study only considers papers in English. Consequently, potential insights from studies published in other languages may have been overlooked; and (4) we cannot rule out that we missed some terms relevant to research gaps; therefore, it is likely that many studies relevant to this work were not included in the analysis.

Fourth, we do not attempt a trend analysis in favor of a compact and concise overview, and we do not yet provide automated means to analyze original research gaps that have been addressed since the publication of a given paper. However, we believe that the presented analysis is still useful as we observe clear “hot topics” that resonate with earlier thematic research areas defined by the WHO which are extremely unlikely to have been researched fully since formulating the gap.

Practical and Research Implications
This study’s aim is not to generate research topics automatically that, when worked on, guarantee impact or publication. Instead, we scope the landscape of research gaps outlined in the literature. This study should therefore be taken as a help to
identify and prioritize “hot” topics that need addressing. This study also does not seek to replace individual literature research in a chosen topic, but it can serve as a guide to formulate specific literature search queries in specific areas related to research questions left as future work by prior publications. Therefore, literature reviews or scoping reviews are still required. Nevertheless, we anticipate that with the advent of more advanced techniques like Large Language Models, the performance of such approaches will enhance in the foreseeable future, subsequently reducing the necessity for literature reviews or scoping reviews to identify research gaps.

To overcome the above-mentioned limitations of our approach, thereby improving the identification of gaps, future research should use an up-to-date list of studies that are retrieved from the most common databases in the target area (eg, MEDLINE, Scopus, Web of Science, Embase, PsycINFO, IEEE Xplore, and ACM Digital Library). If practical, researchers should also analyze full texts, or at least discussion sections, rather than only abstracts. Moreover, additional terms related to research gaps (eg, limited evidence, inconclusive findings, and insufficient evidence) should be used to identify the relevant studies and sentences appropriate for analysis. Further, there is a need to improve the performance of the BERTopic modeling algorithm in clustering studies, specifically with respect to removing outliers from the clustering. From a technical perspective, this may require further research to combine topic modeling with statistical uncertainty quantification, such as conformal prediction [32,33].

The process of clustering text documents is an essential technique used in the text mining area, as well as in a variety of applications including ML and pattern recognition. Since clustering text documents is an optimization problem, several meta-heuristics (MH) optimization algorithms have been presented as possible solutions to this nondeterministic polynomial-time hard problem. However, while obtaining the best solution, individual optimization MH algorithms may run into serious problems including poor convergence and being stuck in local optima. To address these issues, the hybridization concept was applied to combine the strengths of 2 hybrid search methods (ie, MH algorithms) and so avoid their weaknesses.

From the dominance of topic 6, we see that most research is needed to understand the long-term effects of the pandemic. Especially where COVID-19 “temporary” solutions have become established practice (telehealth, hybrid education, the logistics of resilient food supply, etc), more research is needed to answer questions regarding the efficacy and sustainability of such solutions. Long-term health complications will continue to be a topic for quite some time in the future, given that COVID-19 has only been studied for less than 4 years. As new complications and variants emerge, this subtopic arguably has the highest potential to serve humanity and create a real impact.

**Conclusions**

This paper showed that ML has the potential to identify research gaps in scientific literature. Our proposed method identified research gaps in 21 subtopics that were grouped into 6 topics: virus of COVID-19, risk factors of COVID-19, prevention of COVID-19, treatment of COVID-19, health care delivery during COVID-19, and impact of COVID-19. This study is not intended to replace individual literature research within a selected topic. Instead, it can serve as a guide to formulate precise literature search queries in specific areas associated with research questions that previous publications have earmarked for future exploration. Future research should leverage an up-to-date list of studies that are retrieved from the most common databases in the target area. When feasible, full texts or, at minimum, discussion sections should be analyzed, rather than limiting their analysis to abstracts. Moreover, additional terms related to research gaps should be used to identify the relevant studies and sentences appropriate for analysis. Furthermore, future studies could evaluate more efficient modeling algorithms, especially those combining topic modeling with statistical uncertainty quantification such as conformal prediction.

**Data Availability**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Authors’ Contributions**

AA-A, AJN, and ZS developed the protocol for this paper. Data collection and preprocessing were carried out by ZS. Data analysis was performed by AA-A, AJN, and ZN. DA and WA wrote the introduction. The Methods section was written by ZS and HA. The Results section was written by AA-A and AJN. The Discussion and Conclusion sections were written by all authors. The paper was revised critically for important intellectual content by all authors. All authors approved the manuscript for publication and agree to be accountable for all aspects of the work.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

The code used for analysis of bibliographic data to identify research gaps.

[DOCX File , 12 KB - formative_v8i1e49411_app1.docx ]

**References**


Feasibility and Acceptability of Web-Based Structured Oral Examinations for Postgraduate Certification: Mixed Methods Preliminary Evaluation

Vanessa Burch¹,², MD, PhD; Jessica McGuire², MD; Eric Buch¹,³, MSc, MD; Mike Sathekge¹,³, MD, PhD; Francis M'bouaffou¹, MSc; Flavia Senkubuge¹,³, MD, PhD; Johannes Fagan¹,², MD

¹The Colleges of Medicine of South Africa, Cape Town, South Africa
²The University of Cape Town, Cape Town, South Africa
³University of Pretoria, Pretoria, South Africa

Corresponding Author:
Vanessa Burch, MD, PhD
The Colleges of Medicine of South Africa
17 Milner Rd
Cape Town, 7700
South Africa
Phone: 27 216899533
Email: vanessa.burch@cmsa.co.za

Abstract

Background: The COVID-19 pandemic disrupted postgraduate certification examinations globally. The Colleges of Medicine of South Africa continued hosting certification examinations through the pandemic. This was achieved by effecting a rapid transition from in-person to web-based certification examinations.

Objective: This formative evaluation explored candidates’ acceptability of web-based structured oral examinations (SOEs) hosted via Zoom (Zoom Communications Inc). We also reported the audiovisual quality and technical challenges encountered while using Zoom and candidates’ overall experience with these examinations conducted during the early part of the COVID-19 pandemic. Additionally, performance in web-based certification examinations was compared with previous in-person certification examinations.

Methods: This mixed methods, single-arm evaluation anonymously gathered candidates’ perceptions of web-based SOE acceptability, audiovisual quality, and overall experience with Zoom using a web-based survey. Pass rates of web-based and previous in-person certification examinations were compared using chi-square tests, with a Yates correction. A thematic analysis approach was adopted for qualitative data.

Results: Between June 2020 and June 2021, 3105 candidates registered for certification examinations, 293 (9.4%) withdrew, 2812 (90.6%) wrote, and 2799 (99.9%) passed, and 1525 (54.2%) were invited to a further web-based SOE. Examination participation was 96.2% (n=1467). During the first web-based examination cycle (2020), 542 (87.1%) of 622 web-based SOE candidates completed the web-based survey. They reported web-based SOEs as fair (374/542, 69%) and adequately testing their clinical reasoning and insight (396/542, 73.1%). Few would have preferred real patient encounters (173/542, 31.9%) or in-person oral examinations (152/542, 28%). Most found Zoom acceptable (434/542, 80%) and fair (396/542, 73.1%) for hosting web-based SOEs. SOEs resulted in financial (434/542, 80%) and time (428/542, 79%) savings for candidates. Many (336/542, 62%) supported the ongoing use of web-based certification examinations. Only 169 technical challenges in using Zoom were reported, which included connectivity-related issues, poor audio quality, and poor image quality. The thematic analysis identified 4 themes of positive and negative experiences related to web-based SOE station design and content, examination station environment, examiner-candidate interactions, and personal benefits for candidates. Our qualitative analysis identified 10 improvements for future web-based SOEs. Candidates achieved high pass rates in web-based certification examinations in 2020 (1583/1732, 91.39%) and 2021 (850/1067, 79.66%). These were significantly higher (2020: N=8635; \( \chi^2 = 667; \ P < .001 \); 2021: N=7988; \( \chi^2 = 178; \ P < .001 \)) than the previous in-person certification examination pass rate of 58.23% (4030/6921; 2017-2019).

Conclusions: Web-based SOEs conducted by the Colleges of Medicine of South Africa during the COVID-19 pandemic were well received by candidates, and few technical difficulties were encountered while using Zoom. Better performance was observed...
in web-based examinations than in previous in-person certification examinations. These early findings support the ongoing use of this assessment method.

(JMIR Form Res 2024;8:e40868) doi:10.2196/40868

KEYWORDS
web-based certification examinations; web-based structured oral examinations; medical education; specialist and subspecialist examinations; structured oral examinations; Colleges of Medicine of South Africa

Introduction

Background
The COVID-19 pandemic severely disrupted the postgraduate specialty and subspecialty certification examinations globally [1-5]. The Colleges of Medicine of South Africa (CMSA) continued hosting postgraduate certification examinations during the pandemic [6]. This decision ensured that the national pipeline of specialist workforce would not be disrupted and that the careers of international medical graduates completing postgraduate training in South Africa would not be unduly delayed. Hosting these examinations during the pandemic was achieved by effecting a transition from in-person to web-based written and oral examinations within a period of 8 weeks.

Prepandemic Certification Examinations
CMSA, a nonprofit organization founded in 1954, comprises 29 constituent member colleges, each representing a primary specialty practiced in South Africa, such as College of Pedodontists, College of Surgeons, College of Physicians, College of Family Physicians, College of Obstetricians and Gynecologists, College of Anesthesiologists, and so on [6]. CMSA conducts certification examinations at the beginning (entry level) and end of postgraduate training (exit level) for all specialties and subspecialties registered with the South African statutory medical licensing authority, the Health Professions Council of South Africa [7]. Currently, there are 65 specialties (pediatrics, internal medicine, general surgery, ophthalmology, otorhinolaryngology, anesthetics, etc) and 30 subspecialties (pediatric pulmonology in the specialty of pediatrics, trauma surgery in the specialty of general surgery, gynecological oncology in the specialty of obstetrics and gynecology, etc) practiced in South Africa. For the purposes of this paper, the term, postgraduate certification examinations, denotes all postgraduate certification examinations conducted by CMSA for all specialties and subspecialties practiced in South Africa. These examinations are similar to other international postgraduate certification examinations such as board certification examinations offered by the American Board of Medical Specialties [8] or membership examinations offered by Royal Colleges in the United Kingdom [9]. Before the COVID-19 pandemic, CMSA certification included handwritten examinations, and successful candidates were invited for in-person, patient-based, clinical examinations and in-person, unstructured oral examinations (viva voce).

Transition to Virtual Certification Examinations
The transition occurred during the first 8-week hard lockdown phase of the pandemic, which included closure of all public facilities except for emergency care, food, and health care; complete ban on in-person meetings and any form of intercity travel; and national curfew from 9 PM to 6 AM [10]. Weekly web-based meetings were used to provide staff with information updates and training, supplemented by digital information manuals and web-based training sessions for all examiners and candidates learning to use Zoom (Zoom Communications Inc), an interactive web-based software meeting platform.

Handwritten examinations were replaced by web-based written examinations (short-answer questions and single, best-answer questions) using commercially available software. In-person, patient-based examinations were largely discontinued, except where geographically decentralized in-person clinical examinations could be offered in COVID-19–compliant settings. For all member colleges, a new emphasis was placed on developing web-based structured oral examinations (SOEs) to assess diagnostic reasoning, clinical decision-making, and patient management, which were previously assessed during real patient encounters. It was acknowledged that web-based SOEs could not assess bedside clinical skills (history taking and physical examination), as was previously done. This compromise, accepted as an emergency measure during the pandemic, was predicated on an agreement that implementation of workplace-based assessment in South African postgraduate training would be prioritized after the pandemic [10-12].

Design of Web-Based SOE
Web-based SOEs were designed as case scenarios, each comprising a case description with supplementary information including laboratory test results, photographs, video clips, and radiographic and histopathological images, where appropriate. The case scenarios were prepared by national panels of examiners working remotely via Zoom. Each scenario was prepared as a Microsoft PowerPoint presentation, which was screen shared with candidates during the oral examination hosted via Zoom. Candidates answered the standardized questions posed by examiners working alone or in pairs. Zoom calls were also attended by moderators and trainee examiners (observers). Examiners scored the candidates’ responses independently during Zoom calls using downloaded and printed memoranda, and examiners’ handwritten notes were digitally transcribed and submitted with final examination scores to conveners at the conclusion of the examination proceedings. All Zoom calls were booked and collated on timetables with embedded hyperlinks for individual calls, which were shared with the examiner panels 1 week before the examination events. Individual Zoom calls varied in length from 15 to 60 minutes, depending on the number of case scenarios discussed per call. Candidates undertaking a web-based SOE were typically
examined on 4 to 12 case scenarios, depending on the specialty or subspecialty.

Conducting Virtual Certification Examinations
CMSA set up 14 examination venues (8 in South Africa and 6 elsewhere in southern Africa) equipped with 12-inch laptop computers; 24-inch high-resolution monitors; and internet connectivity using wireless, microwave, and fiber technology. Owing to electrical power supply interruptions (load shedding) in South Africa, all examination venues were equipped with dual power supply arrangements. On-site IT support and on-site proctoring by trained staff were provided at all venues.

Candidates undertook all web-based examinations (written and oral) at an examination venue closest to their home (<3 h travel by road). Overnight accommodation was not permitted. National occupational health and safety COVID-19 protocols were observed at all examination venues [13]. Examiners joined the Zoom calls from their home or place of work, whichever was more convenient and had better bandwidth and a stable electricity supply.

Process of Evaluation
When CMSA implemented web-based SOEs for postgraduate certification examinations in early 2020, there were few reports in the literature. There was little reference to their use for high-stakes postgraduate specialist certification processes [14-17]. Most referred to in-person SOEs for undergraduate students [18-21]. Given the paucity of data, we set up a preliminary evaluation during the initial period of implementation to obtain early insights into candidates’ perceptions about the acceptability of this assessment method and their performance in web-based examinations as compared with in-person certification examinations. Despite its known limitations, we structured our evaluation process on the Kirkpatrick model of program evaluation (reaction, learning, behavior, and results) using four questions to address level 1 (reaction) and level 4 (results) of the model [22-25]:

1. Are web-based SOEs acceptable to examination candidates?
2. What technical challenges did examination candidates encounter while using Zoom?
3. On the basis of candidates’ experiences, how could we improve the web-based SOE experience?
4. Are pass rates for web-based and in-person certification examinations different?

Methods

Design
The evaluation was conducted as a cross-sectional observational study using a mixed methods design.

Participants
The study population included all postgraduate certification examination candidates who were invited to undertake web-based SOEs after successfully completing the written component of the national certification examinations in 2020 and 2021.

Procedure
The first cohort of examination candidates who undertook a web-based SOE between June 1, 2020, and November 15, 2020, was invited to complete the web-based survey immediately after completing their web-based SOE at one of the examination venues. A web-based study information leaflet was provided to potential participants before recruitment. The web-based questionnaire was administered on laptop computers at the examination venues. Participants were recruited to the study before releasing any examination results.

Survey Design
The survey, designed as a Google Form, consisted of 28 items (Multimedia Appendix 1). A total of 23 closed-ended questions focused on candidates’ perceptions of the adequacy, fairness, and quality of web-based SOEs and the technical adequacy and personal time and cost savings of using Zoom to host SOEs. Closed-ended questions required either a binary (yes or no) response or a 5-point Likert-scale response (completely disagree, disagree, neutral, agree, or completely agree). Candidates’ acceptability of web-based SOEs was defined as candidates’ perceptions of the adequacy of the web-based assessment process, overall fairness, and quality of examination material used. Parameters used to determine the adequacy of the assessment process included the following: adequate assessment of clinical reasoning, judgment, insight, and decision-making; appropriate complexity of case scenarios and questions; appropriate duration of examination; appropriate time allocation per case scenario; and preference for in-person patient encounter or in-person examiner. Fairness was explored in terms of perceived overall fairness and the use of more case scenarios than previously during in-person examinations. Quality of examination material referred to clarity of material presented during the Zoom call. Candidates’ acceptability of Zoom as the hosting platform was defined in terms of perceptions of (1) overall acceptability and fairness of the web-based assessment process, (2) audiovisual quality of Zoom, and (3) personal benefits of web-based SOEs to examination candidates (time and financial savings). The survey also included 5 open-ended questions, which required a typed text response. These questions explored candidates’ overall positive and negative experiences regarding the web-based examination process. The survey was administered immediately after the examination proceedings, and we chose to anonymize all the information so as to allay candidate concerns about possible bias when deciding whether to participate in the survey. They only reported about the examination venue attended and certification examination undertaken.

Quantitative Data Analysis
Quantitative data from the web-based questionnaire were exported into a Microsoft Excel spreadsheet before analysis. For questions using a Likert-scale response, answers were reported in 3 categories: agree (strongly agree and agree responses), neutral (neutral responses), and disagree (strongly disagree and disagree responses). Percentages were calculated and rounded to 1 decimal point.
The respective pass rates for the first 2 cycles of web-based certification examinations, 2020 and 2021, were compared with the overall pass rate for the preceding 6 cycles of in-person certification examinations conducted between 2017 and 2019 using the chi-square test for independence with a Yates correction. \( P < .01 \) was taken as the level for significance.

**Qualitative Data Analysis**

Qualitative data from the open-ended questions were captured on the web using Google Forms, exported into an Excel spreadsheet, and subjected to thematic analysis using the 6-step approach to thematic analysis by Braun and Clarke [26]. The qualitative responses were read with the stated question in mind, after which a set of themes was developed by members of the research team (FM and JKM). The responses were reread and assigned to themes, which were further refined as more themes emerged. The responses were reread a third time (JKM), and the themes were reviewed. Discrepancies were presented to a third author (VB) and discussed until consensus was reached.

**Ethical Considerations**

The study was approved by the human research ethics committee of the University of Cape Town (HREC 280/2020), CMSA, South African Committee of Medical Deans, and Health Professions Council of South Africa. Participation was voluntary, and informed consent was obtained before inclusion in the study. This study, which involved human participants, was performed in accordance with the Declaration of Helsinki. All examination candidates gave signed informed consent before participation in the study. All methods were conducted in accordance with relevant guidelines and regulations. Declared consent for publication is not applicable as no identifying images or information was used in the paper.

**Results**

**Participants**

Table 1 shows that during the first 12 months of the COVID-19 pandemic, from June 2020 to June 2021, a total of 3105 candidates registered for postgraduate certification examinations conducted by CMSA. After announcing a transition to web-based examinations, of the 3105 candidates, 351 (11.3%) candidates withdrew (written: 293/351, 83.5%; oral: 58/351, 16.5%); overall participation was 88.7% (2754/3105) during the pandemic. Of the 2754 entrants, 1912 (69.43%) passed the written components of the certification examinations, and 1525 (55.37%) were invited to a web-based SOE, of which 1467 (53.27%) attended the SOE (1467/1525, 96.2% participation).

**Table 1.** Candidates undertaking the web-based certification examinations hosted by the Colleges of Medicine of South Africa.\(^{a,b}\).

<table>
<thead>
<tr>
<th>Total, n (%)</th>
<th>2020, n (%)</th>
<th>2021, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidates who registered for certification examinations</td>
<td>3105 (100)</td>
<td>2005 (64.57)</td>
</tr>
<tr>
<td>Candidates who withdrew from the online written examinations</td>
<td>293 (100)</td>
<td>266 (90.78)</td>
</tr>
<tr>
<td>Candidates who passed the online written examinations</td>
<td>2799 (100)</td>
<td>1732 (61.88)</td>
</tr>
<tr>
<td>Candidates who were invited to the web-based SOE(^c,d)</td>
<td>1525 (100)</td>
<td>646 (42.36)</td>
</tr>
<tr>
<td>Candidates who withdrew from the web-based SOE</td>
<td>58 (100)</td>
<td>24 (41.38)</td>
</tr>
<tr>
<td>Candidates who participated in the web-based SOE</td>
<td>1467 (100)</td>
<td>622 (42.39)</td>
</tr>
<tr>
<td>Candidates who passed the web-based SOE</td>
<td>1159 (100)</td>
<td>497 (42.88)</td>
</tr>
<tr>
<td>Candidates who were admitted (passed the online written and web-based SOEs)</td>
<td>1159 (100)</td>
<td>497 (42.88)</td>
</tr>
</tbody>
</table>

\(^a\) Pass rate of virtual SOE: total=79% (1159/1467), year 2020=79.9% (497/622), year 2021=78.3% (662/845).

\(^b\) Overall pass rate: total=86.5% (2433/2812, year 2020=91% (1583/1739), year 2021=79.2% (850/1073).

\(^c\) SOE: structured oral examination.

\(^d\) Only exit-level certification examinations include virtual SOEs.

**Questionnaire Completion**

The web-based survey was administered to the first cohort of candidates who participated in web-based SOEs between June 1, 2020, and November 15, 2020. Of the 622 potential participants, 542 (87.1%) completed the web-based survey.

**Are Web-Based SOEs Acceptable to Candidates?**

**Candidates' Perceptions of Web-Based SOEs**

Table 2 shows that, broadly speaking, candidates expressed a positive opinions about web-based SOEs. They considered web-based SOEs to be a fair method of examination (374/542, 69%), which tested their clinical reasoning and insight appropriately (364/542, 67.2%), more cases better displayed their knowledge (363/542, 66.9%), and examination material was well presented (369/542, 68.1%). Less than one-third (173/542, 31.9%) felt that real patient encounters would have been preferred, and a small minority (152/542, 28%) would have preferred an in-person oral examination.
Table 2. Candidates’ perceptions of web-based structured oral examinations (SOEs).

<table>
<thead>
<tr>
<th>Adequate method of assessment</th>
<th>Level of agreement (n=542), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agree</td>
</tr>
<tr>
<td>Adequate method of assessment</td>
<td>396 (73.1)</td>
</tr>
<tr>
<td>The exam adequately tested my clinical reasoning, judgement, insight, and decision-making</td>
<td>417 (76.9)</td>
</tr>
<tr>
<td>The case scenarios (examination questions) were appropriate to assess an entry-level specialist or subspecialist</td>
<td>364 (67.2)</td>
</tr>
<tr>
<td>The total length of the examination was appropriate</td>
<td>364 (67.2)</td>
</tr>
<tr>
<td>Having real patients would have improved the quality of the examination</td>
<td>173 (31.9)</td>
</tr>
<tr>
<td>It would have been preferable to have a local examiner present with me to ask the questions</td>
<td>152 (28)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fair method of assessment</th>
<th>Level of agreement (n=542), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agree</td>
</tr>
<tr>
<td>Fair method of assessment</td>
<td>363 (66.9)</td>
</tr>
<tr>
<td>The use of a larger number of case scenarios rather than the historically smaller number of cases gave me a better chance to show my capability</td>
<td>374 (69)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality of the examination material</th>
<th>Level of agreement (n=542), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the examination material</td>
<td>Agree</td>
</tr>
<tr>
<td></td>
<td>369 (68.1)</td>
</tr>
</tbody>
</table>

Candidates’ Perceptions of Zoom for Hosting Web-Based SOEs

The results in Table 3 show that most candidates found the Zoom platform to be acceptable for hosting web-based SOEs (434/542, 80.1%) and considered it to be a fair examination technique (396/542, 73.1%). Most reported that Zoom was technically adequate: 80.1% (434/542) could clearly see and hear examiners, and 69% (374/542) said that video and image quality was adequate. SOEs conducted via Zoom were associated with personal cost-saving (434/542, 80.1%) and time-saving (428/542, 78.9%) benefits because candidates were spared the trouble of traveling. Approximately two-thirds of the participants (336/542, 61.9%) indicated that CMSA examinations should be conducted in the same manner in the future.

Table 3. Candidates’ perceptions of the use of Zoom for hosting web-based structured oral examinations (SOEs).

<table>
<thead>
<tr>
<th>Adequate for hosting web-based SOEs</th>
<th>Level of agreement (n=542), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequate for hosting web-based SOEs</td>
<td>Agree</td>
</tr>
<tr>
<td>I found it acceptable to have examiners conduct the examination using Zoom</td>
<td>434 (80.1)</td>
</tr>
<tr>
<td>Fair method of assessment</td>
<td>Agree</td>
</tr>
<tr>
<td>Conducting oral examinations using Zoom is a fair examination technique</td>
<td>396 (73.1)</td>
</tr>
<tr>
<td>The CMSA(^a) should continue to run the exams using Zoom as opposed to a face-to-face process (preferable)</td>
<td>336 (61.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audiovisual quality</th>
<th>Level of agreement (n=542), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audiovisual quality</td>
<td>Agree</td>
</tr>
<tr>
<td>I could see the examiners clearly on the computer screen</td>
<td>434 (80.1)</td>
</tr>
<tr>
<td>I could hear the examiners clearly on the Zoom call</td>
<td>437 (80.6)</td>
</tr>
<tr>
<td>Images and videos used were of adequate definition or quality to be considered a fair examination</td>
<td>374 (69)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Personal benefits</th>
<th>Level of agreement (n=542), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal cost saving of having the exam locally using Zoom was worth it</td>
<td>434 (80.1)</td>
</tr>
<tr>
<td>The time saved by being able to participate in the exam locally was worth it</td>
<td>428 (78.9)</td>
</tr>
</tbody>
</table>

\(^a\)CMSA: Colleges of Medicine of South Africa.
What Technical Challenges Did Candidates Encounter While Using Zoom?

The first 2 cycles of virtual SOEs required 6258 Zoom calls, all of which were successfully completed on the appointed examination day. All candidates were able to complete their web-based SOE using Zoom. Altogether, 173 technical challenges were reported during the first examination cycle (2020). All (173/173, 100%) were successfully resolved. Of the 173 challenges, a total of 164 (94.8%) were specifically described in the open-ended section of the survey. Overall, 3 major themes were identified: connectivity-related issues with poor-quality connection, disconnection, and laptop battery failure; poor audio quality of Zoom call with sound delay, poor quality or interruption, and low sound intensity from examiners sitting very far from the microphone; and poor image quality on Zoom call with video or photograph either very small or unclear:

Some issues with connectivity in one station. [It] was quickly resolved and another examiner took over [the] questions so I don’t feel disadvantaged due to it. [Candidate 383]

Electricity load-shedding occurred with the generator being switched on at the venue. The volume of the speaker was low and the invigilator increased the volume. [Candidate 263]

Sometimes it would appear as though an examiner had frozen on the screen where their broadband signal had become temporarily weak. This did not disturb the call entirely given that we had the buffer 5 minutes in case of technical glitches. [Candidate 128]

Some audio disturbance, and can hear the exam next door is a bit distraction. [Candidate 414]

On the Basis of Candidates’ Experiences, How Could We Enhance the SOE Experience?

Overview

Thematic analysis of the open-ended questions allowed us to better understand the positive and negative experiences of candidates undertaking web-based SOEs using Zoom. We identified 27 themes, which we categorized into 4 overarching themes, as shown in Table 4.

Table 4. Themes from qualitative analysis of candidates’ responses.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Subthemes</th>
<th>Positive experiences</th>
<th>Negative experiences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examination design and content</td>
<td>• No home-ground advantage</td>
<td></td>
<td>• No preparation time</td>
</tr>
<tr>
<td></td>
<td>• Standardized scenarios and questions</td>
<td></td>
<td>• No station preview</td>
</tr>
<tr>
<td></td>
<td>• Fair questions</td>
<td></td>
<td>• Limited media engagement</td>
</tr>
<tr>
<td></td>
<td>• Expanded case number</td>
<td></td>
<td>• Limited case discussion</td>
</tr>
<tr>
<td></td>
<td>• Less intimidating</td>
<td></td>
<td>• Long examination waiting times</td>
</tr>
<tr>
<td></td>
<td>• Helpful mock examination</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Less examiner-examiner interaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examination environment</td>
<td>• Helpful and friendly proctoring staff</td>
<td></td>
<td>• Variable audiovisual quality</td>
</tr>
<tr>
<td></td>
<td>• On-site IT support</td>
<td></td>
<td>• Obtaining assistance</td>
</tr>
<tr>
<td>Examiner-candidate interaction</td>
<td>• More examiner interaction without patients</td>
<td></td>
<td>• Lack of earphones</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Lack of screen share control</td>
</tr>
<tr>
<td>Personal benefits for candidates</td>
<td>• Travel convenience</td>
<td></td>
<td>• No hard copies of case scenarios</td>
</tr>
<tr>
<td></td>
<td>• Financial saving</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Home comforts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Less interruption</td>
<td></td>
<td>• Examiners visibility on screen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Virtual examiner engagement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Loss of visual cues from examiners</td>
</tr>
</tbody>
</table>

aNo negative personal benefits reported.

Examination Design and Content

Candidates endorsed the lack of a home-ground advantage:

The fact that all candidates are on the same footing. The home candidates have a clear advantage when it comes to clinical OSCE with patients. Those travelling to other provinces do so at great cost and emotional stress. The examination as it was done now should be the way forward. [Candidate 257]

Candidates appreciated the standardization of case scenarios and examination questions:

Excellent standardisation. All candidates had the same examiners, same questions, same experience. Removed all bias from exam. [Candidate 22]

A great number of case scenarios also contributed to a fair examination in the virtual context than in previous in-person oral examinations:
I really enjoyed that the number of cases in the examination was increased as I think that one is able to get a good impression of a candidate if they are examined over broader spectrum rather that certain cases as depicted by the examiners, so that if a candidate does badly in one case they still have a shot at redemption. [Candidate 338]

It was good to have a wider range of cases including paediatric and emergency cases that we normally would not be examined on. I found the experience of virtual presence of the examiners less intimidating. [Candidate 33]

Candidates also felt that virtual oral examinations were fair and more transparent because there was limited examiner-to-examiner interaction:

I think it was also great that examiners where not in the same room and each examiner examined the candidate independently [examiner not aware what another examiner scored for a particular candidate] and that they did not discuss what mark should be given to a particular candidate. Think this brings fairness and transparency. [Candidate 238]

Mock examinations were also perceived favorably. However, the long waiting times before and after the examination were negatively received:

The Zoom Mock exam was reassuring, thank you for giving us clarity. Though it was unpleasant to wait for 2 hours before exams start, I did appreciate the session slots. It was very organized, I commend the team. [Candidate 487]

Unfortunately the duration of the examination process is tedious. If this platform goes ahead it would be preferable to have wrapped up the exams in a shortest time frame. [Candidate 485]

Negative experiences that candidates felt could be improved included an overview of the station before it begins:

Ask examiners to provide a brief overview of the length of the station/number of questions in order to plan answers/timing accordingly. [Candidate 167]

Participants also wanted to add preparation time to each station:

Inadequate time to review slides and prep, many stations required you time to grasp the concepts and integrate the information. It was difficult to do in the time frame provided. Difficult to read the slides and process information and feedback [to examiners]. [Candidate 63]

Candidates wanted better control of the media included in the case scenarios, for example, scrolling through computed tomography imaging studies:

Inability to interact with images and unavailability of dynamic acquisitions. [Candidate 222]

Unable to control imaging myself to make an assessment. [Candidate 386]

They also wanted the option of viewing video clips more than once:

For the video station, to have an option of watching the video twice if you would like to do so because in a real case scenario when examining a patient, you can repeatedly ask the person to do something. Some of the stations should be made realistic, like the emergency station, in a real-life situation, the patient comes in, you don’t get history first the ask to see the patient. [Candidate 481]

Although the candidates enjoyed the breadth of knowledge that was assessed, they wanted an additional way to portray their clinical maturity and insight:

Feeling unable to give a proper account of my knowledge. There is no room for clinical discussion or arguments when answering against a rubric all the time. Maybe one or two stations would be valuable where there’s no rubric and just a discussion around a topic or a case etc. [Candidate 90]

Other negative experiences included a reflection of case scenario design and highlighted the importance of designing appropriate questions for a virtual examination environment:

The examination should have either a model or a device with which the candidate could demonstrate their skills [if the candidate is being asked to demonstrate use of such device]. Adequate resolution of images is also a necessity. [Candidate 70]

Participants suggested administering the case scenarios in a more authentic way to better reflect a real patient encounter:

Because there are no real patients perhaps show pictures of not only the positive findings that the examiner needs to candidate to comment on, but also important negative clinical findings. For example, if the main pathology is in the external ear, it would be helpful to show a picture of a normal TM [tympanic membrane]. When the candidate now interprets the scan, it helps to know that the middle ear was normal. [Candidate 33]

**Examination Environment**

Candidates really appreciated the helpful, compassionate, on-site proctoring staff:

Invigilators were very supportive and accommodating, they gave us 5-star treatment, Thank you very much. Examiners were very empathetic. [Candidate 487]

Moreover, they specified that efficient troubleshooting and reliable technical support were important:

Well organized, flowed well with on-the-spot problem solving, excellently managed. [Candidate 421]

Having an efficient way to call for rapid technical assistance was identified as a possible improvement:

...Have a buzzer or bell in the exam room to enable you to ask for assistance if signal is lost. Or having
a technical person in the room while you are busy. [Candidate 505]

Clear instruction about the conduct of the examination was also considered important:

Inform the candidates how the examination will be conducted. [Candidate 443]

...Explain the role of the technician prior to examination... [Candidate 44]

Familiarity with the virtual platform was highlighted as an important positive finding. Candidates liked that the virtual platform offered an emotionally and professionally neutral venue. They also reported that the virtual examiner-candidate meeting was less anxiety provoking:

I was comfortable having the oral in a Zoom meeting in a neutral venue that is outside the laboratory. I felt less anxiety than for face-to-face interactions... [Candidate 166]
The zoom meeting was less stressful, I am generally anxious, I was less anxious. [Candidate 98]
Less anxiety as examiners are not present with me. [Candidate 78]

Sound quality issues included concerns about struggling to hear examiners:

Not always being able to hear clearly. [Candidate 241]
Concerns about not being heard by examiners were also mentioned:

Worrying about not projecting well/being audible. [Candidate 462]

Earphones to enhance sound quality were suggested:

Using earphones might assist with sound quality... [Candidate 505]

Candidates would have liked paper to write on:

Allow you to have blank paper with you during the question session to better structure your thoughts. [Candidate 349]
They also wanted hard copies of the case scenarios during the virtual SOEs:

Provide a paper copy of the case scenario and questions. Increase the time allowed for answering. [Candidate 60]

**Candidate-Examiner Interaction**

Examiners’ interaction with candidates was an important theme, and both positive and negative encounters were mentioned by candidates.

Candidates found the virtual encounters to be less intimidating:

I found that having the examiners visible but not face-to-face made the exam less intimidating. [Candidate 182]

They also reported that examiners’ engagement with candidates was improved because of the absence of patients:

I felt that the examiners engaged with you more during the zoom meeting because there were no patients. [Candidate 226]

However, the virtual environment required more examiner engagement with candidates to reassure them that they were being heard:

Due to lack of in person examiners, examiners need to be engaging and acknowledge the participants response. Not guide or prompt, but make it clear to the participant that they could be heard. [Candidate 123]

Furthermore, candidates did not approve of examiners being distracted during virtual examinations conducted in their homes:

Examiners should have their phones off and not be distracted by other factors if examining from home. [Candidate 130]

Examiners’ visibility on the screen was viewed both positively and negatively:

I wasn’t actually told I had to have the examiners’ video visible, so I switched it off and found this a much more pleasant experience than what I had been expecting based on colleagues’ related experiences of previous examinations. [Candidate 251]
All examiners must be visible during the exam. [Candidate 459]

The importance of examiners’ eye contact with the camera was specifically mentioned:

Does lose some of the interpersonal contact. Can be distracting if the examiner doesn’t look at the camera/interact. [Candidate 125]

Candidates noted that there was less “positive” examiner cueing because it was not possible to read the examiners’ body language:

Lacking positive reassurance from visual cues given by examiner in a normal setting. Although it is not a given that you would receive that even in a face to face oral. [Candidate 352]
Not being able to read body language. [Candidate 362]
Less able to elicit nonverbal communication/feedback with mock patient and examiner, but this was not a major factor. [Candidate 427]

**Personal Benefits for Candidates**

Candidates found the examination design to be overwhelmingly positive. It was associated with financial savings, reduced levels of anxiety, travel convenience, and availability of home comforts:

No travel. No need to adjust to unknown environment and clinical setting. Costs saved. [Candidate 349]
To be based locally in a huge relief. It alleviates a lot of the stress during an already very stressful time. [Candidate 365]
I liked that it started later, reduced anxiety of traffic and allowed me to sleep better. [Candidate 431]

Home, familiar surroundings, and staff very helpful. [Candidate 373]

Textbox 1. Suggestions from candidates for enhancing the web-based structured oral examination (SOE) experience.

Suggestions from candidates
- Provide candidates with an overview of time allocation per station or desk timer in station
- Allocate preparation time before station begins
- Design case scenarios such that strong candidates can demonstrate their clinical maturity
- Provide paper for candidates to make notes during virtual SOEs
- Allow candidates to control media content, including scrolling and screen share capability
- Allow candidates to use earphones or headphones to improve audio quality
- Provide dual screen capabilities when additional examination material needs to be viewed
- Allow candidates the option of not seeing their examiners (examiners have their video off)
- Train examiners regarding virtual engagement techniques: eye contact and verbal reassurance
- Provide paper for candidates to make notes during virtual SOEs
- Design case scenarios such that strong candidates can demonstrate their clinical maturity
- Allocate preparation time before station begins
- Provide candidates with an overview of time allocation per station or desk timer in station

Are Pass Rates of Web-Based and In-Person Certification Examinations Comparable?

Most candidates (1159/1467, 79%) who undertook web-based SOEs were ultimately successful in passing their certification examinations and gained admission to the respective member colleges of CMSA. The pass rates for 2020 and 2021 were 91.4% (1583/1732) and 79.7% (850/1067), respectively. The average pass rate for postgraduate in-person certification examinations conducted by CMSA 3 years before the pandemic (2017-2019) was 58.23% (4030/6921). This was significantly lower than the pass rate for the first 2 cycles that included web-based certification examinations (2020: N=8635; \( \chi^2 = 667; \chi^2 = 178; P < .001 \); 2021: N=7988; \( \chi^2 = 178; P < .001 \)).

Discussion

Principal Findings

This paper reports about >1500 candidates’ experiences of web-based SOEs conducted via Zoom as part of postgraduate specialty and subspecialty certification examinations hosted by CMSA during the early part of the COVID-19 pandemic. Despite the short time frame in which this transition to web-based examinations was executed, the initiative was well received by candidates. Overall, they perceived web-based SOEs to be a fair and appropriate assessment method, were generally satisfied with the use of Zoom for hosting the examination proceedings, and reported surprisingly few technical challenges. The qualitative data provided a rich analysis of their positive and negative experiences and provided constructive suggestions that could further improve the web-based SOE experience.

An important finding of our study was the observation that candidates undertaking web-based certification examinations performed better than candidates previously undertaking in-person certification examinations. This study did not set out to explore the possible reasons for this observation, but some qualitative observations from candidates may be of interest in this regard. First, candidates reported feeling less intimated by examiners during web-based oral examinations. Second, improved standardization of the examination process using identical case scenarios and standardized questions with marking memoranda may have contributed to fair examination conditions supporting better overall candidate performance. Third, candidates found the overall web-based examination experience to be less stressful, and this may have positively influenced their performance. Fourth, examiners were required to score candidates’ performance independently without conferring before awarding a final score. This limited examiner-examiner interaction may have limited examiner coercion and had a positive impact on the final scores awarded.

Important Improvement Considerations

On the basis of the findings of the thematic analysis, we compiled 10 suggestions that could possibly further improve the web-based SOE experience for candidates. They are listed in Textbox 1.
verbal engagement with candidates to ensure a more authentic experience for candidates.

Findings of Other Studies
There are several prepandemic papers reporting favorably about the use of in-person SOEs in medical education [14-21]. Although these reports were encouraging and supported the use of SOEs for assessment, they did not speak in the context of web-based oral examinations for high-stakes postgraduate certification purposes, which is the topic of this paper. When looking specifically at this context, it is apparent that this is an emerging field that has gained momentum during the pandemic. Currently, there are a few recent reports directly relevant to the key findings of this study. Before the pandemic, McGrath et al [27] reported a study comparing web-based oral examinations with in-person oral examinations for anesthesiology residents. The study randomized 35 residents to testing conditions using an immersive learning environment. Although an immersive learning environment is not the same as a web-based oral examination conducted via Zoom, the paper broadly contributes to the conversation about moving postgraduate oral examinations into a web-based setting. It is worth noting that the authors reported similar academic performance in both groups, which is consistent with our findings of the noninferior performance of candidates undertaking web-based oral examinations. In the study, as also seen in our study, most examination candidates preferred the web-based experience, found it to be less intimidating than in-person oral examinations, and commented about the cost saved by not traveling to testing sites.

Recently, Chaurasia et al [5] conducted a small pilot study of 8 radiation oncology postgraduate trainees and 8 examiners testing the use of web-based oral examinations conducted via Zoom. Candidates were engaged in 8 stations of 25 minutes each, using breakout rooms. Similar to us, they found Zoom to be easy to use, adequate for examination purposes, and free of serious technical difficulties and described it as a fairly seamless experience. They used screen share and annotate for anatomy review, contouring, and treatment plan evaluation for virtual radiation oncology cases. Users described the web-based experiences easier or the same for ease of understanding of the cases and reported preparation to be the same or less time consuming than for in-person oral exams. The authors concluded that a move to web-based oral examinations for postgraduate certification examinations in radiation oncology should be considered as a feasible alternative.

In 2021, a total of 44 senior vascular surgery postgraduate trainees from 17 US training institutions undertook web-based mock oral examinations with 2 remote examiners via Zoom [28]. For each candidate, examiners selected 4 cases from a book of 30 vascular scenarios, and candidate performance was assessed using a standardized scoring sheet. Consistent with our findings, the authors reported no difference in how well the knowledge base of candidates could be examined compared with in-person oral examinations. They also reported that candidates could adequately express their confidence in the web-based setting and concluded that web-based oral examinations are a viable option to consider for examination purposes. They also mentioned the cost saved by remote examinations. The most important considerations highlighted in that paper was the fairness of test grading achieved by using a standardized marking sheet and the equity of test questions achieved by using a standardized set of case scenarios. This speaks to factors that may contribute to favorable candidate performance as already mentioned by us.

In 2022, the Vascular Surgery Board reported about the successful implementation of web-based certification examinations [29]. They reported about the findings of 356 successful candidates who each undertook three 30-minute virtual oral examinations hosted via Zoom. Similar to the findings of our study, the examination process was well received by candidates and found to be technically adequate. Similar to our candidates, they raised concerns about image quality and time constraints to answer all questions but expressed appreciation for good planning and execution, convenience of local examination conditions, and avoidance of travel costs. Overall, candidates were significantly more in favor of continuing web-based examinations as compared with examiners (87% vs 32%; P<.001). The authors make a case for the feasibility and convenience of continuing web-based certification examinations beyond the COVID-19 pandemic. Questions about IT support costs and the cost of remote proctoring are also raised in the paper.

The American Board of Obstetrics and Gynecology also recently published the outcome of their web-based oral certification examinations that were also conducted via Zoom [30]. Between 2021 and 2022, a total of 1491 specialty and 830 subspecialty candidates undertook three 1-hour oral examinations using 3 pairs of examiners per candidate. They found that candidates performed similarly in the web-based certification examinations as compared with previous in-person oral certification examinations. Similar to our findings, they experienced few technical difficulties, and candidate satisfaction with the remote examination was high; however, they expressed anxiety about the use of technology (remotely at home or elsewhere). Despite the success of these remote oral examinations, the issues of remote proctoring and the technical burden placed on candidates being examined in their homes have resulted in a decision to return to testing center–based in-person examinations.

The American Board of Surgery also recently published the findings of their first web-based certification examination for general surgery candidates conducted in 2021 [31]. They report about 306 successful candidates who also completed a web-based satisfaction survey after the examination. They also found that the pass rates for web-based oral examinations were no different from those of in-person oral examinations and that the new examination format was well received by candidates. Audio and video quality were adequate, and 78% of the survey respondents indicated that web-based certification examinations were preferred. The authors argue that the findings support the expanded use of this method of assessment.

The American Board of Ophthalmology also has reported about their success in conducting web-based oral examinations via Zoom for >1000 candidates over the past 12 months. Similar to our study, they took the opportunity to improve the standardization of the examination process and used 2
independent examiners. They described the inordinate preparation and planning required to effect a rapid transition from in-person to web-based certification examinations. They also found that the examination was very well received by both candidates and examiners who felt that the ability to assess knowledge, insight, and judgment using web-based oral examinations was the same as in in-person examinations. Positive opinions regarding the use of the assessment method after the pandemic were expressed by a range of stakeholders. However, they did express some ambivalence over the continued use of web-based oral examinations.

**Strengths and Limitations**

Most current reports in the literature focus on small numbers of candidates. At the time of writing, this was one of the largest published studies reporting about web-based SOEs for postgraduate certification purposes. There are, however, important limitations in this study. We only reported about the initial 12 months of implementation, and a follow-up report is needed. Candidates may have viewed web-based oral examinations more favorably during the pandemic, and this opinion needs to be reviewed. The most significant limitation of the study is that the opinion of examiners was not captured. This was not possible during the pandemic because examiners were overburdened with additional clinical responsibilities and were setting up a new examination method with very little preparation time before the web-based examinations process began. Clearly, their opinion is needed, and if 36 months of implementation, it will contain a wealth of important considerations. Finally, this is a single-country study, and reports from elsewhere are needed.

**Future Directions**

Internationally, the move from in-person to web-based oral examinations, as part of postgraduate specialty certification examination processes, has gathered momentum [2,3,29-32]. There are, however, lively debates about the ongoing use of web-based oral examinations in the postpandemic era, and it seems that opinions are currently divided on the matter [29-32]. Cost and time savings and convenience of arranging and attending are significant factors in favor of remote oral examinations. Major disadvantages of web-based oral examinations include the cost of IT support and remote proctoring [33] and the inability of examiners to assess bedside skills (history taking and physical examination). Increased uptake of workplace-based assessment in postgraduate training programs should effectively address the ongoing concerns about clinical competence. More cost-efficient delivery of IT support and the challenges of remote proctoring require further studies.

**Conclusions**

This study highlights the value of conducting formative research about the use of web-based SOEs as part of postgraduate certification processes. We found that this method of assessment was well received by examination candidates, could be conducted via Zoom with surprisingly few technical challenges, and did not have a negative impact on candidates’ academic performance. Our encouraging findings are consistent with early reports from elsewhere that provide positive preliminary evidence supporting the ongoing use of this novel web-based assessment method in postgraduate medical education.

**Acknowledgments**

The authors wish to thank the candidates for participating in the study, and they extend a special thank you to Bianca van der Westhuizen and Riefda Thomson for the time they spent in assisting the authors with data extraction. No funding was obtained for this study.

**Authors’ Contributions**

VB was involved in the conceptualization of the study, proposal design, study implementation, data analysis, and manuscript preparation and review. JM contributed to literature review, data analysis, and manuscript preparation and review. EB was involved in the conceptualization of the study, proposal design, study implementation, and data analysis. FM was involved in study implementation, data capture, data analysis, and manuscript review. MS contributed to the proposal design and manuscript review. FS contributed to the proposal design and manuscript review. JF was involved in the conceptualization of the study, proposal design and submission, and manuscript preparation and review.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Survey.

**References**


6. The Colleges of Medicine of South Africa. URL: https://www.cmsa.co.za [accessed 2023-12-12]

7. Health Professions Council of South Africa. URL: https://www.hpcsza.co.za [accessed 2023-11-20]

8. American Board of Medical Specialties. URL: https://www.abms.org [accessed 2023-11-20]

9. College of Anaesthesiologists of Ireland. URL: https://www.anaesthesia.ie/ [accessed 2023-11-20]


Abbreviations

CMSA: Colleges of Medicine of South Africa
SOE: structured oral examination

©Vanessa Burch, Jessica McGuire, Eric Buch, Mike Sathekge, Francis M'bouaffou, Flavia Senkubuge, Johannes Fagan. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Development of a Social Risk Score in the Electronic Health Record to Identify Social Needs Among Underserved Populations: Retrospective Study

Elham Hatef¹,², MPH, MD; Hsien-Yen Chang², PhD; Thomas M Richards², MS; Christopher Kitchen², MS, MA; Janya Budaraju², BS; Iman Foroughmand², MD; Elyse C Lasser², DrPH; Jonathan P Weiner², DrPH

¹Division of General Internal Medicine, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States
²Center for Population Health Information Technology, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Elham Hatef, MPH, MD
Division of General Internal Medicine
Department of Medicine
Johns Hopkins School of Medicine
624 N Broadway
Room 502
Baltimore, MD, 21205
United States
Phone: 1 4109788006
Email: ehatef1@jhu.edu

Abstract

Background: Patients with unmet social needs and social determinants of health (SDOH) challenges continue to face a disproportionate risk of increased prevalence of disease, health care use, higher health care costs, and worse outcomes. Some existing predictive models have used the available data on social needs and SDOH challenges to predict health-related social needs or the need for various social service referrals. Despite these one-off efforts, the work to date suggests that many technical and organizational challenges must be surmounted before SDOH-integrated solutions can be implemented on an ongoing, wide-scale basis within most US-based health care organizations.

Objective: We aimed to retrieve available information in the electronic health record (EHR) relevant to the identification of persons with social needs and to develop a social risk score for use within clinical practice to better identify patients at risk of having future social needs.

Methods: We conducted a retrospective study using EHR data (2016-2021) and data from the US Census American Community Survey. We developed a prospective model using current year-1 risk factors to predict future year-2 outcomes within four 2-year cohorts. Predictors of interest included demographics, previous health care use, comorbidity, previously identified social needs, and neighborhood characteristics as reflected by the area deprivation index. The outcome variable was a binary indicator reflecting the likelihood of the presence of a patient with social needs. We applied a generalized estimating equation approach, adjusting for patient-level risk factors, the possible effect of geographically clustered data, and the effect of multiple visits for each patient.

Results: The study population of 1,852,228 patients included middle-aged (mean age range 53.76-55.95 years), White (range 324,279/510,770, 63.49% to 290,688/488,666, 64.79%), and female (range 314,741/510,770, 61.62% to 278,488/448,666, 62.07%) patients from neighborhoods with high socioeconomic status (mean area deprivation index percentile range 28.76-30.31). Between 8.28% (37,137/448,666) and 11.55% (52,037/450,426) of patients across the study cohorts had at least 1 social need documented in their EHR, with safety issues and economic challenges (ie, financial resource strain, employment, and food insecurity) being the most common documented social needs (87,152/1,852,228, 4.71% and 58,242/1,852,228, 3.14% of overall patients, respectively). The model had an area under the curve of 0.702 (95% CI 0.699-0.705) in predicting prospective social needs in the overall study population. Previous social needs (odds ratio 3.285, 95% CI 3.237-3.335) and emergency department visits (odds ratio 1.659, 95% CI 1.634-1.684) were the strongest predictors of future social needs.

Conclusions: Our model provides an opportunity to make use of available EHR data to help identify patients with high social needs. Our proposed social risk score could help identify the subset of patients who would most benefit from further social needs
Addressing social needs and social determinants of health (SDOH) challenges in the health care system has emerged as a key component of addressing health disparities [1]. Patients with unmet social needs continue to face a disproportionate risk of increased prevalence of disease, health care use, higher health care costs, and worse outcomes across a range of health-related domains [2-5]. Thus, health disparities cannot be resolved through traditional clinical interventions in the health care system. Targeted interventions to address social needs and SDOH challenges, especially among minority populations, are necessary to overcome widespread disparities [6].

The use of coding systems such as *International Classification of Diseases, Tenth Revision* (ICD-10) codes for social needs (ie, Z-codes) has increased in recent years, suggesting that clinicians and provider organizations are increasingly aware of social needs and SDOH challenges and the importance of screening for and documenting such needs [7]. Social needs are also extensively documented in unstructured electronic health records (EHRs), such as free-text provider notes [8-12]. Moreover, the rapid adoption of EHRs nationwide and the creation of associated health information technology tools have made it possible to use this growing body of data on social needs and SDOH challenges in risk prediction and adjustment models [13-21].

Some existing predictive models use EHR and administrative claims data on social needs to predict patterns of health care use, cost, and health outcomes [16,18-20,22]. Population-level data on community characteristics is a key component of understanding and addressing SDOH challenges and their impact on health care use, cost, and outcomes [23,24]. Therefore, some EHR-based models have developed linkages to these population-level data to better account for community-level information in their risk predictions [17,21,25-27]. Some existing models have also used the available data on social needs and SDOH challenges to predict health-related social needs [14,15] or the need for various social service referrals [13]. Despite these one-off research and pilot efforts, the work to date suggests that many technical and organizational challenges must be surmounted before SDOH-integrated health information technology solutions can be implemented on an ongoing, wide-scale basis within most US-based health care organizations.

Using both patient- and population-level data, we sought to develop a social predictive risk score based entirely on electronic information readily available within most health care delivery systems. Predictive models such as this could help providers to systematically identify patients at risk of having future social needs, who represent likely targets for further in-depth assessment of their social needs and ultimately potential referral to community-based organizations to address such needs. Using a systematic electronic case-finding screening approach such as this would help the health care system avoid burdensome and inefficient social needs assessment (eg, primary data collection from every patient at every visit).

**Methods**

**Data Sources**

This was a retrospective study using the Johns Hopkins Health System (JHHS) Corporation’s EPIC-based EHR structured data from July 2016 to June 2021. Based on the patient’s home address during the in-scope study periods, we linked community-level data (at the census block group level) from the US Census American Community Survey, 2018 five-year cohort [28]. We developed a prospective model (using current year-1 risk factors to predict future year-2 outcomes) within four such 2-year cohorts (ie, 2016-2017, 2017-2018, 2018-2019, and 2019-2020). Each cohort contained model predictors in the first year (2016 in the first, 2017 in the second, 2018 in the third, and 2019 in the fourth cohort) and model outcomes in the second year (2017 in the first, 2018 in the second, 2019 in the third, and 2020 in the fourth cohort). The overall data were randomly split into training and validation data sets (80% of the data were used for model development while the remaining 20% were used for validation). The final model was applied to the 2020-2021 cohort to evaluate its accuracy.

**Study Population**

Adult patients aged 18 years or older at the time of entering the observation period who were alive at the end of the observation, had at least 1 eligible encounter in the first and second years of each study cohort, and had a valid address for linkage to population-level data were included in this study.

**Variable Selection**

We identified variables with the highest potential impact on the health and social well-being of minority populations through a review of the literature and consultation with minority health, population health, and social needs SDOH experts at JHHS. We also sought input from primary care providers and frontline workers, such as social workers and care managers.
representatives of community-based organizations, and patients and their caregivers.

We identified a comprehensive list of predictors of interest available within the EHR’s structured data, including various patient- and community-level characteristics as well as health care use measures (Table 1) [29,30].

To develop the variable on previous social needs, we obtained any ICD-10 codes presenting social needs using the “Compendium of Medical Terminology Codes for Social Risk Factors” developed by the Social Interventions Research and Evaluation Network [31] or any information on social needs available in the JHHS-EHR Wellness Registry, a data mart table in EPIC storing information related to general patient health, consolidated from many subject areas including social history and risk scores. After reviewing the classification of the ICD-10 codes by Social Interventions Research and Evaluation Network, we developed 13 subdomains and 5 domains of social needs (Figure 1).

We reviewed the ICD-10 codes and mapped each to a unique social need subdomain. We also reviewed available information on social needs in the EPIC Wellness Registry and selected variables corresponding to one of the 13 subdomains of social needs. We collapsed the responses available for each variable to generate a binary variable (“yes” or “no” indicator), suggesting the presence or absence of a social need. We defined previous social needs as a binary variable (“yes” or “no” indicator), suggesting the presence or absence of any corresponding mapped ICD-10 codes or any corresponding social needs identified in the EPIC Wellness Registry to 1 or more of the 13 social needs subdomains. We defined the outcome as a binary indicator of having a social need in the second year of each cohort (using the same logic as for the development of the predictor of social needs).

Table 1. Predictors of interest available within the electronic health record (EHR) structured data for inclusion in the generalized estimating equation model predicting prospective social needs for patients at Johns Hopkins Health System between 2016 and 2021.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>Using the date of birth calculated at the time of entering the study cohort</td>
</tr>
<tr>
<td>Gender</td>
<td>Self-identified and reported at the time of entering the study cohorts</td>
</tr>
<tr>
<td>Race</td>
<td>Self-identified and reported at the time of entering the study cohorts</td>
</tr>
<tr>
<td>Preferred language</td>
<td>N/A</td>
</tr>
<tr>
<td>Need for an interpreter</td>
<td>N/A</td>
</tr>
<tr>
<td>Previous health care use</td>
<td></td>
</tr>
<tr>
<td>In-patient admissions</td>
<td>N/A</td>
</tr>
<tr>
<td>Emergency department visits</td>
<td>N/A</td>
</tr>
<tr>
<td>Previous social needs</td>
<td></td>
</tr>
<tr>
<td>ICD-10 codes</td>
<td>Documented using relevant ICD-10 codes</td>
</tr>
<tr>
<td>EPIC Wellness Registry</td>
<td>Documented in other structured social needs assessment fields, presented in the Wellness Registry Table</td>
</tr>
<tr>
<td>Clinical characteristics (derived from the Johns Hopkins ACG System version 12.0 [29], a widely used population-based predictive modeling and case-finding methodology)</td>
<td></td>
</tr>
<tr>
<td>Number of chronic conditions</td>
<td>N/A</td>
</tr>
<tr>
<td>Medication active ingredients</td>
<td>N/A</td>
</tr>
<tr>
<td>Resource Utilization Band</td>
<td>Represents expected future use based on current morbidities</td>
</tr>
<tr>
<td>Neighborhood characteristics (associated with the person’s residence of longest duration)</td>
<td></td>
</tr>
<tr>
<td>Area deprivation index</td>
<td>A composite measure allowing for the ranking of neighborhoods across the country by their socioeconomic disadvantage, reported at the census block group level [30]</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
bICD-10: International Classification of Diseases, Tenth Revision.
cACG: Adjusted Clinical Group.
The Adjusted Clinical Group (ACG) System Resource Utilization Band represent a simplified population segmentation system based on the overall morbidity burden of each patient. Representing expected future use based on current morbidities, the measure is calculated using all available ICD-10 codes for a person in the EHR during year 1, ranked from low to high according to the expectations of resources used during year 2.

Statistical Analysis

We used a generalized estimating equation (GEE) model to predict prospective social needs, adjusting for the effect of the geographically clustered data as well as the effect of multiple visits for each patient (the records were clustered at the patient and 5-digit zip code level). The model selection was based on the goodness of fit test for GEE modeling, and the final risk score was composed using the variables identified as having the highest impact in the GEE model. We also validated the model using multiple denominators to ensure generalizability and retrained and tested the model for each subpopulation of interest (eg, individuals aged 65 years or older, racial and ethnic minority populations, and those living in the most and least disadvantaged neighborhoods).

Ethical Considerations

The institutional review board of the Johns Hopkins Bloomberg School of Public Health reviewed and approved this study as exempt. The board approved the EHR data extraction for the secondary analysis of deidentified data.

Results

Demographics

The final study population included 1,852,228 patients in total. To be included in the sample, the patients had to be in at least 1 of four 2-year study cohorts (Table 2). The characteristics of patients across the study cohorts were comparable. Study cohorts included mostly middle-aged (mean age range 53.76-55.95 years across study cohorts), White (range 324,279/510,770, 63.49% to 290,688/488,666, 64.79%), and female (range 314,741/510,770, 61.62% to 278,488/448,666, 62.07%) patients from neighborhoods with high socioeconomic status (mean area deprivation index [ADI] percentile range 28.76-30.31).

Between 8.28% (37,137/448,666) and 11.55% (52,037/450,426) of patients across the study cohorts had at least 1 social need documented in the ICD-10 codes or EPIC Wellness Registry, with safety issues and economic challenges (ie, financial resource strain, employment, and food insecurity) being the most common documented social needs (87,152/1,852,228, 4.71% and 58,242/1,852,228, 3.14% of overall patients, respectively; Table S1 in Multimedia Appendix 1 provides details on the social needs domains across study cohorts). Between 18.67% (95,350/510,770) and 21.18% (95,393/450,426) of patients across the study cohorts had high or very high Resource Utilization Band, indicative of having a high disease burden, as reflected by many serious comorbidities.
Table 2. Characteristics of the study population for the development of the social risk score using electronic health record data at Johns Hopkins Health System between 2016 and 2021: overall and by 2-year enrollment cohorts (all characteristics are reported based on the first year of each cohort unless otherwise indicated).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Study cohort</th>
<th>Overall (N=1,852,228)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>53.76 (17.65)</td>
<td>54.63 (17.43)</td>
</tr>
<tr>
<td>Gender (female), n (%)</td>
<td>278,488 (62.07)</td>
<td>274,330 (62.01)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>290,688 (64.79)</td>
<td>285,558 (64.55)</td>
</tr>
<tr>
<td>Black</td>
<td>105,746 (23.57)</td>
<td>103,879 (23.48)</td>
</tr>
<tr>
<td>Other</td>
<td>52,232 (11.64)</td>
<td>52,929 (11.97)</td>
</tr>
<tr>
<td>Preferred language (English), n (%)</td>
<td>430,331 (95.91)</td>
<td>423,396 (95.71)</td>
</tr>
<tr>
<td>Interpreter needed (yes), n (%)</td>
<td>9,356 (2.09)</td>
<td>9,715 (2.02)</td>
</tr>
<tr>
<td>Area deprivation index national rank, mean (SD)(^a)</td>
<td>30.31 (23.6)</td>
<td>29.93 (23.46)</td>
</tr>
<tr>
<td>Health care use, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any in-patient admission</td>
<td>39,364 (8.77)</td>
<td>37,154 (8.4)</td>
</tr>
<tr>
<td>Any emergency department visits</td>
<td>71,737 (15.99)</td>
<td>69,436 (15.70)</td>
</tr>
<tr>
<td>Previous social needs, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1: ICD-10(^b) codes</td>
<td>37,137 (8.28)</td>
<td>50,332 (11.38)</td>
</tr>
<tr>
<td>Year 1: EPIC Wellness Registry</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Year 2: ICD-10 codes</td>
<td>45,272 (10.09)</td>
<td>46,587 (10.53)</td>
</tr>
<tr>
<td>Year 2: EPIC Wellness Registry</td>
<td>0 (0)</td>
<td>3 (0)</td>
</tr>
<tr>
<td>Clinical characteristics, mean (SD)(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Chronic Conditions</td>
<td>2.3 (2.7)</td>
<td>2.4 (2.8)</td>
</tr>
<tr>
<td>Number of Medication Active Ingredients</td>
<td>2.2 (6.1)</td>
<td>2.2 (6.1)</td>
</tr>
<tr>
<td>Resource Utilization Band, n (%)(^d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No or only invalid diagnosis</td>
<td>2,2507 (5.02)</td>
<td>20,273 (4.58)</td>
</tr>
<tr>
<td>Healthy users</td>
<td>44,731 (9.97)</td>
<td>43,408 (9.81)</td>
</tr>
<tr>
<td>Low resource use</td>
<td>59,400 (13.24)</td>
<td>55,221 (12.48)</td>
</tr>
<tr>
<td>Moderate resource use</td>
<td>232,748 (51.88)</td>
<td>231,680 (52.37)</td>
</tr>
<tr>
<td>High resource use</td>
<td>59,979 (13.37)</td>
<td>60,841 (13.75)</td>
</tr>
<tr>
<td>Very high resource use</td>
<td>29,301 (6.53)</td>
<td>30,943 (6.99)</td>
</tr>
</tbody>
</table>

\(^a\)Neighborhood characteristics for the person’s residence of the longest duration are reported as a percentile of national rank [30].

\(^b\)ICD-10: International Classification of Diseases, Tenth Revision.

\(^c\)These clinical measures are derived from the Johns Hopkins Adjusted Clinical Group (ACG) System version 12.0. The Resource Utilization Band represents expected future use based on current morbidities [29].

**GEE Modeling**

Details of the GEE models are presented in Table 3. The GEE model had an area under the curve (AUC) of 0.702 (95% CI 0.699-0.705) in predicting prospective social needs in the overall study population. The strongest predictors of future social needs in the whole population in descending order were social needs documented in the EHR during the previous year period (odds ratio [OR] 3.285, 95% CI 3.237-3.335), ≥1 emergency department visit in the previous periods (OR 1.659, 95% CI 1.634-1.684), and a very high Resource Utilization Band measure indicative of a significant morbidity burden (OR 1.371, 95% CI 1.317-1.427).
Table 3. Generalized estimating equation model predicting prospective social needs for patients at Johns Hopkins Health System using electronic health record data between 2016 and 2021: overall model and models for selected subpopulations.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall Population</th>
<th>Population aged ≥65 years</th>
<th>Racial groups</th>
<th>Neighborhood characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>White</td>
<td>Black</td>
</tr>
<tr>
<td>Area under the curve (95% CI)</td>
<td>0.702 (0.699-0.705)</td>
<td>0.701 (0.696-0.706)</td>
<td>0.689 (0.685-0.693)</td>
<td>0.711 (0.706-0.716)</td>
</tr>
<tr>
<td>Age (years), ORb (95% CI)</td>
<td>0.994 (0.993-0.997)</td>
<td>1.002 (1.001-1.004)</td>
<td>0.994 (0.994-0.999)</td>
<td>0.993 (0.992-0.994)</td>
</tr>
<tr>
<td>Gender (male; reference: female), OR (95% CI)</td>
<td>0.993 (0.981-1.004)</td>
<td>0.922 (0.903-0.941)</td>
<td>0.964 (0.95-0.979)</td>
<td>1.068 (1.046-1.092)</td>
</tr>
<tr>
<td>Race (Black; reference: White), OR (95% CI)</td>
<td>1.125 (1.11-1.141)</td>
<td>1.149 (1.118-1.182)</td>
<td>__b</td>
<td>—</td>
</tr>
<tr>
<td>Preferred language (English; reference: missing, others, or sign language), OR (95% CI)c</td>
<td>1.061 (1.026-1.097)</td>
<td>1.293 (1.221-1.371)</td>
<td>1.034 (0.991-1.08)</td>
<td>1.112 (1.043-1.187)</td>
</tr>
<tr>
<td>Interpreter needed (yes; reference: no or missing), OR (95% CI)</td>
<td>1.179 (1.122-1.238)</td>
<td>1.266 (1.156-1.386)</td>
<td>1.006 (0.908-1.114)</td>
<td>0.816 (0.679-0.981)</td>
</tr>
<tr>
<td>Area deprivation index national rank (percentile), OR (95% CI)c</td>
<td>1.005 (1.005-1.005)</td>
<td>1.001 (1.001-1.002)</td>
<td>1.003 (1.003-1.004)</td>
<td>1.007 (1.006-1.007)</td>
</tr>
<tr>
<td>Health care use, OR (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any in-patient admission</td>
<td>1.017 (0.993-1.042)</td>
<td>1.077 (1.03-1.126)</td>
<td>1.064 (1.03-1.1)</td>
<td>0.987 (0.947-1.029)</td>
</tr>
<tr>
<td>Any emergency department visits</td>
<td>1.659 (1.634-1.684)</td>
<td>1.539 (1.495-1.584)</td>
<td>1.691 (1.656-1.728)</td>
<td>1.627 (1.587-1.668)</td>
</tr>
<tr>
<td>Previous social needs, OR (95% CI)</td>
<td>3.285 (3.237-3.335)</td>
<td>3.043 (2.96-3.128)</td>
<td>3.459 (3.391-3.529)</td>
<td>2.9 (2.824-2.977)</td>
</tr>
<tr>
<td>Clinical characteristics, OR (95% CI)d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of chronic conditions</td>
<td>1.066 (1.064-1.069)</td>
<td>1.08 (1.07-1.08)</td>
<td>1.070 (1.067-1.074)</td>
<td>1.063 (1.058-1.068)</td>
</tr>
<tr>
<td>Number of medication active ingredients</td>
<td>0.997 (0.996-0.998)</td>
<td>0.99 (0.99-1.00)</td>
<td>0.997 (0.995-0.998)</td>
<td>0.997 (0.995-0.999)</td>
</tr>
<tr>
<td>Resource Utilization Band (reference: no or only invalid diagnosis), OR (95% CI)d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy users</td>
<td>0.838 (0.808-0.869)</td>
<td>0.792 (0.734-0.854)</td>
<td>0.836 (0.798-0.876)</td>
<td>0.785 (0.729-0.844)</td>
</tr>
<tr>
<td>Low resource use</td>
<td>0.855 (0.826-0.885)</td>
<td>0.837 (0.778-0.89)</td>
<td>0.856 (0.819-0.895)</td>
<td>0.846 (0.792-0.904)</td>
</tr>
<tr>
<td>Moderate resource use</td>
<td>0.964 (0.935-0.994)</td>
<td>0.935 (0.876-0.997)</td>
<td>0.942 (0.906-0.981)</td>
<td>1.008 (0.952-1.069)</td>
</tr>
<tr>
<td>High resource use</td>
<td>1.259 (1.217-1.302)</td>
<td>1.226 (1.144-1.314)</td>
<td>1.238 (1.184-1.294)</td>
<td>1.328 (1.246-1.415)</td>
</tr>
<tr>
<td>Very high resource use</td>
<td>1.371 (1.317-1.427)</td>
<td>1.218 (1.128-1.315)</td>
<td>1.291 (1.225-1.361)</td>
<td>1.514 (1.408-1.629)</td>
</tr>
</tbody>
</table>

aOR: odds ratio.
bNot available.
cNeighborhood characteristics for the person’s residence of longest duration are reported as a percentile of national rank [30].
dThese clinical measures are derived from the Johns Hopkins Adjusted Clinical Group (ACG) System version 12.0. The Resource Utilization Band represents expected future use based on current morbidities [29].

To help assess bias and applicability to various subpopulations, we identified comparable performance for models of select subgroups, with AUCs of 0.701 (95% CI 0.696-0.706), 0.711 (95% CI 0.706-0.716), and 0.711 (95% CI 0.708-0.714), respectively, for individuals aged 65 years or older, Black patients, and those living in the most disadvantaged neighborhoods.
neighboring. The strongest predictor of future social needs in the study subpopulations remained the previous social needs, with ORs of 3.043 (95% CI 2.96-3.128), 2.9 (95% CI 2.82-2.977), and 3.390 (95% CI 3.334-3.447) among individuals aged 65 years or older, Black patients, and those living in the most disadvantaged neighborhoods, respectively.

To ensure that the most common social needs (ie, safety and economic challenges) were not the main drivers of the model’s performance, we performed a sensitivity analysis and ran the model after excluding patients with any social needs in subdomains of safety and economic challenges. This resulted in a slightly better-performing model with an AUC of 0.768 (95% CI 0.763-0.773). In this instance, previous social needs were by far the most significant predictor (OR 11.857, 95% CI 11.521-12.202), followed by emergency department visits (OR 1.916, 95% CI 1.865-1.969), need for an interpreter (OR 1.528, 95% CI 1.397-1.672), and Black race (OR 1.307, 95% CI 1.273-1.342). Contrary to the analyses performed with all social needs included, patients with higher Resource Utilization Band had slightly less risk of increased social needs, and lower Resource Utilization Band had slightly greater protective value regarding social needs (Table S2 in Multimedia Appendix 1 provides details of the GEE model).

Discussion

Overview

Achieving a comprehensive assessment of a person’s health and addressing health disparities goes beyond just documenting clinical diseases and medical interventions. We must also capture, standardize, analyze, and report reliable information on social needs and SDOH challenges within operational clinical decision support systems that are built into EHRs. Moreover, the rapid change toward “value-based” health care models [32] in the United States has required the incorporation of social needs and SDOH contexts and frameworks to ensure that the health care systems and health plans equitably address the needs of minority and disadvantaged communities [33,34]. For these value-based models to perform well, it is critical that clinical and social interventions are aligned and that no financial disincentives are imposed on providers who disproportionately serve minority and disadvantaged patients [33,34].

To achieve these goals, applied research is needed to identify optimal solutions for the effective collection and application of social needs and SDOH information within EHRs, link provider-based data to community-level data describing the characteristics of patients’ neighborhoods, and anchor such information to the providers’ digital workflow. This approach will provide the vehicle for harnessing social needs and SDOH data to target interventions at the point of care (eg, referrals of an individual); the health delivery system level (eg, hiring a social worker in the clinic); or the community (eg, building or strengthening community-based initiatives) [35]. To avoid burdensome and inefficient social needs assessment and data collection, it is essential to develop automated screening tools using EHR or community-based data to help identify the subset of patients who would most benefit from social needs assessment and data collection.

Several EHR-based screening tools for social needs assessment have been piloted in recent years, and results have shown these tools to be effective in determining social needs and SDOH challenges [36,37]. However, the feasibility of such tools remains unclear, with health care systems needing to dedicate considerable time and budget to train and educate staff and manage workloads [21]. Our proposed social risk score aimed to reduce the burden of this process and increase its accuracy by identifying patients at high risk of having any social needs for more efficient screening.

Comparison With Previous Evidence

Our proposed model was based on a large and diverse data set of patients in the JHHS-EHR. The AUC of our model in predicting prospective social needs was 0.702 (95% CI 0.699-0.705) in the overall study population. This AUC may be the result of many instances of false negatives related to the documentation of social needs in structured EHR data. At the time of completing this study, social needs screening and referral were not common practices at our institutions. Thus, we expect many patients with social needs did not get a proper screening and documentation of such needs. The new mandate established by the Centers for Medicare and Medicaid Services requires hospitals to report to the Inpatient Quality Reporting Program 2 brand new measures of social needs (ie, the number of patients screened for social needs and the number of patients identified with selected social needs) [38]. We expect the health care system to establish more systematic and uniform processes for screening and documentation of social needs. This effort will increase the volume of data on social needs in the EHR, which will result in better performance of our models in the future.

Our findings were comparable with those in the study by Holcomb et al [14], where they predicted health-related social needs using EHR and community-level data and machine learning modeling for Medicare and Medicaid beneficiaries participating in the Accountable Health Communities project. Their models performed relatively well, with AUCs ranging from 0.59 to 0.68 for patients with different domains of social needs. Another notable mention was the study by Byrne et al [15], where they used EHR data, including responses to the Veterans Health Administration’s Homelessness Screening Clinical Reminder Survey, to develop and test predictive models of housing instability and homelessness. All their models performed well, with the random forest models performing better than the logistic regression models for both the housing instability (85.4 vs 78.3) and homeless (91.6 vs 87.1) outcomes.

Lastly, Kasthuriratne et al [13] built random forest decision models to predict the need for social work referrals using clinical and population-level data on SDOH challenges. The performance of the model ranged from an AUC of 0.713 for the model using both clinical and SDOH data to 0.731 for the model using clinical data.

Moreover, our results demonstrated that the most significant predictive factor for having prospective social needs was the documentation of previous social needs. This study also found associations between prospective social needs with previous ED visits and morbidity-related high resource use presented in Resource Utilization Band. Lastly, our model showed a
minimally increased risk of social needs in association with characteristics of the neighborhood of residence, presented as the ADI measure. This finding was similar to the results of the study by Nguyen et al [39], where they identified a small statistically significant association between the ADI and total score on social needs from the Health Leads Social Needs Survey among pediatric patients’ families receiving primary care at a large academic institution. A review by Chen et al [21] also indicated the low success of the integration of population-level data for predictive modeling and risk stratification purposes, including the prediction of social-related service referrals [13], in contrast to the performance of models using individual-level data in referrals to a social worker [40]. Overall, these findings indicated that individual characteristics played a more crucial role in predicting future social needs than neighborhood characteristics.

**Clinical Implications**

The implementation of an EHR-based social risk score such as the one we developed would have many implications for clinicians and practice organizations. At the point of care, our social risk score could be integrated directly with EHR-derived data warehouses; thus, the proposed risk score could be leveraged to allow clinicians to tailor more personalized care and modulate care coordination efforts. This personalized social risk score could also help support tools tailored to the needs of patients, which would empower the navigation of available social services [35]. At the health delivery system level, the linkages of geo-derived databases could improve the assessment of social needs and SDOH challenges for health systems and provide opportunities for longer-term plans to address those factors in current and future care management programs. Additionally, the merged clinical and nonclinical databases could enable providers to follow patients with social needs and SDOH challenges over time through their interaction with the health system. At the community level, the social risk score and the social needs or SDOH data could enhance care coordination efforts by integrating community-level data into clinical decision support tools and coordinated interventions [41].

**Limitations**

Several limitations existed for this study. First, EPIC Wellness Registry data were essentially nonexistent between the years 2016 and 2018, which, in addition to the possibility of ICD-10 codes being underused by providers, might lead to an underrepresentation of social needs in this study’s population. Second, Latino and Hispanic patients were underrepresented in this study’s population, which may have impacted the generalizability of the proposed model. Furthermore, data on racial and ethnic minority groups such as American Indian and Alaskan Native, Native Hawaiian, Other Pacific Islander, and multiracial individuals were limited in our data set. Another important factor that might have affected our results was the potentially more frequent screening of social needs in female individuals, ethnic and racial minority populations, those with higher disease burdens, and superusers of health care services [42], leading to biased results for these individuals. Misclassification and inconsistency in documenting social needs in the EHRs could influence our results. Moreover, the use of EHR data as the sole source of information limited our data to services provided to patients across the JHHS facilities and did not include other services outside JHHS. Also, we used the patient’s home address to link the EHR data to the American Community Survey community-level data. Thus, we did not include patients with a missed or invalid home address. This may have resulted in missing some patients with social needs, such as residential instability. Finally, our final model was applied to the 2020-2021 cohort to evaluate its accuracy, which included patient encounters during the first peak of the COVID-19 pandemic. While some subpopulations of patients experienced more social needs during the pandemic, the challenges that the health care systems faced during this period impacted the screening and documentation of social needs and may have resulted in missed data on social needs for this cohort.

**Conclusion**

Screening for social needs and SDOH challenges using available secondary EHR data represents an important step toward addressing health disparities more efficiently and improving patient care and population health. Our proposed model integrates community-level data with patient-level data to arrive at a social risk score that can be used to systematically identify patients at increased risk of having future social needs and is thus appropriate for in-depth assessment of their social needs and potential referral to community-based organizations to address these needs. Our model identified previous social needs and high morbidity levels as the strongest predictors of future social needs. Missing data in the EHR reflecting past clinician documentation of social needs may have impacted the performance of our proposed model, and the predictive accuracy of models like ours will likely increase as the capture of such information becomes more commonplace. Future studies should focus on developing EHR-integrated clinical decision support tools to make this information available in the providers’ digital workflow and at the point of care.

**Acknowledgments**

This study was supported by a grant from the National Institute on Minority Health and Health Disparities (NIMHD; grant R01MD015844-01). Its contents are solely the responsibility of the authors and do not necessarily represent official NIMHD views.

**Data Availability**

The data sets generated and analyzed during this study are not publicly available due to concerns regarding the privacy of individuals whose electronic health records information were used.

https://formative.jmir.org/2024/1/e54732
Conflicts of Interest

Johns Hopkins University (JHU) holds the copyright to the ACG System and receives royalties from the global distribution of the ACG System. This revenue supports a portion of the authors’ salaries. The authors are members of a group of researchers who develop and maintain the ACG System with support from JHU. After the completion of his involvement in this work, H-YC joined Janssen Scientific Affairs, LLC (a Johnson & Johnson company) as a full-time employee and holds stock in Johnson & Johnson.

Multimedia Appendix 1
Social needs domains across study cohorts and sensitivity analysis.

[DOCX File, 31 KB - formative_v8i1e54732_app1.docx]

References


30. Area Deprivation Index. University of Wisconsin School of Medicine and Public Health. URL: https://www.neighborhoodatlas.medicine.wisc.edu/ [accessed 2023-05-09]


38. Medicare Program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and policy changes and fiscal year 2023 rates; Quality Programs and Medicare Promoting Interoperability Program requirements for eligible hospitals and critical access hospitals; costs incurred for qualified and
non-qualified deferred compensation plans; and changes to hospital and critical access hospital conditions of participation.


Abbreviations

ACG: Adjusted Clinical Group
ADI: area deprivation index
AUC: area under the curve
EHR: electronic health record
GEE: generalized estimating equation
ICD-10: International Classification of Diseases, Tenth Revision
JHHS: Johns Hopkins Health System
OR: odds ratio
SDOH: social determinants of health

©Elham Hatef, Hsien-Yen Chang, Thomas M Richards, Christopher Kitchen, Janya Budaraju, Iman Foroughmand, Elyse C Lasser, Jonathan P Weiner. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Patient-Centered Approaches for Designing Destigmatizing Sexual Pain-Related Web-Based Platforms: Qualitative Study

Abdul-Fatawu Abdulai, PhD; Hasti Naghdali, BSN; Heather Noga, MA; Paul J Yong, MD, PhD

1 School of Nursing, University of British Columbia, Vancouver, BC, Canada
2 Women's Health Research Institute, Vancouver, BC, Canada
3 Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Abdul-Fatawu Abdulai, PhD
School of Nursing
University of British Columbia
T201-2211 Wesbrook Mall
Vancouver, BC, V6T 2B5
Canada
Phone: 1 604 822 7214
Email: fatawu.abdulai@ubc.ca

Abstract

Background: Sexual pain is a common but neglected disorder that affects approximately 3% to 18% of women and an unmeasured number of gender-diverse people worldwide. Despite its wide prevalence, many people feel reluctant to visit conventional health care services or disclose their symptoms due to the fear of stigmatization. To alleviate this stigma, various web-based interventions have been developed to complement and, in some cases, replace conventional sexual health interventions. However, the way these web-based interventions are developed could inadvertently reproduce, perpetuate, or exacerbate stigma among end user patients.

Objective: The purpose of this study was to understand patients’ perspectives on how sexual pain–related web platforms can be designed to alleviate stigma or prevent the unintended effects of stigma among patients who use web-based interventions.

Methods: Individual semistructured interviews were conducted among 16 participants with lived experiences of painful sex in a large urban city in Western Canada. Participants were recruited via social media platforms, newsletters, and a provincial health volunteer website. Using a sample sexual pain website to provide context, participants were interviewed about their experiences of stigma and how they think web platforms could be designed to address stigma. The interviews were conducted via Zoom (Zoom Technologies Inc) and analyzed using thematic analysis.

Results: The findings revealed 4 overarching themes that represented participants’ perspectives on designing web platforms that may alleviate or prevent the unintended effects of stigma. These findings suggested the design of inclusive web platforms, having a nonprovocative and calming user interface, having features that facilitate connections among users and between users and providers, and displaying personal testimonials and experiences of sexual pain.

Conclusions: This study highlighted patient-centered design approaches that could serve as a reference guide in developing web platforms that alleviate or prevent the unintended effects of stigma, particularly among nonheterosexual and gender-diverse people. While this study was conducted in the context of sexual pain, the results might also apply to web platforms on other potentially stigmatizing health-related disorders or conditions.

(JMIR Form Res 2024;8:e53742) doi:10.2196/53742

KEYWORDS

stigma; digital health; sexual pain; destigmatizing; end user patients

Introduction

Sexual pain is a common but neglected problem, predominantly experienced by females, that is estimated to affect approximately 3% to 18% of women globally [1]. The disorder affects multiple aspects of one’s life including absenteeism from work, poor interpersonal relationships, and impaired social functioning with a profound impact on the quality of life [1]. Despite the
negative effects of sexual pain, many women feel reluctant to visit service providers or disclose their symptoms to family and friends due to the fear of stigma [2]. Stigma is defined as an attribute and dynamic social process characterized by widespread social disapproval, blame, rejection, devaluation, and segregation [3]. The extent to which people with sexual pain are affected by stigma depends on the nature of the disorder and whether it is concealed or exposed. According to Tyler and Slater [4], people with invisible conditions or disorders are more likely to be affected by internalized stigma, while those with visible conditions or disorders may be predisposed to public stigma. Public stigma refers to the overt discriminatory practices that are directed at someone, while internalized stigma refers to the acceptance of negative attributes and a reduced sense of self-worth as a result of possessing a supposedly stigmatizing attribute [2]. While hidden disorders like sexual pain may not immediately provoke public stigma, people with such disorders may still be concerned about an impending negative reaction from the public or even stigma from present and future intimate partners if their pain status is eventually revealed [4]. The fear of disclosing a condition can be classified as anticipated stigma, and this is quite common among people with sexual health–related disorders [5]. Therefore, people with sexual pain may face both internalized and public stigma of various proportions at the personal, interpersonal, and societal levels. As stigma becomes more of a dynamic social process [3], the modalities of managing this issue are becoming increasingly complex and challenging, especially for people with hidden disorders like sexual pain.

In response to the negative impact of stigma on people with sexual pain, various web-based platforms have been developed to complement, and in some cases, replace conventional care services [2,6,7]. Web-based platforms on painful sex could range from educational resources, decision support systems, web-based social support platforms, and therapy-based platforms [8–12]. Even though web-based platforms are increasingly used as resources inquiring about and managing sexual pain [13,14], available evidence suggests that sexual health–related web-based platforms could inadvertently foment stigma among end users [15]. Other studies revealed that merely using a sexual health platform in public spaces could be interpreted as having a disorder and could lead to stigma in what is termed stigma by association [6]. Notably, people who use sexual health–related web platforms may likely link the appearance of certain images and content to an existing social stigma, thereby resulting in technology-mediated experiences of stigma [7]. Furthermore, the replacement of face-to-face conversations with web-based tools may further emphasize the feeling of discomfort or stigma associated with in-person interactions as it gives an indication that sexual pain is not something to be discussed but self-discovered using web-based tools.

The likelihood of web-based platforms resulting in technology-mediated experiences of stigma may be attributed to the general lack of attention to the issues of stigma when developing sexual health–related digital platforms [16]. Even though patients’ views and opinions are usually sought when developing sexual health–related digital technologies [17,18], their perspectives on how web-based platforms can be designed to alleviate stigma are usually not considered when developing such interventions. The apparent lack of attention to stigma may not only lead to a mismatch between people’s expectations and web-based platforms but could inadvertently reproduce and perpetuate stigma among patients who may find such platforms undesirable [7]. We attempted to address this gap in our prior studies by developing a set of destigmatizing design guidelines to provide a reference guide for developers of sexual health–related web-based platforms [16,19]. However, these guidelines were based on expert opinions and did not address the salient concerns of patients regarding stigma. With a disorder like sexual pain, patients often withhold sharing their diagnosis publicly, and we believe that many of the stigma-related considerations for designing web-based platforms could only be obtained from people with lived experiences of the disorder.

The purpose of this research was thus to examine the perspectives of people with sexual pain on how web platforms could be designed to alleviate stigma or prevent digital platforms from inadvertently fomenting stigma among end user patients. The findings of this study are expected to produce user-interface design recommendations not just for sexual pain digital platforms but for web-based platforms on other stigmatized conditions or disorders in sexual health and mental health.

### Methods

#### Study Design

A semistructured interview was conducted among 16 people with lived experiences of sexual pain in a large urban city in Western Canada. Participants were recruited via (1) social media platforms, (2) newsletters, and (3) a provincial health volunteer research website. To be eligible, potential participants must have been at least 18 years old, experienced or self-reported sexual pain (alone or partnered), and reported previous or current use of a health-related web-based platform. Due to the limited involvement of male partners in prior sexual health–related studies [8], we decided to recruit people who also identify as biological males and report experiences of painful sex.

#### Interview Guide

A semistructured interview guide (Multimedia Appendix 1) was developed based on the person-centered care model by focusing on the participants’ emotional needs and preferences about how digital health technologies can be developed to be less stigmatizing and emotionally safe. The interview guide was developed to address key areas including demographic information, participants’ experiences with stigma, and their perspectives on how web-based platforms can be designed to alleviate stigma or prevent the unintended effects of stigma. The interview guide was reviewed by our patient advisory team, pretested on 3 patients outside of this study sample, and was revised accordingly.

#### Data Collection

The interviews were conducted by the first author (AFA), with assistance from the second author (HN). Since different web-based platforms exist, we decided to contextualize our study by providing participants with a sample website that we developed in a previous study [10]. A link to this website was
sent to participants at least 3 days before the interviews. Participants, particularly those who have never used a sexual pain web platform, were asked to explore the sample website before the interviews. The interviews were conducted via Zoom with participants’ videos turned off to protect their confidentiality. Participants were asked questions related to the design of web-based content, interactive features, images, colors, sex, gender, and class representations on digital platforms that can help alleviate the stigma of painful sex. Depending on participants’ backgrounds and prior experiences of stigma, probing questions were asked regarding how web platforms can be designed to address the experiences of stigma related to their identity. Data collection occurred from April to July 2023, and each interview lasted approximately 1 hour. All interviews were recorded onto the University of British Columbia’s accredited Zoom Cloud and transcribed using TEMI web-based transcription services (Temi Inc).

Data Analysis
The transcribed data were imported into the NVivo software (version 11; QSR International). Data were analyzed using a thematic analysis approach [20]. The data analysis occurred alongside data collection. The first and second authors familiarized themselves by reading over the first 4 transcripts multiple times. The 2 authors independently developed a coding framework inductively from the first 4 transcripts. We compared the 2 coding frameworks and arrived at a common coding scheme at the end of the fourth interview. Any differences between the coders were resolved by discussing with other members of the research team. We then applied the unified coding scheme to the rest of the data as they emerged from subsequent interviews. As the data analysis proceeded, the coding scheme was revised as new data emerged. The coded data were sorted and organized into categories, and themes were then generated from the categories depending on their overall importance to the research questions. We compared data from people in different age groups, sex, and gender identities to understand if such identities reflected specific preferences on how web platforms can be designed to alleviate stigma.

Ethical Considerations
The study was approved by the University of British Columbia behavioral research ethics board (REB # H23-00273). All participants provided written informed consent before the data collection. All data were anonymous because participants were informed not to state or mention any identifying information during the interview. Each participant was provided with CAD $50, which was equivalent to US $35 honoraria in appreciation of the time spent on the study.

Results

Demographic Characteristics
A total of 16 participants were interviewed. Data saturation occurred after the 14th participant, and we decided to interview an additional 2 participants. Regarding gender identity, 12 participants identified as women, 2 identified as men, and 2 identified as two-spirit. Regarding racial identity, 8 participants identified as White, 3 identified as a person of color, and 2 identified as Indigenous. Participants’ ages ranged from 25 to 63 years, with the majority of them younger than 30 years old. Even though all participants had used health-related websites, 6 of them had additional experiences using sexual pain–related websites. While we did not specifically ask people about their sexual identity, 3 of them openly stated their sexual identity as nonheterosexual. Seven participants stated they are marginalized by being identified as a woman, nonheterosexual, and Indigenous. On one hand, participants who indicated a marginalized identity also reported feeling stigmatized at some point in their lives or a feeling of impending stigma from their experiences of sexual pain. On the other hand, participants who did not express any marginalized identities were not very concerned about stigma, with 1 participant indicating that they do not associate sexual pain with stigma.

Participants’ Perspective on Design Approaches to Alleviate Stigma
The analysis generated 4 main overarching themes that reflected participants’ perspectives on how sexual pain–related web platforms can be designed to alleviate stigma or prevent such platforms from inadvertently fomenting stigma. Some of the broader thematic areas contain subthemes that come together to define each overarching theme.

Theme 1: Inclusive Design of User Interfaces

Overview
Inclusive design emerged as a major theme that reflected participants’ perspectives on how digital platforms could be developed to prevent the unintended effects of stigma. Participants recognized the diversity of people who may have direct or indirect experiences of sexual pain and the need to design digital platforms that satisfy the experiences of people with different racial or cultural backgrounds and sexual and gender identities. Three subthemes emerged under the overarching theme of inclusive design.

Subtheme 1: Design to Include Diverse Sexual Experiences
The participants indicated that web platforms on sexual pain should not just be limited to information on penis–vagina sex but include other forms of sexual experiences. According to the participants, other forms of sexual experiences like anal sex and the use of sex toys could equally elicit pain and therefore must be captured on web platforms.

Sex is not just penis to vagina interaction. There's also sex toys and different ways in which people partake in sex. And so being very inclusive of that sense so that everyone feels safe in that website environment. And touching on the fact that there are different toys and people who are trans, for example, use dilators which can also bring pain.

The participants stated that these other forms of sexual experiences are already stigmatized by society and excluding them from digital platforms could worsen the stigma associated with such sexual experiences. The desire for inclusion of information on diverse sexual experiences was common among younger people. In fact, 2 participants who were older than 60
years were not open to the idea of other sexual experiences beyond penis-vagina sex.

**Subtheme 2: Design to Include Diverse Sexual and Gender Identities**

The participants also recognized that people may have different gender identities apart from being a woman and still experience painful sex. They specifically stated that not all people who identify as a woman may have a vagina and not all people that identify as a man may have a penis. To accommodate gender-diverse people in web-based platforms, participants suggested having inclusive pronouns by having drop-down menus or subsections on websites that account for people with different gender identities. Other participants argued that including pronouns of all gender-diverse people on web platforms may be difficult or impossible to achieve. Thus, they suggested the use of gender-neutral terms as a way of reducing the stigma associated with other stigmatized sexual identities. Participant 8 stated the need for gender-neutral terms by stating:

*Oh my god, my ovaries are partially removed because I had a tumor. And I think that using less gendered language would be very, very, very, desitgramizing for people like me and for people who are trans. Because I love when people actually think of us and it shows that we are not as bad as some people think. I would probably stick with non-gendered wording, like people with uteruses or just like talking about the specific parts that may show pains.*

While participants argued for designs to include diverse gender identities, they agreed that painful sex may be more common among cisgendered people. In addition to the gender-diverse pronouns, participants suggested that web platforms on painful sex should include male partners who may not directly experience pain but may be part of the activity that produces sexual pain. Participant 6 noted the need to include male partners by stating:

*But honing in on the partners aspect, I think is really to make sure that there are resources for partners as well who aren’t actively dealing with the pain. Okay. It’s something that I’ve noticed is missing a lot.*

**Subtheme 3: Design to Reflect Cultural, Religious, and Ethnic Diversities**

The participants also indicated their preference for designs that include people with different cultural, religious, and ethnic identities. This theme occurred across participants who identified as Indigenous or other marginalized identities. Participants expressed the need to see their culture or ethnicity represented when visiting web-based platforms regarding painful sex. Participant 14 specifically stated:

*I was talking to one of my friends and she’s a black woman and she said when she was looking up sexual pain, she only saw white women on the websites. And so, she didn’t think the information applied to her. She didn’t see anyone who looked like herself. So, I think if images are used, it has to be intersectional so people can see themselves on the website.*

According to participants, these minority groups are already marginalized and stigmatized in society, and such stigma may only intensify if their cultural and racial identities are not represented on web platforms. Some participants with religious backgrounds also noted the importance of religious sensitivity and how that can alleviate stigma, particularly stigma relating to traumatic sexual experiences. Participant 7 stated:

*I think that having that cultural sensitivity aspect embodied in a digital platform is important. Whether that be having individuals who have religious trauma associated with sex and who have gone through this experience give a testimonial or talk about it on a video or having religious supports or access to resources on the website, whether it be clickable links, phone numbers, whatever it might be on there.*

**Theme 2: Nonprovocative and Calming User Interfaces**

**Overview**

Participants also reiterated the use of nonprovocative and what they termed “calming user interfaces” as an approach to alleviating stigma via web interfaces. According to participants, people with experiences of painful sex may already be overwhelmed by some form of hidden trauma and stigmatized feelings associated with the disorder. Thus, they suggested that calming and uncluttered interfaces could be an appropriate design approach to alleviating such stigmatized feelings. They specifically argued for the content on the landing pages to provide comfort to website users before they dive into the actual content of the website. The use of (1) nonprovocative and nonexplicit images and (2) subtle and calming colors emerged as subthemes that together defined the overarching theme of uncluttered and calming user interfaces.

**Subtheme 1: Nonprovocative and Nonexplicit Images**

Many participants argued for the use of nonexplicit images. Participants with a health science background thought explicitly displaying anatomical images of sexual organs might help in educating patients on the specific location of pain. Participant 2 stated:

*For me personally, I feel like the more graphic and anatomical images, the better because when I was younger, I literally didn’t know anything about sexual anatomy. it's probably not allowed on the internet, but I think having those images would be helpful.*

However, the majority thought that such images could be uncomfortable and might inadvertently foment stigmatized feelings. Participant 1 stated her abhorrence for seeing explicit content on sexual health–related web platforms and indicated how that made her feel uncomfortable.

*I read on one medium where they have provocative pictures of people, you know, in the lovemaking act. It wasn’t pornographic but certainly explicit and personally, I am uncomfortable with that.*

**Subtheme 2: Subtle and Calming Colors**

Participants’ views on addressing stigma via web platforms extended beyond the use of nonexplicit images to include the
color of the images and the general color scheme of websites. There was unanimity in participants’ preference for colors like blue, green, dark, purple, soft greens, turquoise, and so forth. These colors were perceived to be calming and comforting for an already distressed user population of people with painful sex experiences. On the contrary, participants were generally unfavorable to bright colors like red and pink as such colors were thought to symbolize pain. Participant 9 indicated his desire for calming colors by stating:

so probably not like bright pink and that kind of thing.
You know painful sex comes with its own trauma and having all these bright colors may be a put-off. I liked the one I saw in this interview, it was like really calm colors, and they weren't too noisy, so you really focused on the information on the site and it was like, I think calming colors are maybe a good idea too.

Apart from indicating how these calming colors may be comforting to the eyes, participants were not explicit on how these color schemes could help address stigma when probed further.

### Theme 3: Interpersonal Connection via Web Platforms

#### Overview

As a way of addressing stigma, participants indicated a preference for integrating interactive features on web-based platforms that allow the creation of interpersonal connections among people with similar experiences or with health care professionals. According to participants, being connected to others will be crucial in addressing stigma for a hidden disorder like sexual pain. Two subthemes emerged under this overarching theme.

#### Subtheme 1: Creating Forums to Connect People With Similar Pain Experiences

There was a strong preference for forums, discussion boards, and messaging boards that would enable people with similar sexual disorders to connect, interact, and share experiences with each other while remaining anonymous. Participant 9 compared this to Reddit, and how a Reddit-like function on web platforms can offer a free, easy, and anonymous forum for people to discuss sensitive topics.

I think in that sense; the forums open up the isolation. The experiences and stories of strangers who have the same condition and you know such varied experiences will tell me that I am not alone in this and should not harbor any negative feelings. You know the internet gives us this fantastic gift of connection without obligation.

While participants acknowledged the importance of forum posts in addressing stigma, they also recognized the security and privacy risks associated with forum posts and how they can inadvertently perpetuate stigma. One participant noted how such forums can be used as avenues for harassment, bullying, and other risky behaviors. To prevent digital health platforms from inadvertently fomenting stigma, the participants suggested developing rules and regulations that govern forum posts, the use of pseudonyms and having designated professionals who would moderate web-based platforms.

#### Subtheme 2: Creating Avenues to Connect With Health Care Professionals

In addition to creating avenues on web platforms for people to connect, participants also reiterated the importance of creating avenues to connect with health care professionals, especially if it is a local website. According to participants, connecting with health care professionals via web platforms can “serve as a shield” from being exposed to the public stigma—a situation that participants found helpful during the COVID-19 pandemic. Participant 13 noted the importance of connecting to professionals by stating that:

speaking to a medical professional is more easily accessible and then having kind of the barrier of the screen in between, like not having to talk to someone face to face about it, but be able to type, I think could be helpful and, and perhaps start as a way to reduce the stigma.

According to participants, these forms of web-based interactions via web platforms can facilitate web-based consultations, thereby reducing the stigma people encounter by visiting onsite services.

#### Theme 4: Displaying Personal Accounts and Experiences of Painful Sex

The data revealed that displaying personal experiences on web platforms could help alleviate the stigma of painful sex among web users. The participants stated that seeing or hearing what others have gone through can help them relate to other people’s personal accounts of sexual pain while validating their own experiences. Participant 2 stated:

When you find websites like this sharing information, it makes you feel like, I’m really not that strange. I’m not weird. Like there are a lot of other people who experience this too and that’s why they’re putting forward this information on this platform. So, it kind of makes people realize that it’s more common than they think. And that also helps to alleviate fear surrounding it.

Some participants suggested the use of video testimonials to convey experiences, while others suggested a combination of text and video testimonials but from different sources. According to participants, reading about people’s stories; hearing their voices; and understanding what they went through, how they discovered their diagnosis, and how it was treated will help others learn and understand their situation and clarify myths and misconceptions that often fuel stigma. Some participants also suggested the use of comic images and videos to convey their experiences of painful sex as a way of addressing stigma. Participant 3 stated:

I like the comic thing. Like having people talking about these sorts of things in a comical style and that can make it less terrifying. By making it comical, you make it more entertaining, and people are more likely to pay attention to these sorts of things.
Discussion

Principal Findings
This study examined the perspectives of people with lived experiences of sexual pain on how web-based platforms can be designed to alleviate the unintended effects of stigma. While we acknowledge current practices of involving end users in developing sexual health–related digital interventions [21,22], no study has specifically investigated how such technologies can be designed to address the unintended effects of stigma. To the best of our knowledge, this is the first study that examined patients’ perspectives on how to address stigma through the design of digital health web platforms. Our study identified 4 broad categories that represented patient-centered strategies on how sexual health–related web platforms can be designed to address stigma. These categories include the design of inclusive web technologies, having a nonprovocative and calming user interface, having features that facilitate connections between users, and displaying personal accounts and experiences of sexual pain on web interfaces. Our sample revealed varying demographic information that ranged from marginalized populations including Indigenous, nonbinary, and people of color to White or Western populations. Participants’ indication of a marginalized identity suggests that the stigma of sexual pain may not exist in isolation but may intersect with multiple stigmatized identities to produce what is known as intersectional stigma [23].

These findings may help provide a conceptual or practical user-interface design guide for service providers to design digital health intervention that lessens or avoids the potential of fomenting stigma among people using sexual pain–related web platforms. We believe that these findings are not just applicable to sexual pain web platforms but could be applied to web-based platforms on other stigmatized conditions or disorders. By including male participants, this study also adds another dimension to sexual pain research in what is generally considered to be a woman’s health problem. The inclusion of people who identify as men male partners might have addressed the general lack of men’s involvement in sexual health research [24]. Even though participants who identified as men were not necessarily partners of those who identified as women as suggested in our prior study [10], they nevertheless illuminated the aspects of sexual pain that are specific to men yet often overlooked (eg, pain resulting from men who have sex with men).

Strategies to Addressing Stigma in Web Design
The themes identified from this study reflect patients’ perspectives on how web-based platforms can be designed to alleviate the stigma of sexual pain. The findings highlight the importance of inclusive user interfaces for people with diverse sexual pain experiences; gender-diverse people; and people with different cultural, ethnic, and religious identities. The findings of this study thus extend the idea of inclusive design beyond addressing the usual age, physical, mental, language, and accessibility limitations [25-27] to include specific stigma-related concerns that are often associated with people of diverse sexual and gender identities (two-spirit, lesbian, gay, bisexual, transgender, queer, and questioning, intersex, asexual) as well as diverse sexual preferences (anal sex and sex toys).

It is worthwhile to note that sexual pain is often discussed in relation to heterosexual encounters, people who engage in anal sex or sex using vibrators or sex toys may experience worse pain that, unfortunately, often ends up ignored in sexual pain–related platforms [28]. Designing to include these sexual perspectives is critical to foster inclusivity and may eventually help address some of the myths and misconceptions that often fuel stigma related to nonheterosexual forms of sexual encounters. On the contrary, user-interface design practices that exclude such groups from web platforms could be considered a form of marginalization through design [29]—a practice that may not only places people with nonheterosexual experiences into isolation but also amplifies the stigma associated with such sexual encounters.

The use of nonprovocative images was seen as a good design strategy for addressing stigma. We understand that intervention developers may want to display anatomical images of sexual organs on web platforms for enhanced comprehension and clarity of the information presented [30]. While useful, such practices may need to be reconsidered in sexual health web platforms as they may inadvertently reproduce, perpetuate, or exacerbate stigmatized feelings among people who find such explicit content undesirable. Participants’ suggestion for the display of nonprovocative images supports our previous assertion that provocative and explicit sexual images on web platforms could inadvertently foment stigma among end users [7,15,22]. It further confirms the assertion that the stigma of sexual health issues may be amplified through digital content, particularly digital images [31]. Furthermore, even though there was a general desire for inclusive images of nonheterosexual encounters and gender-diverse people, care should be taken to ensure that such images are not provocatively displayed on digital platforms that may end up stigmatizing an already stigmatized group. A possible solution may involve user testing of such content among the different user populations before they are displayed on web platforms.

Participants did not only recommend the use of nonprovocative images but also expressed a desire for the colors of such images, and the general color schemes of web platforms, to be calming and nondistressing to users. The differences in color preferences might be a reflection of the diverse cultural backgrounds of the participants. Participants’ position on the use of calming colors like green, yellow, and blue confirms Goethe’s theory of colors where he associated certain color categories (eg, yellow and green) with positive emotional responses of warmth and excitement, while other categories (eg, red) were associated with negative arousal [32,33]. Even though participants were not explicit on how the choice of colors can address stigma, the current theories on colors suggest that less distressful color schemes like green and yellow could have a healing effect on people who have experienced the trauma of painful sex [33]. While we identified what is considered as calming colors, the differences in these colors might also pose a new challenge in determining a unified color scheme that is considered calming to a significant majority of web users. Thus, further research is needed to examine how the visual effects and different
combinations of colors on user interfaces could alleviate the stigma of sexual pain.

Interpersonal communication features have been shown to provide informational support, instrumental support, belonging support, self-esteem, and emotional support to people with similar health conditions [34]. Participants’ recommendation for the integration of anonymous and interpersonal communication features (chat rooms, forum posts, and messaging boards) confirms the use of these interventions for people with other stigmatized conditions like HIV and AIDS [35]. The recommendations for networking functions in web platforms also support the use of social networking sites as an effective educational strategy for reaching out to people with diverse sexual experiences who often experience stigma [35]. With the increasing use of digital health technologies, such interpersonal connection features might move a significant part of sexual health services onto web-based platforms (eg, consultation, testing, and prescription). Beyond that, such interaction features could also foster new friendships for isolated individuals while providing opportunities for people to share and validate their personal experiences that are often internalized as negative feelings. These opportunities for interpersonal connection might not only make people feel less isolated but may also normalize their sexual pain experiences, thereby alleviating stigma [36].

Forums could have more effect on stigma reduction when people with internalized negative feelings can read personal accounts and testimonials from other people with the same disorder. Such testimonials may not just be limited to forums or messaging boards but could be displayed on landing pages, particularly if they are coming from popular personalities. Text or video testimonials from “famous” personalities have been shown to be very persuasive and might help elicit healthy emotions and memory [37]. Testimonials from famous personalities are especially important given the significant role of power and power dynamics in stigma formation [38]. Web-based video testimonials have shown effectiveness in alleviating explicit and implied stigma of mental illness among students [39], thus suggesting their usefulness for other stigmatized disorders in sexual health. Despite the importance of interpersonal connections on web platforms, it is important to be aware of the inherent security and privacy risks that could inadvertently expose people to further stigma or restigmatization [40]. The use of anonymous chat rooms and forum posts and moderation of web-based platforms by trained personnel is recommended to prevent the privacy risk associated with such platforms [41].

Strengths and Limitations

Our study relied on web-based data collection methods, which provided some degree of anonymity and might have boosted the confidence of participants to discuss a sensitive topic like painful sex. We believe that the diversity of our sample may have enriched our findings by bringing a nuanced perspective of gender-diverse and nonheterosexual experiences of painful sex and their associated stigmatized feelings. Males engaging in nonheterosexual sexual acts like anal sex or females using sex toys may experience worse forms of stigma and sexual pain, yet are often ignored on web platforms.

Despite these strengths, this study was not without limitations. Even though we included participants who identified as men, two-spirit, and nonheterosexual people, their involvement was limited to 2 and 3 participants, respectively. The relatively lower participation rate could not have allowed us to delve deeper into the unique perspective of two-spirit and people who identify as men on how to address stigma via web-based platforms. Thus, further studies are needed to examine sexual pain stigma specific to each minority group (men, two-spirit, and nonheterosexual people) and highlight ways to incorporate the findings into web-based platforms. By just showing participants only our sample website, we also believed their responses might have been limited to a single website. It is possible that participants’ responses might have differed if they had been shown other websites on sexual pain. Despite this limitation, using a single website helped participants, particularly those who might not have used sexual pain–related websites before to be more articulate on how to develop websites to alleviate stigma.

Conclusions

Web-based platforms on sexual pain have the potential to empower people, promote the development of new web-based friendships, and promote the sharing of personal experiences, thus reducing feelings of isolation. However, the manner in which these platforms are developed has implications by either alleviating or exacerbating the stigma of sexual pain. This study highlighted patient-centered design approaches that could serve as a reference guide in developing destigmatizing web platforms for people with sexual pain, particularly for those who feel stigmatized by their gender, racial, and sexual identity. While this study was conducted in the context of sexual pain, we believe that the results may also be applicable to web platforms on other stigmatized sexual health–related disorders or conditions and to stigmatized issues in mental health. As a next step, intervention studies may be needed to investigate how these guidelines might be applied in practice and whether web platforms resulting from these findings would indeed be less stigmatizing compared to websites that did not adopt such destigmatizing approaches.

Acknowledgments

The funding for this project was received from the UBC Hampton New Faculty Grant Competition (grant GR025485). The funding source has not been involved in the study design, data collection, analysis, writing of the report, or the decision to publish this manuscript.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 48 KB - formative_v8i1e53742_app1.docx]

References


24. Davis AC, Wright CJC, Haller MJ, Hellard ME, Lim MSC. A health education website developed to meet young people’s information needs about web-based pornography and sharing of sexually explicit imagery (SCOPE): usability study. JMIR Form Res 2019;3(3):e12824 [FREE Full text] [doi: 10.2196/12824] [Medline: 31411140]


34. Stetten NE, LeBeau K, Aguirre MA, Vogt AB, Quintana JR, Jennings AR, et al. Analyzing the communication interchange of individuals with disabilities utilizing Facebook, discussion forums, and chat rooms: qualitative content analysis of online disabilities support groups. JMIR Rehabil Assist Technol 2019;6(2):e12667 [FREE Full text] [doi: 10.2196/12667] [Medline: 31573937]


Jinyoung Baek¹, MA; Jonathan Lawson¹, CSPO; Vasiliki Rahimzadeh², MSc, PhD

¹Broad Institute, Cambridge, MA, United States
²Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX, United States

Corresponding Author:
Vasiliki Rahimzadeh, MSc, PhD
Center for Medical Ethics and Health Policy
Baylor College of Medicine
1 Baylor Plaza
Suite 310DF
Houston, TX, 77030
United States
Phone: 1 (713) 798 3500
Email: vasiliki.rahimzadeh@bcm.edu

Abstract

Background: New federal policies along with rapid growth in data generation, storage, and analysis tools are together driving scientific data sharing in the United States. At the same, triangulating human research data from diverse sources can also create situations where data are used for future research in ways that individuals and communities may consider objectionable. Institutional gatekeepers, namely, signing officials (SOs), are therefore at the helm of compliant management and sharing of human data for research. Of those with data governance responsibilities, SOs most often serve as signatories for investigators who deposit, access, and share research data between institutions. Although SOs play important leadership roles in compliant data sharing, we know surprisingly little about their scope of work, roles, and oversight responsibilities.

Objective: The purpose of this study was to describe existing institutional policies and practices of US SOs who manage human genomic data access, as well as how these may change in the wake of new Data Management and Sharing requirements for National Institutes of Health–funded research in the United States.

Methods: We administered an anonymous survey to institutional SOs recruited from biomedical research institutions across the United States. Survey items probed where data generated from extramurally funded research are deposited, how researchers outside the institution access these data, and what happens to these data after extramural funding ends.

Results: In total, 56 institutional SOs participated in the survey. We found that SOs frequently approve duplicate data deposits and impose stricter access controls when data use limitations are unclear or unspecified. In addition, 21% (n=12) of SOs knew where data from federally funded projects are deposited after project funding sunsets. As a consequence, most investigators deposit their scientific data into “a National Institutes of Health–funded repository” to meet the Data Management and Sharing requirements but also within the “institution’s own repository” or a third-party repository.

Conclusions: Our findings inform 5 policy recommendations and best practices for US SOs to improve coordination and develop comprehensive and consistent data governance policies that balance the need for scientific progress with effective human data protections.

(JMIR Form Res 2024;8:e49822) doi:10.2196/49822

KEYWORDS
biomedical research; survey; surveys; data sharing; data management; secondary use; National Institutes of Health; signing official; information sharing; exchange; access; data science; accessibility; policy; policies
Introduction

The rapid growth in human research data, storage, and analysis tools and skilled researchers combined with the declining costs and new federal policies are creating an immense drive for research data sharing [1]. At the same time, triangulating human research data of diverse types can accentuate the risks of reidentification when used for secondary research purposes. It also invites the potential for group and other individual dignitary harms that institutional ethics review committees may be underequipped to properly protect participants against [2] or prevented from considering outright based on statutory interpretation of the US Common Rule [3,4].

The lack of standardized methods for accurately tracking the provenance of data and their permitted uses further complicates the problem [5-7]. Most institutional oversight bodies (eg, institutional review boards [IRBs], privacy boards, and data access committees [DACs]) are also ill-equipped to audit whether and how well secondary data uses align with the terms outlined in participants’ initial consent. These inabilitys hinder opportunities for transparency, recourse, or accountability if uses are malign in indirect and new federal policies are creating an immense drive for research data sharing [1]. At the same, triangulating human research data of diverse types can accentuate the risks of reidentification when used for secondary research purposes.

The responsibility of overseeing the ethical conduct of research involving human subjects traditionally lies with IRBs. However, the sensitivity and complex nature of data governance in genomics research present a new challenge that can exceed the review capacities and expertise of traditional IRB oversight [2,8-10]. While DACs have emerged as potential solutions to this issue [11], they are not yet widely implemented in many research institutions and are still in the early stages of growth and practice [12].

Despite these challenges, the expectation to share data has broadened [13] under new federal policies (eg, National Institutes of Health [NIH] Data Management and Sharing [DMS] [14]) and is aided considerably by cloud storage and analysis platforms [15]. The new NIH DMS policy came into effect in January 2023. It requires, among other things, that scientific data stemming from NIH-supported projects must be made publicly accessible without delay and with few exceptions. The final DMS policy defines scientific data as “The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens” [14]. Emerging cloud computing environments afford new opportunities for tracking, auditing, and enforcing granular permissions [15] for scientific data in ways that were largely impracticable just several years ago. Yet, the broad use of provenance tracking tools continues to lag, creating situations where data stewards impose stricter barriers for access to data once up front, only to loosely permit the secondary use of these same data later in the research data lifecycle [16]. This situation is analogous to a house with a vaulted front door but a wide-open back door.

A clear example is the rigorous process of both a researcher and signing official (SO) agreeing to terms and conditions for the initial secondary access request. Subsequent sharing and use of consented data that transgress the consent, however, would likely go unmonitored by both the DAC approving access and the data recipient’s institution. This reality perversely incentivizes researchers to engage in informal data sharing to sidestep institutional data access review, which risks participants’ rights to privacy and the reputation of the research community. It is our view that limiting informal sharing pathways is consequential for both participant data protection and for maintaining trustworthy data governance practices.

Investing resources into optimizing compliant pathways that are easier and more expedient for data requestors without sacrificing participant terms of use should therefore be a high institutional priority.

Research institutions and repository managers could also confront increased liability burdens from inaccurate, inconsistent, or unspecified data use permissions and restrictions [17]. Misattributing use restrictions can result in regulatory noncompliance and instill a lack of confidence among prospective and current research participants that institutions will respect the terms of their data contributions. SOs, who often serve as signatories on data submission and access request agreements, bear significant responsibility for ensuring the appropriate transfer and use of research data within their institutions in this regard [17].

According to the NIH, SOs have the authority to “legally bind the institution in grant-administration matters by providing signature approval on grant application submissions” [18]. A graphical diagram of where SOs intervene on the data sharing and management processes is depicted in Figure 1. SOs further monitor “grant-related activities within the extramural organization” to ensure compliance with all grant requirements. Although SOs play a leadership role in research compliance, we know very little about the ways in which new data sharing requirements for researchers affect their scope of work and data governance responsibilities in the wake of new agency policies.
Figure 1. Signing officials act as institutional arbiters of data access between oversight bodies such as DACs and researchers requesting access to research data. DAC: data access committee; DUOS: Data Use Oversight System; IRB: institutional review board; PI: principal investigator.

This descriptive survey addresses a significant knowledge gap about the roles of US SOs, their knowledge about where research data from extramurally funded projects are deposited, how these data are accessed, and where data can be found after grant funding sunsets. Better understanding the roles and capabilities of SOs is important for identifying where institutional gatekeepers can intervene along the research data lifecycle to improve data provenance tracking and ethical data reuse in light of new NIH data sharing requirements [14]. We explore issues surrounding data sensitivity, consent tracking, institutional responsibilities, and liability considerations of SOs. As such, we seek to underscore the urgency of improving the distribution and coordination of institutional oversight bodies responsible for research data governance involving humans.

Methods

Study Design

We designed an anonymous survey of institutional SOs (Multimedia Appendix 1). The purpose of this survey study was to establish a baseline of existing institutional policies and practices as well as how these may change in light of new data sharing and management requirements for federally funded research in the United States. Survey items probed where data generated from extramurally funded research are deposited, how researchers outside the institution access these data, and what happens to these data after extramural funding ends.

Recruitment

Survey respondents were recruited from the websites of numerous institutions in the United States that receive funding from the NIH. We identified active SOs and used publicly available email addresses to contact them. Links to participate in the web-based survey were sent via email to each identified contact, and we used the software platform Qualtrics to administer the survey.

Ethical Considerations

This study was exempted by the Broad Institute of MIT and Harvard Office of Research Subject Protection. All survey responses were anonymized. Before advancing to the survey itself, respondents were presented with a detailed description of the study, its purpose, realistic risks and benefits of completing the survey, as well as information regarding the investigators and the funders before advancing to the survey itself. Consent to participate was indicated if respondents proceeded to complete the survey after reviewing the above study information. Respondents were not compensated for their participation.
Results

The survey took approximately 5 minutes to complete and probed SOs about the impacts of the new DMS policy on their institutional tasks and responsibilities. The survey response rate was 19%: a total of 56 participants (of 287 prospective participants who were contacted) completed the survey between February 6 and April 12, 2023.

Of the 56 respondents, 52% (n=29) indicated that they will submit more than 250 grants in the next calendar year (Table 2).

Table 1. Institutional types represented by survey respondents.

<table>
<thead>
<tr>
<th>Institution type</th>
<th>Values (n=56), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic-affiliated research institute</td>
<td>43 (77)</td>
</tr>
<tr>
<td>Academic institution</td>
<td>10 (18)</td>
</tr>
<tr>
<td>Nonprofit research institute</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Health care hospital system</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

Table 2. Estimation of how many NIH\(^a\) grants will be impacted by new DMS\(^b\) policy.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Values (n=56), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many data-generating NIH grants submitted this year would you estimate will be impacted by the DMS policy at your institution?</td>
<td></td>
</tr>
<tr>
<td>1-19</td>
<td>6 (11)</td>
</tr>
<tr>
<td>20-99</td>
<td>7 (13)</td>
</tr>
<tr>
<td>100-249</td>
<td>8 (14)</td>
</tr>
<tr>
<td>Over 250</td>
<td>29 (52)</td>
</tr>
<tr>
<td>I do not know</td>
<td>6 (11)</td>
</tr>
</tbody>
</table>

\(^a\)NIH: National Institutes of Health.

\(^b\)DMS: Data Management and Sharing.

Two-thirds (37/56, 66%) of respondents reported that less than 50% of scientific data generated by NIH funding will be controlled access data. When asked about the percentage of controlled access data generated by NIH funding, 46% (n=26) answered “less than 25%,” 20% (n=11) answered “25% to 49%,” 11% (n=6) answered “50% to 74%,” and 5% (n=3) answered “75% to 100%.” However, there is the potential for additional data sets to be categorized as controlled access data. We found that respondents defaulted to controlled access for data sets when a research participant’s preferences for “open access” versus “controlled access” were unclear in the original consent forms. In total, 27% (13/48) responded that they “always” default to controlled access sharing when the terms of access were unknown, while 52% (25/48) responded they “sometimes” do (Table 3).

Table 3. Nearly 80% (48/56) of survey respondents reported they default to controlled access for data where participant terms of use are unknown or ambiguous.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Values (n=48), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>When research participants’ preferences for open access versus controlled access are unclear in the consent, does your institution default to sharing under controlled access guidelines?</td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>13 (27)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>25 (52)</td>
</tr>
<tr>
<td>Rarely</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Never</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

The respondents reported that most investigators will not only deposit data into “an NIH-funded repository (eg, dbGaP, AnVIL, and BioDataCatalyst)” to meet the DMS requirements but also within the “institution’s own repository” as well as “a third-party repository (eg, Generalist Repository Ecosystem Initiative repositories, Terra, and Data Use Oversight System).” For respondents who reported depositing in an institutional repository, a combined 62% (22/35) reported there was no process for researchers external to the institution to view, request, or access data or that they did not know whether such a process existed (Table 4). Moreover, 77% (10/13) responded that they were unaware if data needed to be downloaded for
Table 4. Survey respondents reported whether they knew of a clear process for how external parties request access to data managed by the institution.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Values (n=35), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is there a clear process for researchers from other institutions to view, request, and access data (if approved)?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (37)</td>
</tr>
<tr>
<td>No</td>
<td>7 (20)</td>
</tr>
<tr>
<td>I do not know</td>
<td>15 (43)</td>
</tr>
</tbody>
</table>

For respondents who reported data deposits into a third-party repository, nearly 69% (20/29) said that their institution will keep a copy of the data (Table 5). In total, 31% (9/29) of the respondents who selected “a third-party repository” as a chosen platform (Table 6) reported that they are unsure about what will happen to the data at the end of the funded storage period paid for by the DMS allocated costs in each grant. The results suggest that there is a lack of clarity on the management and sharing of the data after the funded period for storage ends. This lack of clarity could mean a significant divergence in the life span and accessibility of publicly funded research data beyond the limited funded periods of storage made possible by DMS funding.

Table 5. Signing officials reported that their institution keeps a copy of research data in addition to other data submissions to external repositories.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Values (n=29), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will your institution keep a copy of the data?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20 (69)</td>
</tr>
<tr>
<td>No</td>
<td>2 (7)</td>
</tr>
<tr>
<td>I do not know</td>
<td>7 (24)</td>
</tr>
</tbody>
</table>

Table 6. US signing official perspectives on the types of repositories into which investigators deposit research data to comply with the new DMS policy.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Values (n=115), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What types of repositories will investigators in your institution deposit into to meet DMS policy requirements?</td>
<td></td>
</tr>
<tr>
<td>Your own institution’s data repository</td>
<td>36 (31)</td>
</tr>
<tr>
<td>An NIH repository (eg, dbGaP, AnVIL, and BioDataCatalyst)</td>
<td>48 (42)</td>
</tr>
<tr>
<td>A third-party repository (eg, GREI repositories, Terra, and DUOS)</td>
<td>29 (25)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (2)</td>
</tr>
</tbody>
</table>

Respondents felt that principal investigators (PIs) were best positioned to determine where data should be stored to comply with the DMS policy (Table 7). Nearly 58% (38/66) reported that “individual investigators” should decide on which repositories to submit data, while other parties included “chief compliance officer,” “SOs,” “chief information or technology officer,” and “librarians.” Few mentioned that the decision-making should be a shared responsibility among various stakeholders including PIs, IT officers, and SOs.
Table 7. Signing officials predominately felt that the project principal investigator was best positioned to decide where research data should be deposited.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Values (n=66), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who will decide where data are deposited to meet Data Management and Sharing policy requirements?</td>
<td></td>
</tr>
<tr>
<td>Individual investigators</td>
<td>38 (58)</td>
</tr>
<tr>
<td>Chief compliance officer</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Signing officials</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Chief information or technology officer</td>
<td>7 (11)</td>
</tr>
<tr>
<td>Librarians</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Shared responsibilities among various stakeholders</td>
<td>2 (3)</td>
</tr>
<tr>
<td>I do not know</td>
<td>2 (5)</td>
</tr>
</tbody>
</table>

When asked about who will oversee DMS policy compliance, 33% (19/56) reported that “SOs” should primarily serve that role, while 17% (10/56) mentioned “chief information or technology officer.” 16% (9/56) said “chief compliance officer,” and 9% (5/56) said “PIs.” Approximately 10% (6/56) noted that DMS policy compliance is responsible for multiple institutional offices.

Only 23% (12/52) reported that their institutions track where data from federally funded projects would be deposited, and nearly 37% (19/52) reported they are unaware whether such a tracking system exists. For respondents who reported a lack of a tracking system, 29% (n=6) reported their institutions plan to set up such a system. These data suggest nearly 77% (n=15) of institutions do not have a searchable method for determining where scientific data have been stored to track compliance with the new DMS policies (Table 8).

Table 8. Results from 52 survey respondents indicating that more than 77% (n=15) do not track where research data are deposited or are unaware of such tracking at their institution.

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Values (n=52), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does your institution have a way of tracking into which repositories investigators have deposited data to meet Data Management and Sharing policy requirements?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>12 (23)</td>
</tr>
<tr>
<td>No</td>
<td>21 (40)</td>
</tr>
<tr>
<td>I do not know</td>
<td>19 (37)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

Our survey data suggest that SOs approve data deposit into various repositories to comply with the DMS policy but may be unclear about how data stored within the institution are accessed or analyzed by researchers outside the institution. SOs report that many investigators will deposit data into both an NIH-funded repository to meet the DMS requirements and in their institution’s own repository, as well as sometimes in a third-party repository. Such duplicitous deposit has consequences for data storage, security, and costs, which are underacknowledged in the literature.

Investigators may store data in an institutionally controlled repository to comply with new DMS requirements and institutional data retention policies. However, these data may be more difficult for external investigators working in similar fields to find, access, and share in practice and thus could conflict with the FAIR (findability, accessibility, interoperability, and reusability) principles [19]. Storing the same data set in more than 1 repository or database effectively doubles the storage space and costs without the added benefit of improved ease or efficiency of access insofar as institutions rely on commercial companies to provide storage services.

While storage costs are typically reimbursed only once as a dedicated line item in grant budgets, commercial service providers profit twice: once for services rendered to store data in NIH-controlled repositories and also for storing these same data on institutional servers or in institutionally controlled databases.

This double storage problem also leads to inconsistent standards for data access management. Access requests for data hosted within NIH repositories and knowledge bases undergo more standardized processing by dedicated staff who review applications and verify authorized users. Data hosted on local institutional servers, in contrast, tend to be more liberally distributed to nonauthenticated users and governed post facto, if at all.

Few SOs know what happens to research data at the end of the contract or storage period for data stemming from federally funded research under the new DMS budget allocations. The most frequently cited reason was that legal or contractual obligations for different data types preclude any one uniform procedure for data handling. Sunsetting projects also threaten the availability of data in perpetuity when allowed costs for data management end. The lack of familiarity or knowledge of access processes outside the institution impedes SO’s abilities to promote sustained use of the data resource longer term.
Less than a quarter (12/25, 23%) of SOs work at institutions that track where data from federally funded projects have been deposited to meet DMS policy requirements. Tracking where data are deposited is critical to verifying compliance with the DMS policy as well as ensuring the repository meets participant-defined terms of use. Without knowing where data end up, SOs are unable to communicate to funders about the state of data availability or provide an accurate accounting of institutional resources needed to steward that data.

As a result, SOs defaulted to imposing greater access controls for data where consent was unclear or unspecified. In total, 27% (13/48) of SOs always defaulted to controlled access for data sets where participant consents were unclear or unspecified, which can translate to downstream issues for authorized data access. A precautionary default approach is advantageous for sensitive data sets where access controls are justified and should have been applied in the first place. Applying controlled access defaults for data that should otherwise be open, however, places undue access barriers that hinder compliant research.

SOs are prone to overcontrolling data when there is ambiguity in consent. Data from this survey suggest that more data will undergo controlled access even if permissioned broadly for general research use. Institutional data stewards, including SOs and DACs, will therefore need to manage increased demands for data. Our prior work with DACs showed they see value in testing automated decision support tools and software to improve the efficiency of access requests without sacrificing consistency or quality [16]. SOs participating in this survey were unaware of such tools, creating clear opportunities for targeted outreach and training.

**Textbox 1.** Recommendations for institutional signing officials to promote data provenance tracking, data stewardship, and improved observance of secondary data use permissions.

- Investigators should be encouraged to store data in a centralized location that can be accessible to authorized users at a distance.
- Recognize privileged access rights for original submitters of data for secondary use, provided they comply with data use terms and, where applicable, are approved by an institutional review board.
- Allow principal investigators to negotiate with National Institutes of Health program officers and committees about how research stemming from the use of private data sets should be publicly disseminated.
- Retain a master list of where all research data stemming from funded projects have been deposited.

To prevent paying double for data storage and management costs, investigators should be encouraged to store data in a centralized location that can be accessible to authorized users at a distance. Public-private partnerships with large commercial cloud service providers, including Amazon Web Services and Microsoft Azure, accentuate the immediacy of this problem, considering they are poised to host some of the nation’s largest research data sets in the coming years [23]. This practice would be highly beneficial for cloud-based repositories, for researchers whose preferred repository to deposit data is cloud-based, and for institutions that support researchers who rely on data sets managed in the cloud to secure future research grants. It is said that centralized storage could reduce the risk of data loss if the original files are corrupted. Data storage is also a growing environmental concern, favoring options that lessen the carbon footprint [24].

Facilitating data access in perpetuity is a reasonable benefit to offer researchers who contributed these data in the first place and could, in turn, incentivize more sharing. However, recognizing this privilege also assumes they will always use the data in ethical ways. While researchers could preserve a private copy of the data, a policy that recognizes special rights of access to original submitters of research data could still be considered insofar, as the proposed research uses accords with consent permissions and, where applicable, has been approved by an IRB.

There are many compelling reasons why researchers prefer to use private data sets [25]. Companies protect their proprietary

---

https://formative.jmir.org/2024/1/e49822

Baek et al

JMIR Form Res 2024 | vol. 8 | e49822 | p.2599

(page number not for citation purposes)
interests in the data sets they generate by not sharing them freely with the research community. NIH-funded researchers who nevertheless opt to use private data sets in their work do so with full knowledge of company policies against broad data sharing but should not be absolved of their responsibilities to comply with the DMS policy. The NIH could consider enforcement alternatives that help to close this loophole, possibly suspending funds for investigators who are noncompliant or negotiating the sharing of summary-level data to the extent feasible and possible. The government should be reimbursed for data management and storage costs in cases where the benefits of research knowledge using private data sets cannot be recovered. Research funding agencies, as well as individual institutions, should have a master list of where all data stemming from funded projects have been deposited. Researchers should easily be able to see where data generated from funded projects have been deposited, with accessible links to those repositories to directly request access. For example, a widget on the existing NIH Research Portfolio Online Reporting Tool could facilitate this.

Conclusions
The new NIH DMS policy is a significant step toward responsible data sharing, which will lead to accelerated scientific discoveries and innovation. However, it leaves the details of data sharing and management practices to be determined by individual investigators and their respective institutions. This could create inconsistencies and inefficiencies in data governance, as witnessed in the genomic data sharing ecosystem. Our survey of institutional SOs in the United States demonstrates a lack of clarity in terms of the location and duration of data storage, costs, and data security. Respondents suggested that scientific data generated by NIH funding are stored at multiple locations, including their institutions’ own data repositories, and are often unsure how external researchers gain access. Additionally, our survey shows that there is a pressing need for the development of proper tracking mechanisms to maintain and ensure the integrity of data sets, as many institutions currently do not keep track of where scientific data are stored once funded projects sunset. While more than half of the respondents suggested that PIs would be the sole decision makers on where to deposit data, respondents also acknowledged that compliance with the DMS policy will require institutional support and facilitate coordination between multiple offices.

Acknowledgments
This work was funded by the National Human Genome Research Institute (U24HG011025). The authors wish to thank the respondents for their time and willingness to participate in the study as well as anonymous peer reviewers who significantly helped strengthen earlier versions of this paper. During the preparation of this work, the authors used ChatGPT-4 available through OpenAI [26] to refine the grammatical structure of the text in the abstract and introduction of the paper. The authors take full responsibility for the content and accuracy of the data reported in this publication.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
JB has no competing interests to declare. JL is the senior software engineer for the data use ontology and a member of the Data Use and Research Identities Work Stream of the Global Alliance for Genomics and Health. VR’s work has been funded by the National Institutes of Health and is a member of the Regulatory and Ethics Work Stream of the Global Alliance for Genomics and Health. VR has received compensation for serving on the Ethical, Legal, and Social Issues working group of the Human Pangenome Reference Consortium and on the Bioethics Advisory Panel for the National Aeronautics and Space Administration. VR has also received fees from consulting on bioethics issues for the Coalition Against Childhood Cancer.

Multimedia Appendix 1
Anonymous signing official survey.

References


10. Shabani M, Borry P. "You want the right amount of oversight": interviews with data access committee members and experts on genomic data access. Genet Med 2016;18(9):892-897 [FREE Full text] [doi: 10.1038/gim.2015.189] [Medline: 26795589]


21. Donaldson DR, Koepke JW. A focus groups study on data sharing and research data management. Sci Data 2022;9(1):345 [FREE Full text] [doi: 10.1038/s41597-022-01428-w] [Medline: 35715445]


Abbreviations

DAC: data access committee
DMS: Data Management and Sharing
FAIR: findability, accessibility, interoperability, and reusability
IRB: institutional review board
NIH: National Institutes of Health

https://formative.jmir.org/2024/1/e49822

JMIR Form Res 2024 | vol. 8 | e49822 | p.2601

(page number not for citation purposes)

Please cite as:
Baek J, Lawson J, Rahimzadeh V
JMIR Form Res 2024;8:e49822
URL: https://formative.jmir.org/2024/1/e49822
doi:10.2196/49822
PMID:38506894

©Jinyoung Baek, Jonathan Lawson, Vasiliki Rahimzadeh. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Automated Category and Trend Analysis of Scientific Articles on Ophthalmology Using Large Language Models: Development and Usability Study

Hina Raja¹, PhD; Asim Munawar², PhD; Nikolaos Mylonas³, MS; Mohammad Delsoz¹, MBBS; Yeganeh Madadi¹, PhD; Muhammad Elahi⁴, BS; Amr Hassan⁵, MD; Hashem Abu Serhan⁶, MD; Onur Inam⁷,⁸, PhD; Luis Hernandez⁹, MD; Hao Chen¹,¹⁰, PhD; Sang Tran¹¹, MD; Wuqaas Munir¹¹, MD; Alaa Abd-Alrazaq¹², PhD; Siamak Yousefi¹, PhD

¹Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, TN, United States
²Watson Research Center, IBM Research, New York, NY, United States
³School of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
⁴Quillen College of Medicine, East Tennessee State University, Johnson, TN, United States
⁵Gavin Herbert Eye Institute, School of Medicine, University of California, Irvine, CA, United States
⁶Department of Ophthalmology, Hamad Medical Corporation, Doha, Qatar
⁷Edward S. Harkness Eye Institute, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY, United States
⁸Department of Biophysics, Faculty of Medicine, Gazi University, Ankara, Turkey
⁹Association to Prevent Blindness in Mexico, Ciudad, Mexico
¹⁰Department of Pharmacology, Addiction Science and Toxicology, University of Tennessee Health Science Center, Memphis, TN, United States
¹¹Department of Ophthalmology and Visual Sciences, School of Medicine, University of Maryland, Baltimore, MD, United States
¹²AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar

Corresponding Author:
Hina Raja, PhD
Department of Ophthalmology, University of Tennessee Health Science Center
930 Madison Avenue, Ste. 468
Memphis, TN, 38111
United States
Phone: 1 9016595035
Email: hinaraja65@gmail.com

Abstract

Background: In this paper, we present an automated method for article classification, leveraging the power of large language models (LLMs).

Objective: The aim of this study is to evaluate the applicability of various LLMs based on textual content of scientific ophthalmology papers.

Methods: We developed a model based on natural language processing techniques, including advanced LLMs, to process and analyze the textual content of scientific papers. Specifically, we used zero-shot learning LLMs and compared Bidirectional and Auto-Regressive Transformers (BART) and its variants with Bidirectional Encoder Representations from Transformers (BERT) and its variants, such as distilBERT, SciBERT, PubmedBERT, and BioBERT. To evaluate the LLMs, we compiled a data set (retinal diseases [RenD]) of 1000 ocular disease–related articles, which were expertly annotated by a panel of 6 specialists into 19 distinct categories. In addition to the classification of articles, we also performed analysis on different classified groups to find the patterns and trends in the field.

Results: The classification results demonstrate the effectiveness of LLMs in categorizing a large number of ophthalmology papers without human intervention. The model achieved a mean accuracy of 0.86 and a mean F₁-score of 0.85 based on the RenD data set.

Conclusions: The proposed framework achieves notable improvements in both accuracy and efficiency. Its application in the domain of ophthalmology showcases its potential for knowledge organization and retrieval. We performed a trend analysis that enables researchers and clinicians to easily categorize and retrieve relevant papers, saving time and effort in literature review and
information gathering as well as identification of emerging scientific trends within different disciplines. Moreover, the extendibility of the model to other scientific fields broadens its impact in facilitating research and trend analysis across diverse disciplines.

**KEYWORDS**

Bidirectional and Auto-Regressive Transformers; BART; bidirectional encoder representations from transformers; BERT; ophthalmology; text classification; large language model; LLM; trend analysis

**Introduction**

**Background**

A literature review is an integral component of the research process that involves systematically reviewing, evaluating, and synthesizing existing scholarly publications from databases such as MEDLINE (PubMed), Embase, and Google Scholar. The standard approach for a literature review involves using a bibliographic search engine to conduct an initial comprehensive search. Researchers use relevant keywords and filters, including clinical query filters, to retrieve a wide range of articles. The next steps include manually screening the retrieved articles by reviewing titles; abstracts; and, in most cases, full texts to assess their relevance and inclusion criteria. This combination of automated search and manual screening ensures a thorough review while targeting specific research objectives.

However, a literature review can be a challenging and time-consuming task for researchers, requiring meticulous examination of numerous sources and a critical analysis of their findings. The process demands substantial time and effort to effectively navigate through the vast expanse of scholarly literature from different databases and extract meaningful insights. Artificial intelligence tools have been used to facilitate this search process [1].

**Classical Methods**

Machine learning models have been applied to perform the text classification task based on feature engineering [2-4]. A semiautomated model was proposed for article classification in systemic review articles based on mechanistic pathways [5]. A total of 24,737 abstracts from both the PubMed and Web of Science databases and 861 references were found to be relevant. They evaluated the Naïve Bayes, support vector machines, regularized logistic regressions, neural networks, random forest, LogitBoost, and XGBoost models. The best-performing model achieved a sensitivity and specificity of approximately 70% and approximately 60%, respectively. Kanegasaki et al [6] used long short-term memory networks for the classification of abstracts. They used 2 data sets with 1307 and 1023 articles and achieved 73% and 77% respectively. These machine learning approaches were primarily based on feature engineering, which requires domain expertise.

**Supervised Natural Language Processing Models**

Natural language processing (NLP) applications have significantly advanced in recent years and gained tremendous popularity due to their wide range of applications across various domains. With the increasing availability of large data sets and advancements in computational power, NLP has made remarkable progress, revolutionizing the way we interact with technology [7-10]. In particular, NLP has gained interest in the field of information retrieval. NLP techniques, such as keyword extraction, document clustering, and semantic search, have improved the accuracy and relevance of the search results.

Hasny et al [11] used the Bidirectional Encoder Representations from Transformers (BERT) model for classifying articles into human, animal, and in vivo groups. Ambalavanan and Devarakonda [12] used SciBERT to classify scientific articles into 4 major categories, including format, human health care, purpose, and rigor. The format category included original studies, reviews, case reports, and general articles. The human health care category encompassed all articles discussing human health. The purpose category included articles discussing etiology, diagnosis, prognosis, treatment, costs, economics, and disease-related prediction. Rigor class included the studies that presented design criteria specific to a class purpose. The model achieved an $F_1$-score of 0.753 on the publicly available Clinical Hedges data set. Devlin et al [13] used the BERT model for the classification of scientific articles on randomized controlled trials. The BioBERT variant, trained on titles and abstracts, showed the highest performance of 0.90 in terms of the $F_1$-score.

Another study [14] fine-tuned variants of the BERT model, including BERTBASE, BlueBERT, PubMedBERT, and BioBERT, for the classification of human health studies. They used the abstracts and titles of 160,000 articles from the PubMed database. BioBERT showed the best results and achieved a specificity of 60% to 70% and a recall of >90%. The study [15] proposed a weakly supervised classification of biomedical articles. The model was trained on a weakly labeled subset of the biomedical semantic indexing and question answering 2018 data set based on MeSH (Medical Subject Headings) descriptors. BioBERT was used to generate the embedding for words and sentences, and then the cosine similarity was used to assign labels. The proposed model achieved an $F_1$-score of 0.564 for the BioASQ 2020 data set. BERT and its variant models have shown better performance for the text classification.

**Zero-Shot Learning Methods**

Conventional approaches to text classification have traditionally relied on the assumption that there is a fixed set of predefined labels to which a given article can be assigned. However, this assumption is violated when dealing with real-world applications, where the label space for describing a text is unlimited and the potential labels that can be associated with a text span an infinite spectrum, reflecting the diverse and nuanced nature of textual content. Such complexity challenges the conventional methods and calls for innovative strategies to navigate the expansive and unbounded label space. To address
these issues, zero-shot techniques [16-18] have been developed and are gaining popularity. Zero-shot learning (ZSL) involves classifying instances into categories without any labeled training data [19]. It leverages auxiliary information such as semantic embeddings or textual descriptions to bridge the gap between known and unknown categories. This enables the models to generalize to novel classes and make predictions for unseen categories. Mylonas et al [20] used zero-shot model for classifying PubMed articles into emerging MeSH descriptors. Instead of using the standard n-grams approach, the method exploited BioBERT embeddings at the sentence level to turn textual input into a new semantic space for the Clinical Hedges data set [21]. Unlike traditional models, the ZSL model does not explicitly require the labeled data; however, these models performed well on downstream tasks.

In this study, we have used large language models (LLMs) that include ZSL for categorizing the ophthalmology articles extracted from the PubMed database into different categories based on title and abstract. We fine-tuned the BERT model and its variants BERTBASE, SciBERT, PubmedBERT, and BioBERT for those categories that did not show good results from the ZSL model. Several powerful models, including Decoding-enhanced BERT with Disentangled Attention (mDeBERTa), Bidirectional and Auto-Regressive Transformers (BART), and the recently introduced Llama 2, present competitive alternatives to ChatGPT. However, Llama 2, despite its potential, imposes significant demands on graphics processing unit and memory resources. Even its smaller variant, with 7 billion parameters, requires substantial computational power, posing challenges for users with limited access to high-performance computing resources. To ensure broader accessibility, we prioritized a model with lower resource requirements. Accordingly, we selected the open-source BART model that is executable on central processing unit, thus providing both acceptable performance and enhanced accessibility to broader users. In addition, we performed a trend analysis based on the classified results, providing researchers with insights into emerging trends in the field to stay updated on the latest developments and identifying key areas of interest. Overall, we provide a method that enhances the efficiency, relevance, and interdisciplinary potential of the literature review process.

The rest of the study is organized as follows: the Methods section presents the materials and methods of the proposed framework for text classification and trend analysis. The Results section discusses the results of the different experiments performed for the evaluation of the proposed model. The Discussion section includes the principal findings and concludes the proposed work.

Contributions
Various classification models have been proposed in the literature for biomedical articles to retrieve relevant information [1-21]. Fine-tuned BERT and its variants have been used for text classification. Following are the main contributions of the studies:

- We have explored the ZSL models for the classification of biomedical articles.
- We have developed different use cases targeting the field of ophthalmology.
- To evaluate the model, we have generated a data set that includes 1000 articles related to ocular diseases. The articles were manually annotated by 6 experts into 15 categories.
- The ZSL model BART achieved a mean accuracy of 0.86 and an F1-score of 0.85.
- In addition to the classification of articles, we also performed a trend analysis on different classified groups.
- The model is adaptable to other biomedical disciplines without explicit fine-tuning or training.

Methods

Data Set
There are several annotated data sets available for various NLP tasks in the biomedical domain. However, in the field of ophthalmology, there is a scarcity of publicly accessible data sets for performing NLP tasks. To address this gap, we have taken the initiative to curate a data set focused on ocular diseases. Our retinal diseases (RenD) data set comprises 1000 articles sourced from PubMed, covering various conditions such as diabetic retinopathy (DR), glaucoma, diabetic macular edema, age-related macular degeneration, cataract, dry eye, retinal detachment, and central serous retinopathy. To ensure accurate categorization, we enlisted the expertise of 6 domain specialists who meticulously annotated the articles based on abstracts. To ensure accuracy and reliability in the annotation process, each article in our data set is reviewed and annotated by at least 2 individual annotators (refer to Table S1 in Multimedia Appendix 1 for the guidelines for data annotation). This multiple-annotator approach helps mitigate potential biases and inconsistencies that could arise from a single annotator’s perspective. Once the annotation is completed, the final label for each article is determined based on majority voting. Each article was annotated against 28 labels (Table S1 in Multimedia Appendix 1), and due to not having enough samples in some categories, we dropped those in further classification tasks. Thus, we selected 19 categories and grouped them into 4 categories (Table 1).

We will make this data set publicly accessible to the community for advancing research, facilitating comprehensive analysis, enabling more targeted investigations into ocular diseases, and promoting open science.
Table 1. Description of the data sets.

<table>
<thead>
<tr>
<th>Data set and group</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Retinal disease (N_T^a=1000 and N_C^b=19)</strong></td>
<td>Clinical, experimental, and automated model</td>
</tr>
<tr>
<td>Article type</td>
<td>DR^c, DME^d, AMD^e, glaucoma, dry eye, cataract, CSR^f, and retinal detachment</td>
</tr>
<tr>
<td>Ocular diseases</td>
<td>Screening, diagnosis, prognosis, etiology, and management</td>
</tr>
<tr>
<td>Clinical studies subclass</td>
<td>Image processing techniques, machine learning models, and deep learning model</td>
</tr>
<tr>
<td>Automated studies subclass</td>
<td></td>
</tr>
<tr>
<td><strong>Dry eye (N_T=67 and N_C=6)</strong></td>
<td>Tear film break up time, infrared thermography, lipid layer interface pattern, meibomian gland study, blink study, tear film assessment, and tear meniscus assessment</td>
</tr>
<tr>
<td>Clinical studies subclass</td>
<td></td>
</tr>
<tr>
<td><strong>Glaucoma (DemL; N_T=115 and N_C=2)</strong></td>
<td>Machine learning model and deep learning model</td>
</tr>
<tr>
<td>Automated studies subclass</td>
<td></td>
</tr>
</tbody>
</table>

^aN_T: the total number of articles in each data set.  
^bN_C: the total number of categories in each data set.  
^cDR: diabetic retinopathy.  
^dDME: diabetic macular edema.  
^eAMD: age-related macular degeneration.  
^fCSR: central serous retinopathy.

**Study Design**

Our framework automates the entire literature review process in the ophthalmology domain (Figure 1). More specifically, by incorporating user-defined criteria, including keywords, the number of articles, inclusion criteria, and categories for classification. Our framework performs a systematic retrieval and analysis of relevant articles automatically. Initially, a keyword is fed into PubMed, and the related articles are fetched, including the abstract, title, publication year, and link. The preprocessing step in this context involves the selection of a specific subset of articles from a larger corpus of fetched articles. This selection process is contingent upon the application of predefined inclusion criteria, with a specific temporal constraint. In this study, articles falling within the temporal range spanning from 2015 to 2022 are considered for inclusion in the subsequent analyses. This temporal delimitation serves as a crucial preprocessing measure, narrowing down the data set to a more focused and relevant time frame for the research objectives. The selected articles are then fed into the LLM for classification based on user-defined categories. For LLM, first we targeted using the ZSL models that can classify the text without explicit training. If the ZSL model is unable to achieve better performance, then we fine-tune the BERT model. Finally, after the article classification by LLM, we have performed 2 types of trend analysis.
Figure 1. Flow diagram of the proposed framework. The model takes input (keyword, inclusion criteria, and categories for classification), and articles are fetched from PubMed based on keyword. The inclusion criteria are fed into the preprocessing module to select the desired articles from the fetched data. A large language model classifies the articles based on the predefined categories. Finally, trend analysis is performed on classified categories.

Zero-Shot Classification

Zero-shot classification is an approach to predict the class of instances for categories they have never seen during training, using auxiliary information or semantic embeddings. It enables generalization to unseen classes and expands the classification capabilities beyond the limitations of labeled training data. We have used BART [22], which is pretrained based on a sequence-to-sequence model that combines bidirectional and autoregressive techniques for improved text generation and comprehension. BART has 12 transformer layers with a hidden size of 1024 that was initially trained on Wikipedia and the BookCorpus data set and fine-tuned on Multi-Genre Natural Language Inference tasks. BART is an amalgamation of the bidirectional encoder found in BERT and the autoregressive decoder used in Generative Pretrained Transformer (GPT) (details of architecture can be found in Figure S1 in Multimedia Appendix 1). While BERT comprises approximately 110 million trainable parameters and GPT-3 consists of 117 million parameters, BART, being a combination of the 2, has approximately 140 to 400 million parameters. This larger parameter count in BART accommodates its sequenced structure, which incorporates both encoding and decoding capabilities for a wide range of NLP tasks. The model receives the title and abstract of an article as input and generates probabilities for each class, and if the probability for a specific category is higher than a threshold value, the article is assigned the label corresponding to that category.

In these equations, \( p(i) \) is the probability of an article based on the title and abstract, \( C \) is the total number of classes for a particular category, and \( \xi \) is the threshold.

Fine-Tunning the Classification Model

BERT [13] and its variant models, namely, distilBERT, SciBERT, PubmedBERT, and BioBERT, have been subjected to fine-tuning to address classification tasks for categories in which the ZSL model (BART) is unable to produce more accurate results. The preprocessing stage entails the concatenation of article titles and abstracts, which are subsequently input into the respective BERT model. The model generates probabilities for each class, and if the probability for a specific category is higher than a threshold value, the article is assigned the label corresponding to that category.

Trend Analysis

In addition to the classification of articles, we performed 2 additional analyses as well. More specifically, we performed a technology trend analysis to obtain valuable insights into the distribution of research across different classes, highlighting classes with higher or lower publication frequencies. This information aids in understanding the emphasis and focus of research efforts, enabling resource allocation, and identifying areas that may require further attention or investigation. We also performed an interest trend analysis to provide a comprehensive view of publication trends over specific periods. By identifying the popularity of techniques or topics over time,
this analysis facilitates the detection of emerging trends and the evaluation of long-term patterns. These trend analyses, applicable to all levels of classification categories, contribute to an enhanced understanding of the dynamic nature and evolving landscape of research in the field.

**Ethical Considerations**

We have not included any human and animals in our study.

**Results**

This section presents the experiments we have performed to evaluate the LLMs for classification.

**Experimental Details**

We evaluated our models based on the RenD data set that we annotated. In addition, we evaluated the models to classify categories based on 2 review studies related to dry eye disease and glaucoma (Table 1). We present the results in terms of accuracy, area under the curve (AUC; \( F_1 \)-score, precision, and recall for each data set. For multilabel classification, we evaluated the model in terms of \( F_1 \) micro, \( P \) micro, \( R \) micro, and AUC.

**Ablation Study**

**Overview**

Our study encompasses both multiclass and multilabel classification tasks. To accomplish this, we used the ZSL model and fine-tuned the model for which the ZSL was not performing well. Through a series of ablation experiments, we systematically investigated the impact of different settings (refer to the subsequent sections) on the performance of the model. By modifying and assessing various settings, we gained insights into the individual contributions and effects of each setting, allowing us to refine and optimize our approach accordingly.

**ZSL Model Selection**

In our study, we conducted an evaluation of the ZSL state-of-the-art models and multiple variants of the BART model. On the basis of a comprehensive analysis, we identified the model variant that exhibited the most favorable performance in our specific context (Table 2).

We selected the BART model that showed the best performance. In addition, we performed experiments using different keywords for the different categories. It was observed that for the ZSL model, the prompts should be more descriptive and provide some information about related categories to improve the accuracy (Table 3).

For the category “Clinical, Experimental, and Automated Model,” we have tested various keywords and found that “Clinical finding based on humans,” “Experimental study based on animals,” and “Technical study based on automated model” keywords showed the best results with an accuracy of 0.91 and an \( F_1 \)-score of 0.92. During our evaluation, we investigated the potential of using abstracts and titles for classification across various categories. We discovered that classification solely based on titles closely approximates the results obtained from using abstracts for most of the categories. However, the abstract and title together enhance the efficacy of the classification.

### Table 2. Evaluation of the zero-shot learning classification models for category 1 from the retinal disease (RenD) data set.

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Time (minutes)</th>
<th>Accuracy</th>
<th>( F_1 )-score</th>
<th>AUC(^a)</th>
<th>Precision</th>
<th>Recall</th>
<th>Time (minutes)</th>
<th>Accuracy</th>
<th>( F_1 )-score</th>
<th>AUC</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>BART(^b)-base</td>
<td>34.34</td>
<td>0.08</td>
<td>0.03</td>
<td>0.42</td>
<td>0.86</td>
<td>0.08</td>
<td>3.5</td>
<td>0.01</td>
<td>0.005</td>
<td>0.5</td>
<td>0.006</td>
<td>0.01</td>
</tr>
<tr>
<td>Bart-large</td>
<td>104.5</td>
<td>0.11</td>
<td>0.03</td>
<td>0.46</td>
<td>0.17</td>
<td>0.115</td>
<td>12.24</td>
<td>0.5</td>
<td>0.57</td>
<td>0.39</td>
<td>0.68</td>
<td>0.50</td>
</tr>
<tr>
<td>Bart-large-CNN(^c)</td>
<td>37.63</td>
<td>0.08</td>
<td>0.03</td>
<td>0.42</td>
<td>0.86</td>
<td>0.08</td>
<td>4.2</td>
<td>0.36</td>
<td>0.50</td>
<td>0.58</td>
<td>0.84</td>
<td>0.36</td>
</tr>
<tr>
<td>Bart-mnli(^d)-CNN</td>
<td>231.42</td>
<td>0.74</td>
<td>0.78</td>
<td>0.65</td>
<td>0.84</td>
<td>0.74</td>
<td>17.76</td>
<td>0.09</td>
<td>0.06</td>
<td>0.42</td>
<td>0.2</td>
<td>0.09</td>
</tr>
<tr>
<td>mDeBERT(^e)-a-v3-base</td>
<td>79.28</td>
<td>0.87</td>
<td>0.85</td>
<td>0.74</td>
<td>0.88</td>
<td>0.87</td>
<td>31.06</td>
<td>0.76</td>
<td>0.82</td>
<td>0.80</td>
<td>0.91</td>
<td>0.76</td>
</tr>
<tr>
<td>Bart-large-mnli</td>
<td>141.50</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
<td>0.93</td>
<td>0.91</td>
<td>15.21</td>
<td>0.91</td>
<td>0.82</td>
<td>0.93</td>
<td>0.94</td>
<td>0.91</td>
</tr>
</tbody>
</table>

\(^a\)AUC: area under the curve.

\(^b\)BART: Bidirectional and Auto-Regressive Transformers.

\(^c\)CNN: convolution neural network

\(^d\)Multi-Genre Natural Language Inference.

\(^e\)BERT: Bidirectional Encoder Representations from Transformers.
Table 3. Investigation of prompts for classifying the retinal diseases data set using the Bidirectional and Auto-Regressive Transformers (BART) zero-shot learning model. Articles are explicitly categorized using abstract.

<table>
<thead>
<tr>
<th>Category and prompt</th>
<th>Abstract Accuracy</th>
<th>$F_1$-score</th>
<th>AUC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Precision</th>
<th>Recall</th>
<th>AUC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Precision</th>
<th>Recall</th>
<th>AUC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Precision</th>
<th>Recall</th>
<th>AUC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Precision</th>
<th>Recall</th>
<th>AUC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical study, experimental study, and automated model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Clinical Study,” “Experimental Study,” and “Automated Studies”</td>
<td>0.80</td>
<td>0.82</td>
<td>0.70</td>
<td>0.85</td>
<td>0.80</td>
<td>0.67</td>
<td>0.76</td>
<td>0.80</td>
<td>0.67</td>
<td>0.76</td>
<td>0.80</td>
<td>0.67</td>
<td>0.76</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Clinical Study,” “Experimental Study,” and “Automated Model”</td>
<td>0.80</td>
<td>0.83</td>
<td>0.74</td>
<td>0.86</td>
<td>0.80</td>
<td>0.68</td>
<td>0.77</td>
<td>0.80</td>
<td>0.68</td>
<td>0.77</td>
<td>0.80</td>
<td>0.68</td>
<td>0.77</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Clinical Study,” “Experimental Study based on animals,” and “Technical study based on Automated Model”</td>
<td>0.85</td>
<td>0.87</td>
<td>0.91</td>
<td>0.92</td>
<td>0.85</td>
<td>0.87</td>
<td>0.91</td>
<td>0.92</td>
<td>0.87</td>
<td>0.91</td>
<td>0.92</td>
<td>0.87</td>
<td>0.91</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Clinical Finding based on humans,” “Experimental Study based on animals,” and “Technical study based on Automated Model”</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
<td>0.93</td>
<td>0.91</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Image processing techniques, machine learning models, and deep learning models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Deep learning Model,” “Image processing technique,” and “ONLY Machine learning”</td>
<td>0.65</td>
<td>0.54</td>
<td>0.12</td>
<td>0.47</td>
<td>0.65</td>
<td>0.68</td>
<td>0.55</td>
<td>0.05</td>
<td>0.47</td>
<td>0.68</td>
<td>0.55</td>
<td>0.47</td>
<td>0.68</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Deep learning Model,” “Image processing technique,” and “Classic Machine learning”</td>
<td>0.66</td>
<td>0.57</td>
<td>0.74</td>
<td>0.79</td>
<td>0.66</td>
<td>0.69</td>
<td>0.60</td>
<td>0.73</td>
<td>0.71</td>
<td>0.69</td>
<td>0.60</td>
<td>0.73</td>
<td>0.71</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Deep learning Model,” “Digital Image processing technique,” and “Classic Machine learning”</td>
<td>0.65</td>
<td>0.58</td>
<td>0.68</td>
<td>0.61</td>
<td>0.65</td>
<td>0.66</td>
<td>0.56</td>
<td>0.71</td>
<td>0.76</td>
<td>0.66</td>
<td>0.56</td>
<td>0.71</td>
<td>0.76</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Deep learning Model,” “Digital Image processing technique,” and “Machine learning Model”</td>
<td>0.82</td>
<td>0.82</td>
<td>0.87</td>
<td>0.86</td>
<td>0.82</td>
<td>0.92</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.95</td>
<td>0.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>AUC: area under the curve.

<sup>b</sup>Italicized prompts show the best results for that particular category.

**Hyperparameters for Fine-Tuning BERT**

For the categories in which the ZSL model (BART) provided poor results, we fine-tuned the BERT model and its variants to perform categorization. We conducted hyperparameter tuning based on this to enhance the model’s reliability and significance. By carefully selecting and fine-tuning hyperparameters such as the learning rates, batch sizes, and regularization strengths, we aimed to achieve accurate and meaningful results (Table S2 in Multimedia Appendix 1). For the BioBERT model, we selected a learning rate of 1e-05, a batch size of 8, a maximum length of 400, and a number of epochs of 20.

**Evaluation Results**

**Article Classification Evaluation**

This section presents the results based on the metrics that were selected in the ablation experiments. On the basis of the evaluation, BART demonstrated the best performance among
the tested models. Therefore, further classification tasks were conducted using the BART model to capitalize on its superior performance. Table 4 shows the classification results using BART for the RenD data set for the categories of article type, ocular diseases, clinical studies subclass, and automated studies subclass.

The article type group was classified into 3 subcategories: clinical, experimental, and automated studies. The BART model demonstrated promising performance for the article type group, with an accuracy of 0.91, an $F_1$-score of 0.92, an AUC of 0.91, a precision of 0.93, and a recall of 0.91. For the article group type, classification based on only abstract and only title and combination of both are performing consistent. The automated model group is further categorized into image processing techniques, machine learning models, and deep learning models. For the automated study subclass group, the ZSL model achieved the best performance for title-based classification, with accuracy, $F_1$-score, AUC, precision, and recall of 0.92, 0.95, 0.94, and 0.92, respectively. However, the second-best scores were achieved by classification based on abstract and title.

The clinical studies are further categorized into screening, diagnosis, prognosis, etiology, and management, constituting a multilabel classification scenario. However, ZSL achieved the best score of $F_1$ micro of 0.52, AUC of 0.68, precision micro of 0.49, and recall micro of 0.61. The ocular group is classified into DR, diabetic macular edema, age-related macular degeneration, glaucoma, dry eye, cataract, central serous retinopathy, and retinal detachment. In terms of accuracy, $F_1$-score, AUC, precision, and recall, the classification based on titles yielded the most favorable outcomes. Specifically, the results for the title-based classification were 0.85 accuracy, 0.85 $F_1$-score, 0.92 AUC, 0.89 precision, and 0.86 recall. Following closely were the results for the abstract-based classification, with values of 0.85 accuracy, 0.83 $F_1$-score, 0.91 AUC, 0.87 precision, and 0.85 recall.
Table 4. Classification of scientific articles from retinal disease (RenD) data set into 4 groups: article type, ocular diseases, clinical studies subclass, and automated studies subclass, which are classified into 3, 8, 4, and 3 categories, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Article type (MC)</th>
<th>Ocular diseases (MC)</th>
<th>Clinical studies subclass (ML)</th>
<th>Automated studies subclass (MC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.91&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.85&lt;sup&gt;d&lt;/sup&gt;</td>
<td>—</td>
<td>0.82</td>
</tr>
<tr>
<td>$F_1$-score</td>
<td>0.92</td>
<td>0.83&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.49</td>
<td>0.82</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.91 (0.89-0.92)</td>
<td>0.91&lt;sup&gt;d&lt;/sup&gt; (0.85-0.92)</td>
<td>0.67 (0.64-0.70)</td>
<td>0.87 (0.85-0.90)</td>
</tr>
<tr>
<td>Precision</td>
<td>0.93</td>
<td>0.87&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.33</td>
<td>0.86</td>
</tr>
<tr>
<td>Recall</td>
<td>0.91</td>
<td>0.85&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td><strong>Title</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.91</td>
<td>0.85</td>
<td>—</td>
<td>0.92&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>$F_1$-score</td>
<td>0.92</td>
<td>0.85</td>
<td>0.50</td>
<td>0.92&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.91 (0.87-0.93)</td>
<td>0.92 (0.86-0.94)</td>
<td>0.67 (0.64-0.71)</td>
<td>0.95&lt;sup&gt;d&lt;/sup&gt; (0.79-0.88)</td>
</tr>
<tr>
<td>Precision</td>
<td>0.93</td>
<td>0.89</td>
<td>0.42</td>
<td>0.94&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Recall</td>
<td>0.91</td>
<td>0.86</td>
<td>0.61</td>
<td>0.92&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Probability (abstract+title)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.85</td>
<td>0.78</td>
<td>—</td>
<td>0.90&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>$F_1$-score</td>
<td>0.87</td>
<td>0.73</td>
<td>0.51</td>
<td>0.89&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.91 (0.87-0.94)</td>
<td>0.86 (0.79-0.88)</td>
<td>0.67 (0.79-0.88)</td>
<td>0.93&lt;sup&gt;d&lt;/sup&gt; (0.89-0.94)</td>
</tr>
<tr>
<td>Precision</td>
<td>0.92</td>
<td>0.73</td>
<td>0.42</td>
<td>0.91&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Recall</td>
<td>0.85</td>
<td>0.78</td>
<td>0.61</td>
<td>0.90&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Appending title to abstract</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.91&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.84</td>
<td>—</td>
<td>0.90&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>$F_1$-score</td>
<td>0.91&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.82</td>
<td>0.52</td>
<td>0.89&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.91&lt;sup&gt;d&lt;/sup&gt; (0.86-0.92)</td>
<td>0.90 (0.86-0.92)</td>
<td>0.68 (0.62-0.71)</td>
<td>0.93&lt;sup&gt;e&lt;/sup&gt; (0.90-0.95)</td>
</tr>
<tr>
<td>Precision</td>
<td>0.93&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.86</td>
<td>0.49</td>
<td>0.91&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Recall</td>
<td>0.91&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.84</td>
<td>0.61</td>
<td>0.90&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>MC: multiclass classification.
<sup>b</sup>ML: multilabel classification.
<sup>c</sup>The best results are italicized.
<sup>d</sup>The second-best scores.
<sup>e</sup>Not available.
<sup>f</sup>AUC: area under the curve.

**Trend Analysis**

A category-wise analysis was performed for the article type and ocular disease groups based on the RenD data set (**Figure 2**). Some other results are also reported in Figure S2 in Multimedia Appendix 1.

Category-wise analysis showed more frequent papers on clinical studies compared to experimental- and automated-based studies. For ocular diseases, more studies have discussed DR and glaucoma compared to other ocular diseases. A timewise analysis was conducted for the subgroup of automated studies from 2015 to 2022. The trends indicated that, in the initial years, these studies primarily relied on image processing techniques. However, as time progressed, machine learning gained traction, and eventually, deep learning models became increasingly popular in this field, reflecting how technology is evolving in ophthalmology. For the time-wise analysis of diseases, refer to Multimedia Appendix 1.
Discussion

Principal Findings

We have developed a framework aimed at streamlining the literature review process. This framework involves an automated system that takes user-specified keywords as input. By leveraging these keywords, the system retrieves relevant articles from the PubMed database. In addition, the user specifies the desired categorization for these articles. This approach aims to simplify and expedite the traditionally time-consuming task of conducting literature reviews. We investigated the efficacy of using the LLM model to perform article classification.

LLMs, particularly ChatGPT, have gained immense popularity due to their versatility in performing various tasks, including

![Category-wise Analysis](image)

![Time-wise Analysis](image)

Figure 2. Trend analysis of classified articles: (A) and (B) category-wise analysis for article type and ocular diseases group, respectively, and (C) timewise analysis for automated studies subclass group: image processing techniques, machine, and deep learning models. AMD: age-related macular degeneration; CSR: central serous retinopathy; DME: diabetic macular edema; DR: diabetic retinopathy.
question answering and trend analysis within specific fields. Moreover, these models demonstrate the capability to generate research papers, letters, and other written content, showcasing their potential for creative text generation. However, the generated content is not always entirely authentic, as it can occasionally produce fake references and links, raising concerns about the reliability and accuracy of the information presented. Therefore, we proposed a framework for automating the literature review process and finding different trends in various disciplines. However, a limitation of ChatGPT-3.5 is the fact that it is not equipped with information beyond September 2021; therefore, it may not provide facts or knowledge beyond this date. We target using open-source LLMs for article classification and then performing category-wise and timewise analysis. Other open-source LLMs have become available recently. For instance, mDeBERTa, BART, and the recently released Llama 2 and its variants may outperform ChatGPT. However, using Llama 2 requires significant graphics processing unit and memory resources. Even the small variant of the model, with 7 billion parameters, demands substantial computational power to function effectively. These resource requirements can pose challenges for users with limited access to high-performance hardware.

Hence, our primary goal is to select a model that not only delivers robust performance but also requires fewer computational resources, thereby enhancing accessibility for a broader user base. The BART model aligns with these criteria as it is open source, allowing seamless execution on central processing unit. This choice ensures a balance between efficiency and performance, making advanced NLP capabilities accessible to a wider community. In a direct comparison with alternative models, BART stands out, showcasing superior overall performance across various evaluation metrics and in terms of computational resources.

Categorization, Classification, and Trend Analysis
We used BART as the ZSL classifier, and we used the abstract and title separately for article classification. After obtaining the probabilities from each model, we combined the probabilities and performed classification. In addition, we also appended the title to the abstract and added it to the models.

The BART model showed compromising results for the categories article type, ocular diseases, and automated studies subclass of the RenD data set. The classifications based on abstract and title are nearly similar in performance. For clinical studies subclass grouping, the BART achieved an F1 micro score of 0.52 and an AUC of 0.68. To improve the performance for this class, we fine-tuned BERT and its variant, BioBERT, which performed best with an F1 micro score of 0.67 and an AUC of 0.70. The lower performance in this class is likely due to the class imbalance, which typically affects the model’s ability to generalize and accurately predict instances of the minority class.

We also evaluated the performance of the BART model to classify the articles into different categories for 2 undergoing review studies including DEye and DemL. The articles in both the review studies were annotated by reviewing the entire article. However, we just used the abstract and title, and for DEye, our model achieved an AUC of 0.79 and an F1-score of 0.63. To improve the performance of the BART model, we performed a hierarchical analysis in which the multilabel task is divided into a binary classification. Classification was performed across different thresholds for each class, and the optimal value was chosen based on the best results achieved (Table S3 in Multimedia Appendix 1). The results showed that converting the multilabel problem into binary classification improves the performance of the BART model. In addition to this, we also observed that abstract-based classification and whole article–based annotation provided comparable results for each class (Table 5).

On the basis of the DemL data set, the BART model’s classification is based on the abstract, and the title is as accurate as the whole article–based annotation for the DemL data set.
Table 5. Bidirectional and Auto-Regressive Transformers (BART) model evaluation for classification of the Dry eye and DemL data sets.

<table>
<thead>
<tr>
<th>Data set and category</th>
<th>Classification type</th>
<th>Abstract</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>$F_1$-score</td>
<td>AUC&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Dry eye</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tear film break up time, infrared thermography, lipid layer interface pattern, meibomian gland study, tear film assessment, and tear meniscus assessment</td>
<td>MC&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.63</td>
<td>0.79</td>
</tr>
<tr>
<td>Infrared thermography</td>
<td>BC&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.91</td>
<td>0.73</td>
</tr>
<tr>
<td>Lipid layer interface pattern</td>
<td>BC</td>
<td>0.94&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.77&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Meibomian gland study</td>
<td>BC</td>
<td>0.91</td>
<td>0.54</td>
</tr>
<tr>
<td>Tear film assessment</td>
<td>BC</td>
<td>0.92</td>
<td>0.87</td>
</tr>
<tr>
<td>Tear meniscus assessment</td>
<td>BC</td>
<td>0.98&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Glaucoma (DemL)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine learning model and deep learning model</td>
<td>MC&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

<sup>a</sup>AUC: area under the curve.
<sup>b</sup>MC: multiclass classification.
<sup>c</sup>BC: binary classification.
<sup>d</sup>Second best score.
<sup>e</sup>Best scores are italicized.
<sup>f</sup>MC: multilabel classification.

**Comparative Analysis of Computational Time**

We have performed a comparative analysis of the processing time between the LLM and human annotators, which has unveiled intriguing insights. This analysis delves into the time required for classification based on the abstract, the title, and both the title and abstract of scientific articles. Manually annotating articles is a time-consuming task, as human annotators require a significant amount of time to label each article. The overall process can span over several weeks, depending on the number of articles and the number of categories for annotations. For instance, annotating the abstract of 1 article with 2 categories may take, on average, 4 to 5 minutes. However, automated models take notably less time to complete similar tasks (Table 6).

Notably, using both title and abstract as input led to a slightly increased processing time for the BART model, although it remained significantly faster than human annotation. We conducted 2 types of trend analysis: category wise and timewise. These analyses can be applied to any classified category and can highlight different trends in a concise and quick manner. A report is generated at the end, encompassing user-specified inclusion criteria and other relevant aspects to aid researchers (Figure S3 in Multimedia Appendix 1).
Table 6. Comparative analysis of processing time by large language model (Bidirectional and Auto-Regressive Transformers [BART]) and human annotator.

<table>
<thead>
<tr>
<th>Data set and category</th>
<th>Articles, n (%</th>
<th>Abstract (minutes)</th>
<th>Title (minutes)</th>
<th>Title and abstract (minutes)</th>
<th>Annotation by human (minutes [approximately])</th>
<th>Timeline (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Retinal disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical, experimental, and automated model</td>
<td>1000 (100)</td>
<td>141.50</td>
<td>15.21</td>
<td>194.2</td>
<td>3000</td>
<td>4</td>
</tr>
<tr>
<td>DR, DME, AMD, glaucoma, dry eye, cataract, CSR, and retinal detachment</td>
<td>1000 (100)</td>
<td>274.06</td>
<td>27.30</td>
<td>283.43</td>
<td>4000</td>
<td>4</td>
</tr>
<tr>
<td>Screening, diagnosis, prognosis, etiology, and management</td>
<td>464 (59)</td>
<td>118.71</td>
<td>13.11</td>
<td>120.34</td>
<td>1600</td>
<td>4</td>
</tr>
<tr>
<td>Image processing techniques, machine learning model, and deep learning model</td>
<td>156 (15.6)</td>
<td>21.23</td>
<td>2.45</td>
<td>23.78</td>
<td>400</td>
<td>4</td>
</tr>
<tr>
<td><strong>Dry eye</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tear film break up time, infrared thermography, lipid layer interface pattern, meibomian gland study, tear film assessment, and tear meniscus assessment</td>
<td>67 (100)</td>
<td>36.45</td>
<td>6.45</td>
<td>37.12</td>
<td>2800</td>
<td>2</td>
</tr>
<tr>
<td><strong>Glaucoma (DemL)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep learning model and machine learning model</td>
<td>115 (100)</td>
<td>19.30</td>
<td>1.83</td>
<td>32.23</td>
<td>1000</td>
<td>1</td>
</tr>
</tbody>
</table>

aDR: diabetic retinopathy.
bDME: diabetic macular edema.
cAMD: age-related macular degeneration.
dCSR: central serous retinopathy.

**Limitations**

The limitation of this study is that we have included articles from the PubMed database, which may have resulted in the exclusion of relevant articles related to the chosen keyword. However, future plans involve the integration of additional databases such as Google Scholar, IEEE Xplore, and Springer to address this limitation and ensure a more comprehensive coverage of relevant literature. Articles from various databases can unveil trends and patterns that transcend specific domains, increasing the applicability of the findings.

We used the ZSL model for text classification. As the model is not specifically trained for downstream tasks, there are several potential biases and challenges to consider. The model may not fully comprehend the semantics or nuances of the new task, leading to biases in predictions or misinterpretations. Downstream tasks have different data distributions compared to the original ZSL task, causing the model to struggle with new patterns or biases.

We selected BART as the ZSL model, as it is the latest open-source model. BART pretraining results in the creation of semantic embeddings that capture various linguistic nuances. These embeddings enable the model to understand the semantics of different tasks, even for those tasks it has not been explicitly trained on. Its task-agnostic pretraining allows it to be adapted for various other fields by providing task-specific prompts during inference. Leveraging BART for zero-shot tasks often involves careful prompt engineering. By formulating prompts that guide the model to perform specific tasks or make predictions for unseen classes, users can harness the model’s prelearned linguistic capabilities. This can be addressed by carefully designing the prompt and adding a little description instead of using 1 word for each category, and then the model will perform better (Table 3). A limitation of the ZSL is its potential to exacerbate inequality. If the categories are not carefully designed, ZSL models may unintentionally reinforce inequalities by favoring certain classes or groups. This can also be addressed by prompt engineering. Another limitation we found is that the ZSL model will face difficulties in performing multilabel classification when categories are closely related. This issue can be resolved by dividing the multilabel classification into binary classification for each class.

The computational demands and knowledge cutoff limitations in LLM are also constraints that can be addressed by leveraging reinforcement learning techniques. The implementation of adaptive learning, active learning strategies, and exploration-exploitation balancing is being explored to address computational challenges while minimizing the impact of the knowledge cutoff. In addition, user-driven reinforcement and collaborative efforts within the research community are being incorporated to refine the models. These reinforcement learning strategies are expected to enhance the adaptability, efficiency, and overall performance of LLM, ensuring its continued relevance and effectiveness.
Conclusions
We developed a framework based on BART ZSL for the categorization and trend analysis of articles and demonstrated a proof-of-concept scenario in the field of ophthalmology. We used the ZSL model for the categorization of articles in the field of ophthalmology, but it is extendable to other categories and fields without requiring any additional training. The model can generalize to new classes it has not observed during training, making it adaptable to different domains and applications. The results demonstrated that the model achieved promising outcomes across most categories. In addition to article classification, trend analysis highlighted the evolution of technology in ophthalmology. Accurate and quick classification of scientific papers enables efficient information retrieval, allowing researchers to access relevant studies more quickly and obtain insights into the trend of technology and future directions. Future research directions include exploring more specialized LLMs for further improvement. In addition to this, we also have plans to develop an automated literature review tool. This tool aims to streamline and enhance the literature review process by incorporating advanced algorithms to efficiently analyze and summarize relevant research findings.

Acknowledgments
This work was supported by National Institutes of Health (NIH) Grants R01EY033005 (SY) and R21EY031725 (SY) and Challenge Grant from Research to Prevent Blindness, New York (SY). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability
The data set (RenD) generated during this study is available in the Mendeley Data repository [23].

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary dataset, tables, and figures.
[DOCX File, 456 KB - formative_v8i52462_appl1.docx ]

References


Abbreviations

- AUC: area under the curve
- BART: Bidirectional and Auto-Regressive Transformers
- BERT: Bidirectional Encoder Representations from Transformers
- DR: diabetic retinopathy
- GPT: Generative Pretrained Transformer
- LLM: large language model
- mDeBERTa: Decoding-enhanced Bidirectional Encoder Representations from Transformers with Disentangled Attention
- MeSH: Medical Subject Headings
- NLP: natural language processing
- RenD: retinal diseases
- ZSL: zero-shot learning
Original Paper

Comparison of Self-Tracking Health Practices, eHealth Literacy, and Subjective Well-Being Between College Students With and Without Disabilities: Cross-Sectional Survey

Soyoung Choi¹, PhD
Department of Kinesiology and Community Health, University of Illinois Urbana-Champaign, Urbana, IL, United States

Corresponding Author:
Soyoung Choi, PhD
Department of Kinesiology and Community Health
University of Illinois Urbana-Champaign
272 Freer Hall
906 S. Goodwin Ave
Urbana, IL, 61801
United States
Phone: 1 2173332573
Email: soyoung@illinois.edu

Abstract

Background: College students with disabilities need to transition from pediatric-centered care to adult care. However, they may become overwhelmed by multiple responsibilities, such as academic activities, peer relationships, career preparation, job seeking, independent living, as well as managing their health and promoting healthy behaviors.

Objective: As the use of smartphones and wearable devices for collecting personal health data becomes popular, this study aimed to compare the characteristics of self-tracking health practices between college students with disabilities and their counterparts. In addition, this study examined the relationships between disability status, self-tracking health practices, eHealth literacy, and subjective well-being among college students.

Methods: The web-based questionnaire was designed using Qualtrics for the cross-sectional online survey. The survey data were collected from February 2023 to April 2023 and included responses from 702 participants.

Results: More than 80% (563/702, 80.2%) of the respondents participated voluntarily in self-tracking health practices. College students with disabilities (n=83) showed significantly lower levels of eHealth literacy and subjective well-being compared with college students without disabilities (n=619). The group with disabilities reported significantly lower satisfaction ($t_{411}$=−5.97, $P<.001$) and perceived efficacy ($t_{411}$=−4.85, $P<.001$) when using smartphone health apps and wearable devices. Finally, the study identified a significant correlation between subjective well-being in college students and disability status ($\beta=3.81$, $P<.001$), self-tracking health practices ($\beta=2.22$, $P=.03$), and eHealth literacy ($\beta=24.29$, $P<.001$).

Conclusions: Given the significant relationships among disability status, self-tracking health practices, eHealth literacy, and subjective well-being in college students, it is recommended to examine their ability to leverage digital technology for self-care. Offering learning opportunities to enhance eHealth literacy and self-tracking health strategies within campus environments could be a strategic approach to improve the quality of life and well-being of college students.

(JMIR Form Res 2024;8:e48783) doi:10.2196/48783

KEYWORDS
college students; personal health data; self-tracking; eHealth literacy; well-being; tracking; students; disability; cross-sectional survey; pediatric care; adult care; smartphone health app; application; literacy

Introduction

Youths or young adults with disabilities related to vision, hearing, mobility, and cognitive ability often experience functional limitations when performing self-care and daily activities [1]. Moreover, in addition to physical or sensory challenges, they often encounter limited job opportunities and employment prospects in the early stage of their career, in comparison with their peers without disabilities [2]. Recent
rehabilitation interventions and programs have focused on optimizing their perceived quality of life including a sense of personal control over an individual’s life course and satisfaction with personal development [3]. To establish basic self-care skills, diverse health education programs are offered in community and education settings [4]. According to Orem, self-care refers to the practice of activities that individuals initiate and perform on their own behalf in maintaining life, health, and well-being [5]. To ensure a successful transition from pediatric-centered care to adult care, it is necessary to develop and establish optimized self-care strategies during the initial phase of young adulthood [6].

With the advancement of health information and digital health technology, an increasing number of laypeople engage in self-tracking health and share their personal health data with family members, friends, and health care providers using mobile, Internet-connected devices such as smartphone apps and wearable sensors [7]. Self-tracking health represents the repeated measurement and recording of health-related information about oneself for self-knowledge of health status, health behavior change, and health monitoring [8]. Recent digital health technology allows individuals to gather personal health information in real-life situations and engage in remote health interventions by sharing health data with health care providers [9]. A previous study revealed that young adults (N=16) aged 18 years and older showed a preference for mobile health app features that include setting and tracking health behavior goals, receiving feedback and advice on health behavior change, alerts and reminders, and adequate privacy settings [2]. Virella Pérez et al [10] mentioned that the use of digital health technology has the potential to assist young adults in self-care during the transition to adult care; however, the available evidence on its usability, feasibility, generalizability, and effectiveness is still lacking.

Despite the low cost, convenience, and automated self-tracking health features offered by digital health or mobile technology, we still have an insufficient understanding of its impact on health behavior changes in young adults with disabilities. Since users’ technology adoption and continued use are influenced by their distinct needs, goals, expectations, usefulness, efficacy, perceived ease of use, satisfaction with features, and context [11], it is possible that consumer technology released in the market may not adequately address the functional limitations of this particular group. For instance, the visually oriented market may not adequately address the functional limitations of utilizing such technologies are narrated, emphasizing the promising role they can play in managing chronic conditions [12].

The World Health Organization (WHO) [12] highlighted that health literacy is an important factor for preventing chronic illnesses because it significantly affects one’s capability against chronic illnesses, including addressing risk factors [13]. Health literacy refers to the degree to which individuals can find, understand, and use information and services to inform health-related decisions and actions for themselves and others [14]. Poor health literacy affects negative health outcomes as well as reduced satisfaction with health care services [15]. Recently, in the digital era, the concept of eHealth literacy, also known as digital health literacy, has emerged to assess individuals’ competence in using the Internet to navigate health care resources and access health information [16]. Based on a systematic review conducted on eHealth literacy among college students, a considerable number of adults were extensively engaged with the Internet and expressed significant confidence in their ability to search for electronic health information [17]. Nevertheless, the review also discovered conflicting outcomes, exposing cases of unsophisticated health information-seeking and a lack of proficient critical appraisals while evaluating the searched health information. Recognizing the discordance between perceived and evaluated eHealth literacy among college students, this study examined and compared the eHealth literacy levels of college students, both with and without disabilities.

During the global COVID-19 outbreak, there was a significant increase in the prevalence of mental health issues among college students [18]. Consequently, counseling centers on campuses took proactive measures to provide mental health care services, with the goal of improving the well-being and overall quality of life of college students [19]. Lattie et al [20] conducted a systematic review to investigate the effectiveness of digital mental health interventions for college students experiencing low levels of psychological well-being. The review not only examined the impact of digital health interventions on psychological outcome variables but also discussed the potential advantages of enhancing health care accessibility and cost-effectiveness [20]. Although some scholars have highlighted the negative impact of mobile phone addiction on this younger generation [21,22], strategically planned use of mobile health apps and wearable sensors could enhance the subjective well-being of college students [23].

Given the rise of innovative digital health technology and the increasing interest in personal health data, it is still unclear whether college students with disabilities experience disparities in digital health care and utilization, particularly in the context of mobile health technology. To fill this gap in prior research, the findings of this study were expected to contribute to the current understanding of self-tracking health practices among college students with disabilities. The conclusion of this study involves a discussion on future directions with a focus on leveraging mobile health technology and interventions to enhance subjective well-being and overall health among college students with disabilities. The potential benefits and implications of utilizing such technologies are narrated, emphasizing the promising role they can play in managing chronic conditions and preventing secondary health problems in younger populations with disabilities.

Methods

Survey and Participants

A cross-sectional online survey was conducted to examine the self-tracking health practices of college students, both with and without disabilities, using the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [24]. The web-based...
questionnaire was designed by the author using Qualtrics [25] and was reviewed for readability and clarity by 5 college students—2 college students without disabilities and 3 college students with disabilities—before being revised. The average time to complete the pilot survey was 15.4 minutes. Prospective participants representing college students with and without disabilities were invited to take part via email and social media platforms, including Twitter and Facebook. Upon clicking the URL in the email or social media post, participants were screened for eligibility and asked to provide informed consent. To be eligible for the survey, participants had to (1) be a US resident, (2) be aged between 18 years and 35 years, (3) be currently enrolled in college or university, (5) not have a cognitive impairment or psychiatric disorder, and (6) speak and write English. To assess disability status, the following additional screening question was asked: “Do you have a disability? If yes, please specify the type of disability you have (eg, visual impairment, hearing loss, mobility, learning).” No identifiable data, except for the email addresses of the participants for the compensation procedure, were collected. The online survey was open to respondents from February 20, 2023, to April 15, 2023. Participants who completed the survey were entered into a raffle for compensation in the form of Amazon eCodes. In 2 groups, 30 people each received US $20 worth of Amazon eCodes.

Web-Based Data Collection
In this study, the self-tracking health practices refer to utilizing digital health technologies such as smartphone health apps, wearable sensors (eg, fitness trackers, smartwatches), and smart medical devices to regularly check one or more health-related data point as well as recording health data with manual tools (eg, pencil and notebook). The following questions were included in the survey to understand the self-tracking health practices and preferences of college students with and without disabilities: (1) “Do you currently track your health (eg, monitoring body weight, calculating calories intake, counting steps)?” (2) “What aspects of your health do you routinely track?” (3) “What types of tracking tools do you use to record your health data?” (4) “How frequently do you check your health data?” (5) “What are the reasons that you do not track your health?” (6) “What kinds of health data do you want to track in the future?” (7) “What smartphone apps have you used to manage your health?” (8) “What features do you prefer the health-related smartphone apps to contain?” (9) “What kinds of wearable devices have you used to manage your health?” (10) “Please describe any barriers or challenges you faced when tracking your personal health data.” In this study, 2 self-reported questionnaires were used. First, the level of eHealth literacy was measured using the eHealth Literacy Scale (eHEALS) [26]. The eHEALS was developed to assess eHealth literacy for a wide range of populations [26]. It has 6 core skills such as traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy [27]. This self-report instrument includes 8 items and uses a 5-point Likert scale ranging from “Strongly disagree” to “Strongly agree.” Higher scores indicate higher levels of eHealth literacy. In the original study, which targeted a youth population aged between 13 years and 21 years, the internal reliability (Cronbach alpha) was .88, and the test-retest reliability (r) ranged from 0.40 to 0.68 [26]. In this study, the reliability (Cronbach alpha) was .88. Second, to quantitatively assess the perceived well-being of college students, the Flourishing Scale (FS) [28] was used. The FS is an 8-item, self-report scale and measures respondents’ perceived success in relationships, self-esteem, purpose, and optimism. It uses a 7-point Likert scale ranging from “Strongly disagree” to “Strongly agree.” Higher scores indicate higher functioning in social-psychological prosperity. The reported Cronbach alpha is .87, and the test-retest reliability is (r) 0.71. The reliability (Cronbach alpha) of this study was .91.

Statistical Analysis
The minimum sample size of this study was 92 based on the number of predictors (n=5), targeted adjusted $R^2$ of 0.30 with type 1 error of <.05, medium effect size of 0.3, and power of 0.8. Descriptive statistics and multiple regression analysis were used to investigate the characteristics and relationships between variables using R software (RStudio). To identify any detectable differences between continuous variables, independent t tests, ANOVAs, and Wilcoxon-Mann-Whitney tests were conducted. In the cases of counts or frequencies of categorical variables, chi-square tests were performed. Multiple linear regression analysis was conducted to examine the relationships between the main variables (ie, disability status, self-tracking health practice, eHEALS, and FS). For this regression model, the categorical variables (ie, doing self-tracking health, living with disability) were coded into 2 categories (ie, 0=No, 1=Yes).

Ethical Considerations
Ethical approval for the survey was obtained from the University of Illinois Urbana-Champaign (Protocol ID: 23316), and the study was monitored by the Institutional Review Board of the University of Illinois Urbana-Champaign. Participants gave their informed consent prior to participating in the survey. The study data were anonymized and stored securely at the university. Through a raffle, 30 participants were randomly selected and compensated with a US $20 Amazon e-gift card.

Results
There were no significant differences in gender, age, and race of the participants (n=702) of the study. Most college students (563/702, 80.2%) participated voluntarily in self-tracking their health using smartphones regardless of disability status (Table 1). More than one-half of the college students with disabilities (51/83, 61%) preferred using a laptop or desktop computer after smartphones, whereas their counterparts leaned toward tablet PCs such as iPads (458/619, 65.2%). In both groups, Health Science majors represented the highest proportion. Among the 83 respondents with disabilities, the following conditions were reported: 28 students (25%) reported having multiple disabilities and a learning disability, followed by autism spectrum disorder (19/83, 23%), hearing loss (18/83, 22%), and physical disability (17/83, 21%). Significant statistical differences were found between the 2 groups when comparing the mean scores of the eHEALS ($t_{df}=-2.22, P=.03$) and FS ($t_{df}=-4.54, P<.001$). The
group with disabilities showed lower mean scores for both the eHEALS and FS, when compared with the other group.

The main characteristics of self-tracking health practices among the 563 respondents—72 college students with disabilities (86.8%) and 491 college students without disabilities (79.3%)—who reported engaging in self-tracking health are as follows. In the self-tracking health-related questions with allowed duplicate responses, the group with disabilities (34/72, 47%) and the group without disabilities (351/491, 71.5%) responded that they primarily engaged in self-tracking to monitor exercise. Next, the college students with disabilities (29/72, 40%) reported frequently observing their heart rate, while the college students without disabilities (286/491, 53.8%) reported observing their food intake and calorie counting. The group with disabilities engaged in self-tracking health for their physiological health status or medication adherence, while the other group was interested in adopting healthy lifestyle habits (Table 2).

As illustrated in Table 3, the college students were predominantly using smartphones, online software, and wearable devices including fitness trackers and smartwatch as tools for self-tracking health practices. The use of paper and pen was the lowest, accounting for 6% (4/72) and 16.3% (80/491), respectively.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>College students with disabilities (n=83)</th>
<th>College students without disabilities (n=619)</th>
<th>Total sample (N=702), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-tracking health, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.15</td>
</tr>
<tr>
<td>Yes</td>
<td>72 (86.8)</td>
<td>491 (79.3)</td>
<td>563 (80.2)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>11 (13.2)</td>
<td>128 (20.7)</td>
<td>139 (19.8)</td>
<td></td>
</tr>
<tr>
<td>Digital device use, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Laptop/desktop computer</td>
<td>51 (61.4)</td>
<td>454 (73.3)</td>
<td>505 (71.9)</td>
<td></td>
</tr>
<tr>
<td>Tablet PC</td>
<td>42 (50.6)</td>
<td>416 (67.2)</td>
<td>458 (65.2)</td>
<td></td>
</tr>
<tr>
<td>Smartphone</td>
<td>55 (66.3)</td>
<td>551 (89)</td>
<td>606 (86.3)</td>
<td></td>
</tr>
<tr>
<td>Wearable device/fitness tracker/smartwatch</td>
<td>28 (33.7)</td>
<td>231 (37.3)</td>
<td>259 (36.9)</td>
<td></td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.38</td>
</tr>
<tr>
<td>Male</td>
<td>33 (39.8)</td>
<td>280 (45.2)</td>
<td>313 (44.6)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>47 (56.6)</td>
<td>313 (50.6)</td>
<td>360 (51.3)</td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>24.48 (4.38)</td>
<td>24.31 (3.53)</td>
<td>24.33 (3.64)</td>
<td>.69</td>
</tr>
<tr>
<td>Age (years), range</td>
<td>18-33</td>
<td>18-33</td>
<td>18-33</td>
<td>_a</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.053</td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>36 (43.4)</td>
<td>378 (61.1)</td>
<td>414 (59)</td>
<td></td>
</tr>
<tr>
<td>Black/African American/Hispanic</td>
<td>15 (18.1)</td>
<td>61 (9.9)</td>
<td>76 (10.8)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>32 (38.6)</td>
<td>180 (29.1)</td>
<td>212 (30.2)</td>
<td></td>
</tr>
<tr>
<td>Major, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Computer Science</td>
<td>6 (7.2)</td>
<td>106 (17.1)</td>
<td>112 (16)</td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>3 (3.6)</td>
<td>71 (11.5)</td>
<td>74 (10.5)</td>
<td></td>
</tr>
<tr>
<td>Communications</td>
<td>5 (6)</td>
<td>43 (6.9)</td>
<td>48 (6.8)</td>
<td></td>
</tr>
<tr>
<td>Government/Political Science</td>
<td>5 (6)</td>
<td>39 (6.3)</td>
<td>44 (6.3)</td>
<td></td>
</tr>
<tr>
<td>Health Science</td>
<td>18 (21.7)</td>
<td>105 (17)</td>
<td>123 (17.5)</td>
<td></td>
</tr>
<tr>
<td>Economics</td>
<td>8 (9.6)</td>
<td>93 (15)</td>
<td>101 (14.4)</td>
<td></td>
</tr>
<tr>
<td>English Language and Literature</td>
<td>6 (7.2)</td>
<td>14 (2.3)</td>
<td>20 (2.8)</td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td>12 (14.5)</td>
<td>46 (7.4)</td>
<td>58 (8.3)</td>
<td></td>
</tr>
<tr>
<td>Biology</td>
<td>8 (9.6)</td>
<td>21 (3.4)</td>
<td>29 (4.1)</td>
<td></td>
</tr>
<tr>
<td>Sociology</td>
<td>3 (3.6)</td>
<td>26 (4.2)</td>
<td>29 (4.1)</td>
<td></td>
</tr>
<tr>
<td>Social Work</td>
<td>2 (2.4)</td>
<td>33 (5.3)</td>
<td>35 (5)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>6 (7.2)</td>
<td>22 (3.6)</td>
<td>28 (4)</td>
<td></td>
</tr>
<tr>
<td>Disability, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Visual impairment</td>
<td>6 (7.2)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Hearing loss</td>
<td>18 (21.7)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Deaf</td>
<td>10 (12)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Learning disability</td>
<td>21 (25.3)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>19 (22.9)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Physical disability</td>
<td>17 (20.5)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>18 (21.7)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>eHEALS&lt;sub&gt;b&lt;/sub&gt; (possible scores: 8-40), mean (SD)</td>
<td>27.28 (7.32)</td>
<td>28.65 (6.02)</td>
<td>28.65 (6.02)</td>
<td>.03</td>
</tr>
<tr>
<td>eHEALS (possible scores: 8-40), range</td>
<td>11-40</td>
<td>2-10</td>
<td>11-40</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Comparative analysis of self-tracking purposes between college students with and without disabilities.

<table>
<thead>
<tr>
<th>Self-tracking purpose</th>
<th>College students with disabilities (n=83), n (%)</th>
<th>College students without disabilities (n=619), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercise</td>
<td>34 (41.0)</td>
<td>351 (56.7)</td>
</tr>
<tr>
<td>Diet/calorie tracking</td>
<td>17 (20.5)</td>
<td>286 (46.2)</td>
</tr>
<tr>
<td>Sleep</td>
<td>21 (25.3)</td>
<td>266 (43)</td>
</tr>
<tr>
<td>Water intake</td>
<td>23 (27.7)</td>
<td>193 (31.2)</td>
</tr>
<tr>
<td>Heart rate</td>
<td>29 (34.9)</td>
<td>193 (31.2)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>19 (22.9)</td>
<td>151 (24.4)</td>
</tr>
<tr>
<td>Sedentary time</td>
<td>21 (25.3)</td>
<td>149 (24.1)</td>
</tr>
<tr>
<td>Body weight</td>
<td>9 (10.8)</td>
<td>136 (22)</td>
</tr>
<tr>
<td>Blood glucose</td>
<td>9 (10.8)</td>
<td>114 (18.4)</td>
</tr>
<tr>
<td>Personal health</td>
<td>20 (24.1)</td>
<td>113 (18.2)</td>
</tr>
<tr>
<td>Body temperature</td>
<td>17 (20.5)</td>
<td>93 (15)</td>
</tr>
<tr>
<td>Stress</td>
<td>11 (13.3)</td>
<td>76 (12.3)</td>
</tr>
<tr>
<td>Menstrual cycle</td>
<td>15 (18.1)</td>
<td>74 (12)</td>
</tr>
<tr>
<td>Mood</td>
<td>4 (4.8)</td>
<td>71 (11.5)</td>
</tr>
<tr>
<td>Medication</td>
<td>17 (20.5)</td>
<td>71 (11.5)</td>
</tr>
</tbody>
</table>

Table 3. Comparative analysis of preferred self-tracking tools between college students with and without disabilities.

<table>
<thead>
<tr>
<th>Self-tracking tools</th>
<th>College students with disabilities (n=83), n (%)</th>
<th>College students without disabilities (n=619), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone app</td>
<td>36 (43.4)</td>
<td>324 (52.3)</td>
</tr>
<tr>
<td>Online program</td>
<td>30 (36.1)</td>
<td>244 (39.4)</td>
</tr>
<tr>
<td>Wearable device</td>
<td>25 (30.1)</td>
<td>210 (33.9)</td>
</tr>
<tr>
<td>Computer software</td>
<td>16 (19.3)</td>
<td>182 (29.4)</td>
</tr>
<tr>
<td>Digital scale</td>
<td>25 (30.1)</td>
<td>155 (25)</td>
</tr>
<tr>
<td>Medical device</td>
<td>25 (30.1)</td>
<td>153 (24.7)</td>
</tr>
<tr>
<td>Pen and paper</td>
<td>4 (0.1)</td>
<td>80 (12.9)</td>
</tr>
</tbody>
</table>

The majority of the college students who participated in the survey reported checking their health-related data on a daily or weekly basis. Of the 72 students with disabilities, 21 (25%) engaged in daily health tracking, while 193 of the 218 students without disabilities (31.2%) conducted daily self-tracking of their health.

To examine whether there were significant differences in satisfaction and perceived efficacy with smartphone health apps and wearable devices based on living with a disability, independent t tests were conducted. Satisfaction and perceived efficacy were evaluated on a 10-point Likert scale. The results indicated significant differences based on the presence of a disability in all categories (Figure 1). The satisfaction score for smartphone health apps ($t_{df}=−6.36, P<.001$) was lower for the group with disabilities (mean 6.47, SD 2.58) compared with the nondisabled group (mean 7.92, SD 1.40), and perceived efficacy ($t_{df}=−3.61, P<.001$) was also lower for the group with disabilities (mean 7.24, SD 2.10) compared with the group without disabilities (mean 8.03, SD 7.92). The satisfaction score for wearable devices ($t_{df}=−5.97, P<.001$) was lower for the group with disabilities (mean 6.72, SD 2.48) than the nondisability group (mean 8.14, SD 1.33), and perceived
efficacy ($t_{308} = -4.85, P < .001$) was also lower for the group with disabilities (mean 7.00, SD 2.43) than the group without disabilities (mean 8.19, SD 1.42).

**Figure 1.** Comparative analysis of (A) satisfaction with health apps, (B) efficacy with health apps, (C) satisfaction with wearable devices, and (D) efficacy with wearable devices between college students with (group 1) and without disabilities (group 2).

The survey results regarding the reasons why college students who do not engage in self-tracking health provided the following responses. Among college students with disabilities, the top 5 reasons for not engaging in self-tracking health were “I forgot to track” (30/83, 36%), “I don’t think it is useful” (20/83, 24%), “I don’t want to know or see the results” (18/83, 22%), “It is too difficult to track” (17/83, 21%), and “It is not important to me” (13/83, 16%). On the other hand, among college students without disabilities, the primary reasons were “I forgot to track” (221/619, 35.7%), “I don’t have enough time” (190/619, 30.7%), “It is too difficult to track” (130/619, 21.0%), “It is not important to me” (112/619, 18.1%), and “I don’t think it is useful” (90/619, 14.5%).

Finally, to verify the impact of disability status, self-tracking health, and eHealth literacy on college students’ subjective well-being, a multiple linear regression analysis was conducted (Table 4). The multiple linear regression model showed statistically significant results ($F_{1,667} = 168.038, P < .001$), and the explanatory power of the model was 55.7% ($R^2 = 0.557$, adjusted $R^2 = 0.554$). Meanwhile, the Durbin-Watson statistic showed a value of 1.93, which is close to 2, indicating that there were no issues with the independence assumption of the residuals. The variance inflation factor was also below 10, indicating that there were no problems with multicollinearity. The significance test of the regression coefficients showed that living with a disability ($\beta = 3.81, P < .001$), self-tracking health ($\beta = 2.22, P = .03$), and eHealth literacy levels ($\beta = 24.29, P < .001$) all had significant positive relationships with overall subjective well-being.
learn to use them, and may not recognize any hidden usability issues.

Some college students displayed a reluctance to know or address their health status, reflecting a similar trend observed in older adults who avoid self-tracking health due to a heightened awareness of their illnesses and medical symptoms [30]. This means that, when individuals become aware of their existing health problems, collecting health data solely for the purpose of maintaining chronic conditions rather than promoting overall health can create self-avoidance or excessive burden. Sometimes, excessive self-tracking routines can result in adverse effects by making individuals too sensitive to minor fluctuations in health indicators or encouraging them to overinterpret the changed health data. It can also frequently evoke negative feelings for individuals who are diagnosed with chronic health problems. Therefore, it is advisable to avoid unnecessary data visiting for specific health issues. Rather than relying on generalized feedback, it would be more appropriate to offer personalized data-driven feedback that informs self-trackers when a health problem is expected or when immediate health behavior changes are necessary.

It is important to consider incorporating educational features in health-related apps that help college students learn how to monitor and interpret personal health data from a health literate approach [31]. This can include the use of charts, graphics, and illustrations to enhance comprehension [32]. In the cases with visual impairments, a detailed audio description should be added to the visualized health information [33]. Furthermore, as forgetting to engage in self-tracking health practices was identified as one of the primary reasons why college students find it challenging, incorporating time cueing features in the app that align with their daily routines would be beneficial [34]. To assist college students, reminders can be customized to their campus activities or sent at suitable times to notify them when health data collection is incomplete. The most convenient solution for individuals who forget to capture personal health data would be to automatically collect health data through digital devices that they can wear or carry. However, there is an argument that fully automated tracking requires additional feedback to encourage engagement and manage the high volume

### Table 4. Relationships between disability status, self-tracking health, eHealth literacy, and subjective well-being in college students, with the Flourishing Scale score as the dependent variable.\(^a\)

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>B</th>
<th>SE</th>
<th>t (df)</th>
<th>P value</th>
<th>VIF(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.36</td>
<td>2.52</td>
<td>1.33 (667)</td>
<td>.18</td>
<td>__(^c)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.05</td>
<td>0.07</td>
<td>-0.80 (667)</td>
<td>.42</td>
<td>1.04</td>
</tr>
<tr>
<td>Female gender</td>
<td>-0.29</td>
<td>0.49</td>
<td>-0.60 (667)</td>
<td>.55</td>
<td>1.02</td>
</tr>
<tr>
<td>Living with a disability</td>
<td>-2.86</td>
<td>0.75</td>
<td>-3.81 (667)</td>
<td>&lt;.001</td>
<td>1.03</td>
</tr>
<tr>
<td>eHEALS(^d)</td>
<td>1.08</td>
<td>0.04</td>
<td>24.29 (667)</td>
<td>&lt;.001</td>
<td>1.26</td>
</tr>
<tr>
<td>Self-tracking health</td>
<td>1.48</td>
<td>0.67</td>
<td>2.22 (667)</td>
<td>.03</td>
<td>1.24</td>
</tr>
</tbody>
</table>

\(^a\)F\(_{1,5}\)=168.03 (\(P<.001\)), \(R^2=0.557\), adjusted \(R^2=0.554\), D-W=1.93.

\(^b\)VIF: variance inflation factor.

\(^c\)Not applicable.

\(^d\)eHEALS: eHealth Literacy Scale.
College students with disabilities reported similar levels of daily use and interest in digital health technology as college students without disabilities. However, their eHealth literacy and subjective well-being levels were significantly lower than for their counterparts. This could be attributed to the direct impact of disabilities, but it could also be interpreted as the result of the benefits they are unable to enjoy due to their disabilities. For example, the college students with disabilities in this study reported significantly lower satisfaction and perceived efficacy with using smartphone health apps and wearable devices. These unsatisfying technology experiences can affect their motivation to engage in self-tracking health practices. This poor user experience for college students with disabilities represents a lack of inclusivity in the current distributed digital health technology, similar to how older adults may feel resistant toward new technology [36]. The fact that users with physical, sensory, and learning limitations cannot experience the same level of positive effectiveness and behavioral changes as common users can be understood as a paradoxical effect of technology that contributes to health inequality in the future [37].

Currently, there are very limited efforts to enhance the usability of mobile or digital health technologies for users with disabilities by addressing their unmet needs, challenges, and barriers. One of the important missing pieces in this study is the exploration of accessibility issues faced by college students with disabilities and their adaptive strategies to overcome these challenges. Recently, Lee and colleagues [38] reported on multiple unique accessibility challenges faced by visually impaired individuals in digital self-tracking systems. They also discussed design opportunities to reduce the gap between visually impaired and sighted users. This highlights the need for greater inclusivity in the design of digital health technologies to ensure that all individuals, regardless of ability, have equal access to the benefits of self-tracking health. Given the heterogeneous nature of disability impacts on technology adoption and users’ preferences and values in technology, diverse disability groups’ user experiences should be deeply understood and considered for future product designs. To prevent the digital divide and its impact on health disparities, a mindset of accessibility and inclusivity is necessary for all software developers and health informatics researchers.

Developing healthy lifestyle habits and the ability to sustain them in the long term is particularly important for college students with disabilities. They are more likely to face barriers and challenges in self-care compared with individuals without disabilities [39,40]. College students with disabilities should receive health education that is tailored to their specific needs and provides them with a proper understanding of health and how to model healthy behaviors. This education should also include information about health information accessibility and the use of necessary assistive technologies to ensure that students with disabilities can freely access and utilize health-related information. For instance, for visually impaired female students, it is important to provide health education on cervical cancer vaccination and breast self-examination in an accessible manner. This could include providing detailed audible descriptions or tangible education materials that they can feel and touch. Ensuring their participation in diverse virtual and in-person health education opportunities should be established.

Implications
This study identified positive correlations between disability status, self-tracking health practices, eHealth literacy, and subjective well-being among college students with and without disabilities. This highlights the potential benefits of using digital health technology on the well-being and quality of life of all college students. Based on these findings, higher education institutions can consider offering learning opportunities specifically for college students with disabilities, who face additional challenges in managing their health conditions. These opportunities should aim to teach them how to effectively and systematically use digital health technology to monitor their daily health conditions and manage their health behaviors. Experiencing rewarding outcomes from self-tracking their health using digital health technology could lead to improved quality of life and well-being in adulthood.

Limitations
This cross-sectional, online survey study has certain limitations that warrant careful consideration when interpreting the findings. First, the sampling strategy used in the study was not a sophisticated design. Due to the small sample size of the target population and practical barriers in implementing random or quota sampling, the respondents may not adequately represent the broader population of college students with disabilities. Second, a substantial number of college students majoring in Health Science participated in the survey due to the relevance of the topic to their field of interest, potentially introducing bias to the results. Last, the study did not account for the detailed characteristics of disabilities, such as their type and severity, when interpreting the results. Conducting a subgroup analysis to address this issue would have been ideal; however, the small sample size made it infeasible. Therefore, further research is necessary to obtain study results that reflect the specific characteristics associated with each disability.

Conclusions
This study indicates that college students with disabilities are engaged in digital self-tracking health activities, much like their peers without disabilities. However, they experience poor satisfaction and efficacy from current digital health technology. Software developers and health informatics researchers should not assume that young adults, including those with disabilities, will readily adopt and effectively utilize technology for health management. Rather, it is crucial to closely examine their abilities and accessibility needs to ensure that digital self-tracking health tools are inclusive and promote long-term self-care. In addition, given their lack of eHealth literacy skills, it is important for health educators in campus settings to provide personalized health consultations for both college students with and without disabilities to enhance the effective utilization of self-tracking health. For college students with disabilities, accessible self-tracking tools can be helpful for establishing healthy behavioral patterns in young adulthood and preventing...
chronic illnesses, thereby optimizing their life satisfaction and self-control.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References
Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
eHEALS: eHealth Literacy Scale
Examining and Comparing the Validity and Reproducibility of Scales to Determine the Variety of Vegetables Consumed: Validation Study

Kaya Ominami¹, BSc, RD; Osamu Kushida¹, RD, PhD
Department of Nutrition and Life Sciences, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan

Corresponding Author: Osamu Kushida, RD, PhD
Department of Nutrition and Life Sciences
School of Food and Nutritional Sciences
University of Shizuoka
52-1 Yada
Suruga-ku
Shizuoka, 422-8526
Japan
Phone: 81 542645832
Email: kushida@u-shizuoka-ken.ac.jp

Abstract

Background: Previous studies have reported that vegetable variety reduces the risk for noncommunicable diseases independent of the amount consumed.

Objective: This study aimed to examine and compare the validity and reproducibility of several scales to determine vegetable variety.

Methods: In total, 23 nutrition students in Japan reported their vegetable intake over the past month using a self-administered questionnaire between July and August 2021. Specifically, four scales were used: (1) a single question regarding the number of vegetables consumed (scale A); (2) a scale containing 9 vegetable subgroups included in the brief-type self-administered diet history questionnaire (scale B); (3) a scale containing 19 vegetable items included in a self-administered diet history questionnaire (scale C); and (4) a scale containing 20 vegetable items from the Ranking of Vegetable Consumers in Japan, which was analyzed based on a report on the National Health and Nutrition Survey in Japan (scale D). Scale validity was assessed by correlation with the number of vegetables consumed, which was collected from dietary records for 7 consecutive days. Reproducibility was assessed by test-retest reliability.

Results: Regarding the validity of the 4 scales, significant correlations were found between scales C (r=0.51) and D (r=0.44) with vegetable variety based on dietary records, but scales A (r=0.28) and B (r=0.22) were not significantly correlated. Reproducibility showed a significant correlation in scale B (r=0.45) and strong correlations in scales C (r=0.73) and D (r=0.75).

Conclusions: The scales for vegetable items have acceptable validity and reproducibility compared to the scales that used a single question or vegetable subgroup and, therefore, may determine the variety of vegetables consumed.

(JMIR Form Res 2024;8:e55795) doi:10.2196/55795

KEYWORDS
vegetable; variety; scale; validity; reproducibility; dietary records; nutrition

Introduction

The World Health Organization [1] recommends that adults consume at least 400 g of fruit and vegetables per day to reduce the risk of noncommunicable diseases. In Japan, the Health Japan 21 (the second term) set a consumption target for an average of 350 g of vegetables daily for adults to prevent lifestyle-related diseases [2]. However, according to the recent report of the National Health and Nutrition Survey (NHNS) in Japan, average vegetable intake is below this target at 280.5 g/d with no significant change occurring over the past 10 years [3]. This is also the case in developed countries such as the United States and Australia, and since most populations do not meet...
the recommended daily intake of vegetables [4,5], health policies to increase vegetable intake are being promoted worldwide.

Regarding vegetable variety, the 2020-2025 Dietary Guidelines for Americans recommend weekly intakes for vegetable subgroups [4]. The 2013 Australian Dietary Guidelines also encourage including a variety of vegetables [6]. Previous studies identified positive associations between high vegetable variety and plasma carotenoid concentrations [7] as well as overall diet quality [8]. Furthermore, vegetable variety is reported to reduce the risk of developing noncommunicable diseases such as coronary heart disease [9], lung cancer [10], and type 2 diabetes [11] independent of the amount consumed.

To measure vegetable variety, existing food frequency questionnaires (FFQs) have been used in many previous studies. However, none have examined whether the scales extracted from the FFQs reflect vegetable variety. The scales vary from study to study and are characterized by 3 main constructs: vegetable items, vegetable subgroups, or a single question [12]. Previous studies reported strong inverse associations between an increased variety of vegetable items and overall cancer risk [13] and between an increased variety of vegetable subgroups and lung cancer risk [10]. Furthermore, vegetable variety assessed by a single question has inverse associations with the risk of injurious falls and fractures [14], as well as with mortality from atherosclerotic vascular disease and common carotid artery intima-media thickness [15]. However, studies examining the vegetable variety scale itself are limited to the evaluation of a brief index of Fruit And Vegetable Variety (FAVVA) using an FFQ in Australia [7].

Guidelines set by the United Nations Food and Agriculture Organization recommend that questionnaires assessing dietary diversity are adapted according to the culture, population, or location [16]. It is expected that developing a scale to determine vegetable variety will be important for promoting nutritional improvement in Japan. Therefore, this study aimed to examine and compare the validity and reproducibility of several scales to determine vegetable variety.

**Methods**

**Study Design and Participants**

This was a cross-sectional study using self-administered questionnaires. The participants were students majoring in nutrition at a university in Shizuoka, Japan, and were, therefore, familiar with dietary records. The students were recruited through a bulletin board on campus. Out of a total population of approximately 100 students, the questionnaire was distributed directly to 23 participants who agreed to be surveyed, and 23 copies were collected by mail or in person (recovery rate=100%). As an incentive, participants in this survey were provided with fruit (eg, missing vegetable condiments, seasonings, and cooking oil). That missing information was then obtained from participants in person or via email.

**Measurements**

**Scales to Determine Vegetable Variety**

In total, 4 types of scales were used to capture responses regarding the variety of vegetables consumed in the past month. The first scale consisted of a single question that was answered with a single or double-digit whole number, in response to the following question: “How many different vegetables do you usually eat per day?” which was adopted from an FFQ developed by the Cancer Council of Victoria [14]. This scale is defined as the variety of vegetable consumption assessed by a single question (scale A).

The second scale consisted of 9 vegetable subgroups presented in a brief-type self-administered diet history questionnaire (BDHQ) [18]. The BDHQ is a nonquantitative FFQ that has been validated for assessing vegetable intake throughout the year in urban, rural inland, and rural coastal areas in Japan (Osaka, Nagano, and Tottori) [18]. Scoring was based on the study by Chou et al [19], in which 1 point was added for consumption at least once a week, 0 for less than once a week, and the total score representing the vegetable variety of the vegetable subgroup (scale B).

The third and fourth scales consisted of 19 vegetable items presented in a self-administered diet history questionnaire (DHQ) [18], and 20 vegetable items presented in the Ranking of Vegetable Consumers in Japan report, which was analyzed based on the 2012 NHNS [20]. The DHQ is a semiquantitative FFQ that has been validated for vegetable intake throughout the year in several areas in Japan, similar to the BDHQ [18]. The Ranking of Vegetable Consumers in Japan report was determined to have good internal consistency among the 20 vegetables, according to the authors’ preliminary analyses (Cronbach α=0.78). Previous studies using vegetable items scales [10,21] add 1 point for reported consumption of at least once per 2 weeks. The same scoring was used in this study, with 0 points added for consumption that was reported as once a month or less. The total score was defined as the vegetable variety for each vegetable item (scales C and D, respectively).

According to a previous study that examined the number of days of dietary records required to estimate habitual vegetable variety, we estimated that 7 consecutive days of dietary records would capture approximately 70% of the maximum theoretical number of different vegetables consumed [17]. It is reasonable to assume that an understanding of dietary habits could be achieved with 7 consecutive days of dietary records; therefore, the same number of consecutive days of dietary record collection was chosen for this study. Finally, participants were asked to provide their responses on the same scale after 1 month following their response to the first scale. After each questionnaire was collected, 2 students majoring in nutrition and 1 researcher certified as a registered dietitian checked for missing questionnaire responses and incomplete dietary records (eg, missing vegetable condiments, seasonings, and cooking oil). That missing information was then obtained from participants in person or via email.

The exact wording and scoring methods for each scale are shown in Multimedia Appendix 1.
Dietary Records for 7 Consecutive Days

Participants were provided instructions to complete their dietary records recording all foods and beverages consumed, and weighing whenever possible. Participants recorded the meal start time, a description of the meal situation (home meal, ready-made meal, eating out, or other), the name of the dish, the name of the food, and the amount consumed. The participants were asked to record the name of the product, seller, and store when consuming commercial products and the name of the restaurant when eating out. If participants recorded only the name of the dish, the food and amount consumed were estimated by referring to the top results for recipes identified through an internet search. When the name of a product or restaurant was recorded, an internet search was conducted to identify the food and amount consumed by the participant. The number of vegetables consumed was extracted from the Standard Tables of Food Composition in Japan 2020 (Eighth Revised Edition) [22]. In doing so, vegetables not listed were replaced with similar vegetables in the food composition tables. The Excel add-in software EiyoPlus (Kenpakusha, Co, Ltd) was used to calculate vegetable variety from the dietary records.

In this study, vegetable variety was defined as the total number of different vegetables consumed as recorded by the participant on their dietary records and the number of different vegetables consumed estimated by the methods described above. In a previous Australian study that used 24-hour recall [23], 1 unit of variety of vegetable consumption consisted of ≥50% (≥37.5 g) of a serving of a vegetable in the country. Furthermore, in a Japanese study that examined the frequency of meals including staple, main, and side dishes and nutrient intake [24], side dishes consisting mainly of vegetables were considered to have been consumed if participants consumed at least half a serving (≥35 g) in the Japanese food guide. Therefore, in this study, vegetables were counted as 1 item when the amount consumed was ≥35 g. The method for counting the number of vegetables consumed followed procedures described elsewhere [17]. Briefly, vegetables within the same category but with different food names were counted as different items, while the same vegetable prepared according to different cooking and processing methods was counted as the same item. Vegetables in the same category with different food names but prepared according to different processing methods were counted as the same item (eg, daikon, kiriboshi-dai kon [cut and dried daikon root], and “pickles” [daikon pickled in salty rice bran paste] were counted as the same item). Furthermore, grated ginger and grated garlic were classified as spices and not vegetables.

Demographic Characteristics

Participants were asked about the following characteristics: gender, age, height, weight, location, number of people living together, smoking status, drinking habits, exercise habits, and whether vegetable intake was restricted for medical reasons.

Statistical Methods

Of the 4 scales included in this study, item-total correlation analysis was performed for each of the 3 scales that are scoring systems. Dietary records for 7 consecutive days were used as a reference standard in the validity analysis. In designing a validation study for the FFQ, administering the questionnaire before the participants create dietary records results in the questionnaire responses being related to the diet before the period during which dietary records were created. On the other hand, the intensive effort involved in creating dietary records may lead to artificially improved accuracy if the questionnaire is administered after dietary record creation [25]. To avoid optimistic estimates of the correlation in this study, the first response given by participants was used to examine the validity of each scale. Reproducibility was assessed by test-retest reliability. Since each scale to determine vegetable variety was intended to be ranked, correlation coefficients were calculated. Furthermore, because the recorded variable distributions were determined not to be normally distributed by the histograms, the Spearman correlation coefficient was used to examine the validity and reproducibility. SPSS Statistics (version 27; IBM Corp) was used for analysis. The level of significance was set at P<.05 (2-sided test).

Ethical Considerations

The purpose and methods of this study were explained to participants before this study. Informed consent was obtained in writing from all participants. Parental or guardian written consent was obtained from students aged 18 or 19 years. This study was approved by the Research Ethics Committee of the University of Shizuoka (3-12).

Results

Participants

All 23 respondents were included in the analysis. None of the participants reported restricting their vegetable intake for medical reasons.

Demographic Characteristics

The participants were 91% (n=21) women, with a mean age of 20.0 (SD 1.4) years (Table 1).

### Participants

- All 23 respondents were included in the analysis. None of the participants reported restricting their vegetable intake for medical reasons.
Table 1. Demographic of participants (N=23).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women, n (%)</td>
<td>21 (91)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>20.0 (1.4)</td>
</tr>
<tr>
<td>BMI (kg/m^2), mean (SD)</td>
<td>20.6 (1.2)</td>
</tr>
<tr>
<td>From the local area, n (%)</td>
<td>9 (39)</td>
</tr>
<tr>
<td>Lives alone, n (%)</td>
<td>17 (74)</td>
</tr>
<tr>
<td>Smokes, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Drinks 1-3 times a month or more, n (%)</td>
<td>11 (48)</td>
</tr>
<tr>
<td>Exercises 1-3 times a month or more, n (%)</td>
<td>7 (30)</td>
</tr>
</tbody>
</table>

### Item Analysis

For scale B, only 1 item (“raw vegetables used in salad: lettuce, cabbage, etc”) had a correlation coefficient greater than 0.5 and was significant at the 1% level (Table 2). Correlation coefficients of 0.4 to 0.5 and significant at the 5% level were found for 2 items (“tomatoes, tomato ketchup, boiled tomato, and stewed tomato” and “vegetables used in cooking: cabbage and Chinese cabbage”).

For scale C, 4 items (“broccoli,” “lettuce,” “burdock,” and “lotus root”) had a correlation coefficient greater than 0.5 and were significant at the 1% level. Correlation coefficients of 0.4 to 0.5 and significant at the 5% level were found for 2 items (“carrots” and “pumpkins”).

For scale D, 6 items (“broccoli,” “spinach,” “Welsh onions [green],” “lettuce,” “burdock,” and “ginger”) had a correlation coefficient greater than 0.5 and were significant at the 1% level. Correlation coefficients of 0.4 to 0.5 and significant at the 5% level were found for 5 items (“carrots,” “pumpkins,” “komatsuna,” “Chinese chive,” and “Welsh onions [branching cultivation]”).

Since none of the items were significantly inversely correlated with the 3 scales, none of the items were removed. The validity and reproducibility of scales selected items that significantly correlated with total scores were also examined. The item selection scales were referred to as short scales B, C, and D.
Table 2. Item-total correlations for the 3 scored scales (N=23).

<table>
<thead>
<tr>
<th>Scale B</th>
<th>$P$ value</th>
<th>$\rho^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pickled vegetables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green leafy vegetables</td>
<td>.23</td>
<td>.29</td>
</tr>
<tr>
<td>Other (excluding salted pickled plum)</td>
<td>.23</td>
<td>.29</td>
</tr>
<tr>
<td><strong>Raw vegetables used in salad</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lettuce, cabbage, etc (excluding tomatoes)</td>
<td>.54</td>
<td>.008</td>
</tr>
<tr>
<td>Tomatoes, tomato ketchup, boiled tomato, and stewed tomato</td>
<td>.55</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Vegetables used in cooking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green leafy vegetables (including broccoli and bitter melon)</td>
<td>-0.05</td>
<td>.82</td>
</tr>
<tr>
<td>Cabbage and Chinese cabbage</td>
<td>.45</td>
<td>.03</td>
</tr>
<tr>
<td>Carrots and pumpkins</td>
<td>.23</td>
<td>.29</td>
</tr>
<tr>
<td>Daikon and turnips</td>
<td>.40</td>
<td>.06</td>
</tr>
<tr>
<td>All other root vegetables (including onions, burdock, and lotus root)</td>
<td>.37</td>
<td>.09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale C</th>
<th>$P$ value</th>
<th>$\rho^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrots</td>
<td>0.47</td>
<td>.02</td>
</tr>
<tr>
<td>Pumpkins</td>
<td>0.49</td>
<td>.02</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>0.15</td>
<td>.49</td>
</tr>
<tr>
<td>Sweet peppers</td>
<td>0.23</td>
<td>.30</td>
</tr>
<tr>
<td>Broccoli</td>
<td>0.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Green leafy vegetables</td>
<td>0.41</td>
<td>.05</td>
</tr>
<tr>
<td>Lettuce</td>
<td>0.54</td>
<td>.008</td>
</tr>
<tr>
<td>Cabbage</td>
<td>0.38</td>
<td>.07</td>
</tr>
<tr>
<td>Cucumbers</td>
<td>0.32</td>
<td>.14</td>
</tr>
<tr>
<td>Chinese cabbage</td>
<td>0.18</td>
<td>.40</td>
</tr>
<tr>
<td>Bean sprouts</td>
<td>0.19</td>
<td>.40</td>
</tr>
<tr>
<td>Daikon</td>
<td>0.32</td>
<td>.14</td>
</tr>
<tr>
<td>Onions</td>
<td>0.36</td>
<td>.09</td>
</tr>
<tr>
<td>Cauliflower</td>
<td>__</td>
<td>b</td>
</tr>
<tr>
<td>Eggplants</td>
<td>0.11</td>
<td>.63</td>
</tr>
<tr>
<td>Burdock</td>
<td>0.71</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lotus root</td>
<td>0.53</td>
<td>.009</td>
</tr>
<tr>
<td>Pickled vegetables (excluding salted pickled plum)</td>
<td>0.24</td>
<td>.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale D</th>
<th>$P$ value</th>
<th>$\rho^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrots</td>
<td>0.47</td>
<td>.02</td>
</tr>
<tr>
<td>Pumpkins</td>
<td>0.49</td>
<td>.02</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>0.10</td>
<td>.69</td>
</tr>
<tr>
<td>Sweet peppers</td>
<td>0.34</td>
<td>.11</td>
</tr>
<tr>
<td>Broccoli</td>
<td>0.65</td>
<td>.001</td>
</tr>
<tr>
<td>Spinach</td>
<td>0.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Komatsuna</td>
<td>0.47</td>
<td>.03</td>
</tr>
<tr>
<td>Welsh onions (green)</td>
<td>0.62</td>
<td>.001</td>
</tr>
</tbody>
</table>
For each scale, the correlation coefficients were as follows: scale A, $\rho=0.28$ ($P=0.20$); scale B, $\rho=0.22$ ($P=0.31$); scale C, $\rho=0.51$ ($P=0.01$); and scale D, $\rho=0.44$ ($P=0.03$; Table 3). This indicated that the 2 scales assessed by vegetable items (ie, scales C and D) were significantly correlated with vegetable variety based on the dietary records for 7 consecutive days. Furthermore, the correlation coefficient was $\rho=0.08$ ($P=0.73$) for short scale B, $\rho=0.42$ ($P=0.04$) for short scale C, and $\rho=0.33$ ($P=0.12$) for short scale D, indicating that the correlation decreased with the selection of items for all 3 scales.

**Table 3.** Correlation of each scale with vegetable variety based on dietary records for 7 consecutive days (N=23).

<table>
<thead>
<tr>
<th>Scale</th>
<th>$\rho^a$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale A</td>
<td>0.28</td>
<td>0.20</td>
</tr>
<tr>
<td>Scale B</td>
<td>0.22</td>
<td>0.31</td>
</tr>
<tr>
<td>Scale C</td>
<td>0.51</td>
<td>0.01</td>
</tr>
<tr>
<td>Scale D</td>
<td>0.44</td>
<td>0.03</td>
</tr>
<tr>
<td>Short scale B</td>
<td>0.08</td>
<td>0.73</td>
</tr>
<tr>
<td>Short scale C</td>
<td>0.42</td>
<td>0.04</td>
</tr>
<tr>
<td>Short scale D</td>
<td>0.33</td>
<td>0.12</td>
</tr>
</tbody>
</table>

$^a$Spearman correlation coefficient.

**Validity**

For each scale, the correlation coefficient of the first and second responses for scale A was $\rho=0.24$ ($P=0.27$; Figure 1). Scale B was $\rho=0.45$ ($P=0.03$), scale C was $\rho=0.73$ ($P<0.001$), and scale D was $\rho=0.74$ ($P<0.001$), confirming a correlation between the 2 responses on the vegetable subgroup scale and a strong correlation for the 2 scales using vegetable items. Furthermore, short scale B was $\rho=0.40$ ($P=0.06$), short scale C was $\rho=0.85$ ($P<0.001$), and short scale D was $\rho=0.79$ ($P<0.001$). Although the item selection reduced the reproducibility of the vegetable subgroup scale, the 2 scales for vegetable items were still reproducible.
Figure 1. Scatter plots of the first and second responses on 4 complete scales and 3 scales with selected items (N=23).

Discussion

Principal Results and Comparison With Prior Work

In this study, the validity and reproducibility of several scales to determine vegetable variety was examined and compared. Regarding validity, scales C and D significantly correlated with vegetable variety based on dietary records for 7 consecutive days. Regarding reproducibility, a significant correlation was observed for scale B, and significant strong correlations were observed for scales C and D.

Item-total correlations were conducted to examine the internal consistency of the scales. In this analysis, none of the items in all 3 scales negatively correlated with the total score. Furthermore, although items with significant correlations were selected for each scale, correlations with vegetable variety based on dietary records for 7 consecutive days were lower than the full version of the scale. Therefore, all items removed for the short scale may be necessary for determining vegetable variety, even though their correlations are weak.

For the association with vegetable variety based on dietary records for 7 consecutive days, significant correlations (p>0.4) were found for scales C and D but not for scales A and B. Furthermore, for the assessment of reproducibility, significant correlations were found for scale B, and significantly strong correlations (p>0.7) were found for scales C and D. For validation studies of FFQs, correlation coefficients with validity lower than 0.4 are considered seriously attenuate associations [25]. Compared to the scales for a single question (scale A) and vegetable subgroup (scale B), the scales for vegetable items (scales C and D) demonstrated acceptable validity and reproducibility regarding vegetable variety and, therefore, may be useful for determining the rank of vegetable variety.

Low-intake vegetables used as condiments (eg, chopped Welsh onion as a garnish), which would have little nutritional significance, were removed from the dietary records used as a reference standard in this study with a quantitative criterion for single use, based on a previous study using the 1-day food record [24]. However, other studies that have used 24-hour recalls have not specified the minimum amount of consumption [12]. Since vegetable variety varies depending on the establishment of a minimum amount consumed using either dietary records or 24-hour recalls used as a reference standard, the necessity of establishing a minimum amount and the appropriateness of the cutoff value should be examined in the future. In this study, the scales examined did not establish a minimum amount of vegetable consumption. Nonquantitative FFQs such as the BDHQ referenced in this study, are confirmed to have reasonable validity for vegetable intake using the frequency measurements only [18]. Therefore, there appears to be little need to establish a minimum amount consumed, at least for a scale aimed at determining vegetable variety.

Limitations

This study has some limitations, including the use of a single survey. It has been suggested that data collected on the eating patterns of individuals should include all seasons of the year as well as days in all parts of the week [26]. Indeed, a previous study reported significant seasonal differences in vegetable intake [27]. Since vegetable variety was assessed using data
collected on 7 consecutive days as a reference standard, seasonality may have affected the vegetable variety measured in this study. The BDHQ and DHQ used as a reference for scales B and C have reasonable validity for vegetable intake conducted over 4 seasons [18]. Furthermore, although this study was conducted in July 2021, the Ranking of Vegetable Consumers in Japan based on the NHNS used as a reference for scale D was assessed in November 2012 [20]. Therefore, data from this study suggested that a minimal effect of seasonal differences on the ability of each scale to determine vegetable variety. A second limitation of this study is the small sample size. Since this study included only students majoring in nutrition at 1 Japanese university (mostly women, younger in age, normal or low BMI, and living alone) and interested in this survey, it is not clear to what extent the results can be generalized. The limited number of participants in this study also made it difficult to adjust for confounding factors. Future research should assess the validity of scales in different regions and populations. Although there were no missing days for the dietary records in this study, which likely occurred since the participants were students majoring in nutrition, it is suggested that the usability of the records decreases as the records progress to the seventh day [28]. A third limitation of this study is that, because biochemical indicators such as plasma carotenoid concentrations were not assessed, the effect of underreporting on the association of vegetable variety is not clear. Finally, in cases where participants ate commercial products, ate at a restaurant, or recorded only the name of the dish, the number of vegetables consumed was estimated using information from the internet. Therefore, although the results may not reflect the actual consumption of the participants in this study, generalization could be maintained because the foods that appeared most frequently on the internet were extracted.

Conclusions

Correlation analysis of vegetable variety based on dietary records for 7 consecutive days and test-retest reliability suggests that the 2 scales for vegetable items have acceptable validity and reproducibility compared to the scales that used a single question or vegetable subgroup and, therefore, may determine the variety of vegetables consumed.

Acknowledgments

The authors would like to thank all the participants in this study and Izumi Mochizuki (University of Shizuoka) for her active cooperation. This research was funded by the Japan Society for the Promotion of Science KAKENHI (JP19K14044).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The exact wording and scoring methods for each scale.

[PDF File (Adobe PDF File), 164 KB - formative_v8i1e55795_app1.pdf ]

References


Abbreviations

BDHQ: brief-type self-administered diet history questionnaire
DHQ: diet history questionnaire
FAVVA: Fruit And Vegetable VAriety
FFQ: food frequency questionnaire
NHNS: National Health and Nutrition Survey
Assessing Priorities in a Statewide Cardiovascular and Diabetes Health Collaborative Based on the Results of a Needs Assessment: Cross-Sectional Survey Study

Elizabeth A Beverly, MS, PhD; Sarah Koopman-Gonzalez, PhD; Jackson Wright, MD, PhD; Kathleen Dungan, MPH, MD; Harini Pallerla, MS; Rose Gubitosi-Klug, MD, PhD; Kristin Baughman, PhD; Michael W Konstan, MD; Shari D Bolen, MPH, MD

1Department of Primary Care, Ohio University Heritage College of Osteopathic Medicine, Athens, OH, United States
2Department of Medicine, Case Western Reserve University, Cleveland, OH, United States
3Department of Medicine, Ohio State University College of Medicine, Columbus, OH, United States
4Department of Family and Community Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, United States
5Department of Family and Community Medicine, Northeast Ohio Medical University, Rootstown, OH, United States
6Center for Health Care Research and Policy, MetroHealth Medical Center, Cleveland, OH, United States
*all authors contributed equally

Corresponding Author:
Elizabeth A Beverly, MS, PhD
Department of Primary Care
Ohio University Heritage College of Osteopathic Medicine
1 Ohio University
122 Medical Education Center
Athens, OH, 45701
United States
Phone: 1 7405934616
Fax: 1 7405932205
Email: beverle1@ohio.edu

Abstract

Background: The Ohio Cardiovascular and Diabetes Health Collaborative (Cardi-OH) unites general and subspecialty medical staff at the 7 medical schools in Ohio with community and public health partnerships to improve cardiovascular and diabetes health outcomes and eliminate disparities in Ohio’s Medicaid population. Although statewide collaboratives exist to address health improvements, few deploy needs assessments to inform their work.

Objective: Cardi-OH conducts an annual needs assessment to identify high-priority clinical topics, screening practices, policy changes for home monitoring devices and referrals, and preferences for the dissemination and implementation of evidence-based best practices. The results of the statewide needs assessment could also be used by others interested in disseminating best practices to primary care teams.

Methods: A cross-sectional survey was distributed electronically via REDCap (Research Electronic Data Capture; Vanderbilt University) to both Cardi-OH grant-funded and non–grant-funded members (ie, people who have engaged with Cardi-OH but are not funded by the grant).

Results: In total, 88% (103/117) of Cardi-OH grant-funded members and 8.14% (98/1204) of non–grant-funded members completed the needs assessment survey. Of these, 51.5% (53/103) of Cardi-OH grant-funded members and 47% (46/98) of non–grant-funded members provided direct clinical care. The top cardiovascular medicine and diabetes clinical topics for Cardi-OH grant-funded members (clinical and nonclinical) were lifestyle prescriptions (50/103, 48.5%), atypical diabetes (38/103, 36.9%), COVID-19 and cardiovascular disease (CVD; 38/103, 36.9%), and mental health and CVD (38/103, 36.9%). For non–grant-funded members, the top topics were lifestyle prescriptions (53/98, 54%), mental health and CVD (39/98, 40%), alcohol and CVD (27/98, 28%), and cardiovascular complications (27/98, 28%). Regarding social determinants of health, Cardi-OH grant-funded members prioritized 3 topics: weight bias and stigma (44/103, 42.7%), family-focused interventions (44/103, 42.7%), and adverse childhood events (57/103, 35.9%). Non–grant-funded members’ choices were family-focused interventions (51/98, 52%), implicit bias...
(43/98, 44%), and adverse childhood events (39/98, 40%). Assessment of other risk factors for CVD and diabetes across grant- and non–grant-funded members revealed screening for social determinants of health in approximately 50% of patients in each practice, whereas some frequency of depression and substance abuse screening occurred in 80% to 90% of the patients. Access to best practice home monitoring devices was challenging, with 30% (16/53) and 41% (19/46) of clinical grant-funded and non–grant-funded members reporting challenges in obtaining home blood pressure monitoring devices and 68% (36/53) and 43% (20/46) reporting challenges with continuous glucose monitors.

Conclusions: Cardi-OH grant- and non–grant-funded members shared the following high-priority topics: lifestyle prescriptions, CVD and mental health, family-focused interventions, alcohol and CVD, and adverse childhood experiences. Identifying high-priority educational topics and preferred delivery modalities for evidence-based materials is essential for ensuring that the dissemination of resources is practical and useful for providers.

(JMIR Form Res 2024;8:e55285) doi:10.2196/55285

KEYWORDS
health collaborative; cardiovascular disease; type 2 diabetes; needs assessment

Introduction

Background
Cardiovascular disease (CVD) is the leading cause of death in the United States [1] and worldwide, representing 32% of all global deaths or 17.9 million deaths annually [2]. In 2021, the age-adjusted CVD mortality rate was 204.7 per 100,000 Ohio residents, which is 17.8% higher than the national average [3]. Numerous risk factors contribute to Ohio’s higher rates of CVD, including tobacco use, hypertension, and diabetes. For example, in 2023, Ohio had the fourth highest smoking rate in the United States, with 21% of the population (2.4 million people) smoking [4]. Among adults with CVD in Ohio, 73% also have hypertension, and 69% have hyperlipidemia [5]. Similarly, data from 2016 show that adults with CVD in Ohio are more likely to have diabetes, cancer, kidney disease, asthma, and chronic obstructive pulmonary disease compared to those without CVD [5]. For these reasons, statewide approaches that address modifiable cardiometabolic risk factors are needed to facilitate the dissemination and implementation of evidence-based best practices for health care teams.

In 2017, the Ohio Department of Medicaid (ODM) funded the development of the Ohio Cardiovascular and Diabetes Health Collaborative (Cardi-OH) [6]. Cardi-OH is a statewide health collaborative that unites the 7 medical schools in Ohio to improve cardiovascular and diabetes health outcomes and reduce disparities in Ohio’s Medicaid population. Health collaboratives mobilize community and public health partnerships to identify and solve health problems [7]. Specifically, Cardi-OH brings together primary care physicians, specialists, pharmacists, physician assistants, nurse practitioners, nurses, dietitians, and social scientists affiliated with the 7 medical schools as well as representatives from the ODM and Ohio’s Medicaid Managed Care plans to share their expertise to accelerate learning and the implementation of best practices. The ODM also separately funds statewide quality improvement projects focused on paired implementation in which Cardi-OH resources may be used [8,9].

Objectives
To prioritize the dissemination of topics each year, Cardi-OH conducts an annual needs assessment to identify educational foci for CVD risk reduction and type 2 diabetes management.

Previous literature on needs assessments by regional and statewide health collaboratives has focused broadly on factors that influence health in the community or on a review of cardiovascular health or diabetes statistics for a particular region [10,11]. Our needs assessment differs by eliciting primary care team members’ most important considerations for implementing evidence-based best practices for both CVD risk reduction and type 2 diabetes. We present findings from the 2021 to 2022 needs assessment survey that identifies the prioritized topics, use of selected screening practices, and selected barriers to the dissemination and implementation of evidence-based best practices, which were used to develop the Cardi-OH materials for the 2022-2023 academic year. These findings may provide guidance for health care systems, professional organizations, payers, community-based organizations, and other health collaboratives looking to prioritize activities for CVD risk reduction and type 2 diabetes management.

Methods

Ethical Considerations
Ethics approval for the needs assessment was obtained from the Case Western Reserve University Institutional Review Board (study STUDY20180486). The research was classified as exempt and, therefore, did not require signed informed consent. In accordance with the Belmont principle of respect for persons, all Cardi-OH members were given the opportunity to choose whether to participate in the assessment. Furthermore, in compliance with federal, state, and local laws and regulations for human participants, we ensured that our research met the requirements set forth in the regulations on public welfare in Part 46 of Title 45 of the Code of Federal Regulations, the principles set forth in the Belmont Report, and the Helsinki Declaration of 1975. All study data were deidentified before analysis. Respondents received no human participant compensation for taking part in the needs assessment.

Research Design
We conducted a descriptive cross-sectional needs assessment with all members of the Cardi-OH community. Specifically, we administered a confidential electronic survey to identify important clinical topics, screening practices, policy changes for home monitoring devices and referrals, and preferences for
the dissemination modality of evidence-based best practice materials. We elected to conduct a needs assessment because it is a systematic approach used to identify the priorities of a group and determine its capacity to address the needs of the population being served.

**Participants**

During the 2021-2022 academic year, Cardi-OH included a total of 1321 members. Of these 1321 members, 117 (8.86%) were grant-funded members affiliated with 1 of the 7 Ohio medical schools. For this noncompetitive statewide grant, grant-funded members included direct care providers and public health professionals specializing in CVD and type 2 diabetes. These grant-funded members were responsible for identifying, producing, and disseminating the latest evidence-based cardiovascular and diabetes best practices. The remaining 91.14% (1204/1321) of members were non–grant-funded members. Non–grant-funded members included community providers or stakeholders who engaged with Cardi-OH by registering through Cardi-OH–sponsored events (ie, statewide webinars and Cardi-OH Extension for Community Healthcare Outcomes [ECHO] clinics), registering through the Cardi-OH website, or engaging with Cardi-OH via alignment efforts with other statewide partners such as regional professional associations and community-based organizations. There were no exclusion criteria for joining Cardi-OH as a non–grant-funded member. Both grant-funded and non–grant-funded Cardi-OH members were tracked in a REDCap (Research Electronic Data Capture; Vanderbilt University) [12,13] database to record membership, event attendance, and program evaluation data.

**Measures**

Cardi-OH consists of 5 large teams: the executive team (ie, principal investigators from the 7 medical schools), team best practices (ie, the team that reviews and synthesizes evidence and national guidelines for dissemination), team ECHO (ie, the team that develops curricular content for 12-week web-based case-based learning series), the communications team (ie, the team responsible for branding, disseminating materials, and maintaining the website), and the data and evaluation team (ie, the team responsible for evaluating the success and effectiveness of the events and materials produced by Cardi-OH). The data and evaluation team comprises 12 experts in quantitative and qualitative methodologies representing 5 of the 7 medical schools. The goal of the data and evaluation team is to establish a set of metrics designed to measure the effectiveness of Cardi-OH. Every year, the data and evaluation team solicits input from all teams for CVD- and type 2 diabetes–related topics for an annual needs assessment. We have conducted needs assessments in previous years; however, this was the first year we administered the needs assessment to both grant-funded and non–grant-funded Cardi-OH members. Demographic questions are carried over from one year to the next, but topical questions are new with each needs assessment. These questions include lists of suggested topics that Cardi-OH members rank as high-priority educational topics to develop events and materials for in the upcoming year. The 2021 to 2022 needs assessment survey consisted of 24 questions; clinical members had to answer all 24 (100%) questions whereas nonclinical members had to answer only 5 of the 24 (21%) questions. For nonclinical members, the needs assessment took approximately 3 to 5 minutes to complete; for clinical members, the needs assessment took approximately 10 to 15 minutes to complete. Clinical members answered more questions because the survey included questions about the direct care provided, screening practices, comfort level with topics, and perceived difficulty ordering remote monitoring devices. While the focus of the needs assessment is clinical, we also include nonclinical members’ priorities to promote alignment with public health professionals across the state. Effective collaboration between clinical and public health professionals is essential to improve cardiovascular and type 2 diabetes health outcomes.

To establish face and content validity, the data and evaluation team reviewed and rated each question to determine whether it was necessary, useful, and relevant to be included in the needs assessment. The needs assessment was then piloted on February 24, 2022, with 2 primary care physicians to identify any weak or irrelevant questions; no questions were removed after the pilot test. Reliability and validity testing was not conducted because no specific constructs (eg, knowledge, attitudes, and beliefs) were measured using this assessment.

**Data Collection**

The 2022 Cardi-OH needs assessment was disseminated electronically via REDCap to all Cardi-OH members. REDCap is a secure electronic data capture program designed for collecting survey data [12,13]; the program was hosted by Case Western Reserve University. The survey opened on March 29, 2022, and closed on April 22, 2022. Participation in this study was voluntary. All Cardi-OH members were sent 2 email reminders.

**Statistical Analysis**

Descriptive statistics were used to describe the survey participants. The frequencies of individual question responses were calculated by clinical and nonclinical status as well as by grant- and non–grant-funded status. For the purposes of the analysis, clinical providers were defined as any Cardi-OH member who provided direct clinical care to patients. Statistical significance was defined as P<.05. All analyses were conducted using SPSS statistical software (version 28.0; IBM Corp).

**Results**

**Cardi-OH Survey Participants**

A total of 88% (103/117) of grant-funded members completed the needs assessment (Table 1). Of these, 51.5% (53/103) provided direct clinical care, of whom 79% (42/53) identified as physicians. More than half (30/53, 57%) of clinical providers practiced in a primary care setting, and 43% (23/53) worked in internal medicine. Providers estimated that 41.5% (SD 19.8%) of patients were enrolled in Medicaid.
Table 1. Ohio Cardiovascular and Diabetes Health Collaborative (Cardi-OH) grant- and non–grant-funded members’ demographic characteristics.

<table>
<thead>
<tr>
<th>Question</th>
<th>Grant-funded members (n=103)</th>
<th>Non–grant-funded members (n=98)</th>
<th>Chi-square (df) or Fisher exact test</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cardi-OH members: in what sector are you employed? (Select all that apply), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>88 (85.4)</td>
<td>27 (27.6)</td>
<td>68.7 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Health care</td>
<td>48 (46.6)</td>
<td>56 (57.1)</td>
<td>2.2 (1)</td>
<td>.14</td>
</tr>
<tr>
<td>Health plan or insurer</td>
<td>1 (1)</td>
<td>6 (6.1)</td>
<td></td>
<td>.06</td>
</tr>
<tr>
<td>Pharmaceutical or manufacturing</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td></td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Nonprofit</td>
<td>9 (8.7)</td>
<td>15 (15.3)</td>
<td></td>
<td>.19</td>
</tr>
<tr>
<td>Philanthropic</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td>.11</td>
</tr>
<tr>
<td>Private</td>
<td>0 (0)</td>
<td>3 (3.1)</td>
<td></td>
<td>.99</td>
</tr>
<tr>
<td>Public</td>
<td>18 (17.5)</td>
<td>20 (20.4)</td>
<td></td>
<td>.60</td>
</tr>
<tr>
<td>Research or policy or not academic</td>
<td>10 (9.7)</td>
<td>2 (2)</td>
<td>0.2 (1)</td>
<td>.03</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>2 (2)</td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>All Cardi-OH members: do you provide direct clinical care to patients as part of a clinical team? Yes, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.67</td>
</tr>
<tr>
<td>Yes</td>
<td>53 (51.5)</td>
<td>46 (46.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>48 (46.6)</td>
<td>52 (53.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Cardi-OH members: what is your clinical role?, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
</tr>
<tr>
<td>Physician</td>
<td>42 (79.2)</td>
<td>24 (52.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse practitioner</td>
<td>1 (1.9)</td>
<td>5 (10.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician assistant</td>
<td>2 (3.8)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNc</td>
<td>1 (1.9)</td>
<td>6 (13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPNd</td>
<td>0 (0)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical assistant</td>
<td>0 (0)</td>
<td>2 (4.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registered dietitian</td>
<td>1 (1.9)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social worker</td>
<td>1 (1.9)</td>
<td>3 (6.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychologist</td>
<td>0 (0)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacist</td>
<td>4 (7.5)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Cardi-OH members: in what setting do you practice?, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td>Primary care</td>
<td>30 (56.6)</td>
<td>33 (71.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialty care</td>
<td>20 (37.7)</td>
<td>9 (19.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3 (5.7)</td>
<td>4 (8.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Cardi-OH members: in what specialty area do you work?, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.64</td>
</tr>
<tr>
<td>Family medicine</td>
<td>15 (28.3)</td>
<td>17 (37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal medicine</td>
<td>23 (43.4)</td>
<td>14 (30.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pediatrics</td>
<td>3 (5.7)</td>
<td>5 (10.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geriatrics</td>
<td>1 (1.9)</td>
<td>1 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>10 (18.9)</td>
<td>9 (19.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Cardi-OH members: how would you describe the geographic setting where you primarily practice? Urban, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Urban</td>
<td>42 (79.2)</td>
<td>30 (65.2)</td>
<td>2.4 (1)</td>
<td></td>
</tr>
<tr>
<td>Suburban</td>
<td>18 (34)</td>
<td>16 (34.8)</td>
<td>0.0 (1)</td>
<td>.93</td>
</tr>
</tbody>
</table>
A total of 8.14% (98/1204) of non–grant-funded members completed the needs assessment. Of these, 47% (46/98) reported providing direct clinical care, of whom 52% (24/46) identified as physicians. More than two-thirds (33/46, 72%) of the non–grant-funded providers practiced in a primary care setting, and 37% (17/46) worked in family medicine. Providers estimated that 51.7% (SD 29.3%) of their patients were enrolled in Medicaid.

**Top-Rated Cardiovascular Medicine Topics**

Cardi-OH members were asked to choose their top topics of interest for cardiovascular medicine (Multimedia Appendix 1). For all grant-funded members (n=103; clinical and nonclinical), the top cardiovascular medicine topics were lifestyle prescriptions (50/103, 48.5%; Multimedia Appendix 1), atypical diabetes (38/103, 36.9%), COVID-19 and CVD (38/103, 36.9%), mental health and CVD (38/103, 36.9%), and alcohol and CVD (32/103, 31.1%). Among clinical grant-funded members, the highest-rated topics were atypical diabetes (29/53, 55%), calcium scoring (26/53, 49%), and special cases of CVD and diabetes (24/53, 45%).

For all non–grant-funded members (n=98), the top cardiovascular medicine topics were lifestyle prescriptions (53/98, 54%), mental health and CVD (39/98, 40%), alcohol (27/98, 28%), and cardiovascular complications (27/98, 28%). The top 3 topics among clinical non–grant-funded members were lifestyle prescriptions (27/46, 59%), mental health and CVD (22/46, 48%), and alcohol and CVD (20/46, 43%).

**Top-Rated Topics Related to Social Determinants of Health**

Regarding social determinants of health (SDoHs), grant-funded members (clinical and nonclinical) selected their top topics: (1) weight bias and stigma (44/103, 42.7%; Multimedia Appendix 2), (2) family-focused interventions for CVD and diabetes (40/103, 38.8%), (3) adverse childhood experiences and their association with CVD and diabetes (37/103, 35.9%), and (4) implicit bias and CVD (37/103, 35.9%). For clinical grant-funded members, the highest-rated topics were family-focused interventions for CVD and diabetes (24/53, 45%), weight bias and stigma (22/53, 42%), and peer support interventions for CVD and diabetes (19/53, 36%).

Non–grant-funded members selected (1) family-focused interventions for CVD and diabetes (51/98, 52%), (2) implicit bias and CVD (43/98, 44%), and (3) adverse childhood experiences and their association with CVD and diabetes (39/98, 40%). Among clinical non–grant-funded members, the highest-rated topics were adverse childhood experiences and their association with CVD and diabetes (19/46, 41%), gender disparities in CVD (17/46, 37%), and implicit bias and CVD (17/46, 37%).

**Screening Practices for SDoHs**

More than half (28/53, 53%) of clinical grant-funded members felt “very” or “extremely” comfortable screening for SDoHs (Table 2). Almost half (21/53, 40%) screened for SDoHs when their patients appeared to have barriers, and 25% (13/53) screened at every visit. Only 23% (12/53) of clinical grant-funded members were “very” or “extremely” confident that they could address their patients’ SDoHs. The most common methods for screening for SDoHs were verbal (30/53, 57%), web (21/53, 40%), paper (8/53, 15%), and others (2/53, 4%).

---

### Table 1: Questionnaire Results Among Cardi-OH Members

<table>
<thead>
<tr>
<th>Question</th>
<th>Grant-funded members (n=103)</th>
<th>Non–grant-funded members (n=98)</th>
<th>Chi-square (df) or Fisher exact test</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rural</strong></td>
<td>6 (11.3)</td>
<td>8 (17.4)</td>
<td>0.7 (1)</td>
<td>.39</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>1 (1.9)</td>
<td>0 (0)</td>
<td>0.8 (1)</td>
<td>.35</td>
</tr>
<tr>
<td><strong>Clinical Cardi-OH members: approximately what percentage of the patients you serve are enrolled in Medicaid? (%)</strong></td>
<td>41.5 (19.8)</td>
<td>51.7 (29.3)</td>
<td>−1.9 (87)</td>
<td>.06</td>
</tr>
</tbody>
</table>

### Table 2: Clinical Cardi-OH members: how often do you provide clinical care?, n (%)^b^  

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Grant-funded members (n=53)</th>
<th>Non–grant-funded members (n=46)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;75% of the time</td>
<td>9 (17)</td>
<td>19 (41.3)</td>
<td></td>
</tr>
<tr>
<td>50%-75% of the time</td>
<td>16 (30.2)</td>
<td>8 (17.4)</td>
<td></td>
</tr>
<tr>
<td>25%-49% of the time</td>
<td>11 (20.8)</td>
<td>8 (17.4)</td>
<td></td>
</tr>
<tr>
<td>&lt;25% of the time</td>
<td>17 (32.1)</td>
<td>11 (23.9)</td>
<td>.06</td>
</tr>
</tbody>
</table>

^aFisher exact test.  
^bGrant-funded members: n=53; non–grant-funded members: n=46.  
^cRN: registered nurse.  
^dLPN: licensed practical nurse.
Table 2. Ohio Cardiovascular and Diabetes Health Collaborative grant- and non–grant-funded members’ responses to the questions on screening.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Clinical grant-funded members (n=53), n (%)</th>
<th>Clinical non–grant-funded members (n=46), n (%)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions on screening for SDoHs&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How comfortable do you feel screening for SDoHs in adults with type 2 diabetes?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely comfortable</td>
<td>13 (25)</td>
<td>12 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very comfortable</td>
<td>15 (28)</td>
<td>11 (24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately comfortable</td>
<td>18 (34)</td>
<td>11 (24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly comfortable</td>
<td>4 (8)</td>
<td>5 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all comfortable</td>
<td>2 (4)</td>
<td>7 (15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often do you screen for SDoHs?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annually</td>
<td>6 (11)</td>
<td>12 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every visit</td>
<td>13 (25)</td>
<td>11 (24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only when the patient appears to have barriers</td>
<td>21 (40)</td>
<td>10 (22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5 (9)</td>
<td>7 (15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5 (9)</td>
<td>2 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not know</td>
<td>2 (4)</td>
<td>2 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How confident are you that you can address the SDoHs of your patients (eg, referrals)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely confident</td>
<td>7 (13)</td>
<td>6 (13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very confident</td>
<td>5 (9)</td>
<td>13 (28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately confident</td>
<td>23 (43)</td>
<td>13 (28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly confident</td>
<td>12 (23)</td>
<td>9 (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all confident</td>
<td>3 (6)</td>
<td>5 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What method do you use to screen for SDoHs? (Select all that apply)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper</td>
<td>8 (15)</td>
<td>18 (39)</td>
<td>7.4 (1)</td>
<td>.007</td>
</tr>
<tr>
<td>Web</td>
<td>21 (40)</td>
<td>20 (43)</td>
<td>0.2 (1)</td>
<td>.70</td>
</tr>
<tr>
<td>Verbal</td>
<td>30 (57)</td>
<td>25 (54)</td>
<td>0.2 (1)</td>
<td>.82</td>
</tr>
<tr>
<td>Other</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>1.8 (1)</td>
<td>.18</td>
</tr>
<tr>
<td>Do not know</td>
<td>5 (9)</td>
<td>5 (11)</td>
<td>0.1 (1)</td>
<td>.81</td>
</tr>
<tr>
<td>Questions on screening for behavioral health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you routinely screen patients for depression?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All patients are screened</td>
<td>36 (68)</td>
<td>35 (76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certain populations are screened</td>
<td>5 (9)</td>
<td>3 (7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>5 (9)</td>
<td>3 (7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not know</td>
<td>7 (13)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>0 (0)</td>
<td>5 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often do you screen patients for depression?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annually</td>
<td>14 (26)</td>
<td>17 (37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every visit</td>
<td>20 (38)</td>
<td>19 (41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only when the patient appears depressed</td>
<td>4 (8)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3 (6)</td>
<td>1 (2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Among the clinical non–grant-funded members, 50% (23/46) felt “very” or “extremely” comfortable screening for SDoHs. A total of 22% (10/46) screened for SDoHs when their patients appeared to have barriers, and 24% (11/46) screened at every visit. A total of 41% (19/46) of clinical non–grant-funded members were “very” or “extremely” confident that they could address their patients’ SDoHs. The most common methods for screening for SDoHs were verbal (25/46, 54%), web (20/46, 43%), and paper (18/46, 39%).

The only significant difference between clinical grant- and non–grant-funded members pertained to the screening methods for SDoHs. Specifically, clinical non–grant-funded members used paper more frequently than did grant-funded members ($\chi^2 = 7.4; P = .007$). No other differences were observed in screening practices for SDoHs.

### Screening Practices for Depression

For depression screening, 68% (36/53) of clinical grant-funded members indicated that all patients were screened, 9% (5/53) screened only certain populations, 13% (7/53) selected not applicable, and 9% (5/53) did not screen (Table 2). Regarding the frequency of depression screening, 38% (20/53) screened at every visit, 26% (14/53) screened annually, 8% (4/53) screened when the patient appeared depressed, and 6% (3/53) selected other. Only 13% (7/53) of clinical grant-funded members specified that they were always able to connect their Medicaid patients with needed behavioral health services, 30% (16/53) reported that they were able to do so often, 36% (19/53) reported that they were able to do so sometimes, 2% (1/53) reported that they were able to do so rarely, 4% (2/53) did not know, and 15% (8/53) selected not applicable.
Among clinical non–grant-funded members, 76% (35/46) indicated that all patients were screened for depression, 7% (3/46) screened only certain populations, and 11% (5/46) selected not applicable. Regarding the frequency of depression screening, 41% (19/46) screened at every visit, 37% (17/46) screened annually, and 2% (1/46) selected other. Approximately one-quarter (12/46, 26%) of clinical non–grant-funded members specified that they were always able to connect their Medicaid patients with needed behavioral health services, 33% (15/46) reported that they were able to do so often, 26% (12/46) reported that they were able to do so sometimes, 4% (2/46) reported that they were able to do so rarely, and 11% (5/46) selected not applicable.

No differences were observed in behavioral health screening practices between clinical and nonclinical grant-funded members (Table 2).

Screening Practices for Substance Abuse

For substance abuse screening, 23% (12/53) of clinical grant-funded members screened at every visit, 23% (12/53) screened annually, 21% (11/53) screened when patients showed signs of substance abuse, 6% (3/53) did not know, 6% (3/53) did not screen, 13% (7/53) selected not applicable, and 9% (5/53) selected other (Table 2). Regarding their ability to connect Medicaid patients with needed treatment for substance abuse disorders, 9% (5/53) reported being able to do so always, 13% (7/53) reported being able to do so often, 32% (17/53) reported being able to do so sometimes, 8% (4/53) reported being able to do so rarely, 9% (5/53) did not know, and 26% (14/53) selected not applicable.

No differences were observed in substance abuse screening practices between clinical and nonclinical grant-funded members (Table 2).

Perceived Difficulty Obtaining Home Monitoring Devices and Making Referrals

Clinical Cardi-OH members were asked about their perceived difficulty obtaining specific home monitoring devices and making referrals because these were areas that Medicaid was specifically trying to address. Of the 53 clinical grant-funded members, 16 (30%; Table 3) perceived obtaining home blood pressure monitors as “moderately,” “very,” or “extremely” difficult, followed by 7 (13%) who perceived obtaining glucometers as “moderately” or “very” difficult and 36 (68%) who perceived obtaining continuous glucose monitoring (CGM) devices as “moderately,” “very,” or “extremely” difficult. Finally, 34% (18/53) of clinical grant-funded members found referring a Medicaid patient to diabetes self-management education and support (DSMES) “moderately,” “very,” or “extremely” difficult.
### Table 3. Ohio Cardiovascular and Diabetes Health Collaborative grant- and non–grant-funded members’ responses to questions on obtaining home monitoring devices and making referrals.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Clinical grant-funded members (n=53), n (%)</th>
<th>Clinical non–grant-funded members (n=46), n (%)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How difficult is it for you to obtain home BP(^a) monitors for your Medicaid patients with hypertension?</strong></td>
<td></td>
<td></td>
<td>4.0 (5)</td>
<td>.55</td>
</tr>
<tr>
<td>Extremely difficult</td>
<td>2 (4)</td>
<td>2 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very difficult</td>
<td>4 (8)</td>
<td>5 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately difficult</td>
<td>10 (19)</td>
<td>12 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly difficult</td>
<td>13 (25)</td>
<td>8 (17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all difficult</td>
<td>7 (13)</td>
<td>10 (22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A(^b)</td>
<td>17 (32)</td>
<td>9 (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>How difficult is it for you to obtain glucometers for your Medicaid patients with type 2 diabetes?</strong></td>
<td></td>
<td></td>
<td>1.8 (4)</td>
<td>.77</td>
</tr>
<tr>
<td>Extremely difficult</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very difficult</td>
<td>2 (4)</td>
<td>1 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately difficult</td>
<td>5 (9)</td>
<td>3 (7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly difficult</td>
<td>18 (34)</td>
<td>12 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all difficult</td>
<td>17 (32)</td>
<td>20 (43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>11 (21)</td>
<td>10 (22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>How difficult is it for you to obtain CGM(^c) devices for your Medicaid patients with type 2 diabetes?</strong></td>
<td></td>
<td></td>
<td>16.4 (6)</td>
<td>.01</td>
</tr>
<tr>
<td>Extremely difficult</td>
<td>6 (11)</td>
<td>3 (7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very difficult</td>
<td>9 (17)</td>
<td>11 (24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately difficult</td>
<td>21 (40)</td>
<td>6 (13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly difficult</td>
<td>2 (4)</td>
<td>9 (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all difficult</td>
<td>1 (2)</td>
<td>4 (9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>13 (25)</td>
<td>13 (28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>How difficult is it for you to refer a Medicaid patient to DSMES(^d)?</strong></td>
<td></td>
<td></td>
<td>3.5 (6)</td>
<td>.75</td>
</tr>
<tr>
<td>Extremely difficult</td>
<td>2 (4)</td>
<td>2 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very difficult</td>
<td>3 (6)</td>
<td>2 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately difficult</td>
<td>13 (25)</td>
<td>7 (15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slightly difficult</td>
<td>9 (17)</td>
<td>6 (13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all difficult</td>
<td>14 (26)</td>
<td>17 (37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>12 (23)</td>
<td>11 (24)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)BP: blood pressure.

\(^b\)N/A: not applicable.

\(^c\)CGM: continuous glucose monitoring.

\(^d\)DSMES: diabetes self-management education and support.

Of the 46 clinical non–grant-funded members, 19 (41%; Table 3) found obtaining home blood pressure monitors “moderately,” “very,” or “extremely” difficult, followed by 4 (9%) who perceived obtaining glucometers as “moderately” or “very” difficult and 20 (43%) who perceived obtaining CGM devices as “moderately,” “very,” or “extremely” difficult. Finally, 24% (11/46) of clinical non–grant-funded members perceived referrals to DSMES as “moderately,” “very,” or “extremely” difficult.

Clinical grant-funded members differed from clinical non–grant-funded members in perceived difficulty ordering CGM devices \(X^2=16.4; P=.01; \text{ Table 3}\). No other differences were observed in perceived difficulty obtaining home monitoring devices and referring to DSMES.

### Preferences for Dissemination of Evidence-Based Materials

Finally, both grant- and non–grant-funded members were asked about their preferences for the delivery modality of evidence-based best practice materials. The Cardi-OH best
practices team produces a variety of materials, including a monthly newsletter, capsules (1-page summaries of best practices ready for implementation in clinical care), currents (half-page summaries of recent peer-reviewed articles describing the latest advances in medicine), best practice documents (web-based tools and resources), Cardi-OH ECHO clinic didactic recordings, and podcasts. All Cardi-OH members were asked to select their top 3 delivery modalities; thus, the percentages exceed 100% (Table 4).

Table 4. Ohio Cardiovascular and Diabetes Health Collaborative (Cardi-OH) grant- and non–grant-funded members’ preferences for the delivery modality of evidence-based best practice materials.

<table>
<thead>
<tr>
<th>Question: from the list of Cardi-OH materials, SELECT UP TO THREE that are of most interest to you</th>
<th>Grant-funded members (n=103), n (%)</th>
<th>Non–grant-funded members (n=98), n (%)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newsletters (monthly updates highlighting new best practice content and other timely information about the collaborative, including links to web documents, capsules, and currents)</td>
<td>72 (69.9)</td>
<td>57 (58.2)</td>
<td>3 (1)</td>
<td>.08</td>
</tr>
<tr>
<td>Capsules (brief 1-page summaries of best practices that are ready to be implemented in clinical care)</td>
<td>68 (66)</td>
<td>47 (48)</td>
<td>6.6 (1)</td>
<td>.01</td>
</tr>
<tr>
<td>Currents (brief half-page summaries of recent articles describing the latest advances in medicine or clinical practice related to cardiovascular health)</td>
<td>49 (47.6)</td>
<td>42 (42.9)</td>
<td>0.4 (1)</td>
<td>.50</td>
</tr>
<tr>
<td>Best practice documents (web-based tools and resources to help clinicians aid patients in managing cardiovascular health)</td>
<td>49 (47.6)</td>
<td>65 (66.3)</td>
<td>7.2 (1)</td>
<td>.007</td>
</tr>
<tr>
<td>Cardi-OH ECHO&lt;sup&gt;a&lt;/sup&gt; clinic didactic recordings (brief videos of presentations to improve content knowledge and share evidence-based best practices)</td>
<td>17 (16.5)</td>
<td>24 (24.5)</td>
<td>1.9 (1)</td>
<td>.16</td>
</tr>
<tr>
<td>Podcasts (audio recordings highlighting national, state, and local leaders discussing timely topics for primary care clinicians)</td>
<td>35 (34)</td>
<td>24 (24.5)</td>
<td>2.1 (1)</td>
<td>.14</td>
</tr>
</tbody>
</table>

<sup>a</sup>ECHO: Extension for Community Healthcare Outcomes.

Grant-funded members preferred these materials in the following order: newsletters (72/103, 69.9%), capsules (68/103, 66%), currents (49/103, 47.6%), best practice documents (49/103, 47.6%), podcasts (35/103, 34%), and ECHO didactic recordings (17/103, 16.5%). Non–grant-funded member preferences were as follows: best practice documents (65/98, 66%), newsletters (57/98, 58%), capsules (47/98, 48%), currents (42/98, 43%), podcasts (24/98, 24%), and ECHO didactic recordings (24/98, 24%).

Grant-funded members preferred capsules more than non–grant-funded members (68/103, 66% vs 47/98, 48%; χ² = 6.7; P=.01), and non–grant-funded members preferred best practice documents more than grant-funded members (65/98, 66% vs 49/103, 47.6%; χ² = 6.7; P=.01). No other differences were observed in preferences for delivery modality.

Best Practice Events and Materials

Events and materials were created based on the following top clinical topics: lifestyle prescriptions (website document: Implementing Lifestyle Prescriptions in Primary Care), CVD and mental health (website document: Mental Health and Chronic Conditions: Treating the Whole Patient to Improve Self-Care), and alcohol and CVD (podcast: Addressing Unhealthy Alcohol Use: Strategies for Primary Care). The top SDoH topics included weight bias and stigma (ECHO didactic recording: Obesity: Bias and Discrimination), adverse childhood experiences (capsule: Adverse Childhood Experiences and Cardiovascular Disease Risk), and family-focused interventions (capsule [Tips to Improve Family Support for Heart-Healthy Living] and website document [Family Support as a Key Component of Cardiovascular Disease Prevention and Care]).

To address preferences for dissemination, these topics were presented in a variety of modalities. Additional events and materials were created to cover the following topics that were rated as being of moderate interest in the needs assessment: COVID-19 (statewide webinar), sleep disorders (statewide webinar and podcast), smoking cessation (capsule), atypical diabetes (capsule), supplements (website document and podcast), the role of clinical pharmacists (website document), lipids and statin use (podcast), complications (current), and disability (capsule and podcast). Importantly, all the events and materials are publicly available on the Cardi-OH website.

Discussion

Principal Findings

In this cross-sectional, descriptive needs assessment, we surveyed clinical and nonclinical grant- and non–grant-funded members of the Cardi-OH community. The purpose of the needs assessment was to identify high-priority topics for the dissemination of evidence-based best practices to Ohio’s health care providers. We also evaluated our clinical members’ screening practices regarding critical topic areas important to health equity and CVD care, perceived difficulties obtaining home monitoring devices and referrals to DSMES as Medicaid payers were actively working to address these barriers, and preferences for the dissemination of evidence-based best practice materials. Overall, both clinical grant- and non–grant-funded members prioritized the following CVD-related topics: lifestyle prescriptions, CVD and mental health, and alcohol and CVD. For SDoH-related topics, clinical grant- and non–grant-funded members were asked to select their top 3 delivery modalities; thus, the percentages exceed 100% (Table 4).
members prioritized family-focused interventions and adverse childhood experiences. Regarding screening for SDoHs, half of clinical grant- and non–grant-funded members felt “very” or “extremely” comfortable screening for SDoHs, and they used a variety of modalities to screen. In addition, 68% (36/53) of clinical grant-funded members and 76% (35/46) of clinical non–grant-funded members screened all patients for depression, with most screening at every visit. Regarding substance abuse screening, 23% (12/53) of clinical grant-funded members and 28% (13/46) of clinical non–grant-funded members screened at every visit. In all screening categories, connecting with resources to address these areas was challenging. When asked about difficulty with ordering home monitoring devices, clinical grant- and non–grant-funded members perceived ordering home blood pressure monitors (16/53, 30% vs 19/46, 41%, respectively) and CGMs (36/53, 68% vs 20/46, 43%, respectively) as “moderately,” “very,” or “extremely” difficult. Finally, Cardi-OH grant- and non–grant-funded members’ top 3 delivery modalities for evidence-based materials were newsletters, capsules, and best practice documents. Interestingly, grant-funded members were more likely to prefer capsules, and non–grant-funded members were more likely to prefer best practice documents.

Comparison With Prior Work

Dissemination and implementation of evidence-based best practices for CVD and diabetes are critical to address the leading cause of morbidity and mortality worldwide. Best practices in CVD care include a wide range of factors such as preventive management [14,15]. Needs assessments help organizations identify problems, gaps in programming or content, and strategies to prioritize resources. Our needs assessment identified specific needs and gaps for each Cardi-OH team to inform their development of evidence-based best practice events and materials for the following year. For example, team best practices used the findings to create a website document on lifestyle prescriptions and a capsule on adverse childhood experiences. Because medicine is an information-based science and clinical care requires frequent information seeking, understanding the needs and preferences of providers is essential [16]. Preferences for dissemination were noted, and prioritized topics were covered in website documents and capsules, 2 of the most preferred delivery modalities, although we often try to use multiple modalities due to differences in learning styles and preferences. These preferences align with previous research examining health care providers’ information needs. Specifically, accessible web-based resources and summaries that synthesize evidence-based materials facilitate providers’ information-seeking behaviors [16]. Thus, asking what Cardi-OH members prefer regarding both topics and dissemination modalities may lead to increased engagement with the collaborative. Over time, we anticipate that increased engagement with the collaborative will increase the dissemination of evidence-based best practices to health care providers across the state. Increased dissemination of best practices improves the quality of care [17], standardizes care across providers and settings [18], reduces health expenditures [19], and increases efficiency in health care [20]. Best practices achieve these outcomes by reducing complications, decreasing hospitalizations, and preventing mortality [21,22].

Health collaboratives such as Cardi-OH are partnerships between health care providers, health care organizations, academic institutions, health plans and insurers, public health, government agencies, and other public and private stakeholders that work together to improve the Quadruple Aim: (1) improving the health of the population, (2) improving the patient experience, (3) reducing costs, and (4) improving care team well-being. Health collaboratives achieve this aim through collective learning, shared data, and diffusion of innovation [23,24]. Importantly, health collaboratives can take many forms and focus on different issues. For example, they may focus on a specific health condition, such as Cardi-OH’s focus on CVD and diabetes, or on care delivery or patient safety. Importantly, the success of a health collaborative hinges on members working together to leverage their collective expertise, resources, and authority to address complex challenges in health care. The key to leveraging expertise, resources, and authority is identifying the priorities of a collaborative and determining the group’s capacity to address the needs of the population being served. One way to identify a collaborative’s strengths, challenges, and priorities is through an annual needs assessment [25]. Importantly, annual needs assessments should be updated to capture context changes, needs, and priorities. Finally, needs assessments can build leadership and group cohesion and facilitate community involvement with health collaboratives [25].

The findings of our needs assessment can be used as a guide or template for other health collaboratives to identify high-priority educational topics in their region. We outline an approach that involves engaging with both clinical and nonclinical stakeholders to ensure that our collaborative’s priorities are aligned with our state’s needs. Similarly, our prioritized topics may be of interest to other organizations and primary care providers given the high prevalence of CVD and type 2 diabetes in the United States and worldwide. Furthermore, the events and materials we produce are publicly available to anyone at no cost. Sharing resources such as those created by Cardi-OH promotes additional partnerships to leverage the expertise and capabilities of multiple organizations.

Limitations

Limitations include the sample size, limited use of inferential statistics, participant self-selection, self-reported data, and lack of patient perceptions. The response rate among Cardi-OH grant-funded members was high (103/117, 88%); however, among non–grant-funded members, it was only 8.14% (98/1204). The low response rate among non–grant-funded members introduced a bias between responders and nonresponders that we could not control for in the analysis. Future work should include efforts to increase the response rate of non–grant-funded members. Second, the self-reported findings may be susceptible to selection and social desirability biases. To minimize bias, the researchers informed participants that their responses were confidential. Furthermore, the researchers emphasized the voluntary nature of participation and explicitly informed the respondents that their responses had
Acknowledgments
Funding for Cardi-OH was provided by the Ohio Department of Medicaid’s Medicaid Technical Assistance and Policy Program, with supplemental funding (cost share) from the universities that Ohio’s 7 schools of medicine are part of—Case Western Reserve University, Northeast Ohio Medical University, Ohio University, Ohio State University, University of Cincinnati, University of Toledo, and Wright State University—as well as their associated affiliates: MetroHealth Medical Center, University Hospitals Cleveland Medical Center, and Ohio State University Wexner Medical Center. The authors gratefully acknowledge the administrative support of the Ohio Colleges of Medicine Government Resource Center. The views expressed in this manuscript are solely those of the authors and do not represent the views of the State of Ohio or any federal Medicaid program.

Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
EAB, SKG, JW, KD, HP, RGK, KB, MWK, and SDB made substantial contributions to conception and design, acquisition of data, data analysis, and interpretation of the data; drafted the manuscript; revised the manuscript critically for important intellectual content; and gave final approval to the version of the manuscript to be published. EAB, SKG, JW, KD, HP, RGK, KB, MWK, and SDB agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest
KD has served as a consultant for Eli Lilly, Dexcom, Insulet, and Oppenheimer, and research support from Dexcom, Insulet, Abbott, Sanofi, ViaCyte Honoraria: Uptodate, Elsevier, Cardiometabolic Health Congress, Med Learning Group, Medscape” to the disclosures. JW has served as a consultant at Medtronic and conducted research with the National Institute on Minority Health and Health Disparities (NIMHD), Ohio Department of Medicaid, and Agency for Healthcare Research and Quality (AHRQ). All other authors have nothing to report.

Multimedia Appendix 1
Ohio Cardiovascular and Diabetes Health Collaborative members’ top-rated cardiovascular disease (CVD)–related topics by clinical and nonclinical grant- (n=103) and non–grant-funded (n=98) members.

Multimedia Appendix 2
Ohio Cardiovascular and Diabetes Health Collaborative members’ top-rated social determinant of health–related topics by clinical and nonclinical grant- (n=103) and non–grant-funded (n=98) members.

References


**Abbreviations**

- **Cardi-OH**: Ohio Cardiovascular and Diabetes Health Collaborative
- **CGM**: continuous glucose monitoring
- **CVD**: cardiovascular disease
- **DSMES**: diabetes self-management education and support
- **ECHO**: Extension for Community Healthcare Outcomes
- **ODM**: Ohio Department of Medicaid
- **REDCap**: Research Electronic Data Capture
- **SDoH**: social determinant of health

©Elizabeth A Beverly, Sarah Koopman-Gonzalez, Jackson Wright, Kathleen Dungan, Harini Pallerla, Rose Gubitosi-Klug, Kristin Baughman, Michael W Konstan, Shari D Bolen. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Integrating Explainable Machine Learning in Clinical Decision Support Systems: Study Involving a Modified Design Thinking Approach

Abstract

**Background:** Though there has been considerable effort to implement machine learning (ML) methods for health care, clinical implementation has lagged. Incorporating explainable machine learning (XML) methods through the development of a decision support tool using a design thinking approach is expected to lead to greater uptake of such tools.

**Objective:** This work aimed to explore how constant engagement of clinician end users can address the lack of adoption of ML tools in clinical contexts due to their lack of transparency and address challenges related to presenting explainability in a decision support interface.

**Methods:** We used a design thinking approach augmented with additional theoretical frameworks to provide more robust approaches to different phases of design. In particular, in the problem definition phase, we incorporated the nonadoption, abandonment, scale-up, spread, and sustainability of technology in health care (NASSS) framework to assess these aspects in a health care network. This process helped focus on the development of a prognostic tool that predicted the likelihood of admission to an intensive care ward based on disease severity in chest x-ray images. In the ideate, prototype, and test phases, we incorporated a metric framework to assess physician trust in artificial intelligence (AI) tools. This allowed us to compare physicians' assessments of the domain representation, action ability, and consistency of the tool.

**Results:** Physicians found the design of the prototype elegant, and domain appropriate representation of data was displayed in the tool. They appreciated the simplified explainability overlay, which only displayed the most predictive patches that cumulatively explained 90% of the final admission risk score. Finally, in terms of consistency, physicians unanimously appreciated the capacity to compare multiple x-ray images in the same view. They also appreciated the ability to toggle the explainability overlay so that both options made it easier for them to assess how consistently the tool was identifying elements of the x-ray image they felt would contribute to overall disease severity.

**Conclusions:** The adopted approach is situated in an evolving space concerned with incorporating XML or AI technologies into health care software. We addressed the alignment of AI as it relates to clinician trust, describing an approach to wire framing and prototyping, which incorporates the use of a theoretical framework for trust in the design process itself. Moreover, we proposed that alignment of AI is dependent upon integration of end users throughout the larger design process. Our work shows the importance and value of engaging end users prior to tool development. We believe that the described approach is a unique and
valuable contribution that outlines a direction for ML experts, user experience designers, and clinician end users on how to collaborate in the creation of trustworthy and usable XML-based clinical decision support tools.

(JMIR Form Res 2024;8:e50475) doi:10.2196/50475

KEYWORDS
explainable machine learning; XML; design thinking approach; NASSS framework; clinical decision support; clinician engagement; clinician-facing interface; clinician trust in machine learning; COVID-19; chest x-ray; severity prediction

Introduction

Though much research has been published on the applications of machine learning (ML) in clinical contexts, few studies have proceeded to deployment for patient care [1]. Barriers to adoption in health care include data quality; data bias; and lack of proper validation, reproducibility, and transparency [2]. Particularly with respect to transparency, black box models, which are increasingly used for prediction tasks in clinical contexts, do not provide the rationale behind the prediction in order to justify a clinical decision [3,4]. In fact, studies found that “physicians need to understand artificial intelligence (AI) methods and systems sufficiently to be able to trust an algorithm’s predictions—or know how to assess the trustworthiness and value of an algorithm—as a foundation for clinical recommendations” [5].

Explainable machine learning (XML) is a field focused on developing techniques to help end users understand the predictions made by complex models [6]. Indeed, we followed Rudin [7] in adopting the definition of XML as the use of additional post-hoc models to explain a primary black-box model. Such black-box models are in contrast to interpretable models. This includes information concerning the underlying data and performance of the model [8]. However, the effectiveness of various approaches for explainability are dependent upon well-designed and highly usable user interfaces [9], and Abdul et al [10] pointed out that much of the work within the domains of AI and ML has not focused on usability or practical interpretability. Indeed, as Liao et al [8] discussed, current work provides limited guidance on actualizing guidelines in user interfaces.

As discussed by Schwartz et al [11], clinician involvement in the design of ML clinical decision support has primarily been used to validate the clinical accuracy of underlying models developed by the researchers. A recent review by Chen et al [12] of explainable AI and ML medical imaging design found no evidence of end-user clinical involvement in the design of explainability models and a highly limited number of articles that documented an empirical assessment of explainability claims with end users. These findings mirror our previous unpublished work that looked at the broader state of XML in clinical decision support and the same low engagement of end users in the empirical assessment of XML decision support applications.

Our study used a design thinking [13,14] approach to explore how constant engagement of clinician end users could provide insights on how to improve the alignment of XML decision support to actual end-user needs and address challenges related to presenting explainability in a decision support interface. To this end, we identified a relevant ML decision support tool targeted toward COVID-19 via clinician focus groups. We then developed a clinician-facing interface for the quantification of COVID-19 severity from chest x-ray images with XML. We tested the resulting prototype via structured interviews with clinicians to verify the domain-appropriate representation, potential actionability, and consistency of the tool.

Methods

Ethical Considerations

Ethical approval for this study was granted by the Centre intégré universitaire de santé et de services sociaux (CIUSSS) of West Central Montreal psychosocial research ethics committee (Project 2022-2838) and by the NRC research ethics board (Project 2021-101). Informed consent was received from all participants. In all analysis and research documents, participant-identifying data were replaced by a code. No compensation was offered to any participants of the study.

Design Thinking Approach

We chose a design thinking approach to optimize clinician involvement in the creation of an XML-based clinical decision support system (CDSS). Design thinking is a process for solving complex problems that emphasizes iteration and rapid prototyping to maximize end-user involvement in generating a usable solution. Stanford University Design School describes 5 key phases of the design thinking approach, namely, empathize, define, ideate, prototype, and test. Table 1 provides an overview of the work presented in this manuscript according to design thinking phases. For each phase, we define the objective, associated research activities, end-user involvement, and supplementary theoretical frameworks used to add robustness to our work.

To simplify the structure of the paper, we have chosen to report the majority of research activities conducted in the empathize, define, and ideate phases in the Methods section. The Results section is primarily focused on the outputs of the prototype and test phases.
Table 1. Overview of the design thinking phases and research activities conducted during each phase.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Empathize</th>
<th>Define</th>
<th>Ideate</th>
<th>Prototype</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase objective</td>
<td>Consult experts to better understand the design challenge, and engage and empathize to understand motivations and experiences</td>
<td>Analyze and synthesize observations to identify and define core problems</td>
<td>Brainstorm approaches to achieve solutions</td>
<td>Production of a scaled-down version to iterate different solution ideas with users</td>
<td>Testing of the best solutions</td>
</tr>
<tr>
<td>Research activities</td>
<td>Rapid review, focus groups, and scoping review</td>
<td>Data synthesis, analysis, and decision on tool function</td>
<td>Iterative design of the tool user experience; Identification of an appropriate explainability approach</td>
<td>Paper prototype testing; Implementation and testing of different explainability approaches; Development of an interactive working prototype</td>
<td>Software prototype testing; Analysis of results</td>
</tr>
<tr>
<td>Physician end-user involvement</td>
<td>Share opinions and experiences in focus groups</td>
<td>Select the potentially most useful tool; Prioritize features</td>
<td>Consult on early design concepts</td>
<td>Provide iterative feedback on paper prototypes</td>
<td>Formally assess a working prototype</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.

<sup>b</sup>NASSS: nonadoption, abandonment, scale-up, spread, and sustainability of technology in healthcare.

**Empathize Phase**

The objective of the empathize phase was to better understand the motivation and experiences of potential end users and to consult with experts on the problem in question. In this phase, we conducted three key research activities: (1) a rapid review to identify clinical use cases for ML or AI that could benefit from explainability and be useful to clinicians in the context of the COVID-19 pandemic; (2) focus groups with physicians designed to review the output of the rapid review and to elicit data that would help the team better understand the scope and nature of a tool that would be most useful to physicians in an integrated health care network, responding to the COVID-19 pandemic; and (3) a scoping review to better understand the existing XML CDSS in health care, the associated design frameworks used for explainability, and the research methods used to study end-user perceptions.

During our rapid review, we searched Scopus, the World Health Organization (WHO) COVID-19 publication database, and the Dialog Proquest COVID-19 database to identify systematic reviews, literature reviews, or surveys of AI or ML technologies used to support clinicians in a pandemic (COVID-19). We identified 65 review articles, of which the 7 most pertinent were used to separate the cited papers within the reviews into 8 broad categories of applications. We selected 4 of these categories to present to stakeholders as candidate applications based on assessment of their clinical need, applicability to hospital settings, relevance to our current research field, interest to clinicians, and feasibility in the chosen clinical setting. The selected categories were large-scale COVID-19 screening; detection, diagnosis, or prognosis of COVID-19; predicting recovery, mortality, or severity of COVID-19 patients; and hospital resource management.

After performing a nonexhaustive scan of additional publications that fit these categories, we retained a total of 37 articles that were peer reviewed and that described ML implementations considering the following criteria: techniques where explainability would be beneficial and associated data or codes were available for implementation. These articles were explored to further select 3 themes, each with its own clinical use case for an ML application, which cross-cut the previously described categories. The first theme was screening. It involved algorithms for tools that may help with COVID-19 screening by predicting the risk of COVID-19 in undiagnosed patients through the analysis of text-based telehealth notes or triage notes in the emergency room. The second theme was prognosis. It involved algorithms for tools that may help predict the severity of COVID-19 infections and the prognosis and risk of intensive care unit (ICU) admission through the analysis of chest x-ray images. The third theme was long COVID. It involved algorithms for tools that may predict the likelihood of long-term implications (long COVID) resulting from COVID-19 through patient-reported outcomes.

Two focus groups were conducted with 7 physicians to identify which of the 3 use cases were suitable for use in clinical decision support. Participants represented a broad range of medical specialties and had experience providing COVID-19–related care in a variety of venues (Textbox 1).
Textbox 1. Physician representation in focus groups.

Medical subspecialties
- Emergency medicine
- Intensive care
- Palliative care
- Cardiology
- Family medicine
- Diagnostic medicine
- Internal medicine

COVID-19 care venues
- Long-term care facilities
- Family medicine centers
- Emergency departments
- Intensive care units
- COVID-19 acute care wards

For each category of possible tools (screening, prognosis, and long COVID), we used the following interview guide questions to seed the discussions: (1) How might a clinical decision support tool focused on \(\text{[insert tool type]}\) be useful in the context of our health care sites? (additional prompts: Could you describe what you see as the value of this type of tool for clinicians [doctors, nurses, and others]? Could you describe what you see as the value of this type of tool for patients?); (2) If we assume that we can access the required data to make the tool work, what additional challenges might a tool like this be associated with? (additional prompts: Any specialized additional clinical knowledge needed? Would it require dramatic changes to existing care protocols or workflows? Any special characteristic of our patient population?); and (3) What types of information or data points would be most crucial in an explanation of the prediction being discussed?

All data from the focus groups were transcribed and loaded into NVivo software (QSR International) for thematic coding.

In parallel, a scoping review of the use of XML for decision support in health care was conducted, using the methods proposed by Levac et al [16]. We generally found very few studies that described testing or methods to collect end-user perceptions of explainability, and even fewer studies that referenced any design theory or framework in the development of decision support tools.

Define Phase

The objective of the define phase was to synthesize the findings from the work done in the empathize phase and formally define the scope of the problem. For focus group data, we applied a framework developed to study the nonadoption, abandonment, scale-up, spread, and sustainability of technology in health care (NASSS), to analyze physician feedback on the possible tools for development. The NASSS framework is composed of the following 7 domains: condition, technology, value proposition, adopters, organizations, wider system, and embedding over time [15]. The NASSS framework, while traditionally used to analyze technology implementations, can be used to “generate a rich and situated narrative of the multiple influences on a complex project” [17] and assess in advance whether certain technology will be adopted in a health care setting. According to Fereday and Muir-Cochrane [18], a hybrid deductive or inductive thematic analysis was used. An initial deductive coding framework based a priori on the domains of the NASSS technology implementation framework was completed by one of the researchers. During the coding, 2 members of the research team met frequently to review challenges in the coding process and identify new subcodes for each domain. A detailed summary of the NASSS framework domains and sample coded participant data are provided in Multimedia Appendix 1. The objective of this phase was to better understand the suitability of each tool for adoption within the organization. We present a brief summary of the analysis in Table 2.

Physicians felt that the characteristics of long COVID were still unclear and difficult to define, and thus, it would be inappropriate to develop a long COVID tool. There was strong disapproval for a screening tool that makes predictions based on analysis of free text in the patient medical records given the possibility of incorrect information, inconsistent completion of records, missing information, and false reporting by patients. Physicians were more receptive to the prognosis tool. Their familiarity with x-ray images and more trust in the image data source increased support for this tool. Physicians not only considered this a more useful application but also considered the warning of the impending prognosis to be important.
<table>
<thead>
<tr>
<th>NASSS domain</th>
<th>Description</th>
<th>Screening tool</th>
<th>Prognosis tool</th>
<th>Long COVID tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition domain</td>
<td>Is the nature of the condition or illness (eg, symptoms, diagnosis, and therapeutics) relevant to the organization?</td>
<td>+/−</td>
<td>+</td>
<td>−/−</td>
</tr>
<tr>
<td>Technology domain</td>
<td>Can the technology of the tool (eg, underlying algorithm) be supported by quality data sources?</td>
<td>−</td>
<td>+</td>
<td>−/−</td>
</tr>
<tr>
<td>Value proposition domain</td>
<td>Is there potential of the tool to provide some type of business or health system value?</td>
<td>+/−</td>
<td>+</td>
<td>+/−</td>
</tr>
<tr>
<td>Adopters domain</td>
<td>Could the use of the tool result in a change in practice of care providers, and impact patients and their careers?</td>
<td>+/−</td>
<td>+</td>
<td>+/−</td>
</tr>
<tr>
<td>Organizations domain</td>
<td>Are there considerations related to readiness of the institution to innovate, use, and fund new technologies?</td>
<td>+/−</td>
<td>+</td>
<td>+/−</td>
</tr>
<tr>
<td>Wider system domain</td>
<td>Are there political, regulatory, or sociocultural considerations impacting the implementation of the tool?</td>
<td>−</td>
<td>+</td>
<td>+/−</td>
</tr>
<tr>
<td>Embedding and adaptation over time domain</td>
<td>Is there potential for the organization to adapt and evolve the tool over time?</td>
<td>−</td>
<td>+</td>
<td>−/−</td>
</tr>
</tbody>
</table>

aNASSS: nonadoption, abandonment, scale-up, spread, and sustainability of technology in health care.

bPositive sentiment from physicians.
cNegative sentiment from physicians.

Ideate Phase

In this phase, the research team met continuously to build various approaches for both the user experience (UX) implementation and underlying explainability approaches used in the tool. As discussed in the Introduction section, increasing clinician trust in ML-based CDSS applications is seen as a key driver to increasing their use in actual practice. In line with this principle, the team adopted an evaluation framework published by Tonekaboni et al [4], which presents a series of metrics that can help assess clinician trust in an explanation provided by a ML CDSS tool. The framework includes 3 metrics. The first metric is domain appropriate representation, which represents the degree to which the tool provides adequate information to the end user within the context of the specific clinical setting and workflow. The second metric is potential actionability, which represents the degree to which the tool facilitates the taking of appropriate decisions or “next steps” in the care of the patient. The third metric is consistency, which represents the degree to which changes in the tool’s predictions and corresponding explanations can be explored to determine consistency.

These metrics were incorporated as design guidelines in the ideate phase and then used as the primary metrics in the prototype and test phases.

Prototype Phase

The objective of the prototype phase was to develop low-cost physical representations of the tool that allow for more detailed end-user feedback and more opportunities to iterate on the design of the solution.

For domain appropriate representation, we focused on providing a succinct summary of additional COVID-19–relevant information that physicians would likely find relevant in the context of a prognostic prediction based solely on chest x-ray images. This included the addition of a subset of patient vitals, laboratory values, and history, including symptom onset.

For potential actionability, we assumed the most important visual component of the tool would be the chest x-ray and corresponding explainability features. The team chose to implement a heat map–based visualization approach that would highlight areas of the image that most significantly contributed to the x-ray severity score. The technical implementation is explained in more detail below.

Our assumption was that this would quickly provide clinicians with the information they needed to make appropriate decisions concerning the next steps of the patient’s care trajectory. In order to provide context to the severity score, we considered a model of ICU admission to assess risk by defining 3 categories of risk (low, medium, and high) and the associated likelihood of admission.

Finally, for consistency, we planned for the clinician to be able to click on multiple imaging results in the patient timeline such that the clinician can compare the ML predictions across images and make assessments as to the consistency of the predictions. The prototype wireframe in Figure 1 illustrates the basic design and different domain considerations.
ML Model: Developing Explainability

Based on the selected application, we chose to use the algorithm of Cohen et al [19] as the algorithm to include in the prototype application. Their work predicts the level of lung opacity and the geographic extent of disease regions from the input x-ray data, using a deep neural network. The main network consists of a network pretrained on large public non–COVID-19 data sets followed by 2 regression networks, one for each of the opacity and extent outputs. We computed total disease severity as the sum of the 2 neural network outputs. Training of regression networks was performed using COVID-19 data obtained by radiologist scoring of chest x-ray images with the following scores: (1) the extent of involvement of ground-glass opacity for each lung (for a total between 0 and 8), and (2) the degree of opacity for each lung (for a total between 0 and 6).

For our purposes, we were able to leverage their latest publicly available implementation and database [20], thus allowing us to only focus on developing the explainability methods. The multisite data set consisted of posteroanterior deidentified chest x-ray images of patients with varying COVID-19 severity, and each x-ray image had an associated disease severity score obtained by radiologists. Many of the patients had x-ray images from several time points, and the number of time points per patient was not consistent across patients.

We aimed to explore local post-hoc explainability approaches as we needed to explain specific instances of pre-existing models. We focused on model-agnostic methods as they would be useful in broader contexts and be applicable independent of the clinicians’ choice. In particular, we selected LIME (local interpretable model-agnostic explanations) [21] in order to interpret the model output owing to the simplicity in its implementation and perceived intuitiveness of the results. In our case, LIME explained the predicted severity as a linear function of the contributions from image patches that compose the full x-ray image. Each image was subdivided into patches using the Quickshift Segmentation algorithm [24]. We found that the resulting explanation was unintuitive upon visual inspection, as patches indicating high contributions to increasing severity did not appear to be diseased. In addition, the low fidelity prototype displayed importance values for all patches and proved too confusing during iterative testing. The patches showing a negative contribution were cluttering the display and were unnecessary to show highly diseased regions. This was addressed in future iterations.

According to the design thinking approach [25], we met with 2 clinicians who had participated in the focus group sessions, and asked them to provide individual feedback during informal sessions of 30 to 60 minutes. We asked for their overall impression of the tool prototype design, as well as comments on specific design choices made by the research team to improve UX and explainability. These sessions were conducted over Microsoft Teams. Follow-up was conducted via email as the team made iterations on the comments and ideas from physicians.

Low Fidelity Prototype

Using our wireframe and the results from the ML model, a static low fidelity prototype of the tool was developed using FIGMA (Figure 2) [22].

We defined 3 levels of the risk of admission to the ICU considering the severity score (low, medium, and high), using the observed ICU admission rates and associated severity scores in our test data set. We defined severity score ranges for each of the risk levels in such a way that similarity in severity was maximized within each range. The risk was then calculated as the average probability of admission within each category. Our data set contained 950 x-ray images representing 472 unique patients. A subset of 398 x-ray images for 162 unique patients with information on ICU stay was used to calculate the risk of admission.

For the explainability component, we used the LIME Python package [23] to explain the predicted severity as a linear function of the contributions from image patches that compose the full x-ray image. Each image was subdivided into patches using the Quickshift Segmentation algorithm [24]. We found that the resulting explanation was unintuitive upon visual inspection, as patches indicating high contributions to increasing severity did not appear to be diseased. In addition, the low fidelity prototype displayed importance values for all patches and proved too confusing during iterative testing. The patches showing a negative contribution were cluttering the display and were unnecessary to show highly diseased regions. This was addressed in future iterations.
Test Phase

The objective of the test phase was to gain more insight into the working prototype. The high fidelity (software) prototype was developed in Python using the Plotly Dash framework [26], a library that allows for quick prototyping of user interfaces.

The interactive prototype was tested with 5 physicians from the initial focus group sessions. According to Doshi-Velez Kim [27], we used an application grounded evaluation approach in which intended end users used the tool simulating prognostic prediction in the context of a 1-hour semistructured interview. We began the sessions by asking clinicians to explore the application with no guidance and then facilitated a discussion about the various features of the tool by presenting 3 separate patient cases. The questions from the interview guide were as follows: (1) What is your overall impression of what the tool is presenting to you? (overall impression); (2) How well does the tool provide with you with necessary contextual information about the case? (domain representation; additional probes: Is there too much information or is there missing information? Is the information poorly organized or is the presentation of the information confusing?); (3) How does the tool help or hinder your ability to make a treatment decision or take action with the patient? (actionability; additional probes: Is the prediction the tool is making clear? Is the tool adequately transparent with regard to the certainty of the prediction? Are there any complimentary data points you feel are missing for you to make a decision? Do you feel this tool could be shared with a patient in its current state?); and (4) As we show different cases, what are your impressions of how changes in the prediction are explained by the tool? (consistency; additional probes: In cases where the tool shows the progression of predictions for a single patient, does the tool adequately explain the changes? In cases where the tool shows a range of different patients, are the differences in predictions adequately explained by the tool?).

Results

Prototype Phase (Low Fidelity)

In evaluating domain appropriate representation, the major information elements required by physicians were present in the first prototype; however, physicians did mention that the following elements needed to be added: (1) A more prevalent display of the date of COVID-19 diagnosis and date of symptom onset; (2) A more prevalent display for vaccination dates if available; (3) The method and volume of current oxygen in L/min; and (4) Additional laboratory test values for procalcitonin, C-reactive protein, and interleukin-6 (indicators of infection or inflammation that could be treated).

As mentioned above in the ideate phase, our objective with this design was to present the prediction in the most significant quadrant of the screen to ensure potential actionability; however, both our clinician testers commented on how presenting the image first violated the basic approach for clinical assessment and decision-making. While we assumed that the additional information displayed in the interface meant to provide patient content would only be looked at if needed, clinicians told us that this would be relevant and would need to be incorporated.
into their assessment of the quality of the ML prediction. Based on this feedback, we chose to reverse the orientation of the information for the development of the interactive prototype.

In addressing the consistency metric, while we explained to the physicians that the interactive tool would ultimately allow them to move through multiple images to compare the progression of the disease, we could not simulate this in the static image. Physicians did comment that it would be much easier if the multiple images could be displayed in comparison. They also commented on the importance of being able to turn the heat map overlay on and off so as to be able to compare the areas highlighted by the ML model with the actually affected areas on the x-ray image. This suggestion was implemented in the interactive prototype.

Finally, physicians commented that it was unclear how the consolidation in the image contributed to the severity score and how the severity score contributed to the risk of ICU admission. These were addressed in the interactive prototype through the creation of tool tip pop-ups with textual explanations.

**Testing Phase (High Fidelity)**

The final interactive prototype was designed based on all feedback identified from the prototyping sessions and is presented in Figure 3.

**Domain Appropriate Representation**

Clinicians found the design of the prototype elegant, and domain appropriate representation of data was displayed in the tool. The added contextual data (history, vitals, and laboratory values) were deemed necessary, and no items were considered superfluous. Clinicians appreciated that the domain of interest was succinctly represented on a single screen. However, one of the most important issues uncovered was the degree to which physicians erroneously assumed that the additional data present in the tool, namely vitals and laboratory values, were being included in the x-ray severity score. While the tool tips made it explicitly clear that no data other than chest x-ray data were being used in the prediction model, further work needed to be done to visually distinguish information elements used to provide additional domain context from information used in the ML model.

Based on initial feedback during the prototype phase, we specifically probed clinicians on the reorganization of the structure and sequencing of the information in the interface, and the degree to which it supported the standard thought processes clinicians would follow to reach a decision. Physicians responded well to the revised structure and information flow. Moreover, they highlighted the possibility of using the tool as an additional teaching resource with junior staff or nonradiology specialists.

**Potential Actionability**

In order to reduce the clutter in the display of explainability compared to the low fidelity version, we only showed the patches that had a positive contribution to severity in red, where higher values had a darker color. In order to allow the end user to focus on the most predictive areas only, we tested several methods for displaying only a subset of the regions, including displaying a fixed number of patches (eg, 3, 4, and 5) or displaying patches that explained a certain percentage of the final score (eg, 80% and 90%). The selected prototype only displayed the most predictive patches that cumulatively explained 90% of the final score. The rationale behind it is that in some cases where the top contribution comes from many...
patches with low and equivalent values, by only showing a fixed number of patches, we will be omitting several equivalently important patches. The remaining explainable model hyperparameters (maximum distance, color space and image space proximity ratio, and kernel size for the Quickshift Segmentation; regularization coefficient for the explainer’s ridge regression; and width of the explainer’s exponential kernel) were then optimized by maximizing the coefficient of determination via sequential optimization using decision trees [28]. This optimization process resulted in more intuitive and actionable explanations.

We further explored potential actionability through the following three possible case scenarios for the tool in the interviews: (1) Triage of patients presenting at the emergency department; (2) Discharge planning; and (3) Shared decision-making with patients.

In the first scenario (triage), many physicians were quick to point out the evolution of care protocols across the different waves of the pandemic. In the first and second waves, the lack of global disease knowledge caused a large number of pre-emptive ICU admissions; however, this is no longer the case. Physicians did note that they felt the tool could be very helpful in terms of planning for potential ICU admissions over time and helping manage staff resource issues.

In the second scenario (discharge planning), we proposed that the tool could be used as a final check for moving patients from high acuity care to lower levels of acuity, either discharge to home or discharge to virtual hospital care. Physicians generally felt that the tool could be useful as an additional data source to confirm an assessment of low risk.

In the third scenario (shared decision-making), we asked physicians whether they thought the tool could be helpful in shared decision-making, especially in scenarios where there may be some disagreement between a physician and a patient about a proposed next step. Physicians noted that the tool could be useful in explaining escalations in care to patients currently experiencing moderate symptoms.

Consistency
In terms of consistency, physicians unanimously appreciated the capacity to compare multiple x-ray images in the same view. They also appreciated the ability to toggle the explainability overlay so that both options made it easier for them to assess how consistently the tool was identifying elements of the x-ray image they felt would contribute to overall disease severity. Not all physicians agreed with the tool’s assessments, but felt that more exposure to a larger number of predictions would be necessary for them to gauge how much they trusted the tool.

Discussion
Designing for Trust and Decision-Making
As discussed by Wang et al [29], when predictive AI is used in decision support tools, end users seek explanations to help improve their decision-making, and in cases where the tool performs in unexpected ways, explanations are critical for allowing users to identify what elements of the underlying model may be contributing to an unexpected prediction. In our case, we used saliency heat maps to show causal attribution to a severity score, where the highlighted regions represented areas with the greatest contribution to the severity score. Physicians appreciated the ability to toggle the heat map on and off to clearly identify the areas of the image that most contributed. However, multiple physicians did note that trust in the application would be built over longer term use, allowing them to assess the degree to which the application would align or deviate from their own unaided clinical assessments.

As described by Wang et al [29], this may be considered a type of heuristic representative bias, whereby past experience can lead a physician to wrongly associate a current case with similar previous cases. While our design allowed physicians to compare multiple instances of chest x-ray images for a single patient, a further iteration could incorporate features that would help to address this heuristic bias. Specifically, we could include the potential to compare an existing case to similar prototype example cases and use a dissimilarity metric to compare cases.

It is also important to note that there was skepticism that a model based solely on chest x-ray images could provide prediction as good as a model based on multiple inputs. This highlights the design challenge of optimizing domain specific representation and potential actionability in the user interface. Clinicians felt that it is important to see the prediction of chest x-ray images in the context of additional clinical information that feeds into their heuristic framework used for assessing a patient’s disease trajectory. Those additional data points clearly played a role in their likelihood to trust the explanation; however, they were not accounted for in our model. Further exploration of this challenge might include comparing the accuracy of the current model to predictions that use additional key inputs.

Design Thinking and Rapid Evolution
The COVID-19 pandemic evolved rapidly, and as such, the constant engagement with end users allowed the team to improve the potential application of the tool as well as the information displayed. During the prototype phase of development, physicians pointed out that the hospital was rapidly starting up a virtual care service for early discharge of COVID-19 patients to be cared for at home. This allowed the team to realign some of the discussion in the testing phase to assess the suitability of the tool for a unique case that did not exist in the ideate phase of development.

Moreover, it allowed us to modify the scope of information displayed in the tool to bring vaccine-related information into the main display. Again, this information only began to become available in the prototype phase of the project.

Finally, we were able to probe physicians around the applicability of the prototype as a shared decision-making tool to be used with patients, which was suggested informally during the ideate and prototype phases.

The above examples illustrate the importance of agility that is integral to the design thinking approach and represent ways in which the potential applicability and design were improved.
which would not have been addressed in a traditional waterfall development approach.

Combining Design Theory With Additional Frameworks for a More Robust Approach

While design theory provides a well-established approach for continuous engagement with end users, we believe our approach of augmenting design thinking by incorporating additional conceptual frameworks helped to create a more robust collaborative tool design.

First, we used the NASSS framework during the “define” phase of the project to systematically analyze the results of our physician focus groups. This approach helped the team to quickly identify how the potential solutions would align with the various subdomains of the model. We see this as a pragmatic approach and helpful augmentation of the design thinking process to ensure the chosen design direction does not face dramatic sociotechnical barriers to development and potential implementation. Recent research into adoption of ML into clinical practice has used the NASSS framework in a similar manner. Pumplun et al [30] used the NASSS framework to identify 13 specific factors influencing the adoption of ML systems and further proposed a maturity model to be used by health care institutions to assess their readiness to adopt ML-based tools.

Similarly, we augmented the ideate, prototype, and test phases of the project by applying the evaluation metrics proposed by Tonekaboni et al [4]. This approach allowed us to begin the ideation and design process focused on core domains that would impact physician trust (domain appropriate representation, potential actionability, and consistency). It provided a consistent lens to assess both the prototype and test versions of the tool. This consistency in approach led the team to quickly identify design-specific improvements that directly led to the production of a prototype our physicians felt could provide immediate value.

Overall, the approach taken in our work can be situated in an evolving space concerned with incorporating AI technologies into health care software. Anderson et al [31] proposed a framework of 5 lenses from which to view this growing research field. Our work makes contributions to 2 of these lenses (AI as alignment with human values and AI as a design process). First, we addressed the alignment of AI as it relates to clinician trust, describing an approach to wire framing and prototyping that incorporates the use of a theoretical framework for trust in the design process itself. We described how this allows to gauge end-user alignment or trust in AI at multiple stages and optimize designs accordingly.

Second, as described in detail throughout this work, we propose that the alignment of AI is dependent upon integration of end users throughout the larger design process. Our work shows the importance and value of engaging end users prior to tool development, specifically in the process of assessing the broader applicability of a potential AI tool and its eventual use within actual health care environments.

Limitations

There are several limitations associated with this work. First, from a ML perspective, though we can verify the intuitiveness of the explanation, the accuracy of explainability methods has not been properly studied to date. We thus do not know how well an explanation fits the true underlying prediction in spite of its level of intuitiveness. This is of particular concern in the case of additive feature attribution methods like LIME, where a local linear model is used to explain a potentially more complex nonlinear underlying model.

Second, we used a publicly available data set with limited data, and thus, there were several implications. For example, it was difficult to find exemplary samples where an explanation can clearly demonstrate why an algorithm deviated from the ground truth or examples that can shed light on why an algorithm may have behaved in an unpredictable way. The data set contained a lot of missing data and was limited beyond imaging information, and as such, it was challenging to find examples with the full patient state (such as vital signs, multiple time points, etc) to provide end users with the desired contextual information to make a fully informed assessment. Finally, the data set was relatively old considering the rapid evolution of COVID-19 and approaches to its treatment, and this has implications on the likelihood of ICU admission considering the state of a patient.

With regard to our focus group participants, it is important to note that only physicians were represented in this research. While this was intentional in the study design, as it is primarily physicians who will make decisions about ICU admission, our work could have benefited from the inclusion of additional health care providers, such as nurses and respiratory therapists. Indeed as Nalin [32] pointed out, a larger system perspective of the use of the proposed tool could have provided richer data in the focus group phases.

Conclusion

Our work set out to use a design thinking approach to develop an XML-based decision support tool to assist clinicians. We augmented the design thinking approach by using the NASSS framework to help inform the development focus and direction, and added a formal evaluation framework from the report by Tonekaboni et al [4] to continuously focus our design on elements that would improve clinician trust in the tool. This research contributes to the body of health care literature that deeply integrates end users into the design and evaluation of XML in clinical decision support tools. As discussed, clinician trust is seen to be one of the key barriers to larger scale adoption of ML-based clinical decision support tools.

We believe that the approach described in our work is a unique and valuable contribution that outlines a direction for ML experts, UX designers, and clinician end users on how to collaborate in the creation of trustworthy and usable XML-based clinical decision support tools.
Acknowledgments
This research was funded by the National Research Council of Canada as part of the COVID-19 Pandemic Challenge Program. We thank Maryam Bafandkar and Yunong Shuang for their contributions to prototype design development. We thank Samira Rahimi, Amrita Sandhu, and Gauri Sharma for their contributions to the scoping review.

Data Availability
Two separate data sets were used in the completion of this work. The data sets used for developing the underlying disease severity and outcome prediction models are available through the repository of Cohen et al [2, 20]. The focus group data generated and analyzed during the study are not publicly available owing to the consent obtained when the data were originally collected.

Authors’ Contributions
MS, RH, and JH designed the study. RH, JH, and FT completed the rapid reviews. MS and VD conducted the focus groups. RH, JH, and MS developed and designed the prototype. MS, RH, JH, and VD wrote the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Description of nonadoption, abandonment, scale-up, spread, and sustainability of technology in health care (NASSS) categories and example physician comments.

References


23. LIME Python package. GitHub. URL: https://github.com/marcotcr/lime/commit/243ba43ee4d6153aebbe271080ef4cb4c8edf6e4 [accessed 2024-03-10]


32. Nalin K. A distributed decision making process from a systems perspective: following 33 patients at an emergency department. ResearchGate. 2017. URL: https://www.researchgate.net/publication/315729386_A_distributed_decision_making_process_from_a_systems_perspective_following_33_patients_at_an_emergency_department [accessed 2024-03-10]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
ICU: intensive care unit
LIME: local interpretable model-agnostic explanations
ML: machine learning
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability of technology in health care
UX: user experience
XML: explainable machine learning
©Michael Shulha, Jordan Hovdebo, Vinita D’Souza, Francis Thibault, Rola Harmouche. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Recruitment Strategies in the Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization Study (meSH): Multicenter Survey Study

Chinonyelum Nwosu1, BS, MPH‡; Hamda Khan1, MA; Rita Masese2, MSc, MD; Judith M Nocek3*, PhD; Siera Gollan4, BS, MS, PhD; Taniya Varughese5, OTR/L, MSOT; Sarah Bourne6, MPH; Cindy Clesca7, MA; Sara R Jacobs8, PhD; Ana Baumann5, PhD; Lisa M Klesges5, MS, PhD; Nirmish Shah2, MD; Jane S Hankins9*, MS, MD; Matthew P Smeltzer10*, MStat, PhD

1Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN, United States
2Duke University, Durham, NC, United States
3University of Illinois at Chicago, Chicago, IL, United States
4Augusta University, Augusta, GA, United States
5Washington University School of Medicine, St Louis, MO, United States
6Medical University of South Carolina, Charleston, SC, United States
7Icahn School of Medicine at Mount Sinai, New York, NY, United States
8RTI International, Research Triangle Park, NC, United States
9Department of Global Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN, United States
10The University of Memphis, Memphis, TN, United States
‡Sickle Cell Disease Implementation Consortium
*these authors contributed equally

Corresponding Author:
Matthew P Smeltzer, MStat, PhD
The University of Memphis
222 Robison Hall
Memphis, TN, 38152
United States
Phone: 1 901 678 3702
Email: msmltzer@memphis.edu

Abstract

Background: Hydroxyurea is an evidence-based disease-modifying therapy for sickle cell disease (SCD) but is underutilized. The Integration of Mobile Health into Sickle Cell Disease Care to Increase Hydroxyurea Utilization (meSH) multicenter study leveraged mHealth to deliver targeted interventions to patients and providers. SCD studies often underenroll; and recruitment strategies in the SCD population are not widely studied. Unanticipated events can negatively impact enrollment, making it important to study strategies that ensure adequate study accrual.

Objective: The goal of this study was to evaluate enrollment barriers and the impact of modified recruitment strategies among patients and providers in the meSH study in response to a global emergency.

Methods: Recruitment was anticipated to last 2 months for providers and 6 months for patients. The recruitment strategies used with patients and providers, new recruitment strategies, and recruitment rates were captured and compared. To document recruitment adaptations and their reasons, study staff responsible for recruitment completed an open-ended 9-item questionnaire eliciting challenges to recruitment and strategies used. Themes were extrapolated using thematic content analysis.

Results: Total enrollment across the 7 sites included 89 providers and 293 patients. The study acceptance rate was 85.5% (382/447) for both patients and providers. The reasons patients declined participation were most frequently a lack of time and interest in research, while providers mostly declined because of self-perceived high levels of SCD expertise, believing they did not need the intervention. Initially, recruitment involved an in-person invitation to participate during clinic visits (patients), staff meetings (providers), or within the office (providers). We identified several important recruitment challenges, including (1) lack of interest in research, (2) lack of human resources, (3) unavailable physical space for recruitment activities, and (4) lack of...
documentation to verify eligibility. Adaptive strategies were crucial to alleviate enrollment disruptions due to the COVID-19 pandemic. These included remote approaching and consenting (eg, telehealth, email, and telephone) for patients and providers. Additionally, for patients, recruitment was enriched by simplification of enrollment procedures (eg, directly approaching patients without a referral from the provider) and a multitouch method (ie, warm introductions with flyers, texts, and patient portal messages). We found that patient recruitment rates were similar between in-person and adapted (virtual with multitouch) approaches (167/200, 83.5% and 126/143, 88.1%, respectively; P=.23). However, for providers, recruitment was significantly higher for in-person vs remote recruitment (48/50, 96% and 41/54, 76%, respectively, P<.001).

**Conclusions:** We found that timely adaptation in recruitment strategies secured high recruitment rates using an assortment of enriched remote recruitment strategies. Flexibility in approach and reducing the burden of enrollment procedures for participants aided enrollment. It is important to continue identifying effective recruitment strategies in studies involving patients with SCD and their providers and the impact and navigation of recruitment challenges.

**Trial Registration:** ClinicalTrials.Gov NCT03380351; https://clinicaltrials.gov/study/NCT03380351

**International Registered Report Identifier (IRRID):** RR2-10.2196/16319

**(JMIR Form Res 2024;8:e48767)** doi:10.2196/48767

**KEYWORDS**
sickle cell; recruitment; eHealth; multicenter; utilization; strategy; hydroxyurea; mobile health; mhealth; intervention

**Introduction**

Sickle cell disease (SCD) is an inherited blood disorder that affects approximately 100,000 individuals in the United States [1]. SCD causes debilitating health issues, including progressive organ damage, acute unplanned severe vaso-occlusive crises (VOCs), and premature death [2]. Treatment with hydroxyurea increases fetal hemoglobin and has other salutary benefits (eg, reduced inflammation) that together function as protection against organ damage and VOCs and increase the rate of patient survival [3]. Despite the proven efficacy and effectiveness of hydroxyurea as a prophylactic treatment in SCD, it is vastly underutilized, a result of a combination of poor prescribing practices among providers and low adherence behavior among patients [4].

In 2016, the National Heart, Lung, and Blood Institute established the Sickle Cell Disease Implementation Consortium (SCDIC), which was tasked with conducting implementation research and interventional studies tailored toward improving the quality of SCD care [5]. One of the interventional studies, the Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization (meSH) study, addressed the barriers to hydroxyurea uptake at both the provider and patient levels by targeting the multilevel determinants of poor hydroxyurea utilization [6,7]. In meSH, the targeted interventions were delivered using 2 mobile health (mHealth) apps: the patient app, InCharge Health [8], and the provider app, HU Toolbox [8]. In March 2020, only 4 months after the study began recruiting providers and 2 months after the initiation of patient recruitment, the global COVID-19 pandemic emerged in the United States. In response to the pandemic, institutions were forced to modify clinical and research procedures by temporarily suspending all nonessential in-person contact [9]. This rapid procedural change directly impacted all in-person research activities in meSH, which was initially designed to conduct all recruitment in person. The meSH study team quickly adapted recruitment strategies for the ongoing study, eventually meeting enrollment goals.

In general, SCD studies suffer from underrecruitment [10]. SCD disproportionately affects African-Americans, making it imperative that recruitment is intentional. When targeting a historically underrepresented group, there are some recruitment strategies that have been shown to be highly effective, such as having a champion of the study on site to promote recruitment, employing culturally matched staff, using flyers with images that represent and resonate with the group, and providing culturally relevant incentives [11]. These intentional adjustments to the recruitment process help to address some of the common barriers to research for historically underrepresented groups. Recruitment targets must be met to ensure sufficient power to answer research questions, and recruitment strategies can help ensure this goal. However, recruitment strategies in the SCD population are not widely studied [2]. The goal of this study was to evaluate enrollment barriers faced by the meSH study and the impact of modified recruitment strategies for patients and providers in response to a global emergency.

**Methods**

**Recruitment Strategies for the meSH Study**

The meSH study used the InCharge Health mobile app in an implementation study with three aims: (1) improve patient adherence to hydroxyurea, (2) improve provider hydroxyurea prescribing behaviors, and (3) evaluate barriers and facilitators to implementation of mobile health interventions [6,7]. The meSH study design used a staggered site initiation across 7 sites, with the first 2 sites beginning the enrollment process in September 2019 for providers and in November 2019 for patients. The SCDIC consortium includes clinical sites in academic and nonacademic settings and in both rural and urban settings. Development of this cohort used implementation science principles previously described in detail (ClinicalTrials.Gov NCT03380351) [12].

For our evaluation of meSH recruitment strategies, the eligibility criterion was functioning as the lead study coordinator for a site (all lead study coordinators agreed to participate). For the meSH study itself, the eligibility criteria for providers included being
a physician (including a physician in training) or advanced practice provider (ie, nurse practitioner or physician assistant) who cared for at least 1 patient with SCD for an anticipated minimum of 12 months from study enrollment and had access to a smartphone (either Android or iOS) or a computer with internet connectivity (the HU Toolbox app can be accessed via the internet on any device) [6,7]. Eligibility criteria for patients included being aged 15 to 45 years; receiving treatment at or being affiliated with one of the SCDIC sites; speaking English; having a confirmed SCD diagnosis by a hemoglobin fractionation test; owning a smartphone (either Android or iOS); and being treated with hydroxyurea therapy, defined as already receiving hydroxyurea therapy (at least 1 previous prescription for hydroxyurea in the past 3 months) or initiating hydroxyurea therapy (the first prescription must have been written at study enrollment) [6,7].

Enrollment of patients was initially planned to occur in person during routine clinic visits. Enrollment of providers was planned in person, to occur during staff meetings, at clinics, or in the providers’ offices. The method of enrollment, modification of the enrollment method, and refusal reasons were tracked prospectively. Ineligibility was documented during the screening process and captured via CONSORT forms, and reasons for declining were collected by site study coordinators. REDCap (Research Electronic Data Capture; Vanderbilt University), a secure electronic database, was used to administer and store the surveys, survey data, and consent forms.

To better understand the barriers and adaptations to study recruitment, we surveyed the primary research coordinator responsible for recruitment at each of the 7 participating sites with 9 open-ended questions regarding recruitment strategies, adaptations to enrollment methods, and challenges for enrolling patients and providers (Multimedia Appendix 1). The interview guides included were open-ended questions developed to elicit (1) methods used for recruitment, specifically the details of modifications to the original recruitment process listed in the protocol, such as adaptations or deviations from the in-person recruitment process originally planned; (2) temporal variations of each recruitment modality (ie, how long each recruitment strategy was used and when it was changed), (3) reasons for modification of recruitment strategies; and (4) the perception of success for each recruitment strategy. Responses were deidentified and tabulated, and a thematic content analysis was conducted. Responses were first organized inductively by the main domains they addressed and then categorized based on the identified thematic group, allowing for the organic emergence of themes.

We summarized provider and patient participant characteristics (as frequencies and percentages) overall and by site to illustrate site-level differences. Statistical comparisons were made using chi-square tests with an α level of .05. All analyses were conducted using SAS (version 9.4; SAS Institute).

**Ethical Considerations**

For deidentified participant data access, the original data for this study will be made available at BioLINCC [13]. The meSH study attained institutional review board (IRB) approval at all 7 participating sites (St. Jude Children’s Research Hospital, 19-0159; Augusta University IRB, 1528751; UIC IRB, 2020-0087; Duke University IRB, Pro00073506; Mount Sinai IRB, STUDY-20-00303; Medical University of South Carolina [MUSC] IRB, Pro00097832; Washington University at St. Louis IRB, 202003150). All patient participants (or their legal guardians) signed informed consent, and all provider participants gave verbal assent before any study procedure.

**Results**

### Provider Recruitment and Adapted Recruitment Strategies

All eligible providers in all 7 sites were approached to participate. Of the 104 approached, 89 enrolled, an 86% study acceptance rate. Of the 89 providers enrolled, most were female (65/89, 73%), aged 25 to 44 years (53/89, 60%), White (47/89, 53%), and not Hispanic or Latino (84/89, 94%); most considered themselves SCD specialists [12] (54/89, 61%) (Table 1). The provider recruitment and enrollment duration ranged from 2 to 7 months across all 7 sites. Initially, all sites planned to use an in-person recruitment approach. Some opted to introduce the study via a presentation at regularly scheduled staff meetings, followed by in-person individual provider enrollment. With the onset of the COVID-19 pandemic in March 2020, most sites suspended in-person research activities, which posed recruitment challenges [3]. At that time, recruitment to meSH pivoted from an in-person to a virtual approach. Virtual enrollment of providers was conducted by email. An introductory email was sent by the site study coordinators that detailed the meSH study and inquired about interest in participation. This email was followed by reminder emails every 1 to 2 weeks until recruitment ended in March of 2021. In total, 41 of 54 (76%) providers approached virtually accepted, compared to 48 of 50 (96%) providers who were approached in-person and enrolled (P<.001).

The most common challenges site coordinators faced in recruiting providers were a lack of response, a lack of interest due to an impersonal experience or perceived lack of need for intervention, a lack of time in provider schedules, and technical issues with the provider’s phone, such as outdated software that inhibited them from downloading the intervention app.
Table 1. Demographics of providers enrolled in the Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization (meSH) study. The meSH study was a nonrandomized, closed cohort hybrid-effectiveness trial that used staggered site initiation for recruitment. The study was conducted in the sickle cell disease population at 7 academic institutions across the United States from September 2019 to August 2022.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Site 1 (n=15), n (%)</th>
<th>Site 2 (n=27), n (%)</th>
<th>Site 3 (n=3), n (%)</th>
<th>Site 4 (n=8), n (%)</th>
<th>Site 5 (n=12), n (%)</th>
<th>Site 6 (n=17), n (%)</th>
<th>Site 7 (n=7), n (%)</th>
<th>Total (n=89), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2 (13)</td>
<td>10 (37)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>5 (29)</td>
<td>2 (33)</td>
<td>20 (23)</td>
</tr>
<tr>
<td>Female</td>
<td>13 (87)</td>
<td>17 (63)</td>
<td>3 (100)</td>
<td>8 (100)</td>
<td>11 (92)</td>
<td>10 (59)</td>
<td>3 (50)</td>
<td>65 (73)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-44</td>
<td>8 (53)</td>
<td>20 (74)</td>
<td>3 (100)</td>
<td>2 (25)</td>
<td>7 (58)</td>
<td>9 (53)</td>
<td>4 (67)</td>
<td>53 (60)</td>
</tr>
<tr>
<td>45-64</td>
<td>6 (40)</td>
<td>7 (26)</td>
<td>0 (0)</td>
<td>4 (50)</td>
<td>3 (25)</td>
<td>6 (35)</td>
<td>2 (33)</td>
<td>28 (32)</td>
</tr>
<tr>
<td>65 and older</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (25)</td>
<td>2 (17)</td>
<td>1 (6)</td>
<td>0 (0)</td>
<td>5 (6)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Asian</td>
<td>2 (13)</td>
<td>7 (26)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (17)</td>
<td>5 (29)</td>
<td>0 (0)</td>
<td>16 (18)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>5 (33)</td>
<td>3 (11)</td>
<td>2 (67)</td>
<td>2 (25)</td>
<td>5 (42)</td>
<td>5 (29)</td>
<td>2 (33)</td>
<td>24 (27)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>White</td>
<td>7 (47)</td>
<td>17 (63)</td>
<td>1 (33)</td>
<td>6 (75)</td>
<td>6 (50)</td>
<td>6 (35)</td>
<td>4 (67)</td>
<td>47 (53)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (6)</td>
<td>1 (17)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>14 (93)</td>
<td>27 (100)</td>
<td>3 (100)</td>
<td>8 (100)</td>
<td>12 (100)</td>
<td>15 (88)</td>
<td>5 (83)</td>
<td>84 (94)</td>
</tr>
<tr>
<td>Level of SCD expertise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level I: unengaged</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Level II: willing</td>
<td>4 (27)</td>
<td>3 (11)</td>
<td>0 (0)</td>
<td>2 (25)</td>
<td>1 (8)</td>
<td>7 (41)</td>
<td>1 (17)</td>
<td>18 (21)</td>
</tr>
<tr>
<td>Level III: willing high-volume</td>
<td>1 (7)</td>
<td>7 (26)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>4 (24)</td>
<td>1 (17)</td>
<td>14 (16)</td>
</tr>
<tr>
<td>Level IV: expert</td>
<td>9 (60)</td>
<td>17 (63)</td>
<td>3 (100)</td>
<td>6 (75)</td>
<td>10 (83)</td>
<td>5 (29)</td>
<td>4 (67)</td>
<td>54 (61)</td>
</tr>
</tbody>
</table>

Patient Recruitment

The total patient population treated with hydroxyurea across all 7 sites was 2208 patients. Of those, 2157 (97.6%) patients were screened for eligibility. Of the 2157 individuals screened, 611 (28.3%) patients were eligible for the study, and 343 (56%) were approached for enrollment (Figure 1).

Of the 343 eligible patients approached, 293 accepted participation and were successfully enrolled across all 7 sites, an 85.4% participation rate. Among those enrolled, most patients were aged between 26 to 45 years (164/293, 56%), were Black or African American (283/293, 97%), were not Hispanic or Latino (279/293, 95%), were never married (242/293, 83%), had a high school degree or some college or vocational training (176/293, 60%), were not employed (115/293, 39%), and had a household income of $25,000 or less (166/293, 57%) (Table 2). In total, 126 of 143 (88%) patients virtually approached were enrolled, compared to 167 of 200 (83.5%) patients who were approached in person ($P=0.23$) (Figure 1).

Although slightly lower than the projected 46 patients per site (322 total patients), our total patient participant sample size of 293 met the requirements of the a priori statistical design [5]. A total of 1546 patients were screened and deemed ineligible across all 7 sites, with the most common reason for ineligibility being “not treated with hydroxyurea” for 49.8% of patients. A total of 50 patients who were approached declined participation. Reasons for declining participation included not being interested in research, not having time to participate in research, and not being interested in the app because of already having a reminder system or thinking the app would not help them.
Figure 1. A CONSORT (Consolidated Standards of Reporting Trials) diagram of patient recruitment for the Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization (meSH) study, including the total numbers of patients approached and enrolled virtually and in person. The meSH study was a nonrandomized, closed cohort hybrid-effectiveness trial that used staggered site initiation for recruitment. The study was conducted in the sickle cell disease population at 7 academic institutions across the United States from September 2019 to August 2022.
Table 2. Demographics of patients enrolled in the Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization (meSH) study.

<table>
<thead>
<tr>
<th>Site</th>
<th>Site 1 (n=48), n (%)</th>
<th>Site 2 (n=46), n (%)</th>
<th>Site 3 (n=47), n (%)</th>
<th>Site 4 (n=41), n (%)</th>
<th>Site 5 (n=45), n (%)</th>
<th>Site 6 (n=47), n (%)</th>
<th>Site 7 (n=19), n (%)</th>
<th>Total (n=293), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-17</td>
<td>0 (0)</td>
<td>10 (22)</td>
<td>0 (0)</td>
<td>5 (12)</td>
<td>8 (18)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>23 (8)</td>
</tr>
<tr>
<td>18-25</td>
<td>22 (46)</td>
<td>28 (61)</td>
<td>12 (26)</td>
<td>18 (44)</td>
<td>20 (44)</td>
<td>4 (9)</td>
<td>2 (10)</td>
<td>106 (36)</td>
</tr>
<tr>
<td>26-45</td>
<td>26 (54)</td>
<td>8 (17)</td>
<td>35 (74)</td>
<td>18 (44)</td>
<td>17 (36)</td>
<td>43 (92)</td>
<td>17 (90)</td>
<td>164 (56)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>18 (38)</td>
<td>24 (52)</td>
<td>27 (57)</td>
<td>15 (37)</td>
<td>29 (64)</td>
<td>18 (38)</td>
<td>10 (53)</td>
<td>141 (48)</td>
</tr>
<tr>
<td>Female</td>
<td>30 (63)</td>
<td>22 (48)</td>
<td>20 (43)</td>
<td>26 (63)</td>
<td>16 (36)</td>
<td>29 (62)</td>
<td>9 (47)</td>
<td>151 (52)</td>
</tr>
<tr>
<td>Race&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>48 (100)</td>
<td>46 (100)</td>
<td>47 (100)</td>
<td>41 (100)</td>
<td>41 (91)</td>
<td>47 (100)</td>
<td>14 (74)</td>
<td>283 (97)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>4 (9)</td>
<td>0 (0)</td>
<td>3 (16)</td>
<td>8 (3)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (11)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>9 (47)</td>
<td>11 (4)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>48 (100)</td>
<td>46 (100)</td>
<td>46 (98)</td>
<td>41 (100)</td>
<td>44 (98)</td>
<td>45 (96)</td>
<td>10 (53)</td>
<td>279 (95)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married or living as married</td>
<td>6 (13)</td>
<td>1 (2)</td>
<td>10 (21)</td>
<td>0 (0)</td>
<td>4 (9)</td>
<td>5 (11)</td>
<td>4 (21)</td>
<td>30 (10)</td>
</tr>
<tr>
<td>Never married</td>
<td>39 (81)</td>
<td>41 (89)</td>
<td>36 (77)</td>
<td>33 (83)</td>
<td>41 (91)</td>
<td>37 (79)</td>
<td>15 (79)</td>
<td>242 (83)</td>
</tr>
<tr>
<td>Divorced, separated, or widowed</td>
<td>2 (4)</td>
<td>3 (7)</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>2 (5)</td>
<td>0 (0)</td>
<td>3 (6)</td>
<td>11 (4)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>10 (3)</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school or some high school</td>
<td>15 (33)</td>
<td>3 (6)</td>
<td>15 (33)</td>
<td>3 (6)</td>
<td>6 (15)</td>
<td>9 (20)</td>
<td>4 (9)</td>
<td>2 (11)</td>
</tr>
<tr>
<td>High school graduate or some college or vocational training</td>
<td>28 (58)</td>
<td>24 (52)</td>
<td>33 (70)</td>
<td>21 (51)</td>
<td>26 (58)</td>
<td>31 (66)</td>
<td>13 (68)</td>
<td>176 (60)</td>
</tr>
<tr>
<td>College graduate or greater</td>
<td>16 (33)</td>
<td>6 (13)</td>
<td>11 (23)</td>
<td>9 (22)</td>
<td>10 (22)</td>
<td>10 (21)</td>
<td>4 (21)</td>
<td>66 (23)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>5 (12)</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>8 (3)</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>18 (38)</td>
<td>13 (28)</td>
<td>18 (38)</td>
<td>10 (24)</td>
<td>16 (36)</td>
<td>13 (28)</td>
<td>7 (37)</td>
<td>95 (32)</td>
</tr>
<tr>
<td>Not employed by choice (homemaker, student, retired)</td>
<td>16 (33)</td>
<td>23 (50)</td>
<td>3 (6)</td>
<td>8 (20)</td>
<td>14 (31)</td>
<td>7 (15)</td>
<td>4 (21)</td>
<td>75 (26)</td>
</tr>
<tr>
<td>Not employed for other reasons (laid off, sick leave, maternity leave, unemployed, disabled)</td>
<td>14 (29)</td>
<td>9 (20)</td>
<td>26 (55)</td>
<td>18 (44)</td>
<td>15 (33)</td>
<td>25 (53)</td>
<td>8 (42)</td>
<td>115 (39)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>5 (12)</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>8 (3)</td>
</tr>
<tr>
<td>Household income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$25,000 or less</td>
<td>24 (50)</td>
<td>19 (41)</td>
<td>35 (74)</td>
<td>26 (63)</td>
<td>22 (49)</td>
<td>32 (68)</td>
<td>8 (42)</td>
<td>166 (57)</td>
</tr>
<tr>
<td>$25,001-$50,000</td>
<td>6 (13)</td>
<td>18 (39)</td>
<td>9 (19)</td>
<td>7 (17)</td>
<td>8 (18)</td>
<td>9 (19)</td>
<td>3 (16)</td>
<td>60 (20)</td>
</tr>
<tr>
<td>≥$50,001</td>
<td>12 (25)</td>
<td>5 (11)</td>
<td>2 (4)</td>
<td>3 (7)</td>
<td>10 (22)</td>
<td>3 (6)</td>
<td>5 (26)</td>
<td>40 (14)</td>
</tr>
<tr>
<td>Missing</td>
<td>6 (13)</td>
<td>4 (9)</td>
<td>1 (2)</td>
<td>5 (12)</td>
<td>5 (11)</td>
<td>3 (6)</td>
<td>3 (16)</td>
<td>27 (9)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Individuals could select more than one race.
Patient Recruitment Strategy Adaptations

Patient recruitment challenges fell into 3 themes: lack of interest in participating in research, lack of resources (staff and space), and lack of proper documentation to verify study eligibility. Specifically, challenges included difficulty verifying hydroxyurea status, technical issues such as nonworking cell phones or phone storage issues, inability to reach patients due to reduced clinic flow as patients opted for telehealth or telemedicine visits, wrong or disconnected phone numbers, and lack of interest in research participation.

Patient recruitment was initially designed to be conducted fully in person. In-person recruitment involved screening and approaching eligible patients during a regularly scheduled, nonacute clinic visit. Due to the COVID-19 pandemic restrictions for in-person research activities, virtual enrollment was initiated. The sites most affected by clinic shutdowns were the first 2 sites, as previously reported [9]. Although clinic lockdowns were no longer in effect when the remaining 5 sites opened enrollment, these sites were also affected by long-term changes related to the pandemic (eg, staff resignation, remote working). The subsequent sites, however, benefited from institutional adjustments to the ongoing pandemic and amended their protocols to include remote recruitment before initiating enrollment. The approach for virtual enrollment included telephone and Zoom (Zoom Video Communications, Inc) calls. Research coordinators at some sites attempted to join clinicians during telehealth patient visits. However, leveraging telehealth clinical visits to conduct study enrollment was unsuccessful. It was not feasible for some sites to add extra time to telehealth clinical visits for research purposes because of already strained logistic issues with establishing connections and providers feeling overwhelmed with managing a new system. One site used the electronic health record (EHR) to send patients secure messages through the patient portal. This approach was intended to create study awareness among eligible participants and to prime them to expect a recruitment phone call. Other sites that transitioned to virtual enrollment contacted patients via phone calls and text messages.

While most of the challenges coordinators faced were during the enrollment phase, coordinators also expressed challenges during the screening portion of the recruitment phase. One challenge was verifying eligibility due to poor documentation of hydroxyurea use in the patients’ charts. In some cases, coordinators found it difficult to search the EHR to find a definitive answer to whether patients were taking hydroxyurea and, if they were, the start and stop dates. Multiple study coordinators reported having a difficult time finding patients that were interested in participating in the study. When asked, most participants gave vague answers such as “just being uninterested in research.” At the same time, some gave more specific answers such as “not having time for study-related activities” or “participating in too many studies (research fatigue).” To overcome these challenges, some sites modified their screening process to optimize identifying eligible patients. For instance, one site sought the recommendations of the hematologists for participants to enroll and for permission to approach them. However, this strategy of seeking provider input ultimately was not sustainable as it consumed too much time for the research coordinators and providers. To save time, the research coordinators sent courtesy messages to the SCD providers to let them know which patients they planned to approach for enrollment. Providers accepted this approach as they believed it led to the least interference with clinic flow and impact on their time.

During the enrollment phase, some coordinators also mentioned a lack of physical and human resources, such as a shortage of research staff (some resigned during the COVID-19 pandemic) and a lack of space to conduct research, as areas of their clinics and buildings reduced space for research activities due to restrictions on in-person contact. Multiple coordinators mentioned that the biggest hindrance was a need for more staff to help support study implementation. The reasons for staff shortage varied. Still, one common theme was the presence of other studies competing for the time and focus of a limited number of study coordinators. Coordinators adapted their recruitment strategies to overcome challenges. Coordinators with working space restrictions began to rely heavily on remote methods (eg, phone, text message, telehealth) for recruitment instead of in-person methods and successfully sustained recruitment activities. Those with technology issues worked closely with their IT departments to reach feasible solutions, including a tip sheet provided by the IT department to the coordinators for reference for future problems. For sites experiencing staffing shortages and competing studies, coordinators prioritized the enrollment goals of the mSH study while more staff were being hired.

Discussion

In a multicenter study that tested multilevel mHealth interventions to improve hydroxyurea utilization, recruitment strategies had to be modified and adapted to conform to unforeseen circumstances, which included a global pandemic and its downstream effects, including staff shortages and research space limitations. Despite these unexpected barriers, research staff were nimble in adapting and introducing new recruitment strategies, ultimately leading to sufficient study accrual. We found that an in-person recruitment strategy was more successful than a virtual approach for recruiting providers. For recruiting eligible patients, both in-person and virtual strategies worked well.

The study team launched provider enrollment with a study presentation, allowing providers to ask questions and generate interest in the study. Using a regularly scheduled staff meeting to facilitate provider recruitment helped to eliminate the issue of addressing providers’ busy schedules. Virtual enrollment for patients was effective, with multiple touch points being the best perceived method by study coordinators. Sites that were able to reach out to participants in multiple ways to promote awareness of the study and prime them for future recruitment contact reported fewer obstacles in enrolling patients according to study coordinators. Switching to remote enrollment and using multitouch contact with patients allowed for successful recruitment.

Our study demonstrates that both patients and providers can be successfully enrolled to research using remote methods.
However, remote enrollment of patient participants in this study was often paired with additional strategies that offer the familiar human touch. This is one factor that might have aided our successful remote recruitment, since many potential participants were previously engaged and warmly introduced to the study (eg, by receiving a message from the patient portal or being told about the study by a flyer or clinic staff), as opposed to an impersonal contact (ie, an unplanned phone call received from an unfamiliar person). Sites attempted warm contacts in a variety of ways, including text messages, clinic staff, and flyers posted around the clinic.

Compared with other studies that have looked at effective recruitment strategies for enrolling individuals with SCD, our study supports the use of both virtual and in-person research recruitment. Using multiple recruitment strategies could be the most desirable approach to successfully recruit patients [14], such as the multitouch contact approach. With the rise of the pandemic and other factors that may make in-person enrolling difficult, a focus on whether virtual enrollment can be successful has emerged. Our study supports the findings of others, who have found that virtual enrollment can be a successful strategy for recruiting patients with other diseases [15].

A limitation of this paper is that the method used to gather the coordinators’ opinions was not an exhaustive exploration of barriers. In addition, we did not assess facilitators to study participation, which is another important aspect of optimizing participation. Future studies should expound on and further enhance collection of these data using a mixed methods approach to characterize facilitators and barriers, as well as strategies to overcome them, systematically.

Another limitation of this paper is that the coordinators were asked about their perceptions of the benefits and efficiency of the recruitment strategies after the recruitment and enrollment phases were completed; therefore, the responses could be subject to recall bias. The small enrollment goal at each site is also a potential limitation. Enrollment was successful across the sites, but the enrollment goals were relatively modest, making it impossible to conclude that these challenges would have been overcome in larger studies. Additionally, due to the time period, it is not possible to separate our findings from the influence of the burden of responding to COVID-19.

In conclusion, fast adaptation of recruitment strategies was possible in response to the COVID-19 pandemic with remote enrollment combined with multitouch recruitment strategies leading to successful study accrual of patients with SCD. Engaging providers in research can be challenging. Future researchers should continue to identify and test effective recruitment strategies, both in person and remote, for provider participants. Additionally, it will be important to continue identifying effective recruitment strategies in studies involving patients with SCD and their providers and the impact and navigation of recruitment challenges.

**Acknowledgments**

The Sickle Cell Disease Implementation Consortium (Multimedia Appendix 2) has been supported by the US federal government cooperative agreements HL133948, HL133964, HL133990, HL133996, HL133994, HL133997, HL134004, HL134007, and HL134042 from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities of the National Institutes of Health (NIH) and the Bridge grant from the American Society of Hematology. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the views of the NIH.

**Conflicts of Interest**

MPS has been a paid research consultant for the Association of Community Cancer Centers. JSH receives consultancy fees from Global Blood Therapeutics and Forma Therapeutics and an honorarium from UpToDate.

**References**


Abbreviations

**CONSORT**: Consolidated Standard of Reporting Trials
**EHR**: electronic health record
**meSH**: Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization
**mHealth**: mobile health
**REDCap**: Research Electronic Data Capture
**SCD**: sickle cell disease
**SCDIC**: Sickle Cell Disease Implementation Consortium
**VOC**: vaso-occlusive crisis

---

Edited by A Mavragani; submitted 08.05.23; peer-reviewed by S Creary; comments to author 03.10.23; revised version received 02.12.23; accepted 05.12.23; published 16.04.24.

Please cite as:


Recruitment Strategies in the Integration of Mobile Health Into Sickle Cell Disease Care to Increase Hydroxyurea Utilization Study (mSH): Multicenter Survey Study

JMRI Form Res 2024;8:e48767
URL: https://formative.jmir.org/2024/1/e48767
doi:10.2196/48767
PMID:36655729
Development of a Management App for Postviral Fibromyalgia-Like Symptoms: Patient Preference-Guided Approach

Marc Blanchard¹, BSc, MSc; Cinja Nadana Koller¹, BA, MSc; Pedro Ming Azevedo¹, MD, PhD; Tiffany Prétat¹, MD; Thomas Hügle¹, Prof Dr Med

Department of Rheumatology, Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland

Corresponding Author:
Marc Blanchard, BSc, MSc
Department of Rheumatology
Lausanne University Hospital (CHUV)
University of Lausanne
Pierre-Decker 4
Lausanne, 1011
Switzerland
Phone: 41 794266883
Email: marc_blanchard@bluewin.ch

Abstract

Background: Persistent fibromyalgia-like symptoms have been increasingly reported following viral infections, including SARS-CoV-2. About 30% of patients with post–COVID-19 syndrome fulfill the fibromyalgia criteria. This complex condition presents significant challenges in terms of self-management. Digital health interventions offer a viable means to assist patients in managing their health conditions. However, the challenge of ensuring their widespread adoption and adherence persists. This study responds to this need by developing a patient-centered digital health management app, incorporating patient preferences to enhance usability and effectiveness, ultimately aiming to improve patient outcomes and quality of life.

Objective: This research aims to develop a digital health self-management app specifically for patients experiencing postviral fibromyalgia-like symptoms. By prioritizing patient preferences and engagement through the app’s design and functionality, the study intends to facilitate better self-management practices and improve adherence.

Methods: Using an exploratory study design, the research used patient preference surveys and usability testing as primary tools to inform the development process of the digital health solution. We gathered and analyzed patients’ expectations regarding design features, content, and usability to steer the iterative app development.

Results: The study uncovered crucial insights from patient surveys and usability testing, which influenced the app’s design and functionality. Key findings included a preference for a symptom list over an automated chatbot, a desire to report on a moderate range of symptoms and activities, and the importance of an intuitive onboarding process. While usability testing identified some challenges in the onboarding process, it also confirmed the importance of aligning the app with patient needs to enhance engagement and satisfaction.

Conclusions: Incorporating patient feedback has been a significant factor in the development of the digital health app. Challenges encountered with user onboarding during usability testing have highlighted the importance of this process for user adoption. The study acknowledges the role of patient input in developing digital health technologies and suggests further research to improve onboarding procedures, aiming to enhance patient engagement and their ability to manage digital health resources effectively.

International Registered Report Identifier (IRRID): RR2-10.2196/32193

(JMIR Form Res 2024;8:e50832) doi:10.2196/50832

KEYWORDS
digital health; patient preference; user experience; patient-centricity; platform; development; fibromyalgia; self-management; quality of life; patient outcome; musculoskeletal; usability testing; digital health solution
Introduction

Postviral Syndromes, Post–COVID-19, and Fibromyalgia-Like Symptoms

Postviral syndromes have been substantially observed and described in the past. Epstein-Barr virus or Q fever has for example led patients to develop chronic symptoms such as fatigue or pain [1,2]. After the 2003 severe acute respiratory syndrome (SARS) outbreak, about one-third of the infected population developed reduced tolerance to exercise, despite having normal lung function, which is similar to what has been observed in post–COVID-19 syndrome [3,4]. However, the first SARS outbreak only affected 8000 people locally while COVID-19 has affected more than 750 million people worldwide according to the World Health Organization (WHO) [5]. With a prevalence of 54% for hospitalized and 34% for nonhospitalized patients, post–COVID-19 syndrome is of major concern for public health [6].

Post–COVID-19 syndrome or long COVID is a condition where patients experience persisting general unspecific symptoms like fatigue, myalgia, and concentration or sleep disturbance even several months after a SARS-CoV-2 infection [4]. While the mechanisms underlying the persisting symptoms are not all well understood, it has been shown that the neurotropism of SARS-CoV-2 can lead to neural damage and persisting neurologic symptoms including neuropathic pain, transient memory loss, and olfactory dysfunction [7-9]. Very similar to previously described postviral conditions like myalgic encephalomyelitis or chronic fatigue syndrome, the post–COVID-19 symptoms often fulfill classification criteria for fibromyalgia [10,11]. Moreover, recent studies suggest that fibromyalgia could be triggered by COVID-19, thus becoming a postviral chronic condition [12,13].

Digital Health as a Therapeutic Approach

While these chronic syndromes affecting a significant part of the population are known to be difficult to treat with traditional medication, digital health solutions could have great potential. Digital health solutions have recently proven efficacy in chronic pain and fibromyalgia management [14,15]. Self-management plays a big role in the empowerment of these patients, and encouraging studies showed the efficacy of such programs [16,17]. Moreover, cognitive behavioral therapy or mindfulness and education training have shown substantial positive effects in a self-management setting, thus empowering the patients by involving them in the therapeutic process [18-20]. Based on these observations and findings, the development of a self-management platform that includes similar therapeutic approaches for patients with post–COVID-19 syndrome could be valuable.

The Challenge of Adoption

Treatment adherence is a major issue in chronic disease management. Since digital health solutions require greater effort (time and dedication) from patients in comparison to traditional medication, efforts have to be made to ensure high adoption. However, very few efforts are made to improve patient engagement in digital health, where the focus is usually set on the assessment of the therapeutic value through clinical studies. This lack of consideration for adoption and engagement is certainly responsible for the limited adoption of digital health solutions by patients and physicians, despite the plethora of apps available and the increasing compliance of health care systems for these solutions. Difficulties in distinguishing validated therapies from lifestyle solutions play a role in this phenomenon as well [21]. While studies have shown that barriers to adherence were mostly fear of adverse effects, misperception of therapeutic value, forgetfulness, costs, and poor patient-physician interaction in conditions like rheumatoid arthritis, the role of patient preferences emerges as one of the pivotal factors for adherence [22,23].

Patient Centricity in the Development Process

In recent years, the health care industry has increasingly recognized the importance of patient centricity in the development of health care solutions [24]. The successful adoption and use of digital health solutions, particularly for complex and chronic conditions such as fibromyalgia, heavily rely on incorporating patient perspectives and preferences throughout the development process. To address the challenges of adoption and enhance the usability and effectiveness of digital health solutions, it is crucial to engage patients as active participants in the design and evaluation stages. User experience–based approaches, such as usability testing and patient preference surveys, have emerged as valuable tools to gather insights and guide the development of patient-centered digital health solutions [25]. By involving patients in the research process, we can gain a deeper understanding of their needs, preferences, and expectations, ensuring that the resulting product is tailored to meet their unique needs. Therefore, in this study, we aim to explore the integration of patient preference techniques in the development of a digital health solution for patients enduring from fibromyalgia-like symptoms after a viral infection, leveraging user experience research to enhance usability and patient satisfaction.

Regulatory Requirements and Usability

On the regulatory side, requirements play a crucial role in the development of digital health products, particularly when it comes to ensuring usability and facilitating widespread adoption. Formative evaluations and usability testing are essential components of the development process, enabling the identification and mitigation of usability issues early on. For instance, the International Organization for Standardization (ISO) 13485 standard provides guidelines for quality management systems in the medical device industry, emphasizing the importance of usability engineering and human factors [26]. Additionally, the certification process for software as a medical device involves demonstrating compliance with regulatory requirements, including usability considerations. By adhering to these norms and engaging in rigorous formative evaluations and usability testing, developers can not only ensure compliance but also enhance the overall user experience, thereby increasing the likelihood of adherence and adoption of digital health products.
Objectives
This study aims to leverage patient preferences as a guiding framework for the development of a digital health self-management platform tailored specifically for patients with postviral fibromyalgia-like symptoms. By using patient preference surveys and conducting usability testing, we seek to identify the unique needs, preferences, and challenges faced by this patient population. Through a comprehensive understanding of these factors, we aim to develop an app that aligns with patient expectations, fosters engagement, and enhances self-management practices. This research will contribute to the growing field of patient-centered digital health interventions and provide valuable insights for health care professionals, developers, and stakeholders involved in the development and implementation of digital health solutions for chronic conditions. Ultimately, our goal is to improve the adherence and adoption rates of digital health interventions, leading to better patient outcomes and enhanced quality of life for the patients.

Methods
Study Design
This is an exploratory study that aims to actively involve patients in the development process of a digital health app named “POCOS” through patient preference surveys and usability testing. This design has been chosen to gain a comprehensive understanding of the unique needs, preferences, and challenges faced by the target population.

Participants
Recruitment for the patient preference surveys was conducted through a collaboration with the Swiss patient post–COVID-19 association, Long COVID Schweiz [27]. The association’s coordinator conducted a preliminary screening to identify members who met the study criteria—those who had experienced a COVID-19 infection and were currently experiencing persisting fibromyalgia-like symptoms. Interested individuals were then contacted through email, resulting in 53 and 33 participants responding to the first and second surveys, respectively.

For the usability testing, potential participants were identified from the cohort of patients engaged in the multimodal care program at Lausanne University Hospital’s Rheumatology Department, which caters to patients with fibromyalgia. Eligibility requires patients to exhibit fibromyalgia-like symptoms persisting after a viral infection, predominantly COVID-19. Selected patients were approached for participation during their care program, and those who consented underwent usability testing. This process was conducted in a face-to-face setting by the study nurse. The sample consisted of 6 patients (4 women and 2 men, aged between 20 and 59 years) who fulfilled the inclusion criteria and participated in the usability testing.

Data Collection
Survey data were collected using web-based questionnaires administered through the SurveyMonkey (SurveyMonkey Inc) platform. The survey items were developed based on a comprehensive review of existing literature, expert input, and previous studies in the field.

In total, 2 surveys were sent to patients, the first (survey 1, fifty-three respondents) focused on user preferences, design aspects, and symptom reporting, and the second (survey 2, thirty-three respondents) focused on patient expectations regarding content or usability.

Usability testing data have been collected by a trained nurse during interviews with the participants. These data consist of notes taken from observation of the patients using the newly developed primary prototype of the app by the nurse and reported feedback that patients have given to the nurse during and after the use of the app.

Ethical Considerations
This study adhered strictly to ethical standards for human participant research. Ethical review and approval were obtained from the commission cantonale d’éthique de la recherche sur l’être humain (2021-01680), and all participants provided permission for the use of their data in secondary analyses without the need for additional consent. We ensured the privacy and confidentiality of all study participants by anonymizing or deidentifying all personal data before analysis. Participants did not receive any compensation for their involvement in the study. Furthermore, to protect participant identity, no personal identifiers are present in any images within the study or Multimedia Appendix 1.

Results
Patient Preference Survey (Survey 1)
Participants were asked about the patient-reported outcome design they would prefer between a symptoms list with intensity scales from 1 to 10 or an automated chatbot asking for symptoms entry (Figure 1). In total, 48 out of 53 (91%) respondents found the symptoms list more appropriate according to their conditions.

Patients were then asked about the number of symptoms- and activity-reporting questions they would like to answer regularly. For both types of questions, most of the respondents preferred 5-10 questions (n=37, 70% and n=31, 58%, respectively). A total of 15 (28%) and 10 (19%), respectively, would like to answer more than 10 questions while 1 (2%) and 12 (23%) preferred to answer 5 questions or less (number of respondents is 53).

The most relevant symptoms patients would like to regularly report considering their conditions are respectively, among the list, fatigue, concentration level, sleep quality, memory level, pain, and more specifically pain location (Table 1).

Among the most relevant activities they would like to report according to their daily life with post–COVID-19 are physical activities, nonmedical therapies like yoga, massage or relaxation, doctor visits, and painkiller use (Table 2).
Figure 1. Patient-reported outcome designs. (A) Symptoms are listed with sliders and intensity scales from 1 to 10. (B) The automated chatbot regularly asks the user for symptoms entry through text (survey 1, number of respondents is 53).

Table 1. Most relevant symptoms patients would like to regularly report on the platform considering their post–COVID-19 syndrome conditions (survey 1, number of respondents are 53).

<table>
<thead>
<tr>
<th>Symptoms patients would like to regularly report on the platform</th>
<th>Frequency (N=53), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue level</td>
<td>48 (91)</td>
</tr>
<tr>
<td>Concentration level</td>
<td>36 (68)</td>
</tr>
<tr>
<td>Sleep quality</td>
<td>33 (62)</td>
</tr>
<tr>
<td>Memory level</td>
<td>30 (57)</td>
</tr>
<tr>
<td>Pain location</td>
<td>28 (53)</td>
</tr>
<tr>
<td>Ease of activity</td>
<td>26 (49)</td>
</tr>
<tr>
<td>Breathing problems</td>
<td>21 (40)</td>
</tr>
<tr>
<td>Specific pain level</td>
<td>20 (38)</td>
</tr>
<tr>
<td>General health level</td>
<td>18 (34)</td>
</tr>
<tr>
<td>Mood</td>
<td>16 (30)</td>
</tr>
<tr>
<td>Gastrointestinal problems</td>
<td>16 (30)</td>
</tr>
<tr>
<td>Anxiety level</td>
<td>15 (28)</td>
</tr>
<tr>
<td>Overall pain level</td>
<td>14 (26)</td>
</tr>
<tr>
<td>Smelling and taste problems</td>
<td>9 (17)</td>
</tr>
<tr>
<td>Others: tinnitus, brain fog, postural orthostatic tachycardia syndrome, nausea, dizziness, and menstrual problems</td>
<td>15 (28)</td>
</tr>
</tbody>
</table>
Table 2. Most relevant activities patients would like to regularly report on the platform according to their daily life with post–COVID-19 syndrome (survey 1, number of respondents are 53).

<table>
<thead>
<tr>
<th>Activity</th>
<th>Frequency (N=53), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical activity</td>
<td>48 (91)</td>
</tr>
<tr>
<td>Nonmedical therapies</td>
<td>42 (79)</td>
</tr>
<tr>
<td>Doctor visits</td>
<td>28 (53)</td>
</tr>
<tr>
<td>Painkiller use</td>
<td>24 (45)</td>
</tr>
<tr>
<td>Nutrition</td>
<td>19 (36)</td>
</tr>
<tr>
<td>Others: social activities, housekeeping, job activity, psychological visits, and relaxation</td>
<td>15 (28)</td>
</tr>
</tbody>
</table>

Patient Expectations Survey (Survey 2)

Questions about what patients would like to monitor on the platform and what kind of therapeutic content they would expect from it have been asked. While there was a disparity in what patients would like to visualize on the platform, the evolution of their symptoms and energy level were the most frequently chosen answers. In terms of therapeutic content, most of the patients expect to be informed about what has helped people with similar symptoms, to be provided with a therapeutic exercise program, with links to patient communities, and with information and news about their conditions (Table 3).

Finally, more technical questions like their expected frequency of use, time spent on the platform, and on which support they would use it, were asked. Answers showed that most of the patients would use the platform every day, from 5 to 10 minutes on their mobile phones (Table 4).

Table 3. Response frequency to questions by patients with post–COVID-19 syndrome (survey 2).

<table>
<thead>
<tr>
<th>Responses to the questions</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What would the patients like to visualize on the platform (n=26)</strong></td>
<td></td>
</tr>
<tr>
<td>In total, 3 main symptoms and evolution</td>
<td>7 (27)</td>
</tr>
<tr>
<td>Every symptom and evolution</td>
<td>14 (54)</td>
</tr>
<tr>
<td>List of symptoms that get better or worsen</td>
<td>15 (58)</td>
</tr>
<tr>
<td>Energy level evolution</td>
<td>19 (73)</td>
</tr>
<tr>
<td>Disease activity evolution</td>
<td>12 (46)</td>
</tr>
<tr>
<td>Average disease activity score of other users</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Their activity history</td>
<td>11 (42)</td>
</tr>
<tr>
<td>Physical activity evolution</td>
<td>13 (50)</td>
</tr>
<tr>
<td>Body mass evolution</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Others: activities versus symptoms or system versus symptoms intensity charts</td>
<td>2 (8)</td>
</tr>
<tr>
<td><strong>What type of therapeutic content patients expect from the platform (n=27)</strong></td>
<td></td>
</tr>
<tr>
<td>Personalized therapeutic exercises</td>
<td>16 (59)</td>
</tr>
<tr>
<td>Informative content about symptoms and the disease</td>
<td>15 (56)</td>
</tr>
<tr>
<td>What has helped people with similar symptoms</td>
<td>22 (82)</td>
</tr>
<tr>
<td>Training program</td>
<td>16 (59)</td>
</tr>
<tr>
<td>Advice from experts in the field</td>
<td>15 (56)</td>
</tr>
<tr>
<td>Links to communities of patients or patients’ centers</td>
<td>17 (63)</td>
</tr>
<tr>
<td>Information, news, and articles about the disease</td>
<td>16 (59)</td>
</tr>
<tr>
<td>Quizzes and tools</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Discussion forum with the “POCOS” community</td>
<td>13 (48)</td>
</tr>
<tr>
<td>Others</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
Table 4. Frequency at which patients expect to use the platform, time patients expect to spend on the platform per use and support patients expect to use the platform (survey 2, the number of respondents is 33, 33, and 32, respectively).

<table>
<thead>
<tr>
<th>Frequency patients expect to use the platform (n=33)</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than once a week</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Once a week</td>
<td>6 (18)</td>
</tr>
<tr>
<td>3 times a week</td>
<td>3 (9)</td>
</tr>
<tr>
<td>5 times a week</td>
<td>3 (9)</td>
</tr>
<tr>
<td>Everyday</td>
<td>19 (58)</td>
</tr>
<tr>
<td>Several times a day</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time patients expect to spend on the platform per use (n=33)</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 5 minutes</td>
<td>4 (12)</td>
</tr>
<tr>
<td>5 to 10 minutes</td>
<td>19 (58)</td>
</tr>
<tr>
<td>10 to 20 minutes</td>
<td>6 (18)</td>
</tr>
<tr>
<td>20 to 30 minutes</td>
<td>4 (12)</td>
</tr>
<tr>
<td>30 minutes to 1 hour</td>
<td>0 (0)</td>
</tr>
<tr>
<td>More than 1 hour</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Support patients expect to use the platform on (n=32)</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile phone</td>
<td>27 (84)</td>
</tr>
<tr>
<td>Tablet</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Computer</td>
<td>3 (9)</td>
</tr>
<tr>
<td>Television</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Others</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

**Usability Testing**

Out of the 6 participants, only 2 showed a sufficient understanding of the functionalities and managed to navigate through the app without additional help. In total, 2 patients did not even manage to register and start to use the app due to lack of guidance and information, and the 2 other patients showed some difficulties in understanding the workflow and efforts required to navigate. Only 1 patient was satisfied with the app and the content, reported it to be intuitive, easy to use, and recommended it to peers. Several patients reported a lack of comprehension of the app features and found that some of the therapeutic content was not specific enough.

**Discussion**

The patient preference survey results are insightful regarding patients’ needs for symptom and physical activity reporting. Clear preferences emerged for a list-based approach over an automated chatbot for symptom documentation. A majority opted to respond to 5-10 questions, suggesting a moderate yet flexible reporting frequency could be most beneficial. This preference for moderation extends to activity reporting, supporting the need for adaptable digital health tools (Figure 1).

The survey findings emphasize the significance of tracking key health symptoms—fatigue, concentration, sleep quality, memory level, and pain location—which are reported by over half of the participants (as detailed in Table 1). These symptoms align with documented post–COVID-19 syndrome effects in existing literature [4,28,29]. In addition to symptom monitoring, the survey underlines the critical role of physical activity reporting in the context of fibromyalgia-like symptoms management. A substantial majority, exceeding 90%, expressed a desire to consistently log their physical activity, indicative of a widespread recognition of its contribution to overall health and well-being. This observation is supported by research demonstrating the beneficial impacts of physical activity on improved physical function, reduced risk of chronic diseases, and enhanced mental health [30-32]. Regular assessment of physical activity, therefore, could yield insights into patient health, guiding more informed management of postviral fibromyalgia-like symptoms.

Most patients indicated a strong desire to track the progression of their symptoms and energy levels visually. They also expected the platform to offer diverse therapeutic content, with a significant portion wanting information on effective strategies that have aided others with similar symptoms (82%). Additionally, they anticipated the inclusion of community support links, as well as exercise and educational programs (Table 3).

Usability testing identified significant onboarding challenges and revealed gaps in patients’ understanding of the app’s functions. Further research should be performed on the onboarding process to investigate its impact on usability.
adoption, and retention rate by implementing and testing different onboarding strategies.

This study’s insights should be interpreted in the context of certain limitations. The participant pool was drawn exclusively from a Swiss patient post–COVID-19 association, which may not reflect the broader demographic affected by postviral conditions. The modest sample size, especially for usability testing, may limit the extrapolation of our findings to larger populations. Further research with a more diverse and extensive sample would be beneficial to validate and expand on our results.

In conclusion, the integration of patient preferences into the design of digital health apps has been substantiated by the findings of this study. The data derived from patient surveys have been instrumental in shaping the development of an app for managing postviral fibromyalgia-like symptoms (Figure S1 in Multimedia Appendix 1). These insights not only facilitated a tailored app that aligns with user expectations but also highlighted the significance of patient engagement and the potential for improved health outcomes.

The advancement of patient-centered digital health solutions, as evidenced by the app developed in this study, underscores a paradigm shift toward more responsive and adaptive health technology design. This approach is anticipated to enhance patient adherence and the efficacy of digital health interventions. The broader implications of this research advocate for the incorporation of user experience research early in the development cycle, ensuring that digital health technologies are congruent with the needs and preferences of end users.

Data Availability
All data generated or analyzed during this study are included in this published article (and in Multimedia Appendix 1).

Conflicts of Interest
MB and TH are shareholders and board members of ATREON SA. TH has also received speaker fees and unrestricted grants from Pfizer, Novartis, BMS, GSK, Galapagos and Roche.

Multimedia Appendix 1
POCOS user interface (app). Left side: "How are you?" with electronic patient-reported outcomes and activity. Middle: "My result" with monitoring of symptom activity and health conditions. Right side: "Act and Advice" with personalized training program, adapted to the user’s symptoms.

References
5. World Health Organization. URL: https://www.who.int/ [accessed 2024-04-04]

https://formative.jmir.org/2024/1/e50832


**Abbreviations**

**ISO:** International Organization for Standardization  
**SARS:** severe acute respiratory syndrome  
**WHO:** World Health Organization
©Marc Blanchard, Cinja Nadana Koller, Pedro Ming Azevedo, Tiffany Prétat, Thomas Hügle. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Creating High-Quality Synthetic Health Data: Framework for Model Development and Validation

Elnaz Karimian Sichani¹, MSc; Aaron Smith¹, PhD; Khaled El Emam²,³,⁴, BEng, PhD; Lucy Mosquera²,³, MSc

¹Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada
²Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
³Replica Analytics Ltd, Ottawa, ON, Canada
⁴School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

Abstract

Background: Electronic health records are a valuable source of patient information that must be properly deidentified before being shared with researchers. This process requires expertise and time. In addition, synthetic data have considerably reduced the restrictions on the use and sharing of real data, allowing researchers to access it more rapidly with far fewer privacy constraints. Therefore, there has been a growing interest in establishing a method to generate synthetic data that protects patients’ privacy while properly reflecting the data.

Objective: This study aims to develop and validate a model that generates valuable synthetic longitudinal health data while protecting the privacy of the patients whose data are collected.

Methods: We investigated the best model for generating synthetic health data, with a focus on longitudinal observations. We developed a generative model that relies on the generalized canonical polyadic (GCP) tensor decomposition. This model also involves sampling from a latent factor matrix of GCP decomposition, which contains patient factors, using sequential decision trees, copula, and Hamiltonian Monte Carlo methods. We applied the proposed model to samples from the MIMIC-III (version 1.4) data set. Numerous analyses and experiments were conducted with different data structures and scenarios. We assessed the similarity between our synthetic data and the real data by conducting utility assessments. These assessments evaluate the structure and general patterns present in the data, such as dependency structure, descriptive statistics, and marginal distributions. Regarding privacy disclosure, our model preserves privacy by preventing the direct sharing of patient information and eliminating the one-to-one link between the observed and model tensor records. This was achieved by simulating and modeling a latent factor matrix of GCP decomposition associated with patients.

Results: The findings show that our model is a promising method for generating synthetic longitudinal health data that is similar enough to real data. It can preserve the utility and privacy of the original data while also handling various data structures and scenarios. In certain experiments, all simulation methods used in the model produced the same high level of performance. Our model is also capable of addressing the challenge of sampling patients from electronic health records. This means that we can simulate a variety of patients in the synthetic data set, which may differ in number from the patients in the original data.

Conclusions: We have presented a generative model for producing synthetic longitudinal health data. The model is formulated by applying the GCP tensor decomposition. We have provided 3 approaches for the synthesis and simulation of a latent factor matrix following the process of factorization. In brief, we have reduced the challenge of synthesizing massive longitudinal health data to synthesizing a nonlongitudinal and significantly smaller data set.

(JMIR Form Res 2024;8:e53241) doi:10.2196/53241
KEYWORDS
synthetic data; tensor decomposition; data sharing; data utility; data privacy; electronic health record; longitudinal; model development; model validation; generative models

Introduction

Background

Electronic health records (EHRs) are becoming an increasingly important source of detailed information about patients because the successful integration and efficient analysis of EHRs could help solve many health care problems, such as expediting clinical decisions and enhancing patient safety. However, researchers often encounter challenges when trying to obtain high-quality health data for their research, and EHRs need to be appropriately deidentified before being shared with researchers. This process requires both skill and effort.

A method for deidentifying data, including health data, is anonymization. However, recent allegations show successful reidentification attacks on anonymized data [1,2]. Data synthesis is a recently developed approach for data deidentification. Studies suggest that synthetic data do not pose a significant privacy risk [3,4]. Therefore, data synthesis has emerged as an interesting method for producing nonidentifiable health data, which reduces the restrictions on the use and sharing of actual data. This allows researchers to access it more rapidly and with substantially fewer privacy constraints.

Therefore, there has been growing interest in establishing a method to simulate synthetic data that protects patients’ privacy while properly reflecting the data. One method for generating synthetic data is to choose an appropriate model, fit it within privacy constraints, and then simulate new data from the fitted model.

Furthermore, dimensionality reduction converts the high-dimensional description of the data into a low dimension without losing crucial information and phenotypes [5]. In the medical context, phenotypes are used to describe relevant variations and features [6]. EHR-based phenotyping is a process that maps raw EHR data to meaningful medical concepts [7,8]. Moreover, synthesizing longitudinal data in EHRs becomes difficult because patients may have lengthy sequences of events and come from diverse populations. Therefore, using a dimensionality reduction technique in the generative model could be advantageous. However, the main challenge is to develop a model that captures the most important data features and efficiently finds meaningful characterizations.

Classical phenotyping methods, which require medical field experts, can be time-consuming and expensive [9,10] because of the high-dimensional and heterogeneous nature of the EHR data set. Recently, more efficient unsupervised approaches, such as matrix factorization [11], have been emerging. However, matrix factorization does not necessarily detect associations within a data set because the data may not be accurately represented as matrices. Thus, tensor decompositions [12] have drawn growing attention because of their interpretability and flexibility in accommodating high-dimensional data. They also appear to be beneficial and favorable for computational phenotyping [13] and can easily be privatized [14,15]. They have been recognized as a promising method for the analysis of EHRs [16]. Therefore, generating a model based on state-of-the-art tensor factorization for creating synthetic longitudinal health data is a valuable effort.

The goal of this study is to develop a model that generates valuable synthetic longitudinal health data while protecting the privacy of the patients whose data are collected. We propose a model that relies on generalized canonical polyadic (GCP) tensor decomposition, which has demonstrated substantial outcomes in health data analysis [17,18]. We validated this model using samples from the MIMIC-III (version 1.4) data set [19]. MIMIC-III is a large, freely available database that contains deidentified health-related data.

There is a widespread belief that synthetic data present an insignificant privacy risk because there is no unique link or mapping between the records in the synthetic data and the records in the original data [20]. Therefore, this view should also be satisfied in practice. However, there is a one-to-one mapping and direct correspondence between the entries of the GCP model and the entries of the original data. This undermines the most sensible and acceptable concepts of privacy. Inspired by the studies by Ma et al [14] and Schmidt and Mohamed [21], this is fixed by simulating and modeling a latent factor matrix of GCP decomposition associated with patients.

Related Work

Tensor decomposition is an active area of research that has been widely applied to health care data [17,18,22]. It has been found to be a useful method for phenotyping EHRs [13,23]. It also has various applications beyond health data analysis, including recommender systems [24] and signal processing [25]. Thus far, several tensor decomposition techniques have been developed [26]. The most popular technique is called canonical polyadic (CP) tensor factorization, which is also known by 2 different names: canonical tensor decomposition and parallel factor decomposition, which were introduced separately by Carroll and Chang [27] and Harshman [28], respectively.

The CP decomposition approximates a tensor by the sum of rank-1 tensors using squared errors (L2 loss) [26]. Recently, Hong et al [29] developed a GCP model that offers the flexibility to use other loss functions in addition to squared errors. Furthermore, Kolda and Hong [30] have presented stochastic gradient descent as a way to overcome the challenge of fitting generalized CP to large-scale tensors, making it a perfect fit for a massive and heterogeneous EHR data set.

CP factorization [28] and its generalization, GCP decomposition [29], are fundamental tools for tensor analysis. They result in a factorized tensor that contains the most important computational phenotypes, and many studies show that they perform particularly well in phenotyping EHRs [13,17]. In addition, privacy-preserving methods have been widely applied to them in the medical setting [14,15]. Therefore, we developed
a generative model that produces synthetic longitudinal health data using generalized CP decomposition, which captures the most important features.

**Methods**

In this section, we present the methods that we used to generate synthetic longitudinal health data using generalized CP decomposition and various sampling and simulation techniques. We also describe the data, the evaluation method, and the experimental details of our study.

**Notations and Definitions**

We first describe the preliminaries and notations used in this paper. Before we begin, Table 1 shows some basic symbols used for tensor factorization. In addition, a boldface uppercase letter in Euler font represents a tensor, for example, \( \mathbf{A} \). A matrix is represented by a boldface uppercase letter, such as \( \mathbf{A} \). A boldface lowercase letter, such as \( \mathbf{a} \), symbolizes a vector. A lowercase letter, such as \( x \), denotes a scalar.

The GCP decomposition approximates a \( N \)-way observed tensor of size \( n_1 \times n_2 \times \ldots \times n_N \) by the sum of \( R \) rank-1 tensors as model \( \mathbf{X} \), where \( R \) is smaller or equal to the rank of tensor \( \mathbf{X} \), as illustrated in Figure 1. It is presented as follows [29]:

\[
\mathbf{X} = \sum_{r=1}^{R} \mathbf{a}^T_r \mathbf{A}_r \quad \text{for} \quad r=1,\ldots, R \quad \text{and} \quad k=1,\ldots, N, \quad \text{where} \quad \mathbf{A}_k \in \mathbb{R}^{I_k \times R},
\]

(1)

Where \( \mathbf{a}^T_r \) indicates the \( r \)th column of \( \mathbf{A}_k \) for all \( k=1,\ldots, N \) and \( r=1,\ldots, R \), and \( \mathbf{A}_k \) is the \( k \)-mode factor matrix of size \( I_k \times R \), \( k=1,\ldots, N \), consisting of \( R \) latent components or phenotypes vectors, expressed as follows:

\[
\mathbf{A}_k = \begin{bmatrix}
\mathbf{a}_1^T & \mathbf{a}_2^T & \cdots & \mathbf{a}_R^T
\end{bmatrix}
\]

(2)

Table 1. The notations used in this paper.

<table>
<thead>
<tr>
<th>Notations</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>Outer product</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of dimensions (modes) of a tensor</td>
</tr>
<tr>
<td>( R )</td>
<td>Number of ranks</td>
</tr>
</tbody>
</table>

In addition, it is often convenient to express the decomposition with a positive weights vector of \( \mathbf{\lambda} \) as follows:

\[
\mathbf{X} = \sum_{r=1}^{R} \mathbf{a}^T_r \mathbf{A}_r \mathbf{\lambda}_r \quad \text{for} \quad r=1,\ldots, R \quad \text{and} \quad k=1,\ldots, N, \quad \text{where} \quad \mathbf{\lambda}_r \in \mathbb{R}^R.
\]

(3)

The GCP decomposition was carried out by minimizing the loss between \( \mathbf{X} \) and \( \mathbf{\hat{X}} \) as represented by an objective function, which is also called a loss function. This means that finding factor matrices \( \mathbf{A}_k \) for \( k=1,\ldots, N \), with a given \( R \) such that solves the following optimization problem:

\[
\min_{\mathbf{A}_1,\ldots, \mathbf{A}_N} \| \mathbf{X} - \sum_{k=1}^{N} \mathbf{a}^T_k \mathbf{A}_k \|_F^2
\]

(4)

Where \( x_i = x(i_1,\ldots,i_N), m_i = m(i_1,\ldots,i_N) \), and \( \ell \) is the loss function. The choice of the loss function depends on how the original data are generated, which can be found in section S1 in Multimedia Appendix 1 [29].

Therefore, for a 3-way tensor \( \mathbf{X} \), the generalized CP decomposition with weight vector \( \mathbf{\lambda}=\mathbf{I} \) is represented as (for simplicity, \( \mathbf{A}, \mathbf{B}, \mathbf{C} \) notations are used rather than \( \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \)):

\[
\mathbf{X} = \sum_{r=1}^{R} \mathbf{a}^T_r \mathbf{A}_r \mathbf{\lambda}_r \quad \text{for} \quad r=1,\ldots, R \quad \text{and} \quad k=1,\ldots, N, \quad \text{where} \quad \mathbf{\lambda}_r \in \mathbb{R}^R.
\]

(5)

Where \( \mathbf{a}_r \) and \( \mathbf{a}_r \) are the \( r \)th column vectors within the patient factor matrix \( \mathbf{A}_1 \), and nonpatient factor matrices \( \mathbf{B}_2 \) and \( \mathbf{C}_3 \), respectively.

For simplicity, we discussed a 3-way tensor scenario; however, this approach generalizes to \( N \) modes as well. The GCP decomposes the observed longitudinal health data \( \mathbf{X} \) into 3-factor matrices: a patient factor matrix \( \mathbf{A} \) and 2 nonpatient factor matrices \( \mathbf{B} \) and \( \mathbf{C} \).
Data Synthesis Method

Overview

Our goal was to generate synthetic longitudinal health data, which we will refer to as $\mathbf{X}$. As mentioned earlier, relying on the model tensor $\mathbf{X}$ for synthesis is generally a bad idea and unwise because its elements directly approximate elements of the actual tensor $\mathbf{X}$ and thus disclosing privacy. Instead, we focused on the patient factor matrix $\mathbf{A}$ in the model’s latent space, which contains key phenotypes and information about patients.

In addition, the patient factor matrix $\mathbf{A}$ is not a longitudinal data set and is a significantly smaller data set than the model tensor $\mathbf{X}$; hence, its sampling and synthesis would be considerably simpler, and more efficient. Therefore, the aim was to find the optimal sampling or synthesis technique for the patient factor matrix $\mathbf{A}$, which is denoted by $\mathbf{A}$. As determined by our research and study, we may use one of the generative models listed in the following subsections to create $\mathbf{A}$. Thus, $\mathbf{A}$ is viewed as a synthetic form of $\mathbf{X}$, as illustrated in Figure 2:

(6)

For addressing missing observations in the GCP model, Hong et al [29] used an indicator for the missingness of observations by assigning weights 0 if an observation is missing or 1 if it is observed; this approach for handling missing data is essentially the same as the work of Acar et al [31]. In addition, the missing structure of synthetic data must be similar to that of the original data. We assigned weights 0 or 1 to every element of the tensor $\mathbf{X}$, $x_i$:

(7)

Let $\mathbf{w}$ be the missingness indicator or mask tensor made of weights $w_i$; we then treated it as an input tensor and decomposed it using GCP decomposition. Assume $\mathbf{A}$ is its patient factor matrix attained from the GCP decomposition with rank $R'$:

(8)

We proposed combining $\mathbf{A}_w$ and $\mathbf{A}$ to create a new patient factor matrix $\mathbf{A}$, from which we sample and simulate to obtain the synthesized and simulated form of $\mathbf{X}$. Then, we need to restore $\mathbf{A}$ and $\mathbf{w}$ to their source latent space to obtain the synthetic data and its missingness indicator tensor $\mathbf{X}$.

Furthermore, the EHR data might be an irregular tensor, with patients having a varied number of clinical visits. However, the input for CP and generalized CP decompositions must be a regular tensor. Therefore, we proposed converting the irregular observed tensor into the regular one by adding extra missing visits. We then performed the GCP decomposition and added the number of clinical visits as a new variable to the patient factor matrix $\mathbf{A}$ because this feature is not a longitudinal variable and can be directly added to the patient factor matrix. We then sampled or synthesized the $\mathbf{A}$ along with that variable, and obtained the synthetic number of records, and finally, we modified the number of records in the postprocessing. We also recommended applying the same to the baseline attributes. With this generative model, we can simulate different numbers of patients in synthetic data as well.
Figure 2. The generative model in terms of the generalized canonical polyadic decomposition.

**Copula**

This section summarizes the copula principles that we used in this study. Synthetic data generation using copula has recently attracted a lot of attention because some deep learning generative models, such as generative adversarial networks (GANs), require a very large data set for the learning stage and are therefore unproductive for small data sets. Copula models might be the most effective method to describe dependencies and marginal distributions. Furthermore, these models appear to be among the best options for data synthesis based on complex and small actual data sets [32]. Therefore, we selected the Gaussian copula as one of the simulation and synthesizing techniques for the patient factor matrix as it has been represented in several papers [33,34].

We assumed that have marginal distributions $F_1,...,F_R$ parametric or nonparametric, with covariances $\Sigma$. By a Gaussian copula, we completed the following:

1. Generated variables $t_1,...,t_R$ from a multivariate normal distribution with means all equal to 0, variances all equal to 1, and $\Sigma$.
2. Generated the uniform variables such that: the Gaussian copula is the joint distribution of $u_1,...,u_R$
3. Generated samples $\tilde{x}$.

However, the sampling will never be exact, and because we were sampling in the latent space, the correlation of the synthetic tensor might not be what we want. Therefore, we suggest selecting a sample such that the Frobenius norm of the difference between the correlation matrices of the original and synthetic data is less than a threshold $\epsilon$, which can be viewed as an optimization.

Finally, we produced the synthetic patient factor matrix from the obtained samples, such that $\tilde{x}$.

**Sequential Decision Trees**

Another technique that we proposed is to synthesize the patient factor matrix $A$ using sequential decision trees, as presented in [35]. In data synthesis, sequential trees outperform deep learning methods, such as GANs or recurrent neural network models when the data set is not huge. We briefly summarize the sequential data synthesis process of [35]. Let us assume that the variables in the patient factor matrix $A$ are $\tilde{x}$, and their corresponding synthesized versions are $x$. We fit $R-1$ models using a model $M_j$ defined by $\tilde{a}_j$ where $M_j$ and $f$ denote a decision tree model and the tree model fitting function, respectively. Finally, we sampled $\tilde{a}_j$ from $a_j$ and applied $\tilde{a}_j$ to sample the remaining variables.

**Hamiltonian Monte Carlo**

In this section, we sampled from the patient factor matrix $A$ using a Bayesian simulation, with a focus on Hamiltonian Monte Carlo (HMC), which has advantages over other Markov Chain Monte Carlo (MCMC) methods.

We began by assuming that the patient factor matrix variables $\tilde{x}$ are generated from a multivariate Gaussian distribution prior with a known mean vector $\tilde{a}_i$ (which is set to a vector of zeros for simplicity) and an unknown covariance matrix $\Sigma$. A suitable prior for variances would be the Cauchy distribution. However, we must consider that the distribution of patient factor variables depends on the kind of loss function used in the GCP decomposition. We used this method only on the latent space generated by Gaussian loss.

The model block of STAN is provided in section S2 in Multimedia Appendix 1. In this block, $x$ and $x_{\text{sim}}$ represent the patient factor matrix variables $a_i$ and the corresponding simulations $a_{\text{sim}}$, respectively. $N$ indicates the number of patients.

**Data**

This study used the MIMIC-III data set [19]. MIMIC-III is a large, freely available database that contains deidentified health-related data corresponding to >40,000 patients who stayed in critical care units. It was compiled at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2001 and 2012. Therefore, as the largest publicly available EHR data set, it has received significant interest from researchers as an open platform for the development and validation of their research on EHRs.

The continuous data set that we derived from the MIMIC-III data and used to evaluate the performance of our proposed model is a 3-way tensor of laboratory measurements for patients within the hospital who had 36 clinical visits. This is derived from the “LABEVENT” table in MIMIC-III. The resulting tensor consists of 226 patients, 4 laboratory tests (creatinine, potassium, sodium, and the hematocrit), and 36 clinical visits, with 21% of the data missing. The categorical data set used in
our analysis is a tensor consisting of 246 patients, 2 categorical features (admission type and admission location), and 5 clinical visits, with no missing entries. This was obtained from the “ADMISSIONS” table of MIMIC-III. The detailed descriptions of the tables can be found in the study by Johnson et al [19].

**Evaluation Methods**

We describe how we evaluated the utility and privacy risks of the synthetic data set.

The analysis of synthetic data should provide similar statistical inferences and conclusions to those obtained from actual data analysis. Therefore, we evaluated the proposed method on the utility aspect of the generated data [20,36].

We assessed the ability of the developed generalized CP framework to create synthetic data in terms of dependency structure and marginal fitting (univariate distribution similarity) using the following:

- The absolute difference in correlations between variables in the original and synthetic data
- The Hellinger distance between the synthetic and original variables indicates whether they are drawn from the same distribution. The Hellinger distance is a metric in the range of 0 to 1, where 0 indicates no difference between the distributions.
- The root mean square difference between the actual and synthetic correlations (RMSDC), which means the root mean square difference between the correlations of the original variables and those of the associated synthetic variables, is used to measure how well the dependency structure is captured. A lower RMSDC indicates a better capture of the dependency structure.
- Descriptive statistics

The statistical characteristics of a synthetic data set need to match those of the original data. However, a single record in the synthetic data does not relate to or correspond to a single record in the original data set [13]. There is a one-to-one correspondence between the entries of the GCP model and the original data, which violates the most reasonable and accepted concepts of privacy. Therefore, in terms of privacy disclosure, our approach preserves privacy by preventing the direct sharing of patient information and destroying the one-to-one link between the observed and model tensor records.

**Experimental Details**

We evaluated 3 different simulation and sampling techniques on the generalized CP’s patient factor matrix, which contains patient phenotypes. Initially, we arranged the study into trials on dense and scarce continuous data sets and experiments on dense categorical data. Therefore, we started imputing the continuous data set with 21% missing observations using the GCP decomposition. We then used all 3 previously mentioned techniques to synthesize the imputed version of the original tensor, as well as the imputed data containing just the first 5 or 10 clinical visits.

The GCP decomposition was conducted using various loss functions depending on the kind of variables in our trials. These included gamma, β-divergence (similar to gamma), and Gaussian with nonnegativity constraints. Despite the continuous data set being nonnegative, the Gaussian loss (L2 loss) outperformed the others for all 3 approaches. This is because the dependency structure and univariate distributions of the original variables were significantly better preserved in the generated synthetic data. Refer to section S1 in Multimedia Appendix 1 for the choices of the loss function.

Finding the rank of a tensor is necessary for GCP decomposition. Following the selection of the loss function, we attempted to find the rank R by doing several runs with different values of R, where R ≤ min{IJ,JK,JK}, keeping in mind that the maximum rank is the last option. We then selected the rank for which there were no significant changes in the objective function, fit score, or mean square error from that rank to higher ranks. This allows important features and phenotypes to be captured by the model. This method is similar to the “elbow” rule. Although we may end up with the maximum value for R, the privacy constraints are not violated because the latent space simulation prevents overfitting. In addition, we investigated if the latent space structure, such as normalization and standardization, would lead to better results.

The GCP tensor decomposition was conducted using the algorithm of Hong et al [29], which was implemented in the Tensor Toolbox for MATLAB [37]. There is also a C++ programming software for GCP decompositions developed by Sandia National Laboratories [38], known as Genten, which is accessible in GitLab [39].

We applied the copula method using parametric and nonparametric marginals. For the nonparametric marginals, we used the empirical cumulative distribution function (CDF) and kernel smoothing. In addition, we performed the Gaussian copula based on parametric marginals using gamma, beta, and truncated Gaussian distributions, as suggested in the study by Benali et al [32].

We successfully synthesized the patient factor matrix A based on the sequential decision trees [35] through Replica Synthesis software [40]. The STAN model and the rstan package in R were used to perform HMC sampling on the patient factor matrix.

Finally, we applied sequential trees to validate the generative model on the data set with missing and irregular clinical visits. The categorical data trial was accomplished by GCP decomposition using Poisson with log link and Gaussian losses. Following that, the patient factor matrix was sampled using all 3 approaches.

**Ethical Considerations**

This study was approved by the Children’s Hospital of Eastern Ontario (CHEO) Research Institute Research Ethics Board (protocol number 24/18X). The MIMIC-III database is a third-party anonymous public database approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, Massachusetts) and the Massachusetts Institute of Technology (Cambridge, Massachusetts).
Results

Overview

In this section, we report and analyze the results of our experiments on generating synthetic longitudinal health data using the proposed model. We demonstrate that our method is capable of handling different data structures and scenarios. We conducted numerous experiments and considered the following structures for both the original data and the synthetic data to validate our model:

- Synthetic data for dense original data with continuous variables
- Synthetic data with a varying number of patients compared to the original data
- Synthetic data for original data with missing observations
- Synthetic data for original data with irregular clinical visits
- Synthetic data for dense original data with categorical variables

Experiments on Continuous Dense Data

According to the evaluations, β-divergence is not a suitable objective function for synthesizing continuous data in EHR with patient factor matrix simulation using sequential trees or HMC with the multivariate Gaussian distribution model. Refer to section S3 in Multimedia Appendix 1 for a summary of the findings.

We learned through several analyses that standardizing data and using Gaussian loss improves the results significantly. In the following, we present the outcomes of synthesizing a dense continuous data set that contains 226 patients, 4 laboratory test variables, and 5 clinical visits. In the preceding sections, we described how we obtained the data set for our study. The model used GCP decomposition with Gaussian loss and R=20. It can be observed that synthetic data sets generated by all 3 patient factor matrix simulation methods have comparable dependency structures and marginal distributions to the real ones. Here, we present the results of copula, sequential trees, and the HMC.

According to section S4 in Multimedia Appendix 1, the synthetic data generated from the copula using empirical CDF marginals almost preserves the dependency structure and distribution of the original variables. The box plots in Figure 3 represent the variation in the Hellinger distance and the Pearson correlation between the variables in both the synthetic and original data sets. Furthermore, an RMSDC of 0.04 was obtained.

According to the summaries presented in Tables 2 and 3, the maximum value of the variables is slightly lower than that of the original.

The following are the outcomes of sampling the patient factor matrix of the previously mentioned GCP decomposition using the sequential trees approach. According to section S5 in Multimedia Appendix 1, the structure of the original data was preserved in different modes upon synthesis.

Figure 4 of the variation of the Hellinger distance also shows that synthetic variables from sequential decision trees are derived from a similar distribution as the original variables. However, the copula performed slightly better in capturing the correlations between the variables. The RMSDC computed for this experiment was 0.078.

The summary presented in Table 4 shows that the range of variables has significantly improved compared to the previous copula analysis, refer to Table 3 for the summary of the original data.

The following is the result of MCMC method using the HMC algorithm: if the distribution of the HMC model is well specified, the resulting outcome will be more effective.

Section S6 in Multimedia Appendix 1 indicates that the structure of the synthetic data in different modes is comparable to that of the original data. Furthermore, an RMSDC of 0.071 was obtained. We expected the HMC to perform well on the Gaussian latent space, which is defined by our Gaussian model. Defining a proper model distribution for the HMC would significantly enhance the findings. Figure 5 illustrates that HMC was slightly better at capturing the correlations between variables.

On the basis of Table 5, the summary of HMC synthetic variables is similar to that of the other methods. The maximum values of the variables dropped in the synthetic data compared to the actual data in Table 3.

Finally, we present Figure 6 and section S7 in Multimedia Appendix 1 for an easier comparison of the 3 sampling techniques on GCP decomposition with Gaussian loss. In Figure 6, we can observe the dependency structure and univariate distribution of both synthetic and original variables simultaneously.

The findings and figures demonstrate that all 3 synthetic data sets have similar statistical properties in terms of dependency and univariate distributions. However, the copula and sequential trees performed slightly better than the MCMC technique when the HMC algorithm was used. In brief, we need to define a proper distribution that corresponds to the latent space to obtain decent results through HMC sampling. The Gaussian loss is the ideal loss function for simulating from the patient factor matrix using sequential trees, but the observed data must be standardized beforehand. Further analysis, which we have not included in this paper, has shown that it is preferable to use empirical CDF marginals instead of parametric ones when sampling the patient factor matrix by copula. The outcomes of generating synthetic data using β-loss in GCP decomposition can be found in section S3 in Multimedia Appendix 1.

In the following section, we will provide the results of generating synthetic data with an additional number of patients compared to the original data set.
Figure 3. The box plots display the variation of the Hellinger distance and the Pearson correlation between the original variables and the synthetic variables generated by copula. (A) This is the box plot of the absolute differences in bivariate correlations between the real and synthetic data. Smaller values indicate that the bivariate relationships in the data have been greatly preserved during the generation of synthetic data. (B) This is the box plot of the Hellinger distance between the original variables and the synthetic variables. This shows the similarity of the univariate distributions between the real and synthetic data. This is a value between 0 and 1, with lower values indicating similarity between the univariate distributions of the real and synthetic variables.

Table 2. A summary of the copula’s synthetic variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Hematocrit</th>
<th>Sodium</th>
<th>Potassium</th>
<th>Creatinine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimuma</td>
<td>0</td>
<td>110.6</td>
<td>2.23</td>
<td>10.6</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.08 (0.52-2.21)</td>
<td>4.17 (3.77-4.62)</td>
<td>138 (134.5-141.9)</td>
<td>4.17 (3.77-4.62)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.67 (1.58)</td>
<td>4.22 (0.59)</td>
<td>138 (6.2)</td>
<td>1.08 (0.52-2.21)</td>
</tr>
<tr>
<td>Maximumb</td>
<td>14.1</td>
<td>7.32</td>
<td>157.5</td>
<td>48.49</td>
</tr>
</tbody>
</table>

aMinimum: minimum of data.
bMaximum: maximum of data.

Table 3. A summary of the original variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Hematocrit</th>
<th>Sodium</th>
<th>Potassium</th>
<th>Creatinine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimuma</td>
<td>0.2</td>
<td>109</td>
<td>2.5</td>
<td>9.2</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1 (0.7-1.6)</td>
<td>4.1 (3.8-4.51)</td>
<td>138 (135-141.6)</td>
<td>109</td>
</tr>
<tr>
<td>mean (SD)</td>
<td>1.64 (1.6)</td>
<td>4.23 (0.62)</td>
<td>138.4 (6.5)</td>
<td>9.2</td>
</tr>
<tr>
<td>Maximumb</td>
<td>16.2</td>
<td>10</td>
<td>170</td>
<td>52.6</td>
</tr>
</tbody>
</table>

aMinimum: minimum of data.
bMaximum: maximum of data.
Figure 4. The box plots display the variation of the Hellinger distance and the Pearson correlation between the original variables and the synthetic variables generated by sequential decision trees. (A) This is the box plot of the absolute differences in bivariate correlations between the real and synthetic data. Smaller values indicate that the bivariate relationships in the data have been greatly preserved during the generation of synthetic data. (B) This is the box plot of the Hellinger distance between the original variables and the synthetic variables. This shows the similarity of the univariate distributions between the real and synthetic data. This is a value between 0 and 1, with lower values indicating similarity between the univariate distributions of the real and synthetic variables.

![Box plots](image)

Table 4. A summary of the sequential decision trees’ synthetic variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Variables</th>
<th>Creatinine</th>
<th>Potassium</th>
<th>Sodium</th>
<th>Hematocrit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum⁴</td>
<td>0</td>
<td>1.88</td>
<td></td>
<td>116.4</td>
<td>10.09</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.25 (0.67-1.86)</td>
<td>4.18 (3.7-4.71)</td>
<td>138.8 (134.2-143.3)</td>
<td>31.68 (27.01-36.38)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.53 (1.62)</td>
<td>4.23 (0.58)</td>
<td>138.7 (6.6)</td>
<td>31.87 (5.8)</td>
<td></td>
</tr>
<tr>
<td>Maximum⁵</td>
<td>14.84</td>
<td>8.16</td>
<td></td>
<td>177.4</td>
<td>54.8</td>
</tr>
</tbody>
</table>

⁴Minimum: minimum of data.
⁵Maximum: maximum of data.

Figure 5. The box plots display the variation of the Hellinger distance and Pearson correlation between the original variables and the Hamiltonian Monte Carlo synthetic variables. (A) This is the box plot of the absolute differences in bivariate correlations between the real and synthetic data. Smaller values indicate that the bivariate relationships in the data have been greatly preserved during the generation of synthetic data. (B) This is the box plot of the Hellinger distance between the original variables and the synthetic variables. This shows the similarity of the univariate distributions between the real and synthetic data. This is a value between 0 and 1, with lower values indicating similarity between the univariate distributions of the real and synthetic variables.

![Box plots](image)
Table 5. A summary of the Hamiltonian Monte Carlo’s synthetic variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Creatinine</th>
<th>Potassium</th>
<th>Sodium</th>
<th>Hematocrit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum(^a)</td>
<td>0</td>
<td>0.66</td>
<td>117.8</td>
<td>7.22</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.85 (0.6-3.1)</td>
<td>4.19 (3.74-4.7)</td>
<td>138.6 (134.5-142.8)</td>
<td>32.18 (27.85-36.47)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>2.01 (1.4)</td>
<td>4.23 (0.54)</td>
<td>138.5 (5.02)</td>
<td>31.93 (4.8)</td>
</tr>
<tr>
<td>Maximum(^b)</td>
<td>7.24</td>
<td>6.78</td>
<td>156.9</td>
<td>51.63</td>
</tr>
</tbody>
</table>

\(^a\)Minimum: minimum of data.
\(^b\)Maximum: maximum of data.

Figure 6. The plots show the correlation and distribution of the original and synthetic variables. A correlation matrix displays bivariate scatter plots of the adjacent variables below the diagonal, histograms of the data distribution of the respective variables on the diagonal, and the Pearson correlation above the diagonal. Ellipses specify the direction of the correlation. The information regarding the relationship between the 2 selected variables is always perpendicular to each other. (A) This is the plot of the original variables. (B) This is the plot of synthetic variables generated by copula. (C) This is the plot of synthetic variables generated by sequential decision trees. (D) This is the plot of synthetic variables generated by Hamiltonian Monte Carlo.

Experiments on Continuous Data With a Different Number of Patients in Synthetic Data Compared to Original Data

We can generate different numbers of patients in the synthetic data using all 3 patient factor matrix simulations. However, we only applied the sequential trees approach, and the results are as follows: the findings indicate that the dependency and univariate structure of the original variables are well maintained in the synthetic data.

The following are the outcomes of generating 250 patients from the patient factor matrix using sequential trees: the patient factor matrix is derived from the GCP decomposition in the previous experiment. The original data set is dense, consisting of 226...
patients, and is the same data set used in the previous experiment.

Figure 7 indicates that the Pearson correlations and univariate distributions were similar between the synthetic and original data sets. Figure 8 illustrates that the synthetic variables created by sequential decision trees are derived from a distribution similar to that of the original variables. The structure of the generated data in different modes is included in section S8 in Multimedia Appendix 1, which shows similar results to those of the analyses presented here. The RMSDC computed for this experiment was 0.066.

In this scenario, the summary presented in Table 6 also demonstrates that the ranges of the synthetic variables are comparable to those of the original ones in Table 3.

On the basis of the results of this section, we are optimistic that our model may perform even better when generating larger data sets.

**Figure 7.** The plots show the correlation and distribution of sequential decision trees’ synthetic variables as well as the original variables. A correlation matrix displays bivariate scatter plots of the adjacent variables below the diagonal, histograms of the data distribution of the respective variables on the diagonal, and the Pearson correlation above the diagonal. Ellipses specify the direction of the correlation. The information regarding the relationship between the 2 selected variables is always perpendicular to each other. (A) This is the plot of the original variables. (B) This is the plot of the synthetic variables.

**Figure 8.** The box plots display the variation of the Hellinger distance and the Pearson correlation between the original variables and the sequential decision trees’ synthetic variables. (A) This is the box plot of the absolute differences in bivariate correlations between the real and synthetic data. Smaller values indicate that the bivariate relationships in the data have been greatly preserved during the generation of synthetic data. (B) This is the box plot of the Hellinger distance for all variables between the original and synthetic data sets. This shows the similarity of the univariate distributions between the real and synthetic data. This is a value between 0 and 1, with lower values indicating similarity between the univariate distributions of the real and synthetic variables.
Table 6. A summary of sequential decision trees’ synthetic variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Variables</th>
<th>Creatinine</th>
<th>Potassium</th>
<th>Sodium</th>
<th>Hematocrit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0</td>
<td>2.16</td>
<td></td>
<td>108.3</td>
<td>8.97</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.22 (0.76-1.88)</td>
<td>4.18 (3.71-4.72)</td>
<td>138.4 (134.4-142.2)</td>
<td>31.44 (27.37-35.87)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.67 (1.6)</td>
<td>4.26 (0.61)</td>
<td>138.4 (5.6)</td>
<td>31.56 (5.83)</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>16.42</td>
<td>7.94</td>
<td></td>
<td>163.1</td>
<td>53.04</td>
</tr>
</tbody>
</table>

a Minimum: minimum of data.
b Maximum: maximum of data.

Experiments on Continuous Data With Missing Observations

The following is a trial on the continuous data derived from the MIMIC-III data set without imputation. As mentioned earlier, the data set consists of 226 patients, 4 laboratory tests, and 36 clinical visits, with 21% of the observations missing. We have previously described the approach for synthesizing this sort of data. In this experiment, we attempted to sample the patient factor matrix using sequential trees and generate the same sample size of 226 patients as in the original data. We performed GCP factorization with Gaussian loss and \( R = 100 \). Then, we applied the CP via an alternating least square decomposition to the missing tensor with \( R = 50 \), where the loss function was also Gaussian.

The structure of the synthetic and original data sets in different modes is similar, as shown in section S9 in Multimedia Appendix 1. Furthermore, Figures 9 and 10 indicate that the dependency structure is preserved, and the marginal fitting is quite comparable in both the synthetic and original data sets. This demonstrates that our proposed model for EHR synthesis performs well.

According to the summary presented in Tables 7 and 8, the missing percentages in the synthetic and actual data are 46% and 21%, respectively. It appears to be almost double. As the missing tensor is binary, we can try to factorize it to determine if there are any possible improvements in the outcomes.

Figure 9. The plots show the correlation and distribution of sequential decision trees’ synthetic variables as well as the original variables. A correlation matrix displays bivariate scatter plots of the adjacent variables below the diagonal, histograms of the data distribution of the respective variables on the diagonal, and the Pearson correlation above the diagonal. Ellipses specify the direction of the correlation. The information regarding the relationship between the 2 selected variables is always perpendicular to each other. (A) This is the plot of the original variables. (B) This is the plot of the synthetic variables.
Figure 10. The box plots display the variation of the Hellinger distance and the Pearson correlation between the original variables and the sequential decision trees’ synthetic variables. (A) This is the box plot of the absolute differences in bivariate correlations between the real and synthetic data. Smaller values indicate that the bivariate relationships in the data have been greatly preserved during the generation of synthetic data. (B) This is the box plot of the Hellinger distance for all variables between the original and synthetic data sets. This shows the similarity of the univariate distributions between the real and synthetic data. This is a value between 0 and 1, with lower values indicating similarity between the univariate distributions of the real and synthetic variables.

Table 7. A summary of the sequential decision trees’ synthetic variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Creatinine</td>
</tr>
<tr>
<td>Minimum(^a)</td>
<td>0</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.3 (0.85-1.87)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.47 (0.97)</td>
</tr>
<tr>
<td>Maximum(^b)</td>
<td>7.97</td>
</tr>
<tr>
<td>#NA’s(^c)</td>
<td>3893</td>
</tr>
</tbody>
</table>

\(^a\)Minimum: minimum of data.
\(^b\)Maximum: maximum of data.
\(^c\)#NA’s: the number of missing values.

Table 8. A summary of the original variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Creatinine</td>
</tr>
<tr>
<td>Minimum(^a)</td>
<td>0</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1 (0.6-1.6)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.52 (1.67)</td>
</tr>
<tr>
<td>Maximum(^b)</td>
<td>16.2</td>
</tr>
<tr>
<td>#NA’s(^c)</td>
<td>2050</td>
</tr>
</tbody>
</table>

\(^a\)Minimum: minimum of data.
\(^b\)Maximum: maximum of data.
\(^c\)#NA’s: the number of missing values.

Experiments on Continuous Data With Irregular Clinical Visits

To create irregularity in clinical visits, a subset of the continuous data was created by choosing the first 10 observations. Those outcomes are presented in this section, and we have previously elaborated on the method used for addressing this particular scenario. The GCP decomposition was performed with Gaussian loss and \( R=30 \), resulting in a mean square error of approximately 0.004. The experiment was conducted using the sequential decision trees approach to sample the patient factor matrix. The RMSDC was computed as 0.14.
Upon analyzing the results in Figures 11 and 12, we found that the process of generating synthetic data generally maintained bivariate relationships and univariate distributions in the data. When analyzing Tables 9 and 10, it was found that the provided descriptive statistics have remained comparable throughout the process of generating synthetic data.

**Figure 11.** The plots show the correlation and distribution of variables generated by sequential trees and the original ones. A correlation matrix displays bivariate scatter plots of the adjacent variables below the diagonal, histograms of the data distribution of the respective variables on the diagonal, and the Kendall correlation above the diagonal. Ellipses specify the direction of the correlation. The information regarding the relationship between the 2 selected variables is always perpendicular to each other. (A) This is the plot of the original variables. (B) This is the plot of the synthetic variables.

![Figure 11](image)

**Figure 12.** The box plots display the variation of the Hellinger distance and the Kendall correlation between the original variables and the sequential decision trees’ synthetic variables. (A) This is the box plot of the absolute differences in bivariate correlations between the real and synthetic data. Smaller values indicate that the bivariate relationships in the data have been greatly preserved during the generation of synthetic data. (B) This is the box plot of the Hellinger distance for all variables between the original and synthetic data sets. This shows the similarity of the univariate distributions between the real and synthetic data. This is a value between 0 and 1, with lower values indicating similarity between the univariate distributions of the real and synthetic variables.

![Figure 12](image)

**Table 9.** A summary of the sequential decision trees’ synthetic variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Variables</th>
<th>Potassium</th>
<th>Sodium</th>
<th>Hematocrit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (^a)</td>
<td>Creatinine</td>
<td>0</td>
<td>100.3</td>
<td>10.62</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td></td>
<td>0.87</td>
<td>139.8</td>
<td>31.43 (27.02-35.21)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td></td>
<td>10.3</td>
<td>139.5</td>
<td>31.36 (5.7)</td>
</tr>
<tr>
<td>Maximum (^b)</td>
<td></td>
<td>9.35</td>
<td>178.9</td>
<td>58.15</td>
</tr>
</tbody>
</table>

\(^a\)Minimum: minimum of data.  
\(^b\)Maximum: maximum of data.
Table 10. A summary of the original variables.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Variables</th>
<th>Metric</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Creatinine</td>
<td>Potassium</td>
<td>Sodium</td>
</tr>
<tr>
<td>Minimum&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.2</td>
<td>2.6</td>
<td>111.2</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>1.07 (0.76-1.7)</td>
<td>4.15 (3.8-4.5)</td>
<td>139 (135.6-142)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.6 (1.64)</td>
<td>4.21 (0.62)</td>
<td>139 (6.53)</td>
</tr>
<tr>
<td>Maximum&lt;sup&gt;b&lt;/sup&gt;</td>
<td>13.6</td>
<td>7.9</td>
<td>170</td>
</tr>
</tbody>
</table>

<sup>a</sup>Minimum: minimum of data. 
<sup>b</sup>Maximum: maximum of data.

Experiments on Categorical Dense Data

The categorical data contain 2 variables: “admission type” and “admission location.” The GCP decomposition was implemented using 2 different loss functions: the Poisson log link, the results of which we discuss in section S10 in Multimedia Appendix 1, and the Gaussian loss function. Initially, a series of postprocessing steps were conducted. The outcomes of the synthesis, which apply the Gaussian loss function and use HMC, are presented in the subsequent section. In this experiment, \( R = 10 \) was obtained.

The structure of the generated data in different modes is included in section S11 in Multimedia Appendix 1. The Hellinger distance and Kendall correlation were calculated for the categorical variables as shown in Figures 13 and 14. All the findings demonstrate that the generative model is applicable to any sort of variable.

Figure 13. The plots show the Kendall correlation and distribution of the Hamiltonian Monte Carlo’s synthetic variables as well as the original variables. A correlation matrix displays bivariate scatter plots of the adjacent variables below the diagonal, a bar chart of the data distribution of the respective variables on the diagonal, and the Kendall correlation above the diagonal. Ellipses specify the direction of the correlation. (A) This is the plot of the original variables. (B) This is the plot of the synthetic variables.
Figure 14. The box plot shows the variation of Hellinger distance for all variables between the original and the Hamiltonian Monte Carlo synthetic data sets. This shows the similarity of the univariate distributions between the real and synthetic data. This is a value between 0 and 1, with lower values indicating similarity between the univariate distributions of the real and synthetic variables.

Discussion

Summary

Our objective was to develop and validate a generative model that produces synthetic longitudinal health data. We constructed a model by using a GCP tensor decomposition and sampling from its latent factor matrix, which contains factors related to patients.

We applied the GCP decomposition because tensor decompositions offer interpretability and flexibility in handling high-dimensional data, including massive and heterogeneous EHR data sets. However, the most sensible and acceptable privacy concepts were undermined because of the one-to-one mapping and direct correspondence between the entries of the GCP model and the entries of the original data. Thus, by simulating and modeling the latent factor matrix of GCP decomposition associated with patients, we could address privacy concerns.

We proposed 3 methods for synthesizing and simulating the patient’s factor matrix: sequential trees, Gaussian copula, and HMC. These techniques appear to be the best options for data synthesis and simulation, particularly when working with complex and small data sets, such as the patient factor matrix in our model.

The model was validated through several experiments conducted on various data structures. We assessed the similarity between our synthetic data and the real data by conducting utility assessments. The assessments involved evaluating the structure and general patterns present in the data, such as the dependency structure, the descriptive statistics, and the marginal distributions.

Limitations

In this study, we were not able to use the huge data set. In addition, we could not investigate further simulation and sampling techniques for the patient factor matrix due to time constraints. We focused on longitudinal health data in our model.

However, there are also other types of longitudinal data, such as transactions in financial data sets, that occur over time.

Therefore, a future study could use a huge data set for this model and explore other techniques for synthesizing the patient factor matrix, such as GANs and recurrent neural network models. Another possible future work could involve conducting a more rigorous comparison between the original and synthetic data sets to evaluate both the generative model and the superior sampling approach. We could also look at the HMC sampling approach and see if we can improve its results by defining more appropriate distributions. We believe that our model could perform well on various types of longitudinal data sets. It would be interesting and valuable to carry out a future study to assess the effectiveness and feasibility of this model on different types of longitudinal data sets, such as financial data, as the current model has been developed and validated using health data.

Conclusions

There is an increasing demand to access EHRs for secondary analysis. Data synthesis is one method that can address this demand and satisfy privacy concerns simultaneously. The objective of this study was to develop and validate a generative model for producing synthetic longitudinal health data. This was achieved using GCP tensor decomposition and sampling its latent factor matrix, which contains patient factors. All the simulation methods used in the generative model provided the same high level of performance in certain experiments. However, the sequential decision trees performed better when data standardization was used, and the Gaussian loss was used in the generalized CP decomposition. When applied to a non-Gaussian latent space, the copula was preferred. Our approach could also solve the problem of sampling patients from EHRs. This means that we could simulate different numbers of patients in the synthetic data set as well. On the basis of our findings, we highly recommend the standardization and decomposition of EHRs using Gaussian loss. This will ensure that the synthetic data is a true reflection of the original data set.
We successfully addressed the challenge of synthesizing massive longitudinal health data by synthesizing a significantly smaller nonlongitudinal data set instead. Thus, it is encouraging that our generative model could be applied to produce valuable synthetic data in various fields and areas of research.

Tensor decompositions have drawn growing attention because of their interpretability and flexibility in high-dimensional and heterogeneous data sets. In addition, they can easily be privatized. The GCP decomposition is the most popular tensor decomposition technique, which is ideal for large-scale and heterogeneous data sets. It has various applications beyond health data analysis, such as in predicting financial markets. There are significant benefits to banks and financial institutions when it comes to generating and using synthetic data. For instance, they gain the ability to analyze and test data without any cybersecurity or privacy concerns, which is crucial and saves a tremendous amount of time.

Therefore, we believe that our model can efficiently apply to various types of longitudinal data, including generating synthetic longitudinal financial data, such as synthetic transactions.

Acknowledgments

AS is supported by the NSERC Discovery Grant Program.

Conflicts of Interest

KEE is the cofounder and SVP of Replica Analytics Ltd and has equity in this company. LM is a data scientist employed by Replica Analytics Ltd.

Multimedia Appendix 1

Detailed analysis results, loss functions, and algorithm descriptions.

References


Abbreviations

CDF: cumulative distribution function
CP: canonical polyadic
EHR: electronic health record
GAN: generative adversarial network
GCP: generalized canonical polyadic
HMC: Hamiltonian Monte Carlo
MCMC: Markov chain Monte Carlo
RMSDC: root mean square difference between the actual and synthetic correlations

©Elnaz Karimian Sichani, Aaron Smith, Khaled El Emam, Lucy Mosquera. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Momentary Factors and Study Characteristics Associated With Participant Burden and Protocol Adherence: Ecological Momentary Assessment

Allan D Tate1*, PhD; Angela R Fertig2*, PhD; Junia N de Brito3*, PhD; Émilie M Ellis1*, PhD; Christopher Patrick Carr1*, MPH; Amanda Trofholz3*, MPH, RD; Jerica M Berge4*, MPH, PhD, LMFT

1Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA, United States
2Humphrey School of Public Affairs, University of Minnesota, Minneapolis, MN, United States
3Department of Family Medicine and Community Health, University of Minnesota Medical School, University of Minnesota, Twin Cities, Minneapolis, MN, United States
4Department of Family Medicine and Adult and Child Center for Outcomes Research and Delivery Science, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

*all authors contributed equally

Corresponding Author:
Allan D Tate, PhD
Department of Epidemiology and Biostatistics
College of Public Health
University of Georgia
202 Miller Hall, 101 Buck Rd, Health Sciences Campus
Athens, GA, 30602
United States
Phone: 1 706 542 6317
Email: allan.tate@uga.edu

Abstract

Background: Ecological momentary assessment (EMA) has become a popular mobile health study design to understand the lived experiences of dynamic environments. The numerous study design choices available to EMA researchers, however, may quickly increase participant burden and could affect overall adherence, which could limit the usability of the collected data.

Objective: This study quantifies what study design, participant attributes, and momentary factors may affect self-reported burden and adherence.

Methods: The EMA from the Phase 1 Family Matters Study (n=150 adult Black, Hmong, Latino or Latina, Native American, Somali, and White caregivers; n=1392 observation days) was examined to understand how participant self-reported survey burden was related to both design and momentary antecedents of adherence. The daily burden was measured by the question “Overall, how difficult was it for you to fill out the surveys today?” on a 5-item Likert scale (0=not at all and 4=extremely). Daily protocol adherence was defined as completing at least 2 signal-contingent surveys, 1 event-contingent survey, and 1 end-of-day survey each. Stress and mood were measured earlier in the day, sociodemographic and psychosocial characteristics were reported using a comprehensive cross-sectional survey, and EMA timestamps for weekends and weekdays were used to parameterize time-series models to evaluate prospective correlates of end-of-day study burden.

Results: The burden was low at 1.2 (SD 1.14) indicating “a little” burden on average. Participants with elevated previous 30-day chronic stress levels (mean burden difference: 0.8; P=.04), 1 in 5 more immigrant households (P=.02), and the language primarily spoken in the home (P=.04; 3 in 20 more non-English–speaking households) were found to be population attributes of elevated moderate-high burden. Current and 1-day lagged nonadherence were correlated with elevated 0.39 and 0.36 burdens, respectively (P=.001), and the association decayed by the second day (β=0.08; P=.47). Unit increases in momentary antecedents, including daily depressed mood (P=.002) and across-day change in stress (P=.008), were positively associated with 0.15 and 0.07 higher end-of-day burdens after controlling for current-day adherence.

Conclusions: The 8-day EMA implementation appeared to capture momentary sources of stress and depressed mood without substantial burden to a racially or ethnically diverse and immigrant or refugee sample of parents. Attention to sociodemographic attributes (eg, EMA in the primary language of the caregiver) was important for minimizing participant burden and improving
data quality. Momentary stress and depressed mood were strong determinants of participant-experienced EMA burden and may affect adherence to mobile health study protocols. There were no strong indicators of EMA design attributes that created a persistent burden for caregivers. EMA stands to be an important observational design to address dynamic public health challenges related to human-environment interactions when the design is carefully tailored to the study population and to study research objectives.

* JMIR Form Res 2024;8:e49512 doi: 10.2196/49512

**KEYWORDS**

adherence; burden; data quality; ecological momentary assessment; mental health; mHealth; mobile health; participant adherence; public health; stress; study design; survey burden; survey

**Introduction**

**Overview**

Ecological momentary assessment (EMA), a type of mobile health (mHealth), is a class of assessment techniques that uses multiple daily assessments, often delivered through a smartphone or tablet, to a research participant to gain insight into a variety of states and behaviors at the time that they are experienced in a subject’s natural environment [1]. EMA has become a popular observational study design in the field of public health and provides numerous advantages when compared to traditional retrospective survey methods. The rich, intensive longitudinal data allows for harvesting insights from between-participant variation and within-person variation [1-3]. It minimizes recall bias and increases density in data points, which is especially important for momentary exposures with short induction periods, and it allows for data collection during real-world situations [1,4,5]. In some settings, computerized EMA delivery has replaced paper diary methods because it allows for time-stamped entries and efficient delivery to participant devices that they frequently use in their everyday environments (eg, tablets, smartphones, and computers). This has led to improved adherence to EMA protocols and higher data quality compared with paper diaries [6-11].

**Types of Adherence Considerations**

While there are many advantages to electronically delivered EMA, such as frequently measuring momentary events and outcomes, several important adherence-related considerations exist. First, researchers need to be cautious not to overburden participants. Frequent EMA may result in response conditioning and loss of interest in the study, which may significantly hamper adherence if participants experience EMA as intrusive to daily routines (eg, family dinner time or receiving surveys during work commutes) [12-14]. Second, protocol nonadherence (eg, a gap day) is likely nonrandom [11,15] and may be related to momentary exposures that are of interest to the study (eg, when stressors impair the ability of participants to fully engage with the design). Gap days may particularly impair analytic aims concerning across-day, exposure-outcome relationships (eg, rush hour traffic may disrupt meal planning for the following day). Minimizing observation gaps is critical to taking full advantage of EMA methodologies. Third, the characteristics of the study population are important to consider when building an EMA protocol. Young populations may have lower completion rates [16], participants may have complex work commitments that vary over days [17,18] which have been reported in our previous research [18], and respondents may have health conditions that could affect their responsiveness [19].

**Study Design-Related Burden Versus Burden From Momentary Sources**

The determinants of EMA adherence are critical to understand because nonadherence could affect the internal and external validity of the study. Low adherence is regularly interpreted as being the result of high participant burden caused by the study design, and measurement of participant burden has been typically characterized using proxy measures specific to the EMA instrument, including survey length, subjective respondent effort, and frequency of instrument delivery [20]. However, there are many factors occurring in a participant’s day-to-day life that may have a direct impact on adherence (eg, momentary high-stress states could affect adherence) [21]. Separating nonadherence due to momentary factors from the burden associated with objective design features may reveal new approaches for anticipating how EMA implementations can be difficult for participants.

This study examines sociodemographic and tempo-spatial correlates of daily participant-reported burden in interacting with an EMA instrument administered in a racially or ethnically diverse and immigrant or refugee sample over an 8-day period. The aims are (1) to examine population characteristics that may be associated with moderate-to-high participant-reported burden and (2) to examine the role of momentary determinants of burden. The study hypothesis associated with the first aim was that burden would be lower among those with both socioeconomic privileges and low momentary and chronic stress. The hypothesis related to the second aim is that burden caused by external or momentary sources, as evidenced by the report of the adults’ experiences of stressors, their mood, and the contextual environments, will present as a transient burden, and that burden caused by the EMA design will present as a high burden on multiple days in a row. The latter case may also indicate that the current study population may have experienced low momentary variability (ie, persistent or disruptive momentary exposures), making less demanding designs more attractive. The results of this study will inform EMA designs for observational studies and ecological momentary intervention studies by helping researchers anticipate what design elements may be appropriate for their target population to maximize momentary exposure and outcome assessment by minimizing participant burden.
Methods

Recruitment

The current study draws from Phase 1 (2015-2016) of the Family Matters Study [22], which is a 5-year incremental mixed methods (eg, video-recorded tasks, EMA, interviews, or surveys) study designed to identify novel risk and protective factors for childhood obesity in the home environments of racially or ethnically diverse and primarily low-income children. Phase 1 includes a cross-sectional examination of the family home environment of diverse families (n=150) with children aged between 5 and 7 years each from Black, Hispanic, Hmong, Native American, Somali, and White households (n=25 from each household). Phase 2 is a longitudinal cohort study with parent or child dyads (n=1307; children aged between 5 and 9 years) [22]. Participants were provided with the iPad minis (Apple Inc) by the study to take their EMA surveys and were then given them as an incentive for participating in the study along with US $100 in gift cards.

EMA Design

EMA was implemented using an iPad mini, on which parents completed signal, event, and end-of-day contingent EMA surveys multiple times throughout the day through a link on the iPad’s home screen in their preferred language (English, Spanish, Somali, or Hmong) [18]. “Signal-contingent surveys” were researcher-initiated and were used in a stratified random manner so that each parent was prompted to fill out a survey 4 times a day within a 3-hour time block (eg, 7 AM-10 AM, 11 AM-2 PM, 3 PM-6 PM, and 7 PM-10 PM) and were set to expire after 1 hour. Each signal-contingent survey was designed to be completed in under 1 minute to minimize response burden and included 10 questions about parent stress and mood, coping self-efficacy, and child eating behaviors and physical activity. “Event-contingent surveys” included 23 questions about a recently eaten meal (eg, who was present, food served, meal atmosphere, child eating behaviors, and food parenting practices) and were either participant-initiated or added to the beginning of a signal-contingent survey to catch any missed event-contingent surveys. Participants were instructed to complete an event-contingent mealtime survey every time they shared a meal with the study child. “End-of-day signal-contingent surveys” were designed to be completed in under 3 minutes and asked 31 questions about parent modeling of eating, physical activity and sedentary behavior, parent stress and mood levels, primary sources of stress, child eating, physical activity and sedentary behaviors, and the caregiver’s overall experience of burden in responding to EMA that day. Participants were given 6 hours to complete the end-of-day survey. Participants were alerted to signal-contingent and end-of-day surveys by an iPad beep or vibration; participants also had the ability to receive a corresponding reminder SMS text message. All EMA responses were time-stamped, and participants were assigned additional days to complete EMA to obtain a minimum of 8 full days of EMA data if they did not meet the minimum EMA responses per day. Based on best practice for EMA, minimum adherence to EMA was defined as submitting 2 signal-contingent, 1 event-contingent mealtime survey, and 1 end-of-day survey, totaling a minimum of 4 surveys per day. Other details regarding the study design unrelated to EMA are described elsewhere [22].

An analytic conceptual model is presented in Figure 1 to demonstrate how study design and momentary factors may affect early-day adherence that may be related to measured burden at the end of the day. We are not able to examine the relationship between EMA survey characteristics (different survey lengths or varying delivery frequency) and participant-reported burden because all participants experienced the same surveys (top part of Figure 1). Instead, we estimate the associations between momentary stressors and burden and attempt to identify the extent to which momentary stressors (bottom part of Figure 1), as opposed to survey characteristics, are responsible for participant-reported burden. In addition, we examine adherence and burden across time to assess whether burden is transient, indicating that momentary factors rather than the study design led to burden, versus persistently high burden over multiple days, thus signaling that the study design was burdensome.
**Cross-Sectional Survey Measures**

Sociodemographic characteristics of the study population and parent-reported chronic stress were measured by a cross-sectional web-based survey administered after the EMA data collection was complete. These sociodemographic characteristics included caregiver and child age, sex, and objectively-measured anthropometrics (adult BMI and sex- or age-adjusted child BMI percentile), child racial and ethnic group, caregiver nativity (ie, immigrant status), household caregiver structure (single parent with no other adults and with other adults, 2-parent with no other adults and with other adults; 4 categories total), the primary language spoken in the home (English, Spanish, Hmong, or Somali), sources of household income (eg, wages from self, public assistance, wages from other guardians, alimony or child support, and other sources), and average annual household income. Chronic stress was measured by the survey item, “On a scale of 1 to 10, with 1 being not stressed at all and 10 being very stressed, how would you rate your average level of stress in the past 30 days?”

**Participant Burden**

The primary study outcome measure was the overall participant-reported burden on the EMA observation day, measured on the end-of-day survey. The item asked, “Overall, how difficult was it for you to fill out the surveys today?” and was measured on a 5-item Likert scale where the increasing burden was measured as not at all, a little, moderately, quite a bit, and extremely. A trait-level participant burden score was computed to evaluate subpopulations that reported low burden (average was less than or equal to “a little”; n=63) and high burden (average was more than “a little”; n=87).

**Adherence to EMA Over the Course of the Day**

In the context of this study, nonadherent days were characterized as a completed end-of-day survey on which burden is reported but otherwise did not meet the minimum requirements for an adherent day (at least 2 signal-contingent surveys and at least 1 event-contingent mealtime survey). Each participant responded to at least 8 end-of-day surveys, resulting in 1391 participant-day observations, of which 151 (10.9%) days were nonadherent days. Approximately 65/150 (43%) caregivers in the sample had at least 1 nonadherent day. In some analyses, burden on subsequent days was examined, and 1114 day-pair observations (eg, Monday and Tuesday) of EMA burden were available for analysis. Average days to complete the EMA period (8 per protocol day) were computed for each participant to characterize gap-day frequency.

**Momentary Stressors**

Daily stress was measured over the course of the day through signal-contingent surveys (a minimum of 2 and a maximum of 4 each day for an adherent day). On all 151 nonadherent days on which end-of-day burden was measured, participants completed at least 1 daily signal-contingent survey on which momentary stress and mood were reported. Average daily stress (“Overall, how stressful was your day?”) was operationalized as the average of daily surveys on each observation day and could vary on each participant-EMA observation day for the duration of the EMA period. A change score was computed by

---

**Figure 1.** Conceptual model of ecological momentary assessment (EMA) study characteristics and momentary stressors on signal-contingent daily survey adherence and end-of-day participant burden. Sample of person-days restricted to those with at least 1 end of day survey to capture participant burden rating.
differencing the stress level on the following and current days to operationalize across-day stress trajectories. Daily depressed mood (“Overall, how SAD or DEPRESSED did you feel today?”) was operationalized in the same way as the caregiver’s average daily stress level, and both measures shared the same Likert scale as the burden outcome. A change score for depressed mood was not examined as a correlate of burden in inferential procedures to avoid multicollinearity with the primary predictor (stress change between days) and to disaggregate possible dependent measurement errors between the 2 variables. Depressed mood on the current day was used as an important control variable in the regression models. Change scores were not possible to compute for subsequent gap days with no daily signal-contingent surveys, resulting in 1392 available participant-day observations to compute the change in daily stress across days. On 1116 day-pairs, 262 (23.5%) days were measured as declining stress, 253 (22.7%) were increasing stress, and 601 (53.9%) had no change in stress. There were 2 day-pairs where stress change was measurable but end-of-day burden was not measured.

Statistical Analysis

Descriptive statistics were used to examine sociodemographic, stress, and EMA responsiveness correlates of low and high participant burden. A Fisher 2-sided exact statistical test was used for all categorical predictors of high participant burden and equal variance t tests were used to examine continuous variables (average EMA burden, average survey reported chronic stress and depressive symptoms, days to complete EMA, parent and child age, and adult and child BMI percentile).

Conditional fixed effects estimators were used to control for all nontime varying confounders and to examine within-caregiver momentary determinants of EMA burden (bottom part of Figure 1), specifically changes in daily stress level, depressed mood, and adherence that day. Adherence was included in the model to examine how partial completion of the EMA protocol may indicate momentary processes were the plausible causes of burden as opposed to study design factors in the case of a negative association. A second model was parameterized to examine how persistent the association is between an adherent day and burden on current and future days, adjusting for weekend versus weekdays. If persistence is high (eg, an adherent day is strongly associated with burden on current and future days), then the burden likely results from study design characteristics rather than momentary factors. A 2-sided statistical significance test was performed at the .05 level for all inferential analyses.

Burden transition probabilities over subsequent EMA observation days were computed to evaluate the population frequency of high-burden states within respondents. High frequency at low burden states and low frequency of persistent high burden over subsequent days were used to interpret if the reported burden was related to momentary stressors as opposed to the EMA instrument. A sensitivity-stratified analysis was computed among the per-protocol adherent caregivers (n=85) and the caregivers with at least 1 nonadherent day (n=65) to evaluate whether daily burden transitions were different for each group. There was no substantive difference in the across-day transition of burden levels for the 2 groups, and the full sample was combined for analysis. All data management and analysis were conducted in Stata (version 17.0 MP; StataCorp).

Ethical Considerations

The University of Minnesota Institutional Review Board approved Phase 1 of the Family Matters study, and all participants consented in accordance with the Declaration of Helsinki procedures (ID: 1107S02666) [23]. All participant records were deidentified for analysis to ensure participant privacy.

Results

Overview

The study population was majority female caregivers and 35 years old and children were 47% (71/150) female and 6.4 years old on average. The majority of parents 57% (85/150) were born in the United States, and immigrant parents (n=62) spoke a language other than English in the home about 65% (40/62) of the time. More than a quarter of households were single-caregiver homes, and 70% (105/150) of the full sample reported less than US $35,000 in annual household income (Table 1 shows participant demographics).
Table 1. Phase 1 Family Matters Study demographic and survey characteristics. Stratification by high or low ecological momentary assessment (EMA) burden participants in an ecological momentary study, 2015-2016 (N=150).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>None to low levels of EMA burden (n=63)</th>
<th>Moderate to high levels of EMA burden (n=87)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMA burden, mean (SD)</td>
<td>0.4 (0.4)</td>
<td>1.8 (0.5)</td>
<td>&lt;.001²</td>
</tr>
<tr>
<td>Chronic stress over previous 30 days, mean (SD)</td>
<td>3.5 (2.6)</td>
<td>4.3 (2.5)</td>
<td>.04²</td>
</tr>
<tr>
<td>Depressive symptoms in previous 2 weeks, mean (SD)</td>
<td>1.5 (1)</td>
<td>1.4 (0.7)</td>
<td>.65</td>
</tr>
<tr>
<td>Days to complete EMA, mean (SD)</td>
<td>9.6 (7)</td>
<td>11.8 (8)</td>
<td>.08</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>59 (94)</td>
<td>78 (90)</td>
<td>.56</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>34.2 (7.1)</td>
<td>34.7 (7.1)</td>
<td>.66</td>
</tr>
<tr>
<td>BMI, mean (SD)</td>
<td>30.3 (7.5)</td>
<td>31.3 (7)</td>
<td>.43</td>
</tr>
<tr>
<td><strong>Weight status, n (%)</strong></td>
<td></td>
<td></td>
<td>.74</td>
</tr>
<tr>
<td>Nonoverweight</td>
<td>16 (25)</td>
<td>19 (22)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>17 (27)</td>
<td>21 (24)</td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td>30 (48)</td>
<td>47 (54)</td>
<td></td>
</tr>
<tr>
<td>Born in the United States, n (%)</td>
<td>43 (68)</td>
<td>42 (48)</td>
<td>.02²</td>
</tr>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>33 (52)</td>
<td>38 (44)</td>
<td>.32</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>6.4 (0.8)</td>
<td>6.4 (0.8)</td>
<td>.61</td>
</tr>
<tr>
<td>BMI percentile, mean (SD)</td>
<td>76.5 (23.4)</td>
<td>75.5 (23.1)</td>
<td>.79</td>
</tr>
<tr>
<td><strong>Weight status, n (%)</strong></td>
<td></td>
<td></td>
<td>.86</td>
</tr>
<tr>
<td>Nonoverweight</td>
<td>32 (51)</td>
<td>45 (52)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>13 (21)</td>
<td>15 (17)</td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td>18 (29)</td>
<td>27 (31)</td>
<td></td>
</tr>
<tr>
<td><strong>Household characteristics</strong></td>
<td></td>
<td></td>
<td>.13</td>
</tr>
<tr>
<td>Household structure, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 parent (no other adults)</td>
<td>12 (19)</td>
<td>25 (29)</td>
<td></td>
</tr>
<tr>
<td>1 parent (with other adults)</td>
<td>6 (10)</td>
<td>12 (14)</td>
<td></td>
</tr>
<tr>
<td>2 parents (no other adults)</td>
<td>40 (63)</td>
<td>38 (44)</td>
<td></td>
</tr>
<tr>
<td>2 parents (with other adults)</td>
<td>5 (8)</td>
<td>12 (14)</td>
<td></td>
</tr>
<tr>
<td><strong>Family race or ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td>Black</td>
<td>12 (19)</td>
<td>13 (15)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>10 (16)</td>
<td>15 (17)</td>
<td></td>
</tr>
<tr>
<td>Hmong</td>
<td>13 (21)</td>
<td>12 (14)</td>
<td></td>
</tr>
<tr>
<td>Native American</td>
<td>11 (17)</td>
<td>14 (16)</td>
<td></td>
</tr>
<tr>
<td>Somali</td>
<td>4 (6)</td>
<td>21 (24)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>13 (21)</td>
<td>12 (14)</td>
<td></td>
</tr>
<tr>
<td><strong>Preferred language in home, n (%)</strong></td>
<td></td>
<td></td>
<td>.04²</td>
</tr>
<tr>
<td>English</td>
<td>49 (78)</td>
<td>56 (64)</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>7 (11)</td>
<td>9 (10)</td>
<td></td>
</tr>
<tr>
<td>Hmong</td>
<td>4 (6)</td>
<td>3 (3)</td>
<td></td>
</tr>
<tr>
<td>Somali</td>
<td>3 (5)</td>
<td>17 (20)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>0 (0)</td>
<td>2 (2)</td>
<td></td>
</tr>
</tbody>
</table>
The burden assessment is on a 0 to 4 scale, where 0 indicates no reported burden, 2 indicates moderate burden, and 4 indicates extreme burden on each end-of-day EMA assessment. Chronic stress is on a 1-10 scale, where 1 indicates not stressed and 10 indicates very stressed in the past 30 days. Depressive symptoms in the previous 2 weeks were measured on a 4-item Likert scale, where 1 indicated never or rarely and 4 indicated nearly every day feeling down, depressed, or hopeless. Both chronic stress and depressed mood were measured through parent survey self-reports. Family race or ethnicity are self-identified categories that the caregiver best characterized as the identity and culture of household members and the home environment.

### Research Question 1: What are the Sociodemographic Correlates of EMA Burden?

The overall reported EMA burden in the sample was low (Table 1). Parents who reported overall low burden generally reported not at all or a little daily burden of 0.4 (SD 0.4), and those who reported overall moderate-to-high burden generally reported a little burden to moderate daily burden of 1.8 (SD 0.5), which was statistically different at \( P < .001 \). Parent reports of chronic stress and depressed mood were measured through parent survey self-reports. Family race or ethnicity are self-identified categories that the caregiver best characterized as the identity and culture of household members and the home environment.

### Research Question 2: What are the Momentary Determinants of EMA Burden?

Depressed mood levels and changes in stress were examined as predictors of EMA burden after controlling for survey-day adherence (top panel in Table 2). There was strong evidence that increased stress levels from one day to the next (+0.07 burden; \( P =.008 \)) and within-day depressed mood (+0.15 burden; \( P =.002 \)) were each independently associated with elevated EMA burden.

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Burden level ( \beta ) coefficient</th>
<th>95% CIs</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in daily stress level</td>
<td>0.07*</td>
<td>0.02* to 0.13*</td>
<td>.008*</td>
</tr>
<tr>
<td>(current day less previous day)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressed mood level (current day)</td>
<td>0.15*</td>
<td>0.05* to 0.25*</td>
<td>.002*</td>
</tr>
<tr>
<td>Adherent day (current day)</td>
<td>-0.34*</td>
<td>-0.52* to -0.16*</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Participant adherence and burden; conceptual model: survey antecedents of burden</td>
<td>Survey submitted on a compliant day</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current-day adherent</td>
<td>-0.39*</td>
<td>-0.60* to -0.18*</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>1-day lag adherent</td>
<td>-0.36*</td>
<td>-0.56* to -0.16*</td>
<td>.001*</td>
</tr>
<tr>
<td>2-day lag adherent</td>
<td>-0.08</td>
<td>-0.28 to 0.13</td>
<td>.47</td>
</tr>
<tr>
<td>Weekend survey day (reference: weekday)</td>
<td>0.04</td>
<td>-0.08 to 0.16</td>
<td>.51</td>
</tr>
</tbody>
</table>

*Values significant at \( P <.05 \).
Per-protocol adherence was examined as a predictor of burden over 3 days (bottom panel in Table 2). Adherence on the current day was associated with a −0.39 burden (95% CI −0.60 to −0.18; \( P<.001 \)), and current-day adherence was also negatively associated with a next-day −0.36 burden (95% CI −0.56 to −0.16; \( P<.001 \)). This adherence pattern did not extend to the burden reported on the third day (−0.08 burden; \( P=.47 \)), and weekend days were not found to be more or less burdensome (\( P=.51 \)). Within-participant models for repeated measure data were fitted that control for all-time-invariant participant characteristics. Adjustment for time-varying predictors includes weekdays and the composition of participant daily study adherence.

The parent report of survey burden in the current day was −0.39 lower (95% CI −0.60 to −0.18) when the parent was adhering to the study protocol. Adherence was also associated with decreased survey burden on the following day (−0.36; 95% CI −0.56 to −0.16), but the adherence association with reduced burden dissipated by the second day. Weekends were not associated with a higher or lower survey burden compared to weekdays (\( P=.51 \)).

The transition between states of burden was examined in 1114 day-pairs (eg, Monday to Tuesday; Table 3). On 59.4% (662/1114) of days, there was no change in reported burden between days. Over day-pairs, both days were between “none” and “moderate” 82.5% (919/1114) of the time. The subsequent day transitioned from an elevated burden state of “quite a bit” or “extreme” to a lower burden state in 7% (79/1114) of days, and from a low burden state to an elevated burden state in 6.5% (73/1114) of days. Persistently high burden was low, as indicated by only 4% (43/1114) of day-pairs involving “quite a bit” or “extreme” burden on both days.

### Discussion

#### Overview

Findings from the current study overall suggest that the current EMA study was tailored to a racially or ethnically diverse and immigrant or refugee study population, and momentary processes were sufficiently captured (ie, stress responses) over a 1-week study duration. Specifically, there was no evidence that participants reported persistently high levels of burden that would indicate the EMA design elements for this study were burdensome for caregivers (research aim 1). Across-day variability in study burden indicated that human-environment interactions (eg, response to sources of stress) rather than human-instrument interactions (eg, frequency of daily survey delivery) were related to participant burden, suggesting that comparable protocols are appropriate for examination of momentary states in low-income, racially or ethnically diverse, and immigrant or refugee sample populations (research aim 2).

There was also evidence that some subpopulations may have trouble adhering to EMA studies. Participant characteristics that should be considered include chronic stress levels, immigrant status, and the language primarily spoken in the home—specifically non-English speaking households.

#### Importance of Population Characteristics for Protocol Development

Study population attributes continue to be important for the effective deployment of mHealth observational study designs. This study’s results support and extend previous studies that examined smartphone-based data collection among populations with affective disorders [24] and with substance-use dependency [25] in that mHealth studies interested in the relationship between momentary states and health and health behavior outcomes are feasible and pose low burden to participants. New findings from this study showed that factors affecting adherence to study protocol may include the spoken language of participants, indicating the critical importance of instruments that are tailored to the primary or preferred language of respondents, as well as the levels of strain and chronic stressors that a study population may experience during the observation period. Word recognition, for example, has been identified as

<table>
<thead>
<tr>
<th>Current-day burden level</th>
<th>None</th>
<th>A little</th>
<th>Moderate</th>
<th>Quite a bit</th>
<th>Extreme</th>
<th>Total (n=1114)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>315 (28.3)(^a)</td>
<td>44 (4)</td>
<td>58 (5.2)</td>
<td>10 (0.9)</td>
<td>6 (0.5)</td>
<td>433 (38.9)</td>
</tr>
<tr>
<td>A little</td>
<td>44 (4)</td>
<td>89 (8)(^a)</td>
<td>48 (4.3)</td>
<td>18 (1.6)</td>
<td>1 (0.1)</td>
<td>200 (18)</td>
</tr>
<tr>
<td>Moderate</td>
<td>52 (4.7)</td>
<td>44 (4)</td>
<td>225 (20.2)(^a)</td>
<td>26 (2.3)</td>
<td>12 (1.1)</td>
<td>359 (32.2)</td>
</tr>
<tr>
<td>Quite a bit</td>
<td>7 (0.6)</td>
<td>18 (1.6)</td>
<td>31 (2.8)</td>
<td>19 (1.7)(^a)</td>
<td>5 (0.5)</td>
<td>80 (7.2)</td>
</tr>
<tr>
<td>Extreme</td>
<td>8 (0.7)</td>
<td>6 (0.5)</td>
<td>9 (0.8)</td>
<td>5 (0.5)</td>
<td>14 (1.3)</td>
<td>42 (3.8)</td>
</tr>
</tbody>
</table>

\(^a\)Indicates no change in the state of EMA burden between one day and the next. Cells to the left of these indicate the frequency of days in which burden declined the following day, and cells to the right indicate the frequency of increasing burden days.

Interpretation example: Participants were required to submit at least 2 daily surveys, at least 1 mealtime survey, and 1 final survey in which the overall EMA burden was measured. On 59.4% (662/1114) of days, there was no change in reported burden between days, and over day-pairs, both days were between “none” and “moderate” 82.5% (919/1114) of the time. The subsequent day transitioned from an elevated burden state of “quite a bit” or “extreme” to a lower burden state in 7% (79/1114) of days, and from a low burden state to an elevated burden state in 6.5% (73/1114) of days. Persistently high burden was low, as indicated by only 4% (43/1114) of day-pairs involving “quite a bit” or “extreme” burden on both days.
an important determinant of study adherence in other studies
[26]. Although measures were translated for families who spoke
a language other than English, immigrant families experienced
a higher burden, suggesting that other factors and chronic
stressors unique to this subsample, such as acculturative stress,
may influence adherence to protocols [27]. With respect to
chronically stressed populations, research that addresses the
consequences of momentary stress on within- and across-day
behavioral- and health-related relationships should pay careful
attention to longitudinal patterns of stress level and other
measures of psychological distress (eg, depressive symptoms)
that may affect measurement and missingness. For example, in
a previous study, populations with substance-use dependency
were less adherent than nondependent samples, which can affect
both the internal validity and representativeness of findings for
other comparable groups [25]. Cultural barriers and a lack of
overall trust in health research may affect patterns of
missingness in EMA data that could further affect the
representativeness of study data on the intended population
[28,29]. In this study, similar adherence patterns were found
among chronically stressed populations compared to less
chronically stressed respondents. There was little evidence of
persistent, high burden, suggesting that some groups may
experience stressors at greater intensity and frequency than
others; further, the second-day lag, nonadherence-burden
association was observed to dissipate as a predictor of
subsequent burden 2 days later. This reality is why EMA can
be an effective tool for describing momentary
human-environment interactions.

In this study, the length of surveys, study duration, and
frequency of EMA assessments were intentionally limited to
minimize participant burden, and direct measurement of burden
supported that these design decisions were effective. Other
studies have made different decisions about these design factors
and had lower adherence and loss of follow-up. For example,
a 6-month study of food-related behaviors using a smartphone
data collection methodology found that adherence declined over
the course of the study [30]. Waving EMA observation intervals
to allow for resting states (eg, 1 week on, 1 week off, and then
1 week on again for a 2-wave design) may be one effective
strategy to improve confidence in measurement validity,
long-term retention of participants, and overall adherence in
EMA studies. Long study duration and high sampling frequency
were related to poor data quality in a study [31], and in another
study, a 10-week observation period found a similar decline in
adherence over time that was speculated to indicate a waning
interest in the study [32]. In the latter study, adherence improved
following direct interaction with providers and phone contact
with staff during the scheduling of follow-up visits. Attentive
support from the research team may buffer against participant
burden and increase adherence to the study protocol.

Momentary and Study-Related Stress and Burden
Direct measurement of study burden is an important tool for
assessing the feasibility of design elements that may be used in
future studies. Although proxy measures of burden, including
time to complete EMA and time to start a survey following
delivery of a survey notification, can be effective in assessing
measurement validity, more research is needed to assess how
respondents interact with the data collection instrument. In this
study, the receipt of participant feedback at the end of the day
was more effective in quantifying patterns of burden that may
be related to momentary factors, including stress sources and
the quality of caregiver-child interactions, than the study design
itself, which improves the replicability of results in future
implementations. As observed in this study, measured levels of
low-to-moderate burden may be good indicators that respondents
are paying careful attention to survey items in balance with their
other environmental demands. The daily participant-reported
burden was also effective in describing state versus trait patterns
of subpopulations in this study. Evidence of varied states of
burden indicated that momentary processes were captured during
the 1-week observation period. The low frequency of persistent
high burden indicated that this study was reasonably tailored to
the needs of a predominantly low-income, racially or
ethnically diverse and immigrant or refugee population.

Strengths and Limitations
This study had notable strengths, including stratified recruitment
of a racially or ethnically diverse and immigrant or refugee
study population and instruments translated into participants’
preferred language. Although incentives were given for a
complete 8 days of data collection, the daily completeness of
the protocol was not incentivized in this implementation. Bonus
incentives for consecutive-day adherence may increase the
quality of data collected over the course of the study period, as
consecutive-day data allows for better evaluation of real-time
influences on outcomes across days. The greater the frequency
of nonsequential EMA reporting, the greater was possibility
that selection bias may affect the representativeness and
transportability of study findings to other populations. Bonus
incentives may also be critical for minimizing burden and
promoting adherence in EMA studies that are carried out over
long time frames. A request for participant feedback about their
experience of burden increased confidence about differentiating
between momentary and study-related burden. While the 8-day
observation period was adequate for characterizing the low
study burden for the current protocol, future studies may require
longer observation intervals to capture rare momentary
exposures (eg, periodic financial stressors or housing insecurity)
for significant life events. These studies should measure
participant burden to determine if these findings cohere for
long-term observation periods. Replication should also be
performed in larger study samples to determine if comparable
protocols represent typical respondent experiences.

Conclusions
The EMA study burden was found to be low in an 8-day EMA
study of the momentary sources of stress, depressed mood, and
caregiver-child interactions in a racially or ethnically diverse
and immigrant or refugee sample of adult caregivers. Attention
to the sociodemographic attributes of the sample, including
chronic stress levels, nativity, and preferred language, is
important for minimizing participant burden and improving
data quality. Momentary stress is a strong determinant of
participant-experienced burden and may affect adherence to
mHealth study protocols. EMA stands to be an important
observational design for providing critically relevant insight
into human-environment interactions that advances intervention development appropriate for dynamic public health challenges when the design is carefully tailored to the study population and research objectives.

Acknowledgments

Research is supported by grant numbers R01HL126171 and R01HL156994 from the National Heart, Lung, and Blood Institute (PI: Berge). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung and Blood Institute or the National Institutes of Health. Dr de Brito’s research was supported by the National Institute of Arthritis Musculoskeletal and Skin Diseases of the National Institutes of Health under Award Number K12 AR084223. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References


Abbreviations

EMA: ecological momentary assessment
mHealth: mobile health

Edited by A Mavragani; submitted 05.06.23; peer-reviewed by H Li, M Witry, C Baxter; comments to author 02.12.23; revised version received 25.01.24; accepted 19.02.24; published 24.04.24.

Please cite as:
Tate AD, Fertig AR, de Brito JN, Ellis ÉM, Carr CP, Trofholz A, Berge JM
Momentary Factors and Study Characteristics Associated With Participant Burden and Protocol Adherence: Ecological Momentary Assessment
JMIR Form Res 2024;8:e49512
URL: https://formative.jmir.org/2024/1/e49512
doi:10.2196/49512
PMID:38656787
Assessing Electronic Health Literacy in Individuals With the Post–COVID-19 Condition Using the German Revised eHealth Literacy Scale: Validation Study

Matthias Marsall1, Dr; Hannah Dinse2,3, Dr Med; Julia Schröder2; Eva-Maria Skoda2,3, Prof Dr Med; Martin Teufel2,3, Prof Dr Med; Alexander Bäuerle2,3, Dr

1Institute for Patient Safety, University Hospital Bonn, Bonn, Germany
2Clinic for Psychosomatic Medicine and Psychotherapy, LVR–University Hospital Essen, University of Duisburg-Essen, Essen, Germany
3Center for Translational Neuro- and Behavioral Sciences, University of Duisburg-Essen, Essen, Germany

Corresponding Author:
Matthias Marsall, Dr
Institute for Patient Safety
University Hospital Bonn
Venusberg-Campus 1
Bonn, 53127
Germany
Phone: 49 228 287 ext 11595
Email: matthias.marsall@ukbonn.de

Abstract

Background: The eHealth Literacy Scale (eHEALS) is a widely used instrument for measuring eHealth literacy (eHL). However, little is known so far about whether the instrument is valid for the assessment of eHL in persons who are affected by the post–COVID-19 condition. This is particularly important as people with the post–COVID-19 condition are frequently affected by false information from the internet.

Objective: The objective of our study was to evaluate the validity and reliability of the German Revised eHealth Literacy Scale (GR-eHEALS) in individuals with the post–COVID-19 condition.

Methods: A cross-sectional study was conducted from January to May 2022. The self-assessment survey consisted of the GR-eHEALS, health status– and internet use–related variables, sociodemographic data, and (post)–COVID-19–related medical data. Confirmatory factor analysis (CFA), correlational analyses, and tests of measurement invariance were deployed.

Results: In total, 330 participants were included in the statistical analyses. CFA revealed that the 2-factor model reached an excellent model fit (comparative fit index=1.00, Tucker–Lewis index=0.99, root mean square error of approximation=0.036, standardized root mean square residual=0.038). Convergent validity was confirmed by significant positive correlations between eHL and knowledge of internet-based health promotion programs, experience in using these programs, and the duration of private internet use. In addition, a significantly negative relationship of eHL with internet anxiety supported convergent validity. Furthermore, significant relationships of eHL with mental health status and internal health locus of control confirmed the criterion validity of the instrument. However, relationships of eHL with physical health status and quality of life could not be confirmed. The 2-factor model was fully measurement invariant regarding gender. Regarding age and educational level, partial measurement invariance was confirmed. The subscales as well as the overall GR-eHEALS reached good-to-excellent reliability (Cronbach α≥.86).

Conclusions: The GR-eHEALS is a reliable and largely valid instrument for assessing eHL in individuals with the post–COVID-19 condition. Measurement invariance regarding gender was fully confirmed and allows the interpretation of group differences. Regarding age and educational level, group differences should be interpreted with caution. Given the high likelihood that individuals with the post–COVID-19 condition will be confronted with misinformation on the Internet, eHL is a core competency that is highly relevant in this context, in both research and clinical practice. Therefore, future research should also explore alternative instruments to capture eHL to overcome shortcomings in the validity of the GR-eHEALS.

(JMIR Form Res 2024;8:e52189) doi:10.2196/52189

https://formative.jmir.org/2024/1/e52189

JMIR Form Res 2024 | vol. 8 | e52189 | p.2718
(page number not for citation purposes)
KEYWORDS
eHealth literacy; eHEALS; factor analysis; measurement invariance; psychometric properties; infodemic

Introduction

In the current phase of the COVID-19 pandemic, it has become apparent that infection with SARS-CoV-2 can be associated with the experience of prolonged physical and cognitive impairments. According to the clinical case definition by the World Health Organization, the occurrence of specific delayed symptoms about 12 weeks after recovery from COVID-19 infection is referred to as the “post–COVID-19 condition” and is expected to affect about 10%-35% of people with infection [1,2]. Individuals with this condition regularly report physical and psychological symptoms, such as fatigue, muscular weakness, and sleep difficulties, as well as anxiety and depression [3]. Current data from Germany demonstrate that employees on sick leave due to the post–COVID-19 condition were unable to work for an average of 90 days. Moreover, people who needed inpatient treatment more than 7 days due to the post–COVID-19 condition were unable to work for an average of 168 days [4]. Data from Sweden likewise confirmed high incidences of about 13% of individuals with the post–COVID-19 condition resulting in long sick leave periods [5]. Nevertheless, studies have also reported that employees, especially health care workers, return to work despite ongoing symptoms due to the post–COVID-19 condition, which could be associated with negative long-term health effects [5,6]. These findings imply that the post–COVID-19 condition is a major public health issue with a significant impact on entire societies and public health systems. However, the etiology of the post–COVID-19 condition is not conclusively understood, and psychosocial factors may have a significant impact on physical and cognitive impairments experienced after COVID-19 infection [7,8].

In addition to the illness-related burden of the post–COVID-19 condition, individuals affected are faced with a high amount of false or confusing information, mostly spread via social media [9]. This comprises all topics around the COVID-19 pandemic, including symptoms, medication, treatments, and vaccination, as well as denial of the existence of the virus and the disease at all [10-12]. In this context, it has become apparent that individuals with COVID-19–related symptoms are susceptible to encountering misinformation, referred to as the COVID-19 “infodemic” [10,13-16]. This infodemic arose right after the beginning of the worldwide pandemic, and an analysis to investigate the proportion of false information revealed that more than 80% of the information spread via social media was presumably wrong [17]. Education about the existence of COVID-19–related misinformation and the labeling of misinformation may be relevant factors in maintaining a functioning health care system [10,18].

The fact that individuals are confronted with false or dubious health information from the internet is not a new phenomenon [19]. The concept of the competence in dealing with health information from the internet—eHealth Literacy (eHL)—was introduced back in 2006 by Norman and Skinner [20]. The concept and its measurement were based on the assumption that individuals may have competencies to seek, find, understand, and evaluate health information from the internet and to distinguish valid and reliable health information from dubious or wrong information [20]. Namely, eHL comprises people’s skills, knowledge, and competencies required to navigate, evaluate, and effectively use health information from digital sources, such as websites, apps, and online communication platforms, to make informed decisions about health-related issues and engage in self-care, health promotion, and disease prevention activities [20-22]. The measurement instrument developed by the same researchers, the eHealth Literacy Scale (eHEALS) [23], was translated into several languages and validated in many countries worldwide [24-33]. Over the years of research regarding eHL, studies have shown that higher eHL is associated with health-related outcomes, such as better health behaviors and health cognitions in older adults [34,35], more regular sports exercises in students [36], better health behaviors regarding physical exercise and eating behavior in adults [21], and higher physical and mental health in patients after percutaneous coronary intervention [37]. Moreover, studies have shown that individuals with higher eHL show higher adherence to prevention behavior guidelines in the context of the COVID-19 pandemic [38,39]. Vaccination is a critical factor in reducing the infection rates and severity of COVID-19 symptoms and therefore the burden on the public health system [40]. Consequently, initiatives to improve eHL could increase acceptance of vaccinations as part of a human-centered strategy to reduce the burden on the health care system [41].

In summary, humans benefit from high eHL, which empowers them to effectively use health information from the internet to cope with physical or psychological impairments. Due to the high density of information, and especially due to the high proportion of false information regarding COVID-19, people with COVID-19 or the post–COVID-19 condition have particularly high requirements for well-developed eHL. Because when people are exposed to the fact that around 80% of health information is not evidence based, it is even more difficult to identify the correct and evidence-based information from the internet and apply it to their own health situation [17]. In addition, the use of social media for obtaining health information has changed significantly during the COVID-19 pandemic [42,43]. More precisely, people obtain health information more frequently from social media sources rather than from homepages. This has fundamentally changed the way health information is obtained, which raises the question of whether eHEALS is still adequate to capture the concept of digital health literacy in the changed information era. In addition, criticism regarding the measurement is also arising with regard to the consistency of health-related outcomes as well as the actuality of eHEALS in an increasingly networked, digital environment [44-48]. Nevertheless, eHEALS is, to date, the most widely used instrument to assess eHL [44].

Therefore, a validation of eHEALS is needed to investigate the applicability of this assessment in terms of validity and reliability in a changed environment for people with the
post–COVID-19 condition. In this study, we focused on the application of eHEALS in measuring eHL in German-speaking individuals with the post–COVID-19 condition. Specifically, this study aimed to fill the following research gaps:

• First, we evaluated the construct validity and reliability of the German version of eHEALS, the German Revised eHealth Literacy Scale (GR-eHEALS) [49], in a sample of individuals with the post–COVID-19 condition (aim 1).
• Second, we investigated the convergent validity of the GR-eHEALS (aim 2).
• Third, we examined the criterion validity of the GR-eHEALS (aim 3).
• Fourth, we tested the equivalence of the measurement properties of the GR-eHEALS regarding sociodemographic variables of individuals with the post–COVID-19 condition as part of the construct validity (aim 4).

Methods

Ethical Considerations

Before taking part in the survey, all participants received study information. Electronic informed consent was digitally obtained at the beginning of the survey from all participants. There was no compensation for participation in the study. Due to the anonymous study design, no individual information of participants is reported in this paper. The study was executed in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Medical Faculty of the University of Duisburg-Essen (19-89-47-BO).

Study Design and Participants

A cross-sectional online survey study was conducted. Data were collected via the online survey system Unipark (Tivian XI GmbH). Participation in the survey was anonymous, voluntary, and without monetary compensation. The participants of this study were recruited between January and May 2022. Flyer and information materials were laid out in different hospitals (eg, University Hospital Essen) and rehabilitation clinics (eg, MEDIAN clinics) in North Rhine-Westphalia, Germany. Furthermore, self-help communities on different social media platforms (eg, Facebook) were contacted, and online flyers were distributed. To participate in this study, the following inclusion criteria were applied: age ≥ 18 years, internet access, a good command of the German language, confirmed COVID-19 infection in the past, and reporting as currently having post–COVID-19 symptoms. COVID-19 infection was assessed via a self-report of the date of a positive detection of SARS-CoV-2 with a nasopharyngeal swab. Symptoms related to the post–COVID-19 condition were assessed according to the clinical case definition by the World Health Organization [2]. For this purpose, participants were asked about the presence of the following symptoms: sore throat, cough, shortness of breath, headache/pain in the limbs, body temperature above 38°C, olfactory or gustatory disturbances, diarrhea, and other symptoms that could be entered via a text field. Multimedia Appendix 1 shows how frequently each of these symptoms was reported.

Participants took on average 13 (SD 5.43) minutes to complete the survey. Participants were able to stop the survey at any time. 554 people have participated in the study. According to the inclusion criteria, we excluded n=196 (35.4%) participants who reported that they currently have no post–COVID-19 symptoms. Further, we excluded participants with implausible responses regarding age (age < 18 years: n=3, 0.5%; age > 100: n=1, 0.2%) and outliers regarding the survey completion time (n=24, 4.3%). The final sample consisted of 330 participants, reflecting a completion rate of 59.6%.

Measurements

All data were collected using a self-report questionnaire. The following assessments were used in the survey: eHL, constructs to examine convergent validity, constructs to examine criterion validity, and sociodemographic and COVID-19–related variables.

eHealth Literacy

The 8-item GR-eHEALS [23,49] was applied. This instrument reaches good psychometric properties regarding its validity and reliability as a 2-factor model covering the competences information seeking (IS) and information appraisal (IA) [49]. The response scale ranged from 1 for “strongly disagree” to 5 for “strongly agree,” with higher values corresponding to higher eHL.

Constructs to Examine Convergent Validity

We assessed the knowledge of internet-based health promotion programs via 3 self-developed items using a 5-point Likert scale (1 for “strongly disagree” to 5 for “strongly agree”). Higher scores indicated higher knowledge of internet-based health promotion programs. The items covered knowledge of the contents of such programs (“I can certainly imagine something under that.”), how they work (“I know how such programs work.”), and how to find them (“I know how to find such programs.”). Cronbach α was .88. The instrument has been previously used [50,51].

One self-developed item (“Have you already had experience with internet-based health promotion programs?”) was used to assess experience in using internet-based health promotion programs on a 3-point Likert scale (1 for “already used such programs,” 2 for “not used but aware of the possibilities of such programs,” and 3 for “not aware of the possibilities of such programs”). This item was inverted, so higher values represented more experience with internet-based health promotion programs. The assessment has been used in previous studies [50,51].

We obtained information about the duration of daily private internet use through a self-developed single item: “How long do you use the internet for private purposes every day?” The response scale ranged from 1 for “not at all” to 5 for “more than 5 hours.” Previous studies have used this item [50,51].

The construct of internet anxiety was measured on a 5-point Likert scale (1 for “does not apply” to 5 for “applies”) using 3 self-developed items (“I have concerns about using the internet,” “I am afraid that I could make an irreparable mistake when using the internet,” and “The internet is something that worries me.”). Higher scores represented higher internet anxiety.

https://formative.jmir.org/2024/1/e52189
We assessed the instruments and items used are provided in Multimedia Appendix 2. The instruments and items have been used in previous studies [49,53]. Cronbach α was .82. These items were applied previously [49-51]. We expected significant positive relationships between eHL and knowledge of internet-based health promotion programs, experience in using internet-based health promotion programs, and duration of daily private internet use. Further, a significant negative relationship between eHL and internet anxiety was expected.

### Constructs to Examine Criterion Validity

We assessed the internal health locus of control with 3 items on a 5-point Likert scale (1 for “strongly disagree” to 5 for “strongly agree”) via an adapted German version of the Multidimensional Health Locus of Control Scale [52]. Higher scores reflected a higher internal health locus of control. Cronbach α was .76.

Moreover, we measured the physical health status (“On a scale of 0 to 10, how do you rate your physical health [eg, no physical limitations, pain?]”), mental health status (“On a scale of 0 to 10, how do you rate your mental health [eg, no feelings of anxiety, depression?]”), and overall quality of life (“On a scale of 0 to 10, how would you rate your current quality of life?”) with self-developed single items on 11-point Likert scales. These items have been used in previous studies [49,53]. We expected significant positive relationships between eHL and all criterion validity constructs.

### Sociodemographic and COVID-19–Related Variables

We assessed participants’ main sociodemographic variables: age (as an exact number), gender (female, male, diverse), and marital status (married, living in a relationship, single, divorced/widowed, other). Further, participants’ educational degree (no school degree, secondary school certificate [Hauptschule], secondary school certificate [Mittlere Reife], university entrance qualification, university degree, academic degree, other, not specified), current employment status (attending school/study, employed [part-time, full-time], sick leave, retirement/pension, not employed/other), and population of the community they live in (big city, medium city, small city, rural village) were assessed.

Regarding COVID-19, we asked participants for the date of the positive test result of SARS-CoV-2 with a nasopharyngeal swab, as well as symptoms related to the post–COVID-19 condition (inclusion criteria). In addition, we gathered information about whether participants required treatment in a hospital (yes, no) and, in addition, intensive care due to their COVID-19 infection (yes, no). Further, we asked participants for a self-assessment of the severity of their COVID-19 symptoms (1 for “no symptoms” to 4 for “severe symptoms”) as well as their current physical capacity (1 for “still significantly limited” to 3 for “good”).

There was no “I don’t know” or similar option available for any of the items. The instruments and items used are provided in Multimedia Appendix 2.

### Statistical Analysis

We conducted all statistical analyses using R (R Foundation for Statistical Computing) and RStudio (Posit) [54,55] extended with several packages. We report the distribution of sociodemographic as well as COVID-19–related characteristics of the study sample. Further, descriptive statistics were used to determine the item statistics of the GR-eHEALS items.

Next, we followed analytical steps to examine the validity and reliability of the GR-eHEALS according to the study aims: First, we conducted confirmatory factor analysis (CFA) to evaluate the factorial structure and construct validity. In the first model, we tested the 1-factorial structure of all 8 items loading on a single factor, as proposed by the authors of the original instrument [23]. In the second model, we considered the 2-factorial structure with the 2 intercorrelated factors IS and IA, as validated in the GR-eHEALS [49]. For model evaluation, we considered the recommendations of Hu and Bentler [56] and defined a good model fit when reaching a comparative fit index (CFI) and a Tucker–Lewis index (TLI) of about 0.95. Further, we determined that the root mean square error of approximation (RMSEA) should be <0.06, while the standardized root mean square residual (SRMR) should be ≤0.08. For model estimation, we used the robust version of the means-adjusted unweighted least squares (ULSM) estimator as the items were measured on ordinal-scaled levels [57]. We reported Cronbach α to evaluate the reliability of the scales.

Second, regarding aims 2 and 3, we conducted Spearman rank correlation analyses to examine convergent and criterion validity. All analyses considered a significance level of P=.05, and missing values were treated via listwise deletion.

Lastly, tests of measurement invariance were deployed to assess measurement equivalence of the instrument regarding age, gender, and educational level of participants (aim 4). Measurement invariance is a statistical requirement for the correct interpretation of group differences measured by an instrument. Otherwise, if an instrument is not measurement invariant, mean differences found between different groups should not be interpreted, as it may be that these differences arise because the instrument measures differently in different groups [58,59]. To allow the interpretation of mean differences, a scalar level of invariance should be achieved [57]. Therefore, we performed measurement invariance tests for the gradually more restrictive levels of configural invariance, metric invariance, and, finally, scalar invariance [60]. As there is no consensus regarding cutoff criteria for the evaluation of model fits within measurement invariance testing [61], we considered 2 common approaches: First, we compared the change in the CFI between the more restrictive model and the less restrictive model, which should be ≤0.01 [62]. Second, we set the configural level of invariance as the baseline model that should meet the criteria of good model fit, as reported earlier. Subsequently, we performed χ² difference tests to compare the model of metric invariance against the model of metric invariance and then the model of scalar invariance against the model of metric invariance [61]. Significant χ² difference tests would reflect a significant change in the model fit, implying substantial deterioration of the model.
Results

Study Sample Description
The study participants’ mean age was 36.6 (SD 12.9, min.=18, max.=70 years). An overview of the participants’ sociodemographic background is reported in Table 1. Most participants did not need hospital or intensive care treatment but experienced mild-to-severe symptoms due to COVID-19 infection. Most participants also reported limited physical capacities. Table 2 presents COVID-19–related information of the study participants. On average, participants reported to have experienced 2.7 (range 1-7) of the defined post–COVID-19 symptoms.

Table 1. Sociodemographic description of study participants (N=329a).

<table>
<thead>
<tr>
<th>Sociodemographic characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>264 (80.0)</td>
</tr>
<tr>
<td>Male</td>
<td>63 (19.1)</td>
</tr>
<tr>
<td>Diverse</td>
<td>2 (0.6)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>106 (32.1)</td>
</tr>
<tr>
<td>Living in a relationship</td>
<td>107 (32.4)</td>
</tr>
<tr>
<td>Single</td>
<td>100 (30.3)</td>
</tr>
<tr>
<td>Divorced/widowed</td>
<td>11 (3.3)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (1.5)</td>
</tr>
<tr>
<td>Educational degreeb</td>
<td></td>
</tr>
<tr>
<td>Secondary school diploma</td>
<td>77 (23.3)</td>
</tr>
<tr>
<td>University entrance qualification</td>
<td>97 (29.4)</td>
</tr>
<tr>
<td>University degree</td>
<td>148 (44.8)</td>
</tr>
<tr>
<td>Other/not specified</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
</tr>
<tr>
<td>Attending school/study</td>
<td>81 (24.5)</td>
</tr>
<tr>
<td>Employed (part-time, full-time)</td>
<td>118 (35.8)</td>
</tr>
<tr>
<td>Sick leave</td>
<td>91 (27.6)</td>
</tr>
<tr>
<td>Retirement/pension</td>
<td>5 (1.5)</td>
</tr>
<tr>
<td>Not employed/other</td>
<td>34 (10.3)</td>
</tr>
<tr>
<td>Community size</td>
<td></td>
</tr>
<tr>
<td>Big city (&gt;100,000 inhabitants)</td>
<td>142 (43.0)</td>
</tr>
<tr>
<td>Medium city (&gt;20,000 inhabitants)</td>
<td>79 (23.9)</td>
</tr>
<tr>
<td>Small city (&gt;5000 inhabitants)</td>
<td>52 (15.8)</td>
</tr>
<tr>
<td>Rural village (&lt;5000 inhabitants)</td>
<td>56 (17.0)</td>
</tr>
</tbody>
</table>

aMissing: n=1 (0.3%).
bStatements were summarized: secondary school certificate (Hauptschule) and secondary school certificate (Mittlere Reife) as secondary school diploma, and university degree and academic degree as university degree.
Table 2. Post–COVID-19–related medical data of study participants (N=330).

<table>
<thead>
<tr>
<th>Required treatment in hospital</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>18 (5.5)</td>
</tr>
<tr>
<td>No</td>
<td>312 (94.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Required intensive care</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td>No</td>
<td>323 (97.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severity of COVID-19 symptoms</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No symptoms</td>
<td>16 (4.8)</td>
</tr>
<tr>
<td>Mild symptoms</td>
<td>137 (41.5)</td>
</tr>
<tr>
<td>Moderate symptoms</td>
<td>156 (47.3)</td>
</tr>
<tr>
<td>Severe symptoms</td>
<td>21 (6.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current physical capacity</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>58 (17.6)</td>
</tr>
<tr>
<td>Average</td>
<td>106 (32.1)</td>
</tr>
<tr>
<td>Still significantly limited</td>
<td>166 (50.3)</td>
</tr>
</tbody>
</table>

Psychometric Properties of the German Revised eHealth Literacy Scale

Table 3 shows the item statistics of the 8 GR-eHEALS items. The exact item wordings of the GR-eHEALS are presented in the original study [49].

Table 3. Descriptive item statistics of the GR-eHEALS items including means (SDs), skewness, kurtosis, and distribution of responses for each item (N=330).

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean (SD)</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>Response distribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 (strongly disagree)</td>
</tr>
<tr>
<td>GR-eHEALS1</td>
<td>3.93 (0.91)</td>
<td>−0.97</td>
<td>0.82</td>
<td>2</td>
</tr>
<tr>
<td>GR-eHEALS2</td>
<td>4.05 (0.82)</td>
<td>−1.05</td>
<td>1.53</td>
<td>1</td>
</tr>
<tr>
<td>GR-eHEALS3</td>
<td>3.70 (0.97)</td>
<td>−0.60</td>
<td>−0.20</td>
<td>2</td>
</tr>
<tr>
<td>GR-eHEALS4</td>
<td>3.80 (0.93)</td>
<td>−0.70</td>
<td>−0.01</td>
<td>1</td>
</tr>
<tr>
<td>GR-eHEALS5</td>
<td>3.81 (0.84)</td>
<td>−0.75</td>
<td>0.62</td>
<td>1</td>
</tr>
<tr>
<td>GR-eHEALS6</td>
<td>4.03 (0.90)</td>
<td>−1.06</td>
<td>1.04</td>
<td>1</td>
</tr>
<tr>
<td>GR-eHEALS7</td>
<td>3.91 (1.01)</td>
<td>−1.00</td>
<td>0.69</td>
<td>3</td>
</tr>
<tr>
<td>GR-eHEALS8</td>
<td>3.80 (0.96)</td>
<td>−0.78</td>
<td>0.28</td>
<td>2</td>
</tr>
</tbody>
</table>

GR-eHEALS: German Revised eHealth Literacy Scale.

All items had a median of 4 and showed slightly negative skewness, indicating that participants experienced high levels of eHL.

To examine the factorial structure of the GR-eHEALS, 2 rounds of CFA were performed. The results are presented in Table 4.
Table 4. Results of CFA\(^a\) of the GR-eHEALS\(^b\).

<table>
<thead>
<tr>
<th>Model</th>
<th>(\chi^2) (df)</th>
<th>CFI</th>
<th>TLI(^d)</th>
<th>RMSEA(^e)</th>
<th>SRMR(^f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-factor model(^g)</td>
<td>218.14 (20)</td>
<td>0.98</td>
<td>0.97</td>
<td>0.088</td>
<td>0.079</td>
</tr>
<tr>
<td>2-factor model(^h)</td>
<td>59.96 (19)</td>
<td>1.00</td>
<td>0.99</td>
<td>0.036</td>
<td>0.038</td>
</tr>
</tbody>
</table>

\(^a\)CFA: confirmatory factor analysis. 
\(^b\)GR-eHEALS: German Revised eHealth Literacy Scale. 
\(^c\)CFI: comparative fit index. 
\(^d\)TLI: Tucker–Lewis index. 
\(^e\)RMSEA: root mean square error of approximation. 
\(^f\)SRMR: standardized root mean square residual. 
\(^g\)Model with all 8 items loading on 1 common factor. 
\(^h\)Model with 2 subscales, “information seeking” (items 1-4) and “information appraisal” (items 5-8).

The 2-factor model revealed an excellent model fit, whereas the 1-factor model slightly did not meet the recommendations regarding the RMSEA and SRMR. In addition, the \(\chi^2\) difference test, which compared the 2 models based on the nonrobust estimator, confirmed significant model improvement for the 2-factor model over the 1-factor model \((\chi^2 = 33.18, P < .001)\). Consequently, the 2-factor model with the 2 interrelated factors IS and IA was used for subsequent analyses. All items reached high factor loadings (≥.73) and the 2 factors were significantly correlated, with \(r = 0.74\) \((P < .001)\). The item factor loadings are presented in Multimedia Appendix 3. IS (mean 3.87, SD 0.80) and IA (mean 3.89, SD 0.78) reached high reliability (Cronbach \(\alpha = .90\) and .86, respectively).

Due to the high correlation of the subscales, the overall GR-eHEALS with all 8 items was also reported for the investigation of convergent and criterion validity. The overall GR-eHEALS (mean 3.88, SD 0.72) achieved an excellent reliability of 0.91. Therefore, the construct validity and reliability of the GR-eHEALS was confirmed.

Convergent and Criterion Validity of the GR-eHEALS in Individuals With the Post–COVID-19 Condition

To evaluate the convergent validity and criterion validity of the GR-eHEALS, Spearman rank correlation analyses were performed. The results are presented in Table 5.

Table 5. Spearman rank correlations of the GR-eHEALS\(^a\) subscales and the overall GR-eHEALS with convergent and criterion validity variables.

<table>
<thead>
<tr>
<th>Validity and variables</th>
<th>IS(^b) subscale Correlation coefficient, (r)</th>
<th>(P) value</th>
<th>IA(^c) subscale Correlation coefficient, (r)</th>
<th>(P) value</th>
<th>Overall scale Correlation coefficient, (r)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Convergent validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of internet-based health promotion programs</td>
<td>0.45</td>
<td>&lt;.001</td>
<td>0.31</td>
<td>&lt;.001</td>
<td>0.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Experience in using internet-based health promotion programs</td>
<td>0.28</td>
<td>&lt;.001</td>
<td>0.17</td>
<td>&lt;.01</td>
<td>0.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Duration of daily private internet use</td>
<td>0.15</td>
<td>&lt;.01</td>
<td>0.16</td>
<td>&lt;.01</td>
<td>0.16</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Internet anxiety</td>
<td>−0.20</td>
<td>&lt;.001</td>
<td>−0.30</td>
<td>&lt;.001</td>
<td>−0.28</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Criterion validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical health</td>
<td>0.04</td>
<td>.47</td>
<td>−0.01</td>
<td>.86</td>
<td>0.02</td>
<td>.71</td>
</tr>
<tr>
<td>Mental health</td>
<td>0.14</td>
<td>&lt;.05</td>
<td>0.11</td>
<td>&lt;.05</td>
<td>0.14</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Quality of life</td>
<td>0.12</td>
<td>&lt;.05</td>
<td>0.06</td>
<td>.29</td>
<td>0.10</td>
<td>.06</td>
</tr>
<tr>
<td>Internal health locus of control</td>
<td>0.19</td>
<td>&lt;.001</td>
<td>0.18</td>
<td>&lt;.001</td>
<td>0.20</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\)GR-eHEALS: German Revised eHealth Literacy Scale. 
\(^b\)IS: information seeking. 
\(^c\)IA: information appraisal.

IS and IA were both significantly positively associated with knowledge of and experience in using internet-based health promotion programs as well as with the duration of daily private internet use. Moreover, both factors were significantly negatively related to internet anxiety. The results were supported by similar relationships regarding the overall GR-eHEALS. These results confirm the convergent validity of the GR-eHEALS regarding the study aim 2.

Although criterion validity was supported by significantly positively relations of the IS and IA with mental health and the...
internal health locus of control, quality of life was only related to IS, and physical health was not related to any of the 2 factors. The overall GR-eHEALS was similarly significantly related to mental health status and the internal health locus of control but not to physical health and quality of life. Consequently, results regarding aim 3 were inconsistent.

To further understand these results, we conducted correlation analyses between all criterion validity scales. The results revealed that lower physical health is strongly associated with lower mental health (r=0.40, P<0.001), a lower quality of life (r=0.69, P<0.001), and a lower internal health locus of control (r=0.31, P<0.001). The correlation table is presented in Multimedia Appendix 4.

IS and IA were both not associated with any of the post–COVID-19–related characteristics (need for intensive care was not tested due to the small subgroup sample size).

**Measurement Invariance of the GR-eHEALS in Individuals With the Post–COVID-19 Condition**

For testing measurement invariance regarding gender, we excluded participants who indicated their gender as diverse in order to have sufficient group sample sizes for this analysis. Regarding age, we performed a median split to divide the sample at the median into 2 groups (median age 34 years). Regarding educational level, we divided the sample into 2 groups: group 1 consisted of participants with any school certificate, and group 2 consisted of participants holding a university degree. Participants who indicated other educational levels than these were excluded from this analysis. Multimedia Appendix 5 summarizes the results of the tests of measurement invariance of the GR-eHEALS regarding gender, age, and educational level.

All changes in the CFI were below 0.01, indicating measurement invariance for gender, age, and educational level. Nevertheless, $\chi^2$ difference tests showed that the configurational level of invariance for all 3 sociodemographic variables reached good model fit indices, implying that this level of invariance could be confirmed regarding all 3 sociodemographic variables. Regarding gender, subsequent $\chi^2$ difference tests were not significant for metric and scalar levels of invariance, which is evidence for the assumption of measurement invariance of the GR-eHEALS. With respect to age and educational level, the $\chi^2$ difference tests were not significant for the metric levels of measurement invariance but were significant for the scalar levels. This indicated that only partial measurement invariance of the GR-eHEALS regarding age and educational level could be confirmed. Hence, results regarding aim 4 were inconsistent.

**Discussion**

**Principal Findings**

This is the first study to evaluate the validity and reliability of the internationally most common instrument for measuring eHL in individuals with the post–COVID-19 condition. Thus, there is a gap in the knowledge of whether this instrument is still appropriate in measuring eHL in an era of an increasing number of individuals with the post–COVID-19 condition. Our study addresses this research gap and provides the first evidence by examining the measurement properties of the German version of eHEALS in individuals with the post–COVID-19 condition. The CFA results underpin the assumed factorial structure of the GR-eHEALS [49], supporting the construct validity of the instrument (aim 1). The reliability of the GR-eHEALS could be confirmed with high coefficients for both factors, IS and IA, as well as for the overall scale.

The subsequently performed correlation analyses with 4 convergent validity scales showed highly consistent results and confirmed the convergent validity of the GR-eHEALS (aim 2).

Concerning the criterion validity of the GR-eHEALS (aim 3), results did not fully meet our assumptions: physical health was not related to the GR-eHEALS, and the overall quality of life was only related to IS. However, our results are not the first that showed no relation between eHL and the physical health status [35]. In addition, this result could be explained by the fact that the study sample consisted of individuals with an impaired physical health status due to the inclusion criteria. Many of the study participants still experienced limitations in their physical capacity at the time of completing the questionnaire. Further, as shown in Multimedia Appendix 4, quality of life and physical health were strongly interrelated. That could explain why quality of life also did not confirm the expected relation to IA. In summary, most individuals in the study sample still had symptoms of the post–COVID-19 condition, which may not be affected by their eHL. It is possible that people who currently have diseases with physical health limitations cannot compensate for those limitations by high levels of eHL. Nevertheless, we found significant relations between the GR-eHEALS and mental health as well as the internal health locus of control. A high internal health locus of control reflects individuals’ confidence to influence their own health, which is associated with higher mental health status [63], better health behaviors [64], higher levels of physical activity [65], lower perceived stress [66], and higher medication adherence [67]. The fact that the internal health locus of control was strongly associated with eHL implies that this competence may play a central role when individuals are facing ambiguous information from the internet and still experience themselves as competent in handling that information and their health.

Regarding aim 4, the results of the measurement invariance tests of the GR-eHEALS were partly confirmed. Measurement invariance is important to assume that assessments measure the same underlying constructs consistently over different groups and populations [68]. If an assessment is not measurement invariant, differences between groups may be due to a bias in the assessment tool rather than true differences in the construct, which could lead to inaccurate interpretation of study results and potentially misleading recommendations for public health [69]. Measurement invariance is also an important prerequisite for the generalizability of study results [70]. This is particularly important for public health research, since study results are used to inform public policies and for the development of interventions, such as for health promotion [71,72]. Our results indicate that there are no substantial differences in the measurement model of the GR-eHEALS regarding gender. However, regarding age and educational level, measurement
invariance was only partly confirmed but missed achieving scalar invariance. These results are important as studies have revealed that COVID-19 infection and its consequences are determined by gender, age, and educational level [73-77]. With the confirmation of measurement invariance of the GR-eHEALS regarding gender, we provide a valid instrument to assess and interpret differences in gender. With regard to age and educational level, group differences may only be interpreted under consideration of the partly confirmed invariance.

Summarizing, our results underpin the general construct validity, convergent validity, and reliability of the GR-eHEALS in individuals with the post–COVID-19 condition. The criterion validity of the instrument could mainly be confirmed. By partly confirming the measurement invariance, we showed that the GR-eHEALS is a valid instrument to assess and interpret eHL but is limited in terms of the interpretation of differences in age and educational level.

The overall GR-eHEALS showed consistent results in terms of convergent and criterion validity. Provided the focus is on a general examination of eHL, the overall scale can therefore be used appropriately. Nevertheless, depending on the research question, the subscales offer an opportunity to consider the competence domains of eHL in a differentiated manner.

The mean values and factor loadings of the items as well as the interrelation of the GR-eHEALS subscales were similar of those reported in the initial validation of the GR-eHEALS [49]. This is an indication that the GR-eHEALS performs and can be used similarly across different studies and in different populations. Nevertheless, the measurement invariance of the instrument across different populations should be investigated in future studies before comparing the results of different study groups. However, as the factorial structure of eHEALS is not consistent in different languages [23,78,79], further studies are needed to verify the validity of the instrument in the population with the post–COVID-19 condition.

It is important to note that the GR-eHEALS is based on the original eHEALS, which was published back in 2006 [23]. Now, the (digital) world has changed significantly. New information sources have emerged, and social media platforms have become a considerable source for health information gathering [80]. This is particularly relevant because the COVID-19 pandemic has contributed to the fact that people no longer simply obtain health information from the Internet but increasingly use social media (eg, Instagram, YouTube, Twitter) and follow individual influencers to acquire health information [81]. Therefore, the appropriateness of the use of eHEALS should be considered critically and subject to future research. Newer instruments for measuring eHL have appeared claiming to assess this construct more comprehensively and adapt to new environmental conditions [45]. Future studies should therefore establish comparative analyses of the psychometric properties of different instruments, especially in consideration of the social as well as digital environments individuals are facing in the postpandemic world. However, the study results indicate largely the validity and reliability of the GR-eHEALS to assess eHL in people after COVID-19 infection. As eHEALS is still the most widely used instrument worldwide to assess eHL [44], it offers the possibility of comparing results between different countries and populations. eHEALS thus remains an important instrument for the assessment of eHL.

**Limitations**

The study sample covers a wide range of sociodemographic characteristics. Nevertheless, female participants were overrepresented in the sample, and most participants were middle aged, resulting in a small number of individuals who were already retired. This could represent a bias as young people interact with digital media differently than older people [82]. In addition, a high proportion of individuals with a university degree was present in the sample. Our survey was only accessible online, so the sample may be biased toward individuals with higher affinity to the use of digital media. Moreover, responses to the survey may have been biased by participants’ technology skills. These circumstances may limit the generalizability of our results. Exclusively self-report instruments were used, which may have introduced a response bias. As we also recorded via self-report whether a confirmed COVID-19 infection and post–COVID-19 symptoms were present, the inclusion criteria for this study were particularly affected by this restriction. Moreover, this cross-sectional study does not allow for the interpretation of causal relationships. Therefore, all interrelations found only reflect relationships without implying causal directions. The validation methods applied were based on the assumptions of classical test theory, and future studies should consider alternative paradigms, such as item response theory. Further, even if already used in other studies, the instruments used to test convergent and criterion validity were mostly not validated instruments. This could limit the external validity of the results.

**Conclusion**

The validity and reliability of the GR-eHEALS could mainly be confirmed in a sample of individuals with the post–COVID-19 condition. This is the first study to examine eHL in this population, which is particularly vulnerable to misinformation. By partly confirming the measurement invariance of the instrument, we provided evidence that the GR-eHEALS is an important instrument in public health research due to its ability to be interpretable regarding differences in gender and partly regarding age and educational level. However, future research should also explore alternative instruments to capture eHL and thus consider changing behaviors of disseminating and obtaining digital health information, for example, through social media.

**Acknowledgments**

This work was supported by the Open Access Publication Fund of the University of Bonn.
Data Availability

The data sets used and analyzed during this current study are available from the corresponding author upon reasonable request.

Authors' Contributions

AB, HD, and JS conceptualized the study. Project administration and data collection were performed by HD, JS, and AB. MM conducted statistical analyses and interpretation of results. MM and AB wrote the original draft of the manuscript. AB, EMS, and MT supervised the study and contributed to the study design, data collection, and critical revision of the manuscript. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of participants reporting each post–COVID-19 symptom.
[DOCX File] 19 KB - formative_v8i1e52189_app1.docx

Multimedia Appendix 2
Study survey.
[DOCX File] 28 KB - formative_v8i1e52189_app2.docx

Multimedia Appendix 3
Item factor loadings of the German Revised eHealth Literacy Scale (GR-eHEALS).
[DOCX File] 19 KB - formative_v8i1e52189_app3.docx

Multimedia Appendix 4
Correlation table of criterion validity scales.
[DOCX File] 21 KB - formative_v8i1e52189_app4.docx

Multimedia Appendix 5
Results of tests of measurement invariance of the German Revised eHealth Literacy Scale (GR-eHEALS) regarding gender, age, and educational level.
[DOCX File] 22 KB - formative_v8i1e52189_app5.docx

References


26. Marsall et al. JMIR FORMATIVE RESEARCH


Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
eHL: eHealth Literacy
eHEALS: eHealth Literacy Scale
GR-eHEALS: German Revised eHealth Literacy Scale
IA: information appraisal
IS: information seeking
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker–Lewis index
ULSM: unweighted least squares estimator

Edited by A Mavragani; submitted 28.08.23; peer-reviewed by J Graf, S Bacon, K Lavoie; comments to author 28.12.23; revised version received 12.01.24; accepted 05.02.24; published 25.04.24.

Please cite as:
Marsall M, Dinse H, Schröder J, Skoda EM, Teufel M, Bäuerle A
Assessing Electronic Health Literacy in Individuals With the Post–COVID-19 Condition Using the German Revised eHealth Literacy Scale: Validation Study
JMIR Form Res 2024;8:e52189
URL: https://formative.jmir.org/2024/1/e52189
doi:10.2196/52189
PMID:38662429

©Matthias Marsall, Hannah Dinse, Julia Schröder, Eva-Maria Skoda, Martin Teufel, Alexander Bäuerle. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Proactive Identification of Patients with Diabetes at Risk of Uncontrolled Outcomes during a Diabetes Management Program: Conceptualization and Development Study Using Machine Learning

Arash Khalilnejad¹, PhD; Ruo-Ting Sun¹, PhD; Tejaswi Kompala¹, MD; Stefanie Painter¹, DHEd; Roberta James¹, MSTAT; Yajuan Wang¹, PhD
Teladoc Health, Purchase, NY, United States

Corresponding Author:
Stefanie Painter, DHEd
Teladoc Health
2 Manhattanville Road
Purchase, NY, 10577
United States
Phone: 1 3045469968
Email: spainter@teladochealth.com

Abstract

Background: The growth in the capabilities of telehealth have made it possible to identify individuals with a higher risk of uncontrolled diabetes and provide them with targeted support and resources to help them manage their condition. Thus, predictive modeling has emerged as a valuable tool for the advancement of diabetes management.

Objective: This study aimed to conceptualize and develop a novel machine learning (ML) approach to proactively identify participants enrolled in a remote diabetes monitoring program (RDMP) who were at risk of uncontrolled diabetes at 12 months in the program.

Methods: Registry data from the Livongo for Diabetes RDMP were used to design separate dynamic predictive ML models to predict participant outcomes at each monthly checkpoint of the participants’ program journey (month-n models) from the first day of onboarding (month-0 model) up to the 11th month (month-11 model). A participant’s program journey began upon onboarding into the RDMP and monitoring their own blood glucose (BG) levels through the RDMP-provided BG meter. Each participant passed through 12 predicative models through their first year enrolled in the RDMP. Four categories of participant attributes (ie, survey data, BG data, medication fills, and health signals) were used for feature construction. The models were trained using the light gradient boosting machine and underwent hyperparameter tuning. The performance of the models was evaluated using standard metrics, including precision, recall, specificity, the area under the curve, the $F_1$-score, and accuracy.

Results: The ML models exhibited strong performance, accurately identifying observable at-risk participants, with recall ranging from 70% to 94% and precision from 40% to 88% across the 12-month program journey. Unobservable at-risk participants also showed promising performance, with recall ranging from 61% to 82% and precision from 42% to 61%. Overall, model performance improved as participants progressed through their program journey, demonstrating the importance of engagement data in predicting long-term clinical outcomes.

Conclusions: This study explored the Livongo for Diabetes RDMP participants’ temporal and static attributes, identification of diabetes management patterns and characteristics, and their relationship to predict diabetes management outcomes. Proactive targeting ML models accurately identified participants at risk of uncontrolled diabetes with a high level of precision that was generalizable through future years within the RDMP. The ability to identify participants who are at risk at various time points throughout the program journey allows for personalized interventions to improve outcomes. This approach offers significant advancements in the feasibility of large-scale implementation in remote monitoring programs and can help prevent uncontrolled glycemic levels and diabetes-related complications. Future research should include the impact of significant changes that can affect a participant’s diabetes management.

(JMIR Form Res 2024;8:e54373) doi:10.2196/54373
Introduction

Diabetes is a chronic disease that affects 37.3 million individuals living in the United States and requires ongoing management [1]. Common diabetes-related health complications that can be delayed or prevented with glycemic control include heart disease, chronic kidney disease, and neuropathy. Diabetes management is multifaceted and should include some level of provider care, medication management, and self-management. Telehealth interventions have been shown, and are encouraged, to support condition management and the reduction in diabetes-related complications [2]. With the current capabilities and advancements of telehealth in diabetes management, it is critical to help identify individuals with a higher risk of uncontrolled diabetes and provide them with targeted support and resources to help them manage their condition [3,4].

Predictive modeling has emerged as a valuable tool in diabetes management, offering numerous benefits for individuals with diabetes. By engaging early in a remote diabetes monitoring program (RDMP), predictive modeling can capture the user’s interest and motivation, fostering active participation in their own health through suggesting personalized behavior change in a timely manner to improve outcomes. The integration of proactive targeting into routine diabetes care also holds immense potential to transform the lives of millions affected by this chronic condition through preventing or delaying the onset of diabetes-related complications and reducing the burden on the health care system by lowering the cost of care and improving the quality of care. The current literature describing applied or developed predictive models has used insurance claims or electronic medical records to predict varying outcomes related to uncontrolled diabetes; however, there is a lack of literature on the prediction of uncontrolled diabetes management through a large data set of self-monitored blood glucose (SMBG) values and contributing factors, which provides a unique pathway to real-time predictions [5-7].

Machine learning (ML) can analyze a large amount of data and identify meaningful patterns that correspond to the diabetes risk level of individuals and predict their diabetes outcome risk [8-11]. The objective of this study was to conceptualize and develop a novel ML approach to proactively identify participants enrolled in a large-scale RDMP who were at risk of uncontrolled diabetes at 12 months in the program. Therefore, a set of dynamic predictive ML models were designed and trained at specific checkpoints during the participants’ time in the program to proactively identify those at risk and capture participant attributes that could impact the participants’ at-risk status.

Methods

Ethical Considerations

Approval was granted by the Aspire Institutional Review Board (IRB, #520160099), and guidelines outlined in the Declaration of Helsinki were followed. All participants provided consent to participate during enrollment into the Livongo for Diabetes RDMP, and guidelines outlined in the Declaration of Helsinki were followed. All study data were stored in Health Insurance Portability and Accountability Act–compliant secure servers and were deidentified prior to analysis. Participants were not compensated for their participation in the study.

Livongo for Diabetes

Teladoc Health’s Livongo for Diabetes is an RDMP focused on empowering participants with education and tools to self-manage their diabetes through mobile technology. The program offers participants a cellular-enabled, 2-way messaging device that measures blood glucose (BG) levels and delivers personalized insights into their glycemic management; free unlimited BG test strips; real-time support from diabetes response specialists 24 hours a day, 7 days a week, 365 days a year; and access to certified diabetes care and education specialists (CDCESs) for support and goal setting.

Teladoc participants’ BG meter use was captured remotely through the cellular-enabled device. Participants also had access to a web-based app and mobile phone app that tracked historical SMBG readings; provided reminders for SMBG checking, physical activity (PA), and food log tracking; and provided an asynchronous chat with coaches, the ability to schedule private coaching sessions with CDCESs, educational content for diabetes self-management, and the ability to send historical reports of SMBG readings to care providers, family members, and friends.

Diabetes Management Journey

The diabetes management journey (ie, program journey) refers to the ongoing process of monitoring and managing diabetes through the Livongo for Diabetes RDMP. The program journey involves education and support to make lifestyle changes, monitor SMBG levels, take medications as prescribed, and work with health care providers to achieve optimal health. The goal is to support participants with diabetes achieve BG control for complication avoidance and lead healthy, active lives.

Study Design

Registry data from the RDMP were used to design separate ML models to predict participant outcomes at each monthly checkpoint of the program journey (month-n models) from the first day of onboarding (month-0 model) up to the 11th month (month-11 model). The program journey of a participant began upon onboarding into the Livongo for Diabetes RDMP and...
capturing their first BG measurement through a Livongo BG meter. Overall, each participant passed through 12 predicative models through their first year enrolled in the program.

**Measures**

**Hemoglobin A1c and Observability**

Hemoglobin A1c (HbA1c), or glycated hemoglobin, is a critical metric for diabetes management, which provides a long-term picture of an individual’s average BG levels over a 2-3-month period [12]. HbA1c cannot be directly measured with a BG meter, which only measures the current BG level in the blood. To calculate HbA1c from BG meter readings, an algorithm that considers the average BG levels and the frequency of measurements was used: A1c = \text{average glucose (mg/dL)} – 46.7/28.7 [13]. In this algorithm, the more frequent and consistent the BG readings, the more accurate the estimate of A1c (eA1c). A participant was considered “observable” if enough BG readings over 90 preceding days provided statistical confidence to estimate clinically meaningful A1c; otherwise, the participant was considered “unobservable.” Statistical significance was determined by considering both the mean and SD of BG checks. The more variable the set of BG checks for a given participant, the greater the threshold for that participant to be deemed observable.

**Population Selection**

Participants enrolled in the Livongo for Diabetes RDMP between January 1, 2019, and January 1, 2022, with an activated BG meter who met the criteria to be categorized as observable at month 12 in the program were included as study participants (N=200,000). Ground truth labeling as “cases” and “controls” for diabetes management conditions was performed using 12-month eA1c values as follows: (1) participants with month 12 eA1c≥7.5% were labeled as cases and defined as participants at risk of uncontrolled diabetes management outcomes and (2) participants with month 12 eA1c<7.5% were labeled as controls and defined as participants not at risk of uncontrolled diabetes management outcomes. The ratio of cases to controls in the study population was 23.5%-76.5%.

**Model Design: Participant Features and Attributes**

Using diabetes-related available features and attributes of participants through their program journey, each ML model was designed and developed to make a binary prediction of whether the participant will be at risk of uncontrolled diabetes at the end of month 12 in their program journey. The eA1c at month 12 was used to generate binary labels for either month 12 eA1c≥7.5% or month 12 eA1c<7.5% as controls.

During the program journey, participants used the BG meter provided to record BG levels. Next, A1c was estimated using the accumulation of BG levels over a 90-day period. As described previously, clinically meaningful calculation of a member’s eA1c is dependent on the classification of observability. Since A1c has remained the clinical gold standard for indexing chronic glycemia for decades, and eA1c is an essential metric in assessing a participant’s diabetes condition, observability was a critical metric to predict participant outcomes. For unobservable participants, due to the lack of eA1c related to sparsity of BG checks and with the availability of derived program features, other features and attributes played a significantly more important role in the predictive ML models. Therefore, at each month of predictive modeling design, it was essential to train 2 separate models to cover participants with observable and unobservable eA1c values, making a total of 24 trained models.

The goal of an ML model is to find and learn patterns of input features from training data and then use them to make predictions on new, unseen data. Therefore, the quality and relevance of the features used are crucial for the performance of an ML model. We trained a set of ML models that needed to be sequentially compared, which made it critical to keep the structure of features consistent along the program journey to obtain a robust interpretation of features evolving among models.

**Participant Attribute Categories**

Participant attributes, including survey data, BG data, medication fill data, and health signals, were used for feature construction within the ML models.

**Survey Data**

Survey data gathered from the participants at enrollment included demographic information, such as age, gender, race, ethnicity, height, weight, BMI, language, and diabetes type. Participant intention and preference information around the diabetes management style, interest in becoming more active, and interest in healthy eating was also included in survey data. Each of the engagement attributes was encoded for model features as ordinal values. For example, in the case of interest in healthy eating, the options of not important, somewhat important, and very important were encoded as 1, 2, and 3, respectively.

**Blood Glucose Data**

BG data were measured through RDMP-provided BG meters with blood from a finger prick applied to a test strip, and participants were asked to select “feel tags” and “meal tags” from a set of options. Features were constructed from an accumulation of SMBG readings by 30-day aggregates. BG readings broken down by meal and feel tags, and A1c. The 30-day BG check aggregates generated data of the total number of readings and the number of hypoglycemic and hyperglycemic BG levels. Since the ML models were designed in monthly checkpoints, 30-day aggregates were important indicators of a participant’s diabetes pattern change.

BG levels are affected by both diet and mental health; therefore, BG readings were broken down by meal and feel tags to be correlated along each BG reading [14-16]. Meals increase glucose levels, while fasting typically decreases BG levels. BG levels can also be impacted by feelings, such as stress, and PA. Therefore, in the features, BG levels were broken down based on meal and feel tags. The meal tag options include before/after breakfast, before/after lunch, before/after dinner, after snack, and no meal. In addition, the feel tag options include feel fine, feel sick, stressed, ate extra, lightheaded, after exercise, missed medications, increased medications, and other feelings.
A1c is an important feature in diabetes management as it provides a snapshot of an individual’s overall BG control. In addition, variability of A1c depicts the diabetes management condition of an individual over time [17]. In the RDMP, A1c was estimated if a participant was observable. This feature was only available for observable participants, who recorded enough SMBG readings to estimate A1c with statistical significance.

Medication Fill Data
A GPI is a unique identifier assigned to a drug product to distinguish it from similar products and improve identification and tracking in the health care supply chain. Generic product identifiers (GPIs) were tracked over time to capture medicine fills, understand medication adherence, and determine a change in the health status of participants related to their diabetes. Based on participants’ medicine fills, they were categorized into the following diabetes groups using medication use as a proxy for disease progression [18]:

- Group 1: participants with type 2 diabetes who use only lifestyle modification (diet and exercise) and take no medication
- Group 2: participants with type 2 diabetes who take metformin
- Group 3: participants with type 2 diabetes who take metformin and other noninsulin medications
- Group 4: participants with type 2 diabetes who use basal insulin, in addition to other medications
- Group 5: participants with type 2 diabetes who use basal and bolus insulin (also known as multiple daily injections [MDIs] or intensive insulin)
- Group 6: participants with type 1 diabetes

Health Signal Data
A health signal refers to any measurable aspect of an individual’s physical or physiological state that provides information about their health status. Based on availability, 30-day aggregates of health signals were used in the features, such as average systolic and diastolic blood pressure (BP) readings, whether participants were enrolled in the Livongo for Hypertension (HTN) program, and the average weight, average daily PA, and food logs if participants were enrolled in weight management (WM).

Tracking these features that contribute to participant attribute categories over time can assist in personalized effective interventions for at-risk observable and unobservable participants.

Outcome Data
During evaluation of performance of the ML models for identification of participants at risk of uncontrolled diabetes management outcomes, various viewpoints were considered to prevent model shortcomings and imbalanced data: (1) How many at-risk participants were identified and with what precision by each model? (2) How many participants were targeted to achieve the performance metrics? (3) How does the progression of the performance metrics look over a participant’s program journey?

To provide a comprehensive view of model performance and address the aforementioned questions, the following 6 measures commonly used in ML model evaluation were selected [19]:

- Sensitivity or recall: Recall is a measure of how well a model can identify positive (at-risk participants) instances defined as the number of true-positive predictions divided by the total number of positive instances in the data set.
- Precision: Precision is a measure that evaluates the proportion of positive predictions that are correct and defined as the number of true-positive predictions divided by the sum of true-positive predictions and false-positive predictions.
- Specificity: Specificity is a measure that evaluates the ability of a model to correctly identify negative (not-at-risk participants) examples. It is the proportion of true negatives over all negatives.
- Area under the curve (AUC): The AUC represents a model’s ability to distinguish between positive and negative examples. In this study, AUC values ranged from 0 to 1, with a value of 0.5 representing a random guess and a value of 1 representing a perfect model. An AUC of 0.7 or higher is generally considered good, while an AUC of 0.9 or higher is considered excellent.
- \( F_1 \)-score: The \( F_1 \)-score is a performance metric that balances the precision and recall by calculating the harmonic mean of these 2 metrics.
- Accuracy: Accuracy measures the proportion of correct predictions made by a model out of all predictions.

Model Development
For proactive targeting of the RDMP participants’ diabetes management outcomes, it was necessary to train a set of ML models at specific time stamps through each participant’s program journey for the following reasons. First, the diabetes condition, such as severity, complications, and medication use, of a participant can evolve over time and impact the outcome. Similarly, the diabetes management patterns of participants, such as exercise, diet, and SMBG checking patterns, can alter over time and change the outcome trajectory. To capture these changes, new models needed to be trained along the program journey to update the risk prediction using more recent accumulated information. Second, as participants progressed in their program journey, more temporal data were collected from their retrospective diabetes management attributes. This change in the quality of attributes can change their importance and contribution in the ML modeling, indicating the need for new model training along the participant program journey.

As previously mentioned, from the first day of Livongo BG device activation, at each monthly step, a prediction ML model was trained to predict the binary outcome of the diabetes management status of each participant at 12 months in the program. At each month of the program journey, there were 2 separate models to develop based on observability. Figure 1 represents the timeline of modeling work for observable and unobservable segments of the participants’ program journey. Based on the designed framework, a set of 24 ML models were trained to capture the relationship between participants’ input attributes and diabetes management control outcomes.
Figure 1. Diabetes proactive targeting framework at each monthly observable and unobservable segment of each participant’s program journey. The program journey starts from device activation, and at each monthly checkpoint, based on the participant’s observability status, an ML model is trained to predict the outcome of the diabetes condition of the participant at year 1 of the program journey. ML: machine learning.

In modeling, for each model, data were split into 2 subsets. The first subset, which contained a larger portion of the data, was used as training and validation data to train the model and tune the hyperparameters, as well as find the optimal parameters of the model to achieve the highest accuracy. The second subset was used as testing data, which was held separate from the training data and used to assess how well the model could work with new data. Each model randomly split the data into 85% training and 15% testing subsets.

The models were based on light gradient boosting machine (LightGBM), which is a tree-based learning algorithm developed on the randomly selected subset of training data. With 5-fold cross-validation on training data, the Hyperopt Python library was used for tuning the hyperparameters of each LightGBM model, including the number of estimators, learning rate, number of leaves, feature fraction, and bagging fraction [20,21]. The tree-structured Parzen estimator (TPE) algorithm was used to optimize hyperparameter quantization [22]. Finally, the performance of each model was assessed on the unseen test data sets.

Training and testing instances for cases and controls among the data subsets for observable and unobservable participants for each month of the program journey were distributed (see Multimedia Appendix 1). The class imbalance of observable and unobservable participants was on average 78% versus 22% and 69% versus 31%, respectively, for controls and cases. To mitigate this highly imbalanced ratio of controls and cases and increase the focus of the models on less prevalent cases, the models were set to assign class weights inversely proportional to their respective frequencies during the training process.

Bias Mitigation Consideration
Bias reduction in ML models is crucial to ensure fair and equitable outcomes. The following considerations were taken to reduce bias in the ML model: class imbalance handling, feature selection and inclusion, observability consideration, missing imputation, cross-validation, and temporal changes.

Class Imbalance Handling
As previously mentioned, there was a class imbalance in the observable and unobservable participants. To address this, the models were set to assign class weights inversely proportional to their respective frequencies during the training process. This technique helped the models give more importance to the minority class, reducing bias toward the majority class.

Feature Selection and Inclusion
Four categories of participant attributes for feature construction within the models were used: survey data, BG data, medication fills, and health signals. This diverse set of features helped in capturing different aspects of participants’ health and behavior, reducing bias that might have arisen from relying on a limited set of features.

Observability Consideration
The observability of participants, distinguishing between observable and unobservable participants, was used. This factor was explicitly considered during model development, with different models trained for observable and unobservable participants. This approach acknowledged and addressed the potential bias introduced by the availability of certain data for only a subset of participants.

Missing Imputation
For unobservable participants with missing eA1c values, a robust imputation approach was used, which involved a mixture of historical eA1c data interpolation, leveraging their past records, and incorporation of similarity features from other participants, considering factors such as age and diabetes medications. This approach aimed at reducing bias in the imputation process, ensuring a more accurate estimation of missing eA1c values.

Cross-Validation
The models were based on LightGBM and underwent hyperparameter tuning using 5-fold cross-validation on training data. Cross-validation helped in assessing each model’s performance across different subsets of the data, reducing the risk of overfitting and ensuring generalizability.

Consideration of Temporal Changes
The study acknowledged the dynamic nature of diabetes conditions and the importance of capturing temporal changes. The models were trained at specific checkpoints during each participant’s program journey, allowing the models to adapt to evolving patterns and reducing bias introduced by changes over time.
Summary of the Methodology Workflow

The detailed sequential steps of the modeling process are listed next and represented visually in Figure 2:

Figure 2. Sequential steps in the modeling workflow, which included data preprocessing, feature preparation, model training, and model registry at each monthly program journey checkpoint. ML: machine learning; obs.: observed; unobs.: unobserved.

- Step 1 (data collection): Registry data from the Livongo for Diabetes RDMP were collected.
- Step 2 (population selection): Participants with an activated BG meter who met criteria for observability at month 12 were selected.
- Step 3 (ground truth labeling): Participants with eA1c ≥ 7.5% were labeled as cases, and those with eA1c < 7.5% were labeled as controls.
- Step 4 (observability classification/splitting): Participants were classified as observable or unobservable based on the statistical confidence derived from a sufficient number of BG readings over the preceding 90 days.
- Step 5 (feature construction): Features were constructed based on survey data, BG reading breakdown by meal and feel tags, medicine fill data, and health signals.
- Step 6 (feature engineering): Features were transformed to accommodate high performance of ML modeling. These steps included missing imputation using a combination of historical data and relevant participant characteristics with normalization and standardization of numerical features. In addition, for categorical features, ordinal encoding was applied, when possible. When ordinal encoding was not possible, one-hot-encoding was performed.
- Step 7 (class imbalance mitigation): To address imbalance between controls and cases, class weights were assigned inversely proportional to their respective frequencies during the training process.
- Step 8 (ML model development): Separate ML models were designed and trained at specific checkpoints during each participant’s program journey (from month 0 to month 11) for both observable and unobservable segments.
- Step 9 (hyperparameter tuning): Hyperparameter tuning was performed using the Hyperopt Python library, optimizing parameters such as the number of estimators, learning rate, number of leaves, feature fraction, and bagging fraction.
- Step 8 (model performance evaluation): Models were evaluated using standard metrics, including precision, recall, specificity, the AUC, the $F_1$-score, and accuracy.
- Step 10 (register models): Each trained model was registered into the model registry to be used for inference.

Descriptive Analysis

Patterns and trends of static and time series features and their correlation to the outcomes represented the participant breakdown by interest in learning about healthy eating and interest in becoming more active from the participant onboarding survey along with 12-month eA1c values. Assessing mean BG levels by meal and feel tags was crucial for predictive modeling to provide valuable insights into the patterns and factors influencing diabetes management outcomes. Understanding how BG levels vary before and after meals and in relation to different feelings supports the identification of critical points...
in a participant’s daily routine that may contribute to uncontrolled outcomes.

**Statistical Analysis**

To examine the correlation strength between extracted features and the diabetes outcomes (uncontrolled/controlled), Pearson r correlation analysis was used and represented top performers among 4 categories of attributes.

**Results**

**Participant Demographics**

Participant demographics and characteristics at the time of program enrollment are presented in Table 1.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diabetes type, %</strong></td>
<td></td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>92.1</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>7.3</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Insulin use, %</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>78.8</td>
</tr>
<tr>
<td>Yes, once/day</td>
<td>13.1</td>
</tr>
<tr>
<td>Yes, more than once/day</td>
<td>8.0</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td><strong>Race, %</strong></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>47.9</td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>37.1</td>
</tr>
<tr>
<td>Black/African American</td>
<td>7.4</td>
</tr>
<tr>
<td>Asian/Chinese/Japanese/Korean</td>
<td>3.9</td>
</tr>
<tr>
<td>Other</td>
<td>3.1</td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0.4</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>Ethnicity, %</strong></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>47.7</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>46.3</td>
</tr>
<tr>
<td>Hispanic</td>
<td>6.0</td>
</tr>
<tr>
<td><strong>Gender, %</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>51.7</td>
</tr>
<tr>
<td>Female</td>
<td>48.3</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>61.6 (12.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Year since diabetes diagnosis, mean (SD)</strong></td>
<td>9.03 (9.19)</td>
</tr>
<tr>
<td><strong>Self-reported A1c at enrollment, mean (SD)</strong></td>
<td>7.33 (1.54)</td>
</tr>
</tbody>
</table>

RDMP: diabetes remote monitoring program.

**Descriptive and Statistical Analysis**

Patterns and trends of static and time series features and their correlation to the outcomes are illustrated in Figure 3. Figure 3a shows that participants with a higher interest in learning about healthy eating had a lower mean eA1c outcome. Similarly, as shown in Figure 3b, participants with a higher interest in becoming more active had a lower mean eA1c outcome.
Figure 3. Participant breakdown and mean eA1c of survey responses to (left) interest in learning about healthy eating and (right) interest in becoming more active. eA1c: estimate of A1c.

Figure 4 represents the mean BG levels of participants along their 1-year program journey by meal tag and feel tag. As shown in Figure 4a, the BG level tagged after breakfast, lunch, or dinner was on average 20% higher than that before meals. In addition, as illustrated in Figure 4b, BG levels related to feeling fine, lightheaded, and after exercise were the lowest, while those after missed or increased medications and eating extra were the highest.

In the BG data category, the BG check breakdown by feel and meal tags was highly correlated, including BG checks tagged as “feel fine” ($r=-0.18, P<.001$) and before breakfast, lunch, and dinner ($r=0.155, P<.001$). In addition, among survey data, age had the highest correlation ($r=0.164, P<.001$). Among medicine fills, the metformin fills feature was presented as a top feature ($r=0.072, P<.001$). Finally, among health signals, 30-day PA had a correlation of 0.039 ($P<.001$).

Figure 4. Mean BG level of participants along their 1-year program journey by (left) meal tag and (right) feel tag. BG: blood glucose.

Performance Results and Output of the 12 ML Models
Performance results of the 12 ML models for observable participants for the training and testing subsets from month 0 to month 11 of their program journey are presented in Multimedia Appendix 2. It was crucial to understand that due to class imbalance, a random prediction would not result in a theoretical precision or an $F_1$-score of 0.5. The precision obtained from a random prediction would be proportional to the number of cases, and the $F_1$-score would be accordingly impacted. For observable participants, the baseline case was on average 0.22 (range −0.009 to +0.009 across months in the program journey), which was outperformed by the precision of monthly models with an average precision of 61%, ranging from 0.4 at month 0 to 0.88 at month 11, with consistent improvement along the participant’s program journey. Similarly, recall started from 0.7 at month 0 of the program journey and reached 0.94 at month 11. The AUC of the models also improved consistently, ranging from 0.76 to 0.98 across models along the participant’s program journey.

Similar to observable participants, results of the performance of the 12 ML models for unobservable participants along their program journey are shown in Multimedia Appendix 3. For unobservable participants, the baseline case was on average 0.31 (range −0.05 to +0.035 across months in the program journey).
journey), which was outperformed by the precision of monthly models with an average precision of 0.55, ranging from 0.42 at month 0 to 0.61 at month 11 on the testing set.

The recall of the ML models at month 0 was 0.61 and by further progress of the participant’s program journey increased to 0.82, with a precision of 0.61 at month 11. At month 1, there was a significant improvement in performance, with an improved recall and precision increase of +0.17 and +0.11, respectively. This difference was due to the difference in input features of month 0 and other monthly models. At month 0, only survey data and medicine fills were being used as model features; however, for the rest of the models, BG data and health signals were added, causing a drastic improvement in the performance of the models. In addition, the AUC of the models was on average 0.8 (range 0.7-0.87), representing good performance of the models.

Figure 5 represents the comparative performance of the models along the monthly program journey for observable and unobservable participants on the testing sets. In the figure, blue and red lines represent the performance metrics for observable and unobservable participants, respectively. On average, the recall of models for observable participants was 0.09 higher than that of models for unobservable participants, with the difference ranging from 0.04 to 0.16. Similarly, the precision of models for observable participants was on average 0.06 higher than that of models for unobservable participants, with a range of 0.02-0.26. In addition, specificity metrics had the highest difference in these 2 groups, with an average difference of 0.11.

Figure 5. Performance metrics (recall, precision, and specificity) of models along each participant’s program journey for observable (blue) and unobservable (red) subsets.

Discussion

Principal Findings

The ML approach used in this study demonstrated high capability to proactively identify participants enrolled in an RDMP who were at risk of uncontrolled diabetes using participant survey inputs, SMBG data, medical history, and diabetes management engagement signals. Proactive targeting models accurately identified observable participants at risk of uncontrolled diabetes (71%-94%; mean 86%, SD 6%) and unobservable at-risk participants (64%-82%; mean 77%, SD 5%) from month 0 to month 11, with an achieved precision of 40%-88% (mean 62%, SD 12%) and 42%-61% (mean 57%, SD 8%), respectively. As participants progressed through their program journey, the prediction models became more accurate and performed better in identification of those at risk among observable participants compared to unobservable participants. In addition, the performance difference between month 0 and month 1 in both observable and unobservable participants was significantly higher than in the consecutive months, demonstrating the importance of engagement data for predicting long-term clinical outcomes. The most critical feature in the models was the last-available eA1c, followed by diabetes-related medication fills and BG checking patterns with meal and feel tags [23].

Comparison With Prior Work

Various prediction models have been developed and implemented in the literature over the past decade with increased accuracy in predicting the diabetes risk over time, specifically the transition of prediabetes to diabetes [11,24,25]. However, nearly all the literature focuses on health care data sets from hospital patients [10,26,27]. Although these studies have shown that ML models can preserve performance across populations with health care data collected through demographic data, laboratory values, and hospital records, there is a lack of literature implementing ML and proactive targeting in a real-world RDMP predicting an individual with diabetes entering uncontrolled diabetes status using engagement behaviors and SMBG levels. Additionally, prior studies restrict input data in their models to contain specific information, such as baseline A1c, and train models on a specific cohort of uncontrolled baseline A1c [28]. Our study focused on ML models with the ability to perform on limited and sparse data that are indicative of diabetes management in a real-world population.

The data available through the Livongo for Diabetes RDMP supported the ability to develop and achieve significant accuracy of the ML models used in this study. Consistent capture of
participants’ SMBG patterns and frequency and self-monitoring behaviors around PA, diet, and mental health was essential to develop models for both observable and unobservable participants. By conducting sufficient SMBG checks with statistical confidence, a participant’s A1c can be estimated and the participant becomes observable, which was an important parameter in our study. Due to fewer data points of SMBG-related features for unobservable participants, other features became more important in their corresponding predictive models. The result of this study’s approach is applicable to proceeding years of participants’ program journey and other chronic condition remote monitoring program journeys.

The results achieved in this study demonstrate robust and generalizable performance in proactive targeting of at-risk participants enrolled in an RDMP, which can provide significant advancements toward the feasibility of large-scale implementation in the following aspects. First, early identification of at-risk participants will allow for ease in effective interventions to change the trajectory of outcomes and aid in the prevention of potential uncontrolled glycemic levels and diabetes-related complications. Second, provided that participants can become observable or unobservable and at risk or not at risk at varying time points throughout the program journey, proactive targeting will ensure identification of participants with higher probability. Lastly, using proactive targeting, along predictive models, the evolving metrics for diabetes management can also be obtained, such as the diabetes medication stage. These metrics combined with predictive outcomes can provide personalized intervention strategies to improve participant outcomes.

Strengths and Limitations
This study has many strengths, including the use of data collected from participants enrolled in a real-world RDMP. The ML models used in the study also have 2 distinguishing strengths. First, the proactive targeting design enables ongoing tracking of participants’ diabetes management condition toward reaching the desired outcome and early identification of at-risk participants, which allows for identification of factors that may cause a participant to become at risk of uncontrolled diabetes. Second, the models are not limited to the availability of specific data. In fact, the lack of availability and sparsity in each feature can represent an important factor about the participants. For example, the sparsity of BG checks can represent low program engagement, especially for participants with higher diabetes severity. This consideration avoids the self-selection bias of the models toward active participants.

Some limitations also exist related to the complexity of identifying participants at risk of uncontrolled diabetes management outcomes proactively. Due to the signal of SMBG levels being dependent on SMBG checking patterns of participants, the accuracy of prediction decreases by participant inactivity. This limitation can be resolved by using other BG monitoring technologies, such as continuous glucose monitoring devices [29]. In addition, identifying at-risk participants earlier within the program journey would be helpful to develop effective and timely interventions; however, the lower performance of models at earlier time points due to a lack of enriched features limited this objective. In addition, the models may have been influenced by external factors not accounted for in our study, such as individual socioeconomic factors and environmental variables. This issue can be partially diminished by adding extra generalizable features, such as the social determinants of health, to the feature sets [30]. Future research could explore the incorporation of a broader range of contextual features to enhance the models’ robustness and generalizability across diverse populations.

Conclusion
This study explored participants’ temporal and static attributes, the identification of diabetes management patterns and characteristics, and their relationship to predict diabetes management outcomes. Proactive targeting models accurately identified participants at risk of uncontrolled diabetes with a high level of precision that was generalizable through future years within the RDMP. Future research should include the impact of significant changes that can affect a participant’s diabetes management—for example, granular medication changes, such as drug and dosage changes, as well as medication adherence, which are important factors for the determination of diabetes outcomes [31].

Acknowledgments
The study was funded by Teladoc Health, Inc.

Data Availability
Deidentified data from this study are not available in a public archive and will be made available (as allowable according to Institutional Review Board standards) by emailing the corresponding author.

Conflicts of Interest
All authors were employed by Teladoc Health, Inc, during the study period.

Multimedia Appendix 1
Number of participant cases and controls for training and testing the data subset for each monthly checkpoint of the participants’ program journey.
Multimedia Appendix 2
Performance metrics for training and testing the data subset for observable participants at each monthly checkpoint of their program journey. The accuracy metrics for each month in the program journey represent the performance of a stand-alone ML model. ML: machine learning.

Multimedia Appendix 3
Performance metrics for training and testing the data subset for unobservable participants at each monthly checkpoint of their program journey. The accuracy metrics for each month in the program journey represent the performance of a stand-alone ML model. ML: machine learning.

References


Abbreviations

AUC: area under the curve
BG: blood glucose
CDCES: certified diabetes care and education specialists
eA1c: estimate of A1c
GPI: generic product identifier
LightGBM: light gradient boosting machine
ML: machine learning
PA: physical activity
RDMP: remote diabetes monitoring program
SMBG: self-monitoring blood glucose

Edited by A Mavragani; submitted 07.11.23; peer-reviewed by H Zheng, P Dunn; comments to author 22.12.23; revised version received 12.01.24; accepted 20.01.24; published 26.04.24.

Please cite as:
URL: https://formative.jmir.org/2024/1/e54373
doi:10.2196/54373
PMID:38669074
Expanding Youth-Friendly HIV Self-Testing Services During the COVID-19 Pandemic: Qualitative Analysis of a Crowdsourcing Open Call in Nigeria

Onyekachukwu Anikamadu¹, MPH; Oliver Ezechi², MD, PhD; Alexis Engelhart³, MPH; Ucheoma Nwaozuru⁴, PhD; Chisom Obiezu-Umeh⁵, MPH; Ponnmile Ogunjemite⁶, BSc; Babatunde Ismail Bale⁷, OD; Daniel Nwachuku², BSc; Titilola Gbaja-biamila⁹, MBBS; David Oladele³, MBBS; Adesola Z Musa⁵, PhD; Stacey Mason⁵, MPH; Temitope Ojo³, MPH, PhD; Joseph Tucker⁸,⁹, AM, MD, PhD; Juliet Iwelunmor³, PhD

¹Brown School of Social Work, Washington University in St. Louis, St Louis, MO, United States
²Clinical Sciences Department, Nigerian Institute of Medical Research, Lagos, Nigeria
³Division of Infectious Disease, Washington University School of Medicine in St. Louis, St Louis, MO, United States
⁴Department of Implementation Science, Wake Forest School of Medicine, Winston-Salem, NC, United States
⁵Department of Behavioral Science and Health Education, College for Public Health and Social Justice, Saint Louis University, St Louis, MO, United States
⁶Department of Human Anatomy, Federal University of Technology, Akure, Nigeria
⁷Department of Optometry, University of Benin, Benin City, Nigeria
⁸Division of Infectious Diseases, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
⁹Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding Author:
Onyekachukwu Anikamadu, MPH
Brown School of Social Work
Washington University in St Louis
1 Brookings Drive
St Louis, MO, 63130
United States
Phone: 1 314 935 6600
Email: a.onyekachukwu@wustl.edu

Abstract

Background: HIV self-testing (HIVST) among young people is an effective approach to enhance the uptake of HIV testing recommended by the World Health Organization. However, the COVID-19 pandemic disrupted conventional facility-based HIV testing services, necessitating the exploration of innovative strategies for the effective delivery of HIVST.

Objective: This study analyzed the outcomes of a digital World AIDS Day crowdsourcing open call, designed to elicit youth responses on innovative approaches to promote HIVST among young people (14-24 years) in Nigeria during COVID-19 restrictions.

Methods: From November 2 to 22, 2020, a World AIDS Day 2020 crowdsourcing open call was held digitally due to COVID-19 restrictions. The crowdsourcing open call followed World Health Organization standardized steps, providing a structured framework for participant engagement. Young people in Nigeria, aged 10-24 years, participated by submitting ideas digitally through Google Forms or email in response to this crowdsourcing open call prompt: “How will you promote HIV self-testing among young people during COVID-19 pandemic?” Data and responses from each submission were analyzed, and proposed ideas were closely examined to identify common themes. Four independent reviewers (AE, SM, AZM, and TG) judged each submission based on the desirability, feasibility, and impact on a 9-point scale (3-9, with 3 being the lowest and 9 being the highest).

Results: The crowdsourcing open call received 125 eligible entries, 44 from women and 65 from men. The median age of participants was 20 (IQR 24-20) years, with the majority having completed their highest level of education at the senior secondary school level. The majority of participants lived in the South-West region (n=61) and Lagos state (n=36). Of the 125 eligible entries, the top 20 submissions received an average total score of 7.5 (SD 2.73) or above. The panel of judges ultimately selected 3 finalists to receive a monetary award. Three prominent themes were identified from the 125 crowdsourcing open call submissions.
as specific ways that HIVST can adapt during the COVID-19 pandemic: (1) digital approaches (such as gamification, photoverification system, and digital media) to generate demand for HIVST and avoid risks associated with attending clinics, (2) awareness and sensitization through existing infrastructures (such as churches, schools, and health facilities), and (3) partnerships with influencers, role models, and leaders (such as religious and youth leaders and social influencers in businesses, churches, organizations, and schools) to build trust in HIVST services.

Conclusions: The crowdsourcing open call effectively engaged a diverse number of young people who proposed a variety of ways to improve the uptake of HIVST during the COVID-19 pandemic. Findings contribute to the need for innovative HIVST strategies that close critical knowledge and practice gaps on ways to reach young people with HIVST during and beyond the pandemic.

Trial Registration: ClinicalTrials.gov NCT04710784; https://clinicaltrials.gov/study/NCT04710784

(JMIR Form Res 2024;8:e46945) doi:10.2196/46945

KEYWORDS
crowdsourcing; World AIDS Day; HIV; self-testing; young people; COVID-19 pandemic restrictions; Nigeria; HIV self-testing; health promotion; crowdsourcing open call; young adult

Introduction

Nigeria stands as the fourth-highest contributor to the global burden of HIV and has the largest HIV burden in sub-Saharan Africa [1,2]. An estimated 1.9 million Nigerians were living with HIV in 2018, setting the national HIV prevalence among those aged 15 to 49 years at 1.5% [3]. Since then, as of 2020, rates have remained constant, with only 67% of those affected aware of their status and only 53% receiving treatment [3]. Even with strengthened interventions over the past few decades, Nigeria still has a high burden of HIV, and a high number of young people (14-24 years) acquire the virus [3].

HIV self-testing (HIVST) is an important step that can enhance the uptake of essential HIV prevention services and is recommended by the World Health Organization [4]. HIVST is an antibody test that allows individuals to conveniently test themselves at home or in a private setting, yielding results within 20 minutes [5]. Recognized as a safe and effective method for increasing testing rates, especially among young people, HIVST has long been considered an alternative to clinical testing [4]. With the ability to collect their own samples and conduct the test at their own convenience, individuals gain autonomy in the testing process [4]. Although HIVST had been gaining traction prior to the COVID-19 pandemic, COVID-19 regulations hampered HIV testing, hastening the creation and scale-up of HIVST services [6].

The COVID-19 pandemic significantly impacted and disturbed the global health care system to varying degrees, with HIV prevention and treatment programs among those that were not spared [7]. A shift in the health system’s focus impacted all health programs not only in Nigeria but around the world. Many resources were redirected due to the COVID-19 pandemic, causing disruptions in existing clinical and community-based services surrounding prevention, treatment, and curative services, including those of HIV [8,9]. HIVST can play a critical role in decentralizing services, guaranteeing the use of HIV testing services during periods of uncertainties. It allows both beneficiaries and health care providers to adhere to physical distancing guidelines and limits the danger of exposure to transmission. While certain populations of Nigerian young people are at high risk of being infected with HIV, there is minimal evidence on measures to improve HIVST uptake among young people in Nigeria during the COVID-19 pandemic. An understanding of the strategies would be valuable to Nigerian young people and policy makers as well as global researchers implementing interventions in Nigeria or those who desire to adapt interventions to different contexts, particularly during disruptions such as those caused by the pandemic. To address this gap, we launched a digital crowdsourcing open call to elicit ideas for promoting HIVST during the pandemic.

Crowdsourcing is an open innovation approach, often used as a tool to gather ideas, innovations, or information for specific purposes [10]. It is an effective strategy for communicating and designing interventions related to HIV and sexual health [11,12]. The goal of this qualitative study was to uncover recurring themes in a digital World AIDS Day (WAD) crowdsourcing open call for youth responses on how to increase HIVST among Nigerian young people during the COVID-19 pandemic.

Methods

Overview

Our crowdsourcing open call consisted of a 5-step process including digital crowdsourcing open call, web-based submissions, judging, analysis of themes, and common themes identified (Figure 1) throughout the design and implementation, data collection, and data analysis phases.
Digital Crowdsourcing Open Call Design and Implementation

For this study, we used crowdsourcing, which involves a group of individuals working together to solve a problem and then publicly share the solutions [13,14]. The 4 Youth By Youth (4YBY) team announced a crowdsourcing open call to promote HIVST among young people in Nigeria from November 2 to 22, 2020. 4YBY is a team of young people, health professionals, activists, and entrepreneurs from diverse backgrounds, who are united by the shared passion to advance Nigerian youth participation in creating innovative, sustainable HIV prevention services. The “World’s AIDS Day HIV Self-Testing Contest” was held on December 1, 2020, to commemorate the annual WAD celebration. Due to the COVID-19 pandemic, we only used web-based engagement to recruit Nigerian young people between the ages of 14 and 24 years to take part in the crowdsourcing open call. To guarantee that a varied range of Nigerian young people from various backgrounds were involved, purposive sampling procedures were used [15]. The 4YBY team sent the crowdsourcing open call information to youth listservs and advertised on all social media platforms, including Instagram, Facebook, WhatsApp, and a website. The goal of the digital crowdsourcing open call was to provide open and safe areas for Nigerian young people to express their thoughts and ideas about how to promote HIVST during the pandemic. Individual demographic and contact information such as name, number, age, gender identity, education, and location were collected for each entrant.

Web-Based Collection of Submission Data

Participants responded to the prompt “How might you promote HIV self-testing among young people during COVID-19 measures?” using either Google Forms or email. Each submission had a word limit of 150 words to succinctly capture their unique ideas for HIV promotion throughout the pandemic. Consent to participate and demographic information were also obtained from each participant. The requested demographic information included contact information, age, gender, relationship status, current location, highest level of completed education, and occupation. Individuals had the option of not reporting sociodemographic characteristics.

Judging

Two research members (CO-U and UN) checked for the eligibility of each submission. Four independent reviewers (AE, SM, AZM, and TG) rated participant entries based on defined judging criteria, which included desirability, feasibility, and impact (Textbox 1). Each submission received either a 1 (low), 2 (moderate), or 3 (high) for each of the 3 defined judging criteria, with the highest overall combined score being 9. This process was adapted from the human-centered, design thinking framework [15,16]. The first, second, and third place contestants each earned a monetary prize for their outstanding submission: 50,000 Naira (approximately US $139) for third place, 150,000 Naira (approximately US $417) for second place, and 250,000 Naira (approximately US $694) for third place.

Textbox 1. 4 Youth By World AIDS Day crowdsourcing open call judging criteria (based on a 9-point scale: 3-9, with 3 as the lowest and 9 as the highest).

Desirability (1-3): Concept is appealing to young people and is affordable, accessible, and confidential.
Feasibility (1-3): Concept is practical in terms of implementation and resource availability.
Impact (1-3): Concept has the ability to influence young people to self-test for HIV and can reach young people in Nigeria.

Data Analysis

Following the crowdsourcing open call’s conclusion, staff deidentified each entry and used summary statistics to compile the participants’ demographic data. A thematic analysis was conducted using open coding that assigns themes to capture specific ideas and axial coding, which explores linkages between concepts and categories and determines common themes [17,18]. The coding exercises were completed in a Google Spreadsheet by trained staff who were not part of the study team, and then a thematic codebook was constructed and used to assess qualitative data. Each of the 2 staff members (OA and AE) coded the crowdsourcing open call entries separately before meeting to resolve any discrepancies in coding judgments. The
The codebook was then created by consensus between the 2 researchers (CO-U and UN) [18]. Throughout the coding process, recurring themes were identified, and codes were created to meet the HIVST promotion themes. The staff members reread each submission a final time after developing the initial codebook to fine-tune the final codebook. Snapshots of entries that fit in each code were included in the final codebook. This was then forwarded to a third party for final examination and approval.

**Ethical Considerations**

Regulatory approval to conduct the research was received from the Nigerian Institute of Medical Research Institutional Review Board (Project #: IRB 18/028). Informed consent was obtained from all participants involved in the study. Following the conclusion of the crowdsourcing open call, each entry was deidentified to ensure anonymity. These measures were implemented to uphold ethical standards, ensuring participant protection and transparency throughout the research process. Additionally, monetary compensation was awarded to first-, second-, and third-place submissions.

**Results**

**Characteristics of Entrant Submissions**

The 2020 digital WAD crowdsourcing open call received a total of 153 entries from Nigerians aged 14 to 24 years, of which 83.7% (n=128) were through Google Forms, and 16.3% (n=25) were by email. Of the 153 entries, there were 125 unique submissions identified after duplicates (n=24), and those submissions that could not be scored based on our inclusion criteria (n=4) were removed. Characteristics and demographics were collected for the 125 eligible, unique submissions.

The majority of the entrants were male (n=65, 52%). There were 44 (35.2%) female individuals, 1 (0.8%) individual preferred not to say their gender, and 15 (12%) gender responses were left missing. The average age of the entrants was 19.96 (SD 2.73) years, and the median age was also 20 (IQR 24-20) years. The breakdown of the number of individuals per age group is as follows: 48 (38.4%) individuals aged 14-19 years and 64 (51.2%) individuals aged 20-24 years. The highest level of education obtained by participants was mostly senior secondary school (n=61, 48.8%), followed by some tertiary school (n=22, 17.6%), bachelor’s degree (n=17, 13.6%), junior secondary school (n=4, 3.2%), primary school (n=2, 1.6%), and master’s degree (n=2, 1.6%). In total, 15 (12%) individuals did not report their highest level of education. The majority of entrants lived in Lagos state (n=36, 28.8%) and in the South-West region (n=61, 48.8%). Table 1 includes the characteristics of the eligible submissions.

Of the 125 eligible submissions, the mean score of the submissions was 5.42 (1.65). In total, 20 participants were selected as finalists with their submissions scoring as 7.5 or above. Three submissions were selected as the top submissions and were given a prize (Multimedia Appendix 1).

The average word count of Google Forms submissions was 425.27 (SD 392.91) words. Two videos were submitted with lengths of 4.3 and 6.4 minutes. A total of 5 submissions included images, 1 submission consisted of a video and images, 1 entrant submitted a PowerPoint (Microsoft Corp) presentation file, and 4 submissions included a combination of images and text. All other entries were text submissions.
Table 1. Characteristics of submissions (N=125 eligible submissions).

<table>
<thead>
<tr>
<th>Characteristics of submissions</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Forms, n (%)</td>
<td>100 (80)</td>
</tr>
<tr>
<td>Email, n (%)</td>
<td>25 (20)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>65 (52)</td>
</tr>
<tr>
<td>Female</td>
<td>44 (35.2)</td>
</tr>
<tr>
<td>Preferred not to say</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Missing</td>
<td>15 (12)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>14-19 years, n (%)</td>
<td>48 (38.4)</td>
</tr>
<tr>
<td>20-24 years, n (%)</td>
<td>64 (51.2)</td>
</tr>
<tr>
<td>Missing, n (%)</td>
<td>13 (10.4)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>19.96 (2.73)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>20 (24-20)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Junior secondary school</td>
<td>4 (3.2)</td>
</tr>
<tr>
<td>Senior secondary school</td>
<td>61 (48.8)</td>
</tr>
<tr>
<td>Some tertiary school</td>
<td>22 (17.6)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>17 (13.6)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Missing</td>
<td>15 (12)</td>
</tr>
<tr>
<td>Location (by state), n (%)</td>
<td></td>
</tr>
<tr>
<td>Abia</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Abuja</td>
<td>4 (3.2)</td>
</tr>
<tr>
<td>Akure</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Akwa Ibom</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Anambra</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Bauchi</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Bayelsa</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Benue</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Cross River</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Delta</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Ebonyi</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Ekiti</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Enugu</td>
<td>5 (4)</td>
</tr>
<tr>
<td>Imo</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Kaduna</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Kano</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Kogi</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Kwara</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Kwara</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Lagos</td>
<td>36 (28.8)</td>
</tr>
<tr>
<td>Nasarawa</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Characteristics of submissions</td>
<td>Values</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Ogun</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Ondo</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Osun</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Oyo</td>
<td>10 (8)</td>
</tr>
<tr>
<td>Plateau</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Rivers</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Yobe</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Missing</td>
<td>14 (11.2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location (by region), n (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>4 (3.2)</td>
</tr>
<tr>
<td>East-Central</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>North-Central</td>
<td>10 (8)</td>
</tr>
<tr>
<td>North-East</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>North-West</td>
<td>7 (5.6)</td>
</tr>
<tr>
<td>South-East</td>
<td>21 (16.8)</td>
</tr>
<tr>
<td>South-South</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>South-West</td>
<td>61 (48.8)</td>
</tr>
<tr>
<td>Missing</td>
<td>14 (11.2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score, mean (SD)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3.5</td>
<td>28 (22.4)</td>
</tr>
<tr>
<td>4-6.5</td>
<td>68 (54.4)</td>
</tr>
<tr>
<td>7-9</td>
<td>30 (24.0)</td>
</tr>
<tr>
<td>Above 7.5 (semifinalists)</td>
<td>20 (16.0)</td>
</tr>
</tbody>
</table>

Analysis of Themes

Overview

After analyzing all 125 eligible submissions, 3 key themes were identified from the crowdsourcing open call submissions as specific ways that HIVST can adapt to the COVID-19 pandemic (theme 1) and the possibilities as well as facilitators of HIVST during the COVID-19 pandemic (themes 2 and 3). Common strategies to promote HIVST during and after COVID-19 were through (1) digital approaches (such as gamification, photoverification system, and digital media) to generate demand for HIVST and avoid risks associated with attending clinics, (2) awareness and sensitization through existing infrastructures (such as churches, schools, and health facilities), and (3) partnerships with influencers, role models, and leaders (such as religious and youth leaders and social influencers in businesses, churches, organizations, and schools) to build trust in HIVST services (Multimedia Appendix 2).

Theme 1: Digital Approaches to Generate Demand for HIVST and Avoid Risks Associated With Attending Clinics

This was a recurring theme that appeared to be desirable, feasible, and impactful to young people in Nigeria. Because COVID-19 can be spread through airborne transmission and social distancing regulations were put into place, young people saw digital media as a safe method to transmit awareness and information regarding HIV and HIVST. Young people established ideas to educate and promote HIVST through digital media engagement avenues such as Facebook, Twitter, Instagram, WhatsApp, gamification, and mobile apps. For example, entrants proposed creating mobile apps that provide knowledge on HIVST and COVID-19 testing as well as linkage to testing facilities (Figure 2). The top finalist suggested an interactive health gaming app called “Bambam” that allows users to access and monitor their health and HIV status coupled with competitive incentives to keep users engaged and encourage peer synergy. Because many young people in Nigeria have access to the internet and digital devices, participants mentioned the importance of using digital approaches as a medium for HIV and HIVST awareness campaigns to limit social gatherings and limit the influx of people to high-traffic areas such as hospitals. Submission examples are presented in Multimedia Appendix 2.
Theme 2: Awareness and Sensitization Through Existing Infrastructures

This was a common theme that arose even during the COVID-19 pandemic. Many submissions suggested that existing facilities and infrastructures can still be used as resources for HIV information and as testing sites. They emphasized the importance of wearing face masks, applying hand sanitizer, and social distancing while at these places. A few participants mentioned the idea of creating awareness through religious gatherings, as churches and religious bodies were still allowed to be open during the pandemic. Other participants made it clear that holding discussions and seminars in physical locations gives rise to more social and personal connections. However, participants mentioned it is important to avoid large gatherings. Submission examples are presented in Multimedia Appendix 2.

Theme 3: Partnerships With Influencers, Role Models, and Leaders to Build Trust in HIVST Testing Services

This was portrayed as a strategy that would impact young people in Nigeria and influence them to self-test for HIV. Participants mentioned collaborating with people such as social media influencers, religious and community leaders, and physicians to educate and sensitize young people in Nigeria on HIV and link them to self-testing kits and testing services. Participants suggested using celebrities as ambassadors to promote and encourage HIVST among young people (Figure 3). Entrants also proposed the idea of leaders reaching out, through trained personnel and volunteers, to individuals in rural and more local areas, as these individuals may not have access to internet services. Submission examples are presented in Multimedia Appendix 2.

Additionally, a few participants also mentioned similarities between the fight against COVID-19 with that of HIV. Participants recognized the comparability between COVID-19 and HIV and how HIV facility-based testing is stalled due to COVID-19 regulations such as facilities shutting down or social distancing guidelines being initiated. Participants highlighted the accessibility of HIVST, in which self-testing for HIV can be administered in an individual’s own home and privacy.

Figure 2. ViraCare, an example of entrants’ proposed mobile apps that provide knowledge on HIV self-testing and COVID-19 testing as well as linkage to testing facilities. STI: sexually transmitted infection.
Figure 3. CURES, a participant submission that explained the use of 5 major tools to promote HIVST among young people. HIVST: HIV self-testing.

Follow-Up Activities
Following the conclusion of the digital crowdsourcing event, the top-scored ideas generated informed the refinement of a youth-friendly HIVST intervention delivery, particularly focusing on enhancing linkage to care. This refined intervention is to be evaluated in a randomized controlled trial (ClinicalTrials.gov NCT04710784).

Discussion
Principal Findings
This digital WAD crowdsourcing open call administered during the COVID-19 pandemic generated ideas on how HIVST uptake can be increased among Nigerians in the course of the COVID-19 pandemic. The study illustrated efforts to reach young people creatively to facilitate HIVST while being compliant with the COVID-19 pandemic. Our findings also highlight accessible, affordable, and feasible ways to create awareness of HIVST throughout the COVID-19 pandemic and after the pandemic.

Our data suggest that HIVST can be effectively used to meet the demands of young people’s HIV testing during the pandemic. The COVID-19 pandemic created countless barriers that required the public to rethink a way around the norm and develop new strategies to overcome the obstacles, specifically reconsidering how to promote and fulfill HIVST services [19]. For example, across Nigeria, there were many disruptions to health care services [20], including those of HIV and sexually transmitted infections [9,21]. Yet, there was a need for innovative public health and social measures to reduce the disruptions caused by the pandemic [22]. Our crowdsourcing open call illustrated how a digital program can be used as a strategy to promote HIVST services that are compliant with the COVID-19 guidelines carried out in Nigeria [23]. Digital activities such as our crowdsourcing open call allowed for COVID-19 mitigation measures to continue to occur successfully while being compliant with government guidance [24,25].

Participants generated ideas on how HIVST distribution can be modified to adhere to the ongoing restrictions caused by the pandemic through channels that are digital as well as existing infrastructures and partnerships with key players. The ideas proposed can help form future campaigns, specifically during a time of limited physical interaction, and can pave the way toward more innovative and cost-saving techniques.
This study was held completely digitally due to the COVID-19 pandemic. It was a safe way to connect with individuals across Nigeria and administer new ideas that followed COVID-19 regulations. Although a large number of individuals participated in our crowdsourcing open call, it was significantly less than our crowdsourcing open calls in previous years 2018 and 2019 [15]. This could be related to competing COVID-19 demands. Though we gave the options of both web-based or offline submissions, all submissions were received via email or Google Forms. We recognize that technological advancements are continually occurring globally, and many young people have access to a mobile device, laptop, or computer [26]. In Nigeria, 63.8% of the population had internet access in 2020 [27]. However, internet connections are not always reliable and lead to issues during web-based learning or streaming [26]. Because our digital crowdsourcing open call was held via Zoom (Zoom Video Communications), this may have averted individuals from the crowdsourcing open call, knowing that they may have a possibility of internet connection issues.

However, as in previous years, the crowdsourcing open call engaged a wide range of young people and allowed them to voice their ideas on how to promote HIVST through various innovative strategies [15]. Crowdsourcing is not only a strategic way to receive solutions [9,10] but also very unique in the way that it includes individuals with different perspectives and creates meaningful participation among young people [13,14,28]. Our crowdsourcing open call accentuated this with youth participation and the development of ideas to prospectively scale up HIVST among young people and in communities throughout Nigeria. Harnessing young people’s concepts through crowdsourcing open calls captivates young people’s perspectives and potentially enhances mobilization efforts. Young people generated adapted, effective, and original strategies relating to the design, distribution, and education of HIVST kits during the COVID-19 era. We showed how young people can engage, produce innovative approaches, and create in-person or web-based connections.

Limitations

This study has several limitations. First, most ineligible entries were those submitted via Google Forms, with 24 as duplicates and 2 as submissions that could not be scored. Second, many submissions included text that did not specifically answer the crowdsourcing open-call question, such as those that defined HIV and those that described the burden of HIV in Nigeria. These limitations highlight a need for better communication and marketing of the crowdsourcing event so individuals fully understand the objective of the crowdsourcing open call and prompt being asked. Third, this study inevitably had selection bias, in which some individuals may have been more likely selected than others, and the group of participants may not have been equally representative of all ages of young people in Nigeria (14 to 24 years). In addition, the crowdsourcing open call was completely web-based, and submissions were submitted through Google Forms and email, which may have excluded individuals who wanted to participate but had limited or no access to the internet. Fourth, analytical bias may have been introduced during the evaluation or scoring process. Finally, demographic information was more frequently missing in email submissions (ie, age and gender), whereas participants who submitted via Google Forms filled in the age and gender question.

Findings from this study will help inform and improve strategies of HIVST to increase HIVST uptake in Nigeria during the COVID-19 pandemic and for years to come in the postpandemic era. This study implies that crowdsourcing and involvement of young people are valuable in identifying current and unconventional HIVST uptake strategies, particularly during an ongoing pandemic, and should also be considered in other aspects of research.

Conclusions

Though the study has many limitations, the crowdsourcing open call engaged a large, diverse number of young people through digital connections. The entrants suggested a diverse range of innovative techniques to increase the uptake of HIVST for vulnerable young people in Nigeria. Their tailored strategies to promote HIVST during the COVID-19 pandemic indicate that HIVST is a feasible method of testing despite barriers relating to physical interactions and in-person testing facilities. Findings from our crowdsourcing open call will inform future research on promoting HIVST and can contribute to the sustainability of HIVST even in times of unprecedented crises.

Acknowledgments

The authors extend their thanks to the I-TEST team at Saint Louis University, the University of North Carolina at Chapel Hill, and Nigerian Institute of Medical Research; 4 Youth By Youth; Youth Ambassadors; and other groups that helped to organize the challenge. The authors would also like to thank the program officers and members of the Prevention and Treatment Through a Comprehensive Care Continuum for HIV-Affected Adolescents in Resource Constrained Settings Consortium.

Data Availability

The data generated and analyzed in this study are not publicly available due to privacy and confidentiality considerations but are available from the corresponding author on reasonable request.

Authors’ Contributions

OE, JT, and JI led the study’s conceptualization, methodology, project administration, and supervision, as well as reviewed and edited the paper. OA and AE conducted data curation, formal analysis, and visualization, as well as drafted the original paper.
UN, CO-U, TG, DO, and AZM served as investigators and administrators, contributing to the paper review and editing. PO, BIB, DN, SM, and TO also reviewed and edited the paper. All authors have read and approved the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Top 3 submissions identified.

[DOCX File, 17 KB - formative_v81e46945_app1.docx]

Multimedia Appendix 2

Emerging themes and examples of submissions.

[DOCX File, 20 KB - formative_v81e46945_app2.docx]

References


Abbreviations

4YBY: 4 Youth By Youth
HIVST: HIV self-testing
WAD: World AIDS Day

Edited by A Mavragani; submitted 03.03.23; peer-reviewed by T Lillie; comments to author 03.10.23; revised version received 22.01.24; accepted 23.01.24; published 30.04.24.

Please cite as:
Expanding Youth-Friendly HIV Self-Testing Services During the COVID-19 Pandemic: Qualitative Analysis of a Crowdsourcing Open Call in Nigeria
JMIR Form Res 2024;8:e46945
URL: https://formative.jmir.org/2024/1/e46945
doi:10.2196/46945
PMID:38687582

©Onyekachuku Anikamadu, Oliver Ezechi, Alexis Engelhart, Ucheoma Nwaozuru, Chisom Obiezu-Umeh, Ponnmile Ogunjemite, Babatunde Ismail Bale, Daniel Nwachukwu, Titilola Gbaja-biamila, David Oladele, Adesola Z Musa, Stacey Mason, Temitope Ojo, Joseph Tucker, Juliet Iwelunmor. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Real-World Gait Detection Using a Wrist-Worn Inertial Sensor: Validation Study

Felix Kluge¹, PhD; Yonatan E Brand², MSc; M Encarna Micó-Amigo³, PhD; Stefano Bertuletti⁴, PhD; Ilaria D’Ascanio⁵, MSc; Eran Gazit⁶, MSc; Tecla Bonci⁷, PhD; Cameron Kirk³, PhD; Arne Küderle⁸, MSc; Luca Palmerini⁹, ¹⁰, PhD; Anisoara Paraschiv-Ionescu¹⁰, PhD; Francesca Salis⁴, PhD; Abolfazl Soltani¹⁰, PhD; Martin Ullrich⁸, PhD; Lisa Alcock³,¹¹, PhD; Kamiar Aminian¹⁰, PhD; Clemens Becker¹²,¹³, MD; Philip Brown¹⁴, MSc; Joren Buekers¹⁵,¹⁶,¹⁷, PhD; Anne-Elie Carsin¹⁵,¹⁶,¹⁷, MSc; Marco Caruso⁴, PhD; Brian Caulfield¹⁸,¹⁹, PhD; Andrea Cereatti¹⁰, PhD; Lorenzo Chian⁵,⁹, PhD; Carlos Echevarria²,¹⁰, PhD; Bjørn Eskofier⁸, PhD; Jordi Evers²¹, MSc; Judith Garcia-Aymerich¹⁵,¹⁶,¹⁷, PhD; Tilo Hache¹, Dipl-Ing (FH), MBA; Clint Hansen²², PhD; Jeffrey M Hausdorff²,²³,²⁴,²⁵,²⁶, PhD; Hugo Hiden¹⁴, PhD; Emily Hume²², PhD; Alison Keogh¹⁸,¹⁹, PhD; Sarah Koch¹⁵,¹⁶,¹⁷, PhD; Walter Maetzler²⁵, MD; Dimitrios Megaritis²⁷, PhD; Martijn Niessen²¹, PhD; Or Perlman²,²³, PhD; Lars Schwickert¹², PhD; Kirsty Scott⁷, PhD; Basil Sharrack²⁸,²⁹, PhD; David Singleton¹⁸,¹⁹, MSc; Beatrix Vereijken³⁰, PhD; Ioannis Vogiatzis²⁷, PhD; Alison Yarnall¹¹,¹³,¹⁴, PhD; Lynn Rochester³,¹¹,¹⁴, PhD; Claudia Mazzà³, PhD; Silvia Del Din³,¹¹, PhD; Arne Mueller¹, PhD

¹Novartis Biomedical Research, Novartis Pharma AG, Basel, Switzerland
²Department of Biomedical Engineering, Tel Aviv University, Tel Aviv, Israel
³Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom
⁴Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
⁵Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy
⁶Center for the Study of Movement, Cognition and Mobility, Neurological Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel
⁷Department of Mechanical Engineering and Insigneo Institute for In Silico Medicine, The University of Sheffield, Sheffield, United Kingdom
⁸Machine Learning and Data Analytics Lab, Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
⁹Health Sciences and Technologies-Interdepartmental Center for Industrial Research (CIRI-SDV), University of Bologna, Bologna, Italy
¹⁰Laboratory of Movement Analysis and Measurement, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
¹¹National Institute for Health and Care Research (NIHR) Newcastle Biomedical Research Centre (BRC), Newcastle University and The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom
¹²Robert Bosch Gesellschaft für Medizinische Forschung, Stuttgart, Germany
¹³Unit Digitale Geriatrie, Universitätsklinikum Heidelberg, Heidelberg, Germany
¹⁴The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom
¹⁵Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain
¹⁶Universitat Pompeu Fabra, Barcelona, Spain
¹⁷CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
¹⁸Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland
¹⁹School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
²⁰Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom
²¹McRoberts BV, The Hague, Netherlands
²²Department of Neurology, University Medical Center Schleswig-Holstein Campus Kiel, Kiel, Germany
²³Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
²⁴Department of Physical Therapy, Faculty of Medical & Health Sciences, Tel Aviv University, Tel Aviv, Israel
²⁵Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL, United States
²⁶Department of Orthopaedic Surgery, Rush Medical College, Chicago, IL, United States
²⁷Department of Sport, Exercise and Rehabilitation, Northumbria University Newcastle, Newcastle upon Tyne, United Kingdom
²⁸Department of Neuroscience, The University of Sheffield, Sheffield, United Kingdom
²⁹Sheffield NIHR Translational Neuroscience BRC, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom
³⁰Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

Corresponding Author:

https://formative.jmir.org/2024/1/e50035
Abstract

Background: Wrist-worn inertial sensors are used in digital health for evaluating mobility in real-world environments. Preceding the estimation of spatiotemporal gait parameters within long-term recordings, gait detection is an important step to identify regions of interest where gait occurs, which requires robust algorithms due to the complexity of arm movements. While algorithms exist for other sensor positions, a comparative validation of algorithms applied to the wrist position on real-world data sets across different disease populations is missing. Furthermore, gait detection performance differences between the wrist and lower back position have not yet been explored but could yield valuable information regarding sensor position choice in clinical studies.

Objective: The aim of this study was to validate gait sequence (GS) detection algorithms developed for the wrist position against reference data acquired in a real-world context. In addition, this study aimed to compare the performance of algorithms applied to the wrist position to those applied to lower back–worn inertial sensors.

Methods: Participants with Parkinson disease, multiple sclerosis, proximal femoral fracture (hip fracture recovery), chronic obstructive pulmonary disease, and congestive heart failure and healthy older adults (N=83) were monitored for 2.5 hours in the real-world using inertial sensors on the wrist, lower back, and feet including pressure insoles and infrared distance sensors as reference. In total, 10 algorithms for wrist-based gait detection were validated against a multisensor reference system and compared to gait detection performance using lower back–worn inertial sensors.

Results: The best-performing GS detection algorithm for the wrist showed a mean (per disease group) sensitivity ranging between 0.55 (SD 0.29) and 0.81 (SD 0.09) and a mean (per disease group) specificity ranging between 0.95 (SD 0.06) and 0.98 (SD 0.02). The mean relative absolute error of estimated walking time ranged between 8.9% (SD 7.1%) and 32.7% (SD 19.2%) per disease group for this algorithm as compared to the reference system. Gait detection performance from the best algorithm applied to the wrist inertial sensors was lower than for the best algorithms applied to the lower back, which yielded mean sensitivity between 0.71 (SD 0.12) and 0.91 (SD 0.04), mean specificity between 0.96 (SD 0.03) and 0.99 (SD 0.01), and a mean relative absolute error of estimated walking time between 6.3% (SD 5.4%) and 23.5% (SD 13%). Performance was lower in disease groups with major gait impairments (eg, patients recovering from hip fracture) and for patients using bilateral walking aids.

Conclusions: Algorithms applied to the wrist position can detect GSs with high performance in real-world environments. Those periods of interest in real-world recordings can facilitate gait parameter extraction and allow the quantification of gait duration distribution in everyday life. Our findings allow taking informed decisions on alternative positions for gait recording in clinical studies and public health.

Trial Registration: ISRCTN Registry 12246987; https://www.isrctn.com/ISRCTN12246987
International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-050785

KEYWORDS
digital mobility outcomes; validation; wearable sensor; walking; digital health; inertial measurement unit; accelerometer; Mobilise-D

Introduction

Digital mobility outcomes (DMOs) such as walking speed show promise for assessing and predicting clinical outcomes in various medical conditions [1-4]. However, the traditional assessment of gait characteristics in clinical environments is often limited by infrequent, short-duration assessments and artificial measurement conditions [5,6]. Thus, the goal of ongoing research is to transfer gait assessment into the real-world to assess a patient’s everyday walking performance, investigate treatment and medication effects, and monitor fluctuating disease symptoms over long and continuous periods [7]. Typically, waist or lower limb–worn inertial sensors including accelerometers and gyroscopes are used to assess gait impairment, and numerous studies present implementation and validation of respective algorithms [8-12]. However, wrist-worn inertial sensors might be more acceptable to participants than lower back sensors and thus better suitable for large-scale studies over prolonged periods and are largely available due to the advent of smartwatches and fitness trackers [13,14].

Traditionally, wrist-worn sensors have been used to detect everyday life activities, estimate step counts, and quantify time spent in different physical activity levels [13,15,16]. Even though actigraphy allows real-world activity to be assessed as
part of mobility, it might not deliver accurate insight into gait impairment as assessed by spatiotemporal gait parameters. The relevance of investigating real-world gait performance in more detail has been highlighted by recent research [5]. Accordingly, there is also a rising interest in the use of wrist-worn sensors for gait assessment in the real world, ranging from gait and stride detection [17-20] to the estimation of spatiotemporal gait parameters [21,22].

The real-world measurement paradigm promises new insights into everyday movement abilities. Large amounts of data may better represent a patient’s everyday behavior and capture rare but important episodes. An important first step toward assessing gait in real-world settings is the identification of continuous gait sequences (GSs). Those sequences can serve as preselected regions of interest containing gait in long, continuous recordings before more computationally complex algorithms for DMO extraction are applied [23]. Furthermore, the focus on GSs reduces the risk of estimating nonmeaningful DMOs in nongait conditions. Finally, extracted GSs and their duration can potentially differentiate between disease-related and healthy walking behavior [24,25].

Accurate gait sequence detection (GSD) using wrist-worn inertial sensors is, however, challenging due to several reasons. First, compared to other sensor locations, the complexity of arm movements is challenging for the extraction of mobility in general and gait parameters in particular [15]. Upper limbs are complex locations to assess DMOs due to the high movement variability and individual preferences of the amount of arm swing. Second, the use of upper limbs for a wide variety of functions other than gait, movement constraints due to walking with the hands in the pockets or holding a bag or other dual-task walking, upper limb injuries, and walking aid use may confound the data. Finally, validation data sets that include both wrist and reference data for the assessment of real-world concurrent validity in multiple disease conditions have not been available so far. Validation studies with reference data have mostly been restricted to healthy adults [22,26-29].

Various approaches for gait detection also from the wrist position have been proposed [17,21,30,31], and the aim of this study was to identify, compare, and rank available state-of-the-art algorithms for GSD based on wrist-worn inertial sensors using labeled real-world data from diverse disease and healthy groups from the Mobilise-D technical validation study [32]. In addition, the wrist-worn sensor results were compared to the outcomes generated from the best-performing algorithms for the lower back inertial sensor to allow conclusions about GSD accuracy between different sensor positions.

The results of this study can help decision makers in clinical studies and possibly in public health to recommend the use of either wrist or lower back–worn inertial sensors. This could allow for more agnostic data collection protocols to be adopted. Patients will benefit as this technology will facilitate the assessment of gait impairment in real-world conditions that may allow quantifying a meaningful aspect of life.

Methods

Ethical Considerations

Ethics approval was obtained at the individual sites (London-Bloomsbury Research Ethics Committee, 19/LO/1507; Helsinki Committee, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, 0551-19TLV; ethical committee of the medical faculty of The University of Tübingen, 647/2019BO2; and ethical committee of the medical faculty of Kiel University, D438/18; University of Sheffield Research Ethics Committee, 029143). All participants provided written informed consent before participating. The analysis is based on pseudonymized data, and anonymized data will be published by the Mobilise-D consortium. Participants in this study were not compensated.

Participants

Overview

For optimizing and evaluating algorithms for GSD, 2 separate data sets were generated to follow the same experimental protocol as the validation data set. This multicentric observational study with the aim of validating real-world DMOs included different patient and healthy populations. The study’s experimental protocol including all inclusion and exclusion criteria have previously been described in more detail in [32].

Optimization Sample

To optimize algorithms for wrist position, including parameter tuning, a separate optimization data set was used. This data set was obtained during a test run within the Mobilise-D project, distinct from the validation study. As a result, it exclusively included healthy participants. Real-world gait data of 11 young and healthy adults were assessed (Sheffield Teaching Hospitals NHS Foundation Trust and University of Sassari, Italy) as part of the Mobilise-D technical validation study. They were asked to follow the same experimental protocol as the validation data set.

Validation Sample

A convenience sample of 108 participants across 5 different disease groups and 1 control group with healthy older adults (HAs) were recruited. The data of those participants served as validation data set for the final evaluation of algorithm performance. The participant groups included patients with chronic obstructive pulmonary disease, Parkinson disease, multiple sclerosis (MS), proximal femoral fracture (PFF; hip fracture recovery), and congestive heart failure (CHF). Recruitment was performed at 5 sites: the Newcastle upon Tyne Hospitals NHS Foundation Trust, United Kingdom; Sheffield Teaching Hospitals NHS Foundation Trust, United Kingdom (London-Bloomsbury Research Ethics Committee, 19/LO/1507); Tel Aviv Sourasky Medical Center, Israel (Helsinki Committee, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, 0551-19TLV); Robert Bosch Foundation for Medical Research, Germany (ethical committee of the medical faculty of the University of Tübingen, 647/2019BO2); and University of Kiel, Germany (ethical committee of the medical faculty of Kiel University, D438/18).
Protocol

Activities of the participants were assessed during 2.5 hours of real-world living undergoing their normal activities (home or work or community or outdoor). They were also asked to perform a limited number of predefined activities (outdoor walking, walking up and down a slope and stairs, and moving from one room to another), if they felt comfortable to do so [33].

The participants were equipped with an inertial sensor worn at the wrist on the nondominant hand (target sensor from which our analysis data are derived) and a validated multisensor system, the INDIP (inertial module with distance sensors and pressure insoles) as reference [32,34]. In particular, the INDIP system included 2-feet inertial sensors attached to the shoelaces with clips (instep position), 2 distance sensors positioned asymmetrically with Velcro over the ankles, and 2 pressure insoles. GSD from the reference INDIP system has previously been described [34]. Furthermore, the INDIP system has been validated across the same patient and healthy adult groups showing excellent results and reliability in the qualification of mobility outcomes in laboratory and free-living environments [34-38]. The decision to place the sensor on the nondominant hand balances participant comfort, practicality, and data quality as it minimized interference with other daily tasks (such as writing, typing, and handling objects) and ensured consistent data collection.

Lower back data were collected by a McRoberts Dynaport MoveMonitor wearable inertial sensor (sampling frequency: 100 Hz, triaxial acceleration range: ±8g or resolution: 1 mg, triaxial gyroscope range: ±2000 dps or resolution: 70 mdps), which was attached to the lower back (L5) with an elastic belt and Velcro fastening. The INDIP system, the wrist inertial sensor (identical to those incorporated in the INDIP), and the lower back MR device were synchronized using their timestamp (±10 ms) and stored in a standardized and integrated data structure [39].

Selection and Optimization of Gait Detection Algorithms

We identified algorithms from the literature potentially suitable for gait detection from lower back and wrist-worn inertial sensors. Our algorithm selection was based on previous work for gait detection from the lower back [11] and the availability of code of the algorithm. Furthermore, algorithms were only considered if they were able to extract gait (sequences) or strides that could be assembled to GSs as previously described [11,40], that is, strides were only combined to a GS if they were not further apart than 3 seconds.

Wherever possible, algorithm parameters were optimized as follows. First, it was deemed necessary to replace any axis-specific dependency by the 3D accelerometer signal norm, if possible. While the lower back provides a rather constant vertical orientation with respect to the global world coordinate system during walking, the axis orientations of wrist sensors change constantly due to free arm movement. Using the norm as orientation-independent signal was the most natural choice without introducing any other sensor-body alignment process. Second, algorithm-specific parameters were optimized on the optimization data set of 11 young and healthy participants (described earlier). A grid search was used to assess algorithm performance for different algorithm parameter combinations on the optimization data set. The best-performing parameter combination was used for further validation of the algorithms on the validation data set (participants from all the 6 different participant groups). Algorithm performance was evaluated as described below. The Paraschiv-Ionescu (2020) algorithm was not optimized as it contains a data-adaptive threshold. The Brand (2020) algorithm was initially developed specifically for analyzing data from wrist-worn inertial sensors. Consequently, the algorithm remained largely unchanged, except for modifying the training data. Thus, only the optimized version, which involved training the model on the optimization data set, was evaluated.

The performance of the wrist-based algorithms was compared to 3 lower back–based algorithms, which have previously been validated on the same data set [11]. This includes the algorithms Iluz (2014) (denoted GSDA previously [11]) and Paraschiv-Ionescu (2019) (denoted GSD_B and GSD_C previously [11]). The specific lower back algorithm parameters for the lower back are described in Table 1.
Table 1. Algorithm descriptions including overview over default and tuned algorithm parameters. GSDA, GSDB, and GSDC are algorithm versions described and validated for the lower back [11].

<table>
<thead>
<tr>
<th>Domain</th>
<th>Algorithm name (reference)</th>
<th>Description</th>
<th>Original sensor position</th>
<th>Algorithm parameters (default)</th>
<th>Algorithm parameters (optimized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine learning  b</td>
<td>Brand (2022) [17]</td>
<td>Use of deep convolutional neural network to discriminate gait and nongait segments based on accelerometer data.</td>
<td>Wrist</td>
<td>N/A</td>
<td>CNN trained on Mobilise-D</td>
</tr>
<tr>
<td>Time domain e</td>
<td>Gu (2017) [41,42]</td>
<td>This method finds peaks in the summed and squared (RMS) acceleration signal. It uses multiple thresholds to determine if each peak belongs to a step or artifact.</td>
<td>Wrist</td>
<td>sim_thres=0.5</td>
<td>verisense_k=2</td>
</tr>
<tr>
<td>Time domain e</td>
<td>Hickey (2017) [43]</td>
<td>Window-based threshold comparison of combined SD of 3D acceleration signal and vertical acceleration.</td>
<td>Lower back</td>
<td>ThresholdStill=0.2</td>
<td>ThresholdStill=0.2</td>
</tr>
<tr>
<td>Time domain e</td>
<td>Hickey (2017) [43]</td>
<td>Window-based threshold comparison of combined SD of 3D acceleration signal and vertical acceleration.</td>
<td>Lower back</td>
<td>ThresholdUpright=-0.5</td>
<td>ThresholdUpright=-0.5</td>
</tr>
<tr>
<td>Template based</td>
<td>Iluz (2014) (GSD A) [44]</td>
<td>Convolution of input signal with a gait cycle template (sine wave). Detection of local maxima in convolution result to define regions of gait.</td>
<td>Lower back</td>
<td>Vertical and anteroposterior acceleration used (lower back)</td>
<td>Vertical and anteroposterior acceleration replaced by acceleration norm</td>
</tr>
<tr>
<td>Template based</td>
<td>Karas (2019) [31]</td>
<td>Template-based method (considering covariance between a scaled and translated pattern function) for stride detection based on adaptive empirical pattern transformation.</td>
<td>Wrist</td>
<td>sim_MIN=0.85</td>
<td>sim_MIN=0.3</td>
</tr>
<tr>
<td>Time domain b</td>
<td>Kheirkhahan (2017) [45]</td>
<td>Based on ActiGraph activity counts using sliding windows and adaptive thresholds.</td>
<td>Lower back</td>
<td>Walking threshold=0.75</td>
<td>Walking threshold=0.6</td>
</tr>
<tr>
<td>Time domain e</td>
<td>Paraschiv-Ionescu (2019) (GSD B and GSD C) [46]</td>
<td>Locomotion period detection based on detected steps from the Euclidean norm of the accelerometer signal. Consecutive steps are associated to gait sequences.</td>
<td>Lower back</td>
<td>GSDB: th=0.1</td>
<td>Wrist: th=0.35</td>
</tr>
<tr>
<td>Time domain e</td>
<td>Paraschiv-Ionescu (2020)</td>
<td>Extension of Paraschiv-Ionescu (2019). It applies an improved preprocessing strategy for the acceleration norm including an iterative succession of smoothing and enhancement stages. Furthermore, a data-adaptive threshold was introduced.</td>
<td>Lower back</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Frequency domain e</td>
<td>Wavelets (Proprietary, Center for the Study of Movement, Cognition, and Mobility, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel)</td>
<td>Time-frequency analysis using wavelets.</td>
<td>Lower back</td>
<td>Vertical and anterio-posterior acceleration used (lower back)</td>
<td>Vertical acceleration replaced by acceleration norm</td>
</tr>
</tbody>
</table>
Algorithm parameters (optimized)

<table>
<thead>
<tr>
<th>Domain (reference)</th>
<th>Algorithm name (reference)</th>
<th>Description</th>
<th>Original sensor position</th>
<th>Algorithm parameters (default)</th>
<th>Algorithm parameters (optimized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine learning</td>
<td>Willetts (2018) [20]</td>
<td>Activity detection using random forests and hidden Markov models to detect various activity modes. Only the output for “walking” activity was considered.</td>
<td>Wrist</td>
<td>Epoch length: 30 seconds</td>
<td>Epoch length: 1 second</td>
</tr>
</tbody>
</table>

aGSD: gait sequence detection.
bProgramming language is Python (Python Software Foundation).
cN/A: not applicable.
dCNN: convolutional neural network.
eProgramming language is R (R Foundation for Statistical Computing).
fRMS: root-mean-square.
gProgramming language is Matlab (MathWorks).

Gait Detection Validation Metrics Evaluation

The output of the gait detection algorithms yielded start and end times for all GSs. Each 2.5-hour recording (containing a varying number of GSs) was segmented into windows of 0.1 seconds as previously described [11]. Based on the comparison of the algorithm output to the reference system, each window was classified as true positive (TP), false positive (FP), true negative (TN), or false negative (FN) regarding the detection of gait [11]. For each 2.5-hour recording, the following metrics were calculated:

Furthermore, errors of the total duration of all GSs and for the number of detected GSs in each 2.5-hour recording were calculated. The relative and absolute errors were determined as a ratio between the (absolute) errors per GS and the corresponding estimates from the reference system, expressed as a percentage. All metrics were calculated for each participant and for the algorithms applied to both wrist sensor versus reference system as well as to lower back sensor versus reference system. We aggregated the error metrics on a disease group level using the mean.

The intraclass correlation coefficient (ICC$_{2,1}$) [48] was calculated for the total GS duration for each 2.5-hour recording on a participant group level (n=6). Values smaller than 0.50, between 0.50 and 0.75, between 0.75 and 0.90, and larger than 0.90 were indicative of poor, moderate, good, and excellent reliability, respectively [49].

A previously described methodology to combine the above metrics into 1 performance index ranging between 0 (worst) and 1 (best) was used [50]. This index is calculated based on a weighted combination of the above-defined metrics (accuracy, sensitivity, specificity, sensitivity, positive predictive value, ICC, mean GS duration relative absolute error, and mean GS number relative absolute error). Each metric can be considered a cost or benefit metric contributing to the performance index with a specific weight (Multimedia Appendix 1). This enables a direct comparison and ranking of the algorithm performances [11]. The performance index was calculated per disease group (n=6).

Statistical Comparison of Algorithm Performance (Wrist vs Lower Back)

For each algorithm, the optimized version (if available) was compared against a representative algorithm for the lower back [Iluz (2014)] with a 2-sided paired t test on a participant level for each performance metric and adjusted P values for multiple testing using Benjamini and Hochberg procedure [51].

Influence of Walking Aids on Algorithm Performance

As the validation data set includes participants with a potential need to use walking aids during the assessment, the effect of walking aids was investigated on algorithm performance. Information was available about (1) whether a walking aid was used and (2) what type of walking aid (among 1-sided canes or crutches, 2 crutches, rollators, and walkers) was used during the 2.5-hour free-living recording.

Results

Population Overview

Of the 108 recruited participants, 25 participants were excluded from subsequent analysis, as either reference, wrist, or lower back sensor data were missing or incomplete (HA: n=3, MS: n=7, Parkinson disease: n=5, PFF: n=8, and CHF: n=2). Wrist and lower back validation were based on the same set of participants. Thus, 83 participants were included in the validation analysis. Overall, 10 participants used walking aids. Participants’ clinical and demographic characteristics per disease group are shown in Table 2.
Table 2. Demographic and clinical characteristics of the participants included in the real-world analysis. The gait sequence (GS) information is based on the GSs detected by the reference system given per 2.5-hour recording. Gait duration is given as sum over all GSs in one 2.5-hour recording.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Validation sample</th>
<th>Optimization sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HA(^a)</td>
<td>CHF(^b)</td>
</tr>
<tr>
<td>Participants, n (%</td>
<td>17 (18.1)</td>
<td>10 (10.6)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>72.35 (6.00)</td>
<td>68.60 (12.21)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>167.00 (10.91)</td>
<td>174.40 (10.27)</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>74.36 (12.53)</td>
<td>83.75 (18.44)</td>
</tr>
<tr>
<td>Walking aid users, n (%)</td>
<td>0 (0)</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Walking aid types</td>
<td>—</td>
<td>One cane or crutch: 2; rollator: 2</td>
</tr>
<tr>
<td>MoCA(^h) (0-30), mean (SD)</td>
<td>28.18 (1.38)</td>
<td>26.70 (3.06)</td>
</tr>
<tr>
<td>Hoehn and Yahr stage, n</td>
<td>N/A(^i)</td>
<td>N/A</td>
</tr>
<tr>
<td>MDS-UPDRS III(^j) (0-132), mean (SD)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>EDSS(^k) (0-6), mean (SD)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>SPPB(^l) (0-12), mean (SD)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CAT(^m) score (0-40), mean (SD)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>FEV(_1)^n (L), mean (SD)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6MWT(^o) distance (m), mean (SD)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Gait duration (minutes), median (IQR)</td>
<td>27.2 (25.5-30.1)</td>
<td>17.8 (9.9-29.7)</td>
</tr>
<tr>
<td>Number GS, median (IQR)</td>
<td>66 (56-88)</td>
<td>55 (21-71.8)</td>
</tr>
</tbody>
</table>

\(^a\)HA: healthy older adult.
\(^b\)CHF: congestive heart failure.
\(^c\)COPD: chronic obstructive pulmonary disease.
\(^d\)MS: multiple sclerosis.
\(^e\)PD: Parkinson disease.
\(^f\)PFF: proximal femoral fracture.
\(^g\)Not available.
\(^h\)MoCA: Montreal Cognitive Assessment.
\(^i\)N/A: not applicable.
\(^j\)MDS-UPDRS III: Movement Disorder Society Unified Parkinson Disease Rating Scale Part III.
\(^k\)EDSS: Expanded Disability Status Scale.
\(^l\)SPPB: short physical performance battery.
\(^m\)CAT: Chronic Obstructive Pulmonary Disease Assessment Test.
\(^n\)FEV\(_1\): forced expiratory volume in 1 second.
\(^o\)6MWT: 6-minute walking test.

Algorithms

Overall, 10 algorithms were included in this validation study (Table 1). They included algorithms originally used for the lower back as well as wrist-specific algorithms. They can be grouped into different domains: (1) time- or frequency-domain–based, (2) stride template–based, and (3) machine learning algorithms. All implemented algorithms have been adapted to use the 3D accelerometer signal only. Algorithm-specific parameters were optimized on the optimization data set (Table 1).
Performance Results

The performance of most optimized algorithms increased compared to using default algorithm parameters (Figure 1). The optimized versions of the Brand (2022) and Paraschiv-Ionescu (2019) algorithms had a performance index above 0.7 for all groups, with the Brand (2022) algorithm showing the highest performance (Figure 1). In the following, the wrist results for the optimized algorithm versions are reported.

**Figure 1.** Performance of assessed algorithms based on a disease group level (n=6). Individual data points are highlighted for each disease group as an overlay. The names of lower back algorithms are given as defined previously [11] and referred to in Table 1. Boxes indicate lower and upper quartiles; the whiskers correspond to 1.5 IQR. Colors indicate the algorithm version: orange indicates the default algorithm version without optimized parameters, and blue indicates the optimized algorithm (parameter tuning based on the optimization data set). In the “wrist” subplot, shapes indicate the disease group to visualize algorithm performance for each group. CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease; GSD: gait sequence detection; HA: healthy older adults; MS: multiple sclerosis; PD: Parkinson disease; PFF: proximal femoral fracture (hip fracture recovery).

Regarding wrist-based GSD, the performance index of the algorithms Willetts (2018), Iluz (2014), and Kheirkhahan (2017) was between 0.74 and 0.81 for most disease groups, except for the PFF group (Multimedia Appendix 2). In the PFF group, the performance index was 0.66 for the Iluz (2014) and Kheirkhahan (2017) algorithms, while it was 0.57 for the Willetts (2018) algorithm.

For the 5 best-performing algorithms Brand (2022), Paraschiv-Ionescu (2019), Iluz (2014), Kheirkhahan (2017), and Willetts (2018), the mean sensitivity ranged between 0.52 (SD 0.28) and 0.81 (SD 0.09) (when excluding the PFF group), whereas the only algorithm showing mean sensitivity (per disease group) consistently higher than 0.70 was Brand (2022). The specificity for those algorithms was between 0.91 and 0.98. ICC values (for GS duration) ranged between 0.72 and 0.99. For PFF, the performance was consistently lower, with sensitivity ranging between 0.29 and 0.55, specificity between 0.94 and 0.96, and ICC values between 0.08 and 0.83 (Figure 2 and Multimedia Appendix 2).
Figure 2. Sensitivity (left) and specificity (right) for the best-performing wrist algorithms (performance index higher than 0.7 for most disease groups except proximal femoral fracture) based on a participant level (N=83). Colors indicate the algorithm version: orange indicates the default algorithm version without optimized parameters, and blue indicates the optimized algorithm with parameter tuning based on the optimization data set. GSD: gait sequence detection.

The mean relative absolute error of the total estimated gait duration during the 2.5-hour recordings was between 8.9% (SD 7.1) (HA) and 32.7% (SD 19.2) (PFF) for the best-performing algorithm, that is, Brand (2022). The Paraschiv-Ionescu (2019) algorithm showed an error between 22% (HA) and 38% (PFF), while the other algorithms performed worse. The mean relative absolute error regarding the number of detected GSs in the 2.5-hour recording ranged between 22.3% (SD 21.1) (HA) and 44.6% (SD 55.3) (PFF) for the Brand (2022) algorithm and worse for the other algorithms (Multimedia Appendix 2). Figure 3 visualizes the relative errors indicating whether the algorithms under- or overestimate the number and duration of detected GSs.

Figure 3. Relative errors of the estimated number of GSs (left) and of the estimated gait duration (right) per 2.5-hour recording based on a participant level (N=83). The dashed red line represents an error of 0 (optimal result). Negative relative errors indicate that fewer GS were detected or the total GS duration was lower than estimated by the reference system. The figure includes the best-performing algorithms (performance index higher than 0.7 for most disease groups except proximal femoral fracture). Colors indicate the algorithm version: orange indicates the default algorithm version without optimized parameters, and blue indicates the optimized algorithm with parameter tuning based on the optimization data set. GS: gait sequence; GSD: gait sequence detection.

For the reported algorithms applied to the lower back position, sensitivity ranged between 0.71 and 0.91, specificity between 0.96 and 0.99, and ICC values between 0.68 and 1.0 (Multimedia Appendix 3). Overall, algorithms applied to wrist signals resulted in lower performance compared to the lower back position as shown in Figure 1 and quantified as follows. Differences in validation metrics of algorithms applied to either wrist compared to the lower back algorithm Iluz (2014) were statistically assessed (Table 3) based on the validation metrics per participant (Multimedia Appendix 4). For sensitivity, all algorithms for the wrist are different (P<.001) from GSD\textsubscript{A}, with the Brand (2022) algorithm having the smallest difference in mean (−0.126), and Willetts (2018) the largest (−0.317) compared to the lower back algorithm Iluz (2014). For specificity, the Brand (2022), Iluz (2014), and Kheirkhahan (2017) algorithms are not significantly different (P>.10) from GSD\textsubscript{A}, with the Brand (2022) algorithm having the smallest difference (−0.00192). The Brand (2022) algorithm is closest to GSD\textsubscript{A} for relative error in number of detected GSs (P=.021 and a difference in mean of 0.042). No statistical comparison was conducted for the performance index itself, as the index was calculated only per disease group, resulting in a small sample of 6 data points.
Table 3. Statistical results comparing each wrist algorithm and metric (optimized versions) with a representative lower back algorithm with good performance (Iluz (2014) applied to the lower back).

<table>
<thead>
<tr>
<th>Algorithm and metric</th>
<th>Mean difference</th>
<th>P value</th>
<th>Adjusted P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Willetts (2018)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>−0.317</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>−0.029</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ppV&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.258</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Accuracy</td>
<td>−0.064</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative number GS&lt;sup&gt;b&lt;/sup&gt; error</td>
<td>−0.325</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative GS duration error</td>
<td>−0.055</td>
<td>.37</td>
<td>.38</td>
</tr>
<tr>
<td><strong>Brand (2022)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>−0.126</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>−0.002</td>
<td>.49</td>
<td>.49</td>
</tr>
<tr>
<td>PPV</td>
<td>−0.037</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Accuracy</td>
<td>−0.014</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative number GS error</td>
<td>0.042</td>
<td>.19</td>
<td>.21</td>
</tr>
<tr>
<td>Relative GS duration error</td>
<td>−0.131</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Paraschiv-Ionescu (2019)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>−0.277</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>−0.011</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>PPV</td>
<td>−0.115</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Accuracy</td>
<td>−0.041</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative number GS error</td>
<td>0.335</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative GS duration error</td>
<td>−0.216</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Iluz (2014)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>−0.273</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>−0.004</td>
<td>.27</td>
<td>.29</td>
</tr>
<tr>
<td>PPV</td>
<td>−0.067</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Accuracy</td>
<td>−0.035</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative number GS error</td>
<td>−0.500</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative GS duration error</td>
<td>−0.268</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Kheirkhahan (2017)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>−0.299</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specificity</td>
<td>−0.005</td>
<td>.23</td>
<td>.25</td>
</tr>
<tr>
<td>PPV</td>
<td>−0.063</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Accuracy</td>
<td>−0.038</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative number GS error</td>
<td>−0.671</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relative GS duration error</td>
<td>−0.310</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>PPV: positive predictive value.
<sup>b</sup>GS: gait sequence.

**Effect of Walking Aids**

The frequency of walking aid use depended on the disease group (Table 2). Walking aid use influenced the accuracy of wrist-based gait detection (Figure 4). Participants using bilateral walking aids (rollators, walkers, and 2 crutches) exhibited lower sensitivity for gait detection. The gait of 5 participants using unilateral walking aids (CHF: n=2, MS: n=1, and PFF: n=2) was not as affected and could mostly be estimated as accurately as for participants without walking aids. For the unilaterally
used walking aids (1 cane or crutch), 2 of 5 participants wore the sensor on the same side as they used the walking aid. An exception was the PFF group, in which low sensitivity has also been observed for unilateral walking aid use. In total, 3 patients using no walking aid showed a sensitivity below 0.5, and 2 of those participants reported to usually use walking aids (but not in this study), while the third participant showed a low short physical performance battery score of 4, indicating that walking might be strongly impaired.

Figure 4. Effect of walking aids on sensitivity for the algorithm Brand (2022) [17]. Each data point represents 1 participant (N=83), the color indicates the type of walking aid. CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease; HA: healthy older adults; MS: multiple sclerosis; PD: Parkinson disease; PFF: proximal femoral fracture (hip fracture recovery).

Discussion
Principal Findings
This is the most comprehensive study so far, evaluating real-world gait detection performance of various algorithms from a wrist-worn sensor in a heterogeneous population including 5 disease cohorts.

Performance Results
The Brand (2022) and Paraschiv-Ionescu (2019) algorithms exhibited good performance (>0.75) across all disease groups excluding PFF (moderate performance index of 0.71), while the first outperformed the latter algorithm especially for the total estimated walking time and the number of detected GSs. The Iluz (2014) and Kheirkhahan (2017) algorithms also showed high performance, except for the PFF group (performance index of 0.66). A lower performance was generally observed for the PFF group. This has already been previously reported for the lower back position [11] and can be attributed to several factors, which significantly impacts the accuracy of gait detection algorithms. First, patients with PFF may show altered gait patterns due to pain, muscle weakness, and impaired mobility. Second, they may exhibit asymmetrical walking, making it harder for algorithms to identify consistent patterns. Finally, the gait of hip fracture recovery patients may vary more widely even within the same group, which is also reflected, for example, in the range of number of GSs, which was highest for the PFF group in this study (Table 2).

Based on those results, we suggest the use of the Brand (2022) algorithm, which is suitable for gait detection based on wrist-worn sensors across all investigated disease groups.

Sensitivity and specificity were calculated with regard to the agreement of gait detection algorithm results to the reference GSs based on 0.1-second windows for complete 2.5-hour recordings. Sensitivity was generally lower than specificity, indicating that not all GSs were detected by all algorithms. On the other hand, high specificity indicates that only few nongait activities are misclassified as GSs. Further algorithm optimization on a larger data set is required to find the optimal balance between sensitivity and specificity. If the goal is to subsequently characterize DMOs, a high specificity is needed to exclude nonwalking periods (including transitions and shuffling of gait), but at the same time accept a portion of missed walking periods.

Comparison to Lower Back
Due to the high movement variability of the arm during walking, a performance drop is expected when comparing algorithms applied to a wrist-worn versus a lower back–worn inertial sensor. This performance drop is evident in this study. Sensitivity is lower for the wrist position (sensitivity was between 0.32 and 0.13 smaller compared to (Iluz 2014) applied on lower back data, Table 3), which can most likely be attributed to nonperiodical arm swing with differences in amplitude during walking. However, specificity is comparably high (>0.7) for both sensor positions (Figure 2), indicating the general reliability of correctly rejecting nongait activities.

Algorithm Parameter Optimization
The design of some of the algorithms focused on the lower back position initially (Table 1). However, in this study, we focused on implementing methods initially developed for lower back acceleration signals based on time or frequency methods to wrist acceleration signals. Where possible, algorithm parameters were optimized on the optimization data set. Default and
optimized algorithm parameters differed, and optimization allowed for achieving higher performances of lower back algorithms at the wrist position.

A strong advantage of all investigated algorithms is that they use the 3D accelerometer signal only and do not depend on gyroscopic sensors, sensors that have high energy consumption. Thus, they can potentially be used more ubiquitously for other wearable inertial sensors that acquire accelerometer data only, allowing for energy-efficient gait detection systems and, thus, longer assessment periods. In addition, in future work, we will evaluate the effect of lower accelerometer sampling rates (eg, 30 Hz instead of 100 Hz) on GSD performance. The use of inexpensive, low-sampling consumer-grade watches in public health projects may justify reduced performance as observed in this study.

Walking Aids
Walking aid users move differently due to several factors. Gait impairment can be observed in various diseases including the groups assessed within this study; in addition, the use of walking aids may lead to compensatory gait changes and can influence gait parameters directly [52,53].

Furthermore, biomechanical constraints when using walking aids affect wrist-based gait assessment. Bilateral walking aids such as walkers or rollators may significantly affect the arm movement and thus the acquired accelerometer signals. On the one hand, this can be used to construct specific algorithms for gait assessment when walking aids are used [54]. This, in turn, may lead to deteriorated performance of algorithms that are not fit for the purpose of walking aid–based gait assessment.

The results of this study demonstrate that special care must be taken when defining inclusion and exclusion criteria in studies based on a wrist-worn sensor for gait assessment. Participants using rollators, walkers, or 2 crutches may be separately considered in wrist-based gait assessment. However, the actual use of walking aids in real-world environments can hardly be predicted. Unilateral walking aids can potentially be used on either side and switched during the assessment. Participants may also not use walking aids continuously but only when they feel unsafe (depending on the environment) and may also use other everyday objects (eg, furniture) for increased security, which may affect the interpretation of sedentary or activity levels.

Strengths and Limitations
The focus of this study was to investigate the performance of gait detection algorithms on real-world data, in which full reference information from the sensor-based INDIP system including pressurized insoles was available. We see the use of this multimodal reference system as a unique advantage compared to data sets used in previous studies, as it allows not only to assess gait detection very accurately but also to extract other spatiotemporal gait parameters. The accuracy of this system has previously been assessed against an optical motion capture system and has shown excellent absolute agreement (ICC>0.95) within a laboratory setting [34]. We thus considered the INDIP system as a reliable method for acquiring reference data in real-world environments. One can argue that the 2.5-hour assessment used for validation might not fully represent the full variability of real-world walking. Nevertheless, our data set is one of the largest available ones containing full reference information for a variety of disease indications. Future work could use longer validation periods. Overall, a diverse set of disease areas were represented including orthopedic, pulmonary, cardiovascular, and neurological diseases. Future studies could extend this work to other disease groups.

It is worth noting that the optimization set used for this study was relatively small and comprised a population that differed from the validation set in terms of age and health condition. It was based on a healthy young adult group that did not rely on the use of walking aids, which might bias the optimal parameter choice. Algorithm performance could likely be improved using disease-specific samples including walking aid users or tuning even based on individual participants. Future studies should, thus, focus on optimizing gait detection algorithms specifically tailored for participants with gait disturbances related to the disease groups of interest. The methodology of this paper can serve as a reference for achieving this.

Real-world data are naturally imbalanced, with a significantly larger number of nongait segments (majority class) compared to gait segments (minority class). This inherent imbalance can introduce bias in supervised models, resulting in low sensitivity. Consequently, further analyses should focus on addressing this problem by using techniques such as upsampling from the minority class or generating artificial samples.

We acknowledge that the list of included algorithms might not be exhaustive. Our choice was driven by practical considerations including code availability and applicability on wrist-worn accelerometer data. However, the algorithms cover a broad spectrum of different domains using time and frequency domain, template matching, and machine learning methods. Future work may compare further, also proprietary, algorithms to the presented results. In addition, the Mobilise-D technical validation data set as well as the used validation methodology might provide a blueprint for future validation studies.

This paper only validated the first step of a complete gait analysis pipeline. Future work will need to show whether subsequently extracted DMOs such as cadence, stride length, and walking speed can reliably be estimated for those identified GSs in comparison to the lower back.

Conclusions
To conclude, we identified algorithms that can extract GSs based on a wrist-worn sensor using accelerometer data. In general, the performance for detecting GSs as regions of interest of further gait parameter extraction and quantification of gait duration is lower than for the lower back position. However, the omnipresence of wrist-worn sensors and their easier operationalization and better ergonomics in longitudinal clinical trials may justify some level of lower gait quantification performance for the sake of higher acceptance and more data. Identifying GSs in continuous long-term inertial sensor recordings is the first step that will allow extracting additional DMOs (eg, spatiotemporal parameters such as walking speed in disease cohorts [21,22]).
Our work is a step toward quantifying the limitations of wrist-worn devices for digital mobility analysis and contributes to the evidence needed by researchers, clinical trial teams, and health care professionals in deciding if a lower back inertial sensor is required or a wrist-worn sensor is sufficient. The data presented here should be considered as one part of further opportunities offered by wrist-worn inertial sensors. To assess a comprehensive movement picture of patients, different algorithms can, for example, measure further DMOs related to mobility analysis, including spatiotemporal parameters and physical activity. Quantification of continuous GSs may be a DMO on its own that can be explored in diseases with reduced physical performance.

Acknowledgments
The authors acknowledge all the members of the Mobilise-D Work Package 2 for continuous discussion and critical input. They are particularly grateful to the participants in the study for their time and enthusiastic contribution, especially during the COVID-19 pandemic. This work was supported by the Mobilise-D project that has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (IMI2 JU: grant 820820). The IMI2 JU receives support from the European Union’s Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations. SDD, LR, and AY were also supported by the IMI2 JU project Identifying Digital Endpoints to Assess Fatigue, Sleep and Activities in Daily Living (grant 853981). LA, LR, AY, and SDD were also supported by the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre (BRC) based at The Newcastle upon Tyne Hospital NHS Foundation Trust, Newcastle University and the Cumbria, Northumberland and Tyne and Wear NHS Foundation Trust. JMH and YEB are supported in part by the National Institutes of Health (grant R01AG79133). LA, LR, AY, and SDD were also supported by the NIHR/Wellcome Trust Clinical Research Facility infrastructure at The Newcastle upon Tyne Hospitals NHS Foundation Trust. This study was also supported by the NIHR through the Sheffield BRC (grant IS-BRC-1215-20017). ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation through the “Centro de Excelencia Severo Ochoa 2019-2023” program (CEX2018-000806-S) and from the Generalitat de Catalunya through the Centres de Recerca de Catalunya program. All opinions are those of the authors and not the funders. The content in this publication reflects the authors’ view, and neither Innovative Medicines Initiative nor the European Union, European Federation of Pharmaceutical Industries and Associations, National Health Service, NIHR, Department of Health and Social Care, or any associated partners are responsible for any use that may be made of the information contained herein.

Data Availability
Example data sets generated and analyzed during this study are available in the Zenodo repository [55]. The full data set will be made available by the Mobilise-D consortium by June 2024.

Authors' Contributions
SDD, CM, LR, AC, AM, and BV contributed to project development and study design. Data collection and preprocessing of the data was performed by TB, FS, KS, LA, PB, SB, FS, MC, KS, EG, CH, LP, ID, and LS. BS, WM, CB, JMH, IV, EH, DM, AY, and LR were involved in participant recruitment and clinical oversight. AK, AP-I, AS, MU, EG, YEB, and FK developed and implemented the algorithms and analysis pipeline. Data management platform and analysis were provided by HH, DS, AK, CK, and MU. Data and statistical analyses were done by FK, MEM-A, and AK. FK prepared figures and tables. FK, YEB, MEM-A, SDD, and AM interpreted the data and results. FK, YEB, and AM drafted the paper. All authors have provided critical intellectual input during the study and revision of the paper. All authors have reviewed the paper and approved the submitted version.

Conflicts of Interest
AM and FK are employees of and may hold stock in Novartis. BE reports consulting activities with adidas AG, Siemens AG, Siemens Healthineers AG, and WSAudiology GmbH outside of the study. BE is a shareholder in Portabiles HealthCare Technologies GmbH. In addition, BE holds a patent related to gait assessment. LP and LC are cofounders and own shares of mHealth Technologies. LS and CB are consultants of Philipps Healthcare, Bosch Healthcare, Eli Lilly, and Gait-up. MN is an employee of McRoberts. SD and JMH report consultancy activity with Hoffmann-La Roche Ltd outside of this study.

Multimedia Appendix 1
Weights used for calculating the performance index.
[XLSX File](Microsoft Excel File), 11 KB - formative_v8i1e50035_app1.xlsx]

Multimedia Appendix 2
Gait sequence detection performance metrics and overall performance index for the wrist sensor position (optimized versions, if available). Values are provided as mean and [5%, 95%] quantiles or as mean and limit of agreement (LoA).
[XLSX File](Microsoft Excel File), 22 KB - formative_v8i1e50035_app2.xlsx]
Multimedia Appendix 3
Gait sequence detection performance metrics and overall performance index for the lower back sensor position. Values are provided as mean and [5%, 95%] quantiles or as mean and limit of agreement.
[XLSX File (Microsoft Excel File), 14 KB - formative_v81le50035_app3.xlsx ]

Multimedia Appendix 4
Gait sequence detection metrics for each 2.5 h recording per sensor position, algorithm, and participant.
[XLSX File (Microsoft Excel File), 351 KB - formative_v81le50035_app4.xlsx ]

References


Abbreviations

CHF: congestive heart failure
DMO: digital mobility outcome
FN: false negative
FP: false positive
GS: gait sequence
GSD: gait sequence detection
HA: healthy older adult
ICC: intraclass correlation coefficient
INDIP: inertial module with distance sensors and pressure insoles
**MS:** multiple sclerosis

**PFF:** proximal femoral fracture

**TN:** true negative

**TP:** true positive
Regulatory Issues in Electronic Health Records for Adolescent HIV Research: Strategies and Lessons Learned

Sara Shaw Green¹, MSW‡; Sung-Jae Lee², PhD; Samantha Chahin¹, MPH; Meardith Pooler-Burgess¹, DrPH; Monique Green-Jones³, MPH; Sitaji Gurung⁴, MD, MPH, PhD; Angulique Y Outlaw⁵, PhD; Sylvie Naar¹, PhD

¹Center for Translational Behavioral Science, Department of Behavioral Sciences and Social Medicine, Florida State University, Tallahassee, FL, United States
²Division of Population Behavioral Health, Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, United States
³College of Liberal Arts and Sciences, Wayne State University, Detroit, MI, United States
⁴Department of Health Sciences, New York City College of Technology, New York, NY, United States
⁵Division of Behavioral Health, Department of Family Medicine and Public Health Sciences, College of Medicine, Wayne State University, Detroit, MI, United States

‡ATN 162b Protocol Team

Corresponding Author:
Sara Shaw Green, MSW
Center for Translational Behavioral Science
Department of Behavioral Sciences and Social Medicine
Florida State University
2010 Levy Ave
Building B, Suite B0266
Tallahassee, FL, 32310
United States
Phone: 1 8506442334
Email: sara.green@med.fsu.edu

Abstract

Background: Electronic health records (EHRs) are a cost-effective approach to provide the necessary foundations for clinical trial research. The ability to use EHRs in real-world clinical settings allows for pragmatic approaches to intervention studies with the emerging adult HIV population within these settings; however, the regulatory components related to the use of EHR data in multisite clinical trials poses unique challenges that researchers may find themselves unprepared to address, which may result in delays in study implementation and adversely impact study timelines, and risk noncompliance with established guidance.

Objective: As part of the larger Adolescent Trials Network (ATN) for HIV/AIDS Interventions Protocol 162b (ATN 162b) study that evaluated clinical-level outcomes of an intervention including HIV treatment and pre-exposure prophylaxis services to improve retention within the emerging adult HIV population, the objective of this study is to highlight the regulatory process and challenges in the implementation of a multisite pragmatic trial using EHRs to assist future researchers conducting similar studies in navigating the often time-consuming regulatory process and ensure compliance with adherence to study timelines and compliance with institutional and sponsor guidelines.

Methods: Eight sites were engaged in research activities, with 4 sites selected from participant recruitment venues as part of the ATN, who participated in the intervention and data extraction activities, and an additional 4 sites were engaged in data management and analysis. The ATN 162b protocol team worked with site personnel to establish the necessary regulatory infrastructure to collect EHR data to evaluate retention in care and viral suppression, as well as para-data on the intervention component to assess the feasibility and acceptability of the mobile health intervention. Methods to develop this infrastructure included site-specific training activities and the development of both institutional reliance and data use agreements.

Results: Due to variations in site-specific activities, and the associated regulatory implications, the study team used a phased approach with the data extraction sites as phase 1 and intervention sites as phase 2. This phased approach was intended to address the unique regulatory needs of all participating sites to ensure that all sites were properly onboarded and all regulatory components were in place. Across all sites, the regulatory process spanned 6 months for the 4 data extraction and intervention sites, and up to 10 months for the data management and analysis sites.
Conclusions: The process for engaging in multisite clinical trial studies using EHR data is a multistep, collaborative effort that requires proper advanced planning from the proposal stage to adequately implement the necessary training and infrastructure. Planning, training, and understanding the various regulatory aspects, including the necessity of data use agreements, reliance agreements, external institutional review board review, and engagement with clinical sites, are foremost considerations to ensure successful implementation and adherence to pragmatic trial timelines and outcomes.

(JMIR Form Res 2024;8:e46420) doi:10.2196/46420

KEYWORDS
electronic health record; HIV; pragmatic trial; regulatory; EHR; pre-exposure prophylaxis; retention; attrition; dropout; legal; regulation; adherence; ethic; review board; implementation; data use; privacy

Introduction

Electronic health records (EHRs) can serve as a useful tool to inform various outcomes related to HIV prevention and care among minoritized youth. For example, Dark et al [1] assessed the feasibility of using EHRs to examine and monitor the HIV care continuum and costs associated with implementation science studies for youth living with HIV. In addition, Butame et al [2] highlighted barriers and facilitators to the collection of aggregate HIV-related EHR data. In the field of HIV prevention, Dunn et al [3] used EHRs to capture primary HIV prevention outcomes. In the field of sexually transmitted infections, Jackman et al [4] leveraged EHRs to reduce disparities in the rates of sexually transmitted infections among university students.

EHRs can provide a foundation for innovative, large-scale simple trials at a lower cost than traditional randomized controlled trials and ultimately achieve a learning health care system [5]. While an explanatory (or efficacy) trial seeks to determine if an intervention is effective under ideal conditions to maximize its favorable effects, a pragmatic trial aims to demonstrate that an intervention can be effective in real-world settings. Pragmatic trials often enroll patients most in need, test treatments within settings and with staff who are typical to most clinical situations, and address issues important to clinicians, policy makers, and patients [6]. Furthermore, embedding pragmatic trials within clinical care settings can facilitate the investigation of treatment effects in a more cost-efficient manner than trials conducted without the clinical infrastructure provided by such a setting [7]. These types of trials are perfectly suited for clinic-level interventions focusing on retention in pre-exposure prophylaxis (PrEP) and HIV treatment services.

The Adolescent Trials Network (ATN) for HIV/AIDS Interventions Protocol 162b (ATN 162b) sought to pilot a pragmatic approach to evaluate mobile health (mHealth) retention in a care intervention for PrEP and antiretroviral therapy (ART) using EHRs to determine improvements in on-site appointments arranged across patients without having to obtain consent and collect data from individual patients. Additional EHR variables were collected as a pilot for continua monitoring in future trials as evidence-based guidelines include recommendations for systematic monitoring of HIV care continuum variables such as linkage and retention in care and ART adherence [8,9] and monitoring of the PrEP continuum [10]. These studies also served as a pilot for routine EHR downloads to monitor HIV prevention and treatment continua at ATN sites.

Despite the simplification of the consent process within pragmatic trials [11], there are significant regulatory requirements and associated challenges that may emerge in such trials, including the establishment of reliance and data use agreements [12] ensuring deidentification of individual intervention and outcome data [13], responsiveness to changing regulatory requirements as researchers and institutions respond to everchanging technology and related bioethical concerns [14], and compliance with sponsor data sharing practices [15]. Experts have identified multiple key ethical and regulatory challenges unique to pragmatic trials, including the necessity for harmonization among collaborating institutional review boards (IRBs) [16], and the need to streamline the regulatory oversight process [14,16]. Based on these identified challenges, efforts have been made in recent years to enhance the collaborative nature of multisite clinical trial research such as the policy issued regarding the requirement of a single IRB for National Institutes of Health–funded projects [17]; however, these policy changes are still in their infancy and contain inconsistencies, resulting in confusion regarding their implementation and delays that may impact study timelines. The objective of this study is to describe the unique regulatory processes, challenges, strategies, and lessons learned in the implementation of a multisite pragmatic trial with EHR outcomes in the ATN 162b to contribute useful recommendations for future researchers conducting such trials.

Methods

Overview

The ATN 162b used a quasi-experimental longitudinal design to collect annual EHR via data downloads, supplement with auxiliary data, and curate and harmonize the data using a common data model to support the measurement of retention in care and viral suppression for pilot-testing an mHealth intervention. Para-data, such as usage metrics from components of the intervention app, were also collected to assess the feasibility and acceptability of the intervention [18]. A PrEP site and an HIV treatment site participated in the interventions and another PrEP site and HIV treatment site served as controls. All youth aged 18-24 years enrolled at the 2 sites were eligible for participation.

The process for EHR data extraction required significant attention to the types of data to be extracted and transferred,
establishment of the necessary regulatory agreements, a detailed protocol to be developed, and participating sites to be trained. A limited data set (deidentified except for dates of service) was extracted from 4 ATN participant recruitment venues longitudinally, representing 2 full years (2018-2019). Demographics, behavioral measures, clinical data, prevention services data, and linkage- and retention-in-care data were extracted from EHRs and electronic appointment systems supplemented with remote data entry with all identifiers removed except for dates of service. Data were managed and analyzed in a secure cloud database. Clinic-level data were used as pragmatic outcomes for a pilot study of mHealth apps prevention and treatment, and not to analyze how the intervention impacted health or behavioral outcomes in human participants; hence, the study did not need to be registered as a clinical trial [19].

Sites were trained in a standard protocol to electronically download data, which reduces the risk of data entry error, particularly for complex variables (eg, ICD-10 [International Statistical Classification of Diseases, Tenth Revision] codes). Exceptions for manually extracted variables (eg, laboratory findings from other sites or other paper records) were made so that sites could enter data directly into a supplemental database using an appropriate electronic data collection system (eg, Conserved Domain Architecture Retrieval Tool) [20]. Participant recruitment venues were provided training and an extraction procedural package that included a study summary, data explanation form, protocol adherence checklist, launch email, database user account information, and the data files request link for each site.

Data were directly transferred by a trained site coordinator annually (2018 and 2019) to the coordinating center in accordance with established file transfer protocols and stored using HIPAA (Health Insurance Portability and Accountability Act)-compliant cloud computing services. Once securely stored, data were curated and harmonized by the study’s data management team. To prevent protocol violations and the inadvertent upload of patient data that were not covered as part of the protocol, each site submitted a protocol adherence checklist to the protocol team, who approved it before any data uploads occurred. This mandatory checklist served to ensure multiple reviews of the final data set before transmitting data. An existing evidenced-based youth HIV adherence intervention, Motivational Enhancement System for Adherence (MESA) [21], was refined for adherence to PrEP appointments and HIV treatment appointments. MESA is built on a novel, low-cost intervention platform, Computerized Intervention Authoring Software, which allows investigator-initiated intervention development without programming, thus dramatically reducing the time and cost of reducing ideas to a usable, editable, and shareable application. MESA uses a digital counselor, an animated avatar of the youth’s choosing, to deliver motivational interviewing and goal setting for a specific target behavior. Intervention content was refined on the basis of existing ATN studies and the ATN’s youth advisory groups. The interventions were delivered via a tablet during the clinic visit, at baseline, and at the time of the next scheduled usual care appointment. To ensure anonymity for this partially deidentified data-driven pragmatic feasibility trial, youth will use a “burner” app that provides a second phone number to receive text messages. The site coordinator approached all eligible youth, completed onboarding, assisted participants in downloading the “burner” app with notifications turned on, and assisted them with accessing the web-based intervention on the provided tablet. Youth also received disposable earbuds and then independently completed a single 15-20–minute MESA session while at the clinic. At the end of the session, the site coordinator entered the youths’ next clinic appointment into the intervention application and provided them US $10 in the form of cash or a gift card. Participants then received monthly automated motivational reinforcement and visit text message reminders for their next appointment. At the next appointment (usually quarterly at adolescent sites), the site coordinators guided youth to complete a brief 5-minute follow-up MESA session and receive a second, final US $10 payment.

Ethical Considerations

This study was reviewed and approved by the Florida State University (FSU) IRB (STUDY00000619), the single IRB of record (sIRB). Informed consent was obtained from those individuals who participated in the intervention activities, as informed consent was not required for acquiring the limited data set in line with the HIPAA privacy rule for a covered entity [22]. Acquisition of the limited data set was secured by executing data use agreements, and data were shared as password-protected files on a secure cloud server. As stated above, those participating in the intervention activities were eligible to receive up to US $20 either as cash or gift cards.

Measures

Prevention Variables

Descriptive variables obtained from EHRs with regard to HIV prevention among adolescents in the ATN 162b were the visit date; current age; date of birth; gender; sexual orientation; race; ethnicity; 3 levels of PrEP; indicators of PrEP eligibility, acceptance, and status; behavioral risk factors; smoking status; substance abuse; behavioral health status; mental health; sexual risk; results of tests for sexually transmitted infections; other medical statuses; height and weight; BMI; and results of pregnancy testing.

Descriptive variables obtained from EHRs with regard to HIV care services for adolescents in the ATN 162b include HIV testing; risk assessment; HIV care referral; PrEP referral; a Centers for Disease Control and Prevention–mandated checklist for PrEP providers (such as providing basic education about PrEP; obtaining past medical history including kidney and liver [ie, hepatitis] disease, bone disease, and fractures, as well as assessment of pregnancy desires for women; assessing for possible symptoms of acute HIV infection; performing laboratory tests including HIV tests to rule out acute HIV infection, tests for sexually transmitted diseases, tests for serum creatinine to determine creatinine clearance, hepatitis B surface antigen and antibody, and hepatitis C virus antibody, body weight for determining creatinine clearance, and pregnancy tests; providing prescriptions for Truvada [30 tablets]; providing PrEP education or counseling; assessing substance abuse and...
mental health needs and make referrals as appropriate; reviewing the importance of regular clinic follow-up and determining the best method of communication for reminders [eg, call, email, or SMS text message]; scheduling follow-up visits for 1 month and providing an appointment card; and starting hepatitis B vaccine series and administering meningococcal and human papillomavirus vaccines as appropriate); follow-up visits (30 days after the initial visit and quarterly thereafter; including PrEP adherence monitoring during provider visits [with dried blood spots at provider visits]); behavioral health referral; behavioral health visits; outreach, advocacy, and case management visits; testing for sexually transmitted infections; and pregnancy testing.

**Care Variables**

Descriptive variables obtained from EHRs with regard to HIV care among adolescents in the ATN 162b include the visit date; current age; date of death; gender; sexual orientation; race; ethnicity; diagnostic characteristics; age at HIV diagnosis; date of HIV diagnosis; mode of transmission; 3 levels of ART; indicators of indicators of ART eligibility, acceptance, and status; HIV treatment; date of entry into care; names of antiretroviral medications; viral load results; CD4 T lymphocyte count; names of other non-HIV medications; behavioral risk factors; smoking status; substance abuse; behavioral health status; mental health; sexual risk; results of tests for sexually transmitted infections; other medical statuses; ICD-10 diagnosis codes or other indicators of medical mental health or substance abuse; height and weight; BMI; blood pressure; cholesterol panel; blood glucose; and pregnancy test results.

Variables obtained from EHRs with regard to HIV care services for adolescents in the ATN 162b include CPT-4 (Current Procedural Terminology), HCPCS (Healthcare Common procedure Coding System), or other indicators of medical and behavioral provisions; risk assessment; ART referral; medical visits attended (for retention in care); medical visits missed (for visit consistency); behavioral health referrals (eg, mental health and substance abuse treatment); behavioral health visits; outreach, advocacy, and case management visits; testing for sexually transmitted infections (Neisseria gonorrhoeae and Chlamydia trachomatis in urine, the rectum, and pharynx, and rapid plasma reagin); and pregnancy testing.

**Results**

Our results focus on the unique regulatory processes implemented by the ATN 162b team, which are related to the (1) development and coordination of the necessary regulatory infrastructure across multiple academic and clinic sites, (2) preparing for and initiating site-specific training of study personnel, (3) ongoing monitoring of regulatory components, and (4) ensuring compliance with institutional and sponsor requirements.

Regulatory aspects of developing a multisite study focused on the extraction and analysis of limited data sets require significant attention to regulatory compliance. Regarding the split approach of protocol development, consent for the limited data set was not required from the study participants, as the protocol included limited data sets only and was considered to be of minimal risk; however, HIPAA authorization was required for extracting these data and encompassed all 4 participating sites, in line with the US Department of Health and Human Services guidelines related to the use of Protected Health Information for the purpose of research [9]. For sites participating in the study intervention, informed consent was obtained through a standard process with site-specific documents. Based on these distinctions, to accomplish the regulatory components, study personnel at FSU, as the sIRB, were tasked with the development of reliance agreements and data use agreements that differed depending on the site’s data versus intervention activities. The timing of the development and execution of these agreements posed challenges to the initiation of the study protocol, as clinical sites had varying requirements related to their internal regulatory processes and authorized signatory needs. The process for regulatory development began in the fall of 2019, with the submission of the study to the overseeing sIRB at FSU. As indicated in Multimedia Appendices 1 and 2, three of the 4 sites that participated in the pragmatic trial and in collecting EHRs and those that only collected EHRs were initiated (light gray) in June 2020, and there was considerable variance in the full execution (dark gray). Data use agreements were initiated and executed with data analysis sites as data were not intended to be transferred from the data management site until the completion of the study phases.

Once the protocol was submitted and approved as a multisite study, the project team worked to finalize the participating sites’ locations and worked with the sIRB to determine the appropriate regulatory components for the study (eg, reliance and data use agreements). Ultimately, the sIRB determined that all sites were considered to have been engaged in human subjects research, regardless of whether they were intervention or nonintervention and data extraction sites. Hence, the study team chose to separate the protocol into 2 parts; the first part focused on intervention-only sites, and the second part addressing data extraction separately. This split was intended to allow for the development of informational and instructional materials tailored to each type of clinical site to ensure adequate training of site personnel and to avoid any overlap or confusion among the pragmatic trial intervention and nonintervention sites. Furthermore, this distinction allowed the study team to ensure that proper monitoring of protocol-specific activities could be conducted, and necessary modifications or reports of information to the IRB were specific to the type of activities, as distinguished between the intervention and nonintervention sites. In addition to intervention and nonintervention sites, reliance and data use agreements were executed with external collaborators on the data management team to ensure compliant access to the limited data set. Ongoing monitoring of all relevant protocols and site activities was the responsibility of the sIRB site (FSU), including the submission of modifications, as well as close collaboration with the ATN Coordinating Center, to ensure that data storage and management aligned with the provisions of the data use agreements and complied with the sponsors’ requirements related to data sharing.
Discussion

Principal Findings

This paper describes the regulatory processes, challenges, and strategies to navigate them, and lessons learned in the ATN 162b multisite embedded pragmatic trial within HIV testing and prevention clinical sites to evaluate clinical-level outcomes related to HIV treatment or PrEP services.

The regulatory processes of the project described in this paper, such as ceding of IRB oversight to the sIRB, development and coordination of the regulatory infrastructure, training of personnel, and execution of necessary regulatory agreements, were not without their challenges. For example, different sites had different requirements when executing compliance agreements, often requiring a full submission to their own internal IRB for review before compliance agreements could be executed. Specifically, although the sIRB determined that informed consent is not required in accordance with the US Department of Health and Human Research policy [15], the intervention sites determined that informed consent was needed on the basis of their own internal review, which added additional regulatory components (site specific consent forms) and subsequent regulatory processes in the establishment of the necessary agreements. Based on these external processes, the study team was faced with significant delays at the start of intervention and data extraction training and had to address site-specific requirements related to HIPAA authorization, consent, etc. Additionally, these inconsistencies with site-specific determinations and subsequent delays impacted the timeliness of personnel training across the participating sites, further delaying the implementation of the intervention’s activities.

Strategies and Lessons Learned

The execution of this pilot yielded several regulatory lessons that allowed for the successful and compliant implementation of a pilot pragmatic trial and paved the way for a successful full-scale trial. Strategies used by the study team to navigate regulatory challenges include (1) adopting a phased approach to protocol development and review of regulatory processes to allow for more effective multisite collaboration and ensure compliance with the study’s timeline; (2) coordinating appropriate training of multisite personnel specific to their role on the team, which varied greatly based on site-specific activities; and (3) ensuring that all regulatory aspects were coordinated and monitored by a specialized team member at the sIRB, who served as a primary point of contact across all sites and provided consistent, clear, and timely information related to all regulatory components and requirements.

The regulatory lessons learned from this multisite study indicate a need for investigators to prepare as early as possible owing to the collaborative nature of multisite human subjects research—such as early as the grant preparation phase—and gain a general understanding of all participating sites’ unique regulatory requirements to adequately prepare for entering into the necessary agreements to begin the research. Having this understanding as part of the grant proposal phase will undoubtedly assist in the development and execution of these agreements upon receiving a notice of award, and reduce the likelihood of overall study timeline delays caused by delayed establishment of the regulatory infrastructure. Additionally, we recommend having skilled individuals appointed to the project who are tasked specifically with the regulatory management of multisite trials. While ultimately, it is the responsibility of the study’s principal investigator to oversee all regulatory components, having dedicated personnel who are knowledgeable in the nuances of regulatory aspects of clinical trial research and experienced in developing multisite agreements and establishing multisite infrastructure in line with these agreements would further prevent any delays in the ability to begin study activities.

Contribution to the Existing Literature

Much of the existing literature that describes similar studies more often focuses on aspects related to the study design and methodology of embedded pragmatic trials and addresses ethical considerations through blanket statements that include IRB review, compensation, informed consent, etc [7,23,24]. Further, literature focused on providing recommendations to researchers or regulatory bodies (or both) on the challenges and recommendations related to multisite research [12,25] and pragmatic trials [14,16] has been effective in implementing new approaches in regulatory processes, such as central IRB review and oversight and providing standardized methods for promoting multisite collaborations; such literature is more often generalizable, rather than providing practical examples within the context of a specific pragmatic trial in detail. Thus, the literature on regulatory elements required for conducting multisite pragmatic research and timelines for ensuring regulatory components is scarce, particularly after the implementation of the sIRB policy enacted in 2016 [17]. When considering existing literature that focuses on the ethical considerations of pragmatic trial research as it relates to EHR data, much of this work tends to focus on the larger ethical implications of data sharing and informed consent rather than the regulatory aspects required of researchers undertaking this work [24,26]. Thus, this paper uniquely describes the procedures and the lessons learned from the perspective of a specific project that may serve as a framework for future studies.

Limitations

The regulatory outcomes of the study are not without their limitations. First, the regulatory elements specific to this project are undoubtedly not specific to pragmatic trials. Further, as this study was part of the ATN, these regulatory findings may not be generalizable to other multisite trials that do not operate within a larger federally funded network structure, with their own data sharing and management expectations and policies. Additionally, the protocols were impacted by the onset of the COVID-19 pandemic. These delays resulted in modifications of EHR data downloads, limiting the data to 2018 and 2019, as opposed to the original plan to download 2018-2021 EHR data. In addition, the pragmatic trial’s timeline was significantly impacted by the COVID-19 pandemic, particularly for those intervention sites where direct participant engagement occurred, wherein the sIRB required modifications to the study protocol that included COVID-19 mitigation practices as they relate to participant safety.
Conclusions

The necessary regulatory components of conducting multisite embedded pragmatic trials require considerable planning on behalf of researchers from the grant proposal stage through study completion. Researchers must consider realistic timelines, necessary regulatory infrastructure associated with the use of EHR data, understanding the requirements for complying with guidance from federal sponsors (e.g., the National Institutes of Health) related to data sharing and management, and adequate preparation and training of project personnel. Future pragmatic trials should provide further details for regulatory processes, challenges, and timelines to expand on the literature in this domain.

Acknowledgments

The ATN 162b (Adolescent Trials Network for HIV/AIDS Interventions Protocol 162b) was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5U24HD089880-05).

Data Availability

Data associated with the ATN 162b (Adolescent Trials Network for HIV/AIDS Interventions Protocol 162b) are held by the primary study site (Florida State University) and will be published in the main study manuscript (under preparation). Researchers may contact the study’s principal investigator to make specific data requests.

Conflicts of Interest

None declared.

Multimedia Appendix 1
[DOCX File, 15 KB - formative_v8i1e46420_app1.docx ]

Multimedia Appendix 2
Timeline for multisite data use agreements for electronic health records for adolescent HIV research.
[DOCX File, 14 KB - formative_v8i1e46420_app2.docx ]

References


Abbreviations

ART: antiretroviral therapy
ATN: Adolescent Trials Network
CPT-4: Current Procedural Terminology
EHR: electronic health record
FSU: Florida State University
HCPCS: Healthcare Common procedure Coding System
HIPAA: Health Insurance Portability and Accountability Act
ICD-10: International Statistical Classification of Diseases, Tenth Revision
IRB: institutional review board
MESA: Motivational Enhancement System for Adherence
mHealth: mobile health
**PrEP:** pre-exposure prophylaxis

**sIRB:** single institutional review board of record
ChatGPT as a Tool for Medical Education and Clinical Decision-Making on the Wards: Case Study

Anthony Skryd¹, MD; Katharine Lawrence², MD

¹Department of Medicine, NYU Langone Health, New York City, NY, United States
²Department of Population Health, NYU Grossman School of Medicine, New York City, NY, United States

Corresponding Author:
Anthony Skryd, MD
Department of Medicine
NYU Langone Health
550 1st Avenue
New York City, NY, 10016
United States
Phone: 1 646 929 7800
Email: anthonyskryd@nyulangone.org

Abstract

Background: Large language models (LLMs) are computational artificial intelligence systems with advanced natural language processing capabilities that have recently been popularized among health care students and educators due to their ability to provide real-time access to a vast amount of medical knowledge. The adoption of LLM technology into medical education and training has varied, and little empirical evidence exists to support its use in clinical teaching environments.

Objective: The aim of the study is to identify and qualitatively evaluate potential use cases and limitations of LLM technology for real-time ward-based educational contexts.

Methods: A brief, single-site exploratory evaluation of the publicly available ChatGPT-3.5 (OpenAI) was conducted by implementing the tool into the daily attending rounds of a general internal medicine inpatient service at a large urban academic medical center. ChatGPT was integrated into rounds via both structured and organic use, using the web-based “chatbot” style interface to interact with the LLM through conversational free-text and discrete queries. A qualitative approach using phenomenological inquiry was used to identify key insights related to the use of ChatGPT through analysis of ChatGPT conversation logs and associated shorthand notes from the clinical sessions.

Results: Identified use cases for ChatGPT integration included addressing medical knowledge gaps through discrete medical knowledge inquiries, building differential diagnoses and engaging dual-process thinking, challenging medical axioms, using cognitive aids to support acute care decision-making, and improving complex care management by facilitating conversations with subspecialties. Potential additional uses included engaging in difficult conversations with patients, exploring ethical challenges and general medical ethics teaching, personal continuing medical education resources, developing ward-based teaching tools, supporting and automating clinical documentation, and supporting productivity and task management. LLM biases, misinformation, ethics, and health equity were identified as areas of concern and potential limitations to clinical and training use. A code of conduct on ethical and appropriate use was also developed to guide team usage on the wards.

Conclusions: Overall, ChatGPT offers a novel tool to enhance ward-based learning through rapid information querying, second-order content exploration, and engaged team discussion regarding generated responses. More research is needed to fully understand contexts for educational use, particularly regarding the risks and limitations of the tool in clinical settings and its impacts on trainee development.

(JMIR Form Res 2024;8:e51346) doi:10.2196/51346

KEYWORDS
ChatGPT; medical education; large language models; LLMs; clinical decision-making
Introduction

Large language models (LLMs) are computational artificial intelligence (AI) systems that are trained on large volumes of content from the internet and other sources to create natural, human-like written communications, images, and other outputs [1]. Popular general-use commercial LLMs include ChatGPT (OpenAI), Palm (Google), and LLaMA (Meta); health care–specific LLMs also in development but less widely available include Med-PaLM (Google), PMC-LLaMA (Meta), and BioGPT (Microsoft Corp) [2]. LLMs’ advanced natural language processing capabilities provide a unique opportunity to enhance medical education by providing students and educators with interactive, real-time access to and feedback on a vast amount of medical knowledge. This dialogue can be personalized to the level of the learner and leveraged to answer clinical questions, provide differential diagnoses, generate resources, and facilitate the assimilation of complex medical concepts [3]. Studies have documented facilities of ChatGPTs and others with standardized examinations, academic abstracts, and clinical documentation [4-6]. At the same time, general concerns have been raised regarding the practical usability, clinical practice implications, and overall ethics of using LLMs in health care [7,8]; for medical education, issues of overreliance, plagiarism, misinformation, bias, and inequity are particularly acute [9]. As a result, the adoption of LLM technology into medical education and training has been varied, and comprehensive guidelines for its systematic application in learning contexts remain underdeveloped.

In this case study, we present the use of a publicly available LLM (ChatGPT) as a real-time interactive educational tool for attending teaching rounds on an inpatient resident medicine service at a large urban academic medical center. We identify select ChatGPT use cases and qualitatively evaluate the tool’s impact on real-time clinical learning, diagnostic reasoning, and medical decision-making for medical residents and teaching attendings. We further explore the perceived advantages, limitations, and ethical considerations of ChatGPT in real-world clinical and teaching contexts and consider the future implications of generative AI conversational technology as a tool for medical education.

Methods

Context

We conducted a brief, single-site pilot study of the publicly available ChatGPT-3.5 (OpenAI) by implementing the tool in the attending rounds workflow of an inpatient general internal medicine service in a large urban academic medical center in New York.

Attending-style rounds were conducted for 1.5-2 hours daily in the mornings over the course of the 7-day rotation, consisting of patient presentations, case reviews, case-based learning, and didactics. ChatGPT was integrated into rounds via both structured and organic use, using the web-based “chatbot” style user interface to interact with the LLM through conversational free-text and discrete queries. ChatGPT was prompted via a zero-shot approach, without the use of prior training sets, data, or examples.

Before initiating the pilot study, the team established a code of conduct for ChatGPT’s use based on a shared understanding of its potential, general risks, and implications for patient care (Textbox 1). This code of conduct guided the tool’s use throughout the test period.

Over the 7 days of the pilot study, the team established a standardized implementation method and ensured adherence to the code of conduct for use. All relevant ChatGPT outputs were independently verified by team members using validated resources (eg, PubMed, UpToDate, and medical society guidelines). Any paper citations or references provided by ChatGPT were also reviewed.

Textbox 1. Team-developed code of conduct for ChatGPT use on attending rounds.

We acknowledge the potential risks and harms of using technology like ChatGPT in our clinical work and training.

As a team, we agree:

- Not to use any specific or identifiable patient information in our interactions with ChatGPT.
- To independently verify and validate any answers ChatGPT produces, and not make medical decisions based on ChatGPT’s outputs unless we can confirm their accuracy.
- To be honest with our patients and one another when we use ChatGPT in our clinical work.
- To be open to the possibility that ChatGPT is more intelligent than we are.

Overall, we commit to placing patient care, safety, and trust above any educational or research use of this technology. We further commit to abiding by our institution’s information technology policies and practices.

Analysis

A qualitative approach using phenomenological inquiry was used to identify key insights related to the use of ChatGPT for real-time ward-based educational contexts. Analysis was conducted in two phases: (1) an in situ review and validation of the ChatGPT outputs by the clinical team and (2) a retrospective rapid qualitative analysis using phenomenological inquiry performed by the coauthors (AS and KL) [10]. In phase 1, ChatGPT conversation logs and outputs were group-reviewed by the clinical team and qualitatively assessed for factualness, quality, relevance, and usefulness in clinical contexts. The team also conducted a group debrief at the culmination of the clinical block via a semistructured group interview led by the attending, which explored primary uses, perceptions, and learning from the experience and perceived impacts on medical education. In
phase 2, the senior author (KL) used a pragmatic phenomenological approach to concisely review the ChatGPT conversation logs and associated short-hand attending notes from the clinical sessions and identify both emergent major, minor, and outlier themes and themes specifically related to the use of ChatGPT for real-time ward-based educational contexts; these themes were reviewed and revised between both authors (KL and AS) [10,11].

**Ethical Considerations**

This study was approved as part of a quality improvement initiative under the NYU Grossman School of Medicine institutional review board. Study data do not include any personal health information related to any care provision conducted during the course of the investigation.

**Results**

**Overview**

The team was comprised of 7 members: 1 attending internal medicine physician, 1 senior medicine resident, 2 interns, 2 medical students, and 1 physician assistant student. All team members expressed baseline familiarity with ChatGPT, with most knowledge derived from social and general media information, particularly around controversies in its use. No team member had received formal ChatGPT training or guidance on its use from their educational program or was actively using the technology in their clinical work. No team member had prior experience with prompt engineering, specific coding, or data management skills to interact with the tool at a more advanced level.

Over the course of the 7-day pilot study, ChatGPT was queried 17 times, representing a combination of single-question queries and longer bidirectional interchanges. ChatGPT prompts were generated by all members of the team. The types of ChatGPT use during attending rounds were identified (Textbox 2).
### Discrete Medical Knowledge Inquiries

- Review of common illnesses
- Uncommon diagnoses
- Clinical aids and diagnostic calculators
- Medication interactions and side effects
- Mechanism of action of medications

### Building Differential Diagnoses and Engaging Dual-Process Thinking

- Initial and expanded differential diagnoses
- Rare diagnoses
- Validating clinical reasoning

### Challenging Medical Axioms

- Validating existing knowledge
- Challenging teaching points

### Cognitive Aids in Acute Care Scenarios

- Initial evaluation
- Further testing considerations
- Other cognitive aids
- Debriefing tools

### Facilitating Conversations with Subspecialties

- Identifying specialty-specific best practices to cite when discussing a complex case with multiple specialties involved
- Patient advocacy resources

### Other Topics

- Engaging in difficult conversations
- Ethical challenges and general medical ethics
- Personal continuing medical education resources
- Development of teaching tools
- Clinical documentation
- Productivity and task management support

---

**Discrete Medical Knowledge Inquiries**

Attending rounds generated numerous discrete medical knowledge questions, often based on the knowledge gaps of team members or inquiries from specific patient cases. During this pilot, ChatGPT was substituted for other frequently used web-based resources (eg, UpToDate) to answer many discrete medical inquiries; this represented a combination of first-order (factual information seeking) and second-order (process and reasoning seeking) questions (Textbox 2 and Figures S1-S3 in Multimedia Appendix 1), with first-order questions frequently leading to additional second-order queries. In general, using ChatGPT for discrete knowledge inquiries often led to further questions and additional prompting, in turn generating team discussion and knowledge sharing. This process was identified by both the attending and the junior team members as an effective way to gain additional knowledge regarding a topic of interest without considerable extra time or effort.

**Building Differential Diagnoses and Engaging Dual-Process Thinking**

Crafting a comprehensive differential diagnosis is essential to clinical reasoning in patient evaluation [12]. A well-known method to engage complex reasoning is the dual-process approach [13], whereby system 1 reflexive, intuitive thinking is complemented by system 2 rational and more cognitively intensive analytic thought [14,15]. During the pilot study, ChatGPT was regularly queried to provide differential diagnoses for general patient presentations, with additional prompting to provide expanded differentials, including uncommon diagnoses. The team reviewed these differentials and discussed their plausibility, likelihood in the case context, and completeness.
compared to the team-generated differentials. In most instances, the differentials provided by ChatGPT mirrored those of the medical team; in some cases, differentials collaboratively produced by team members were more expansive and included likely diagnoses not provided by ChatGPT (Figure S4 in Multimedia Appendix 1), while in fewer instances, ChatGPT provided novel diagnoses that prompted additional queries and resulted in new learning for the team (Figures S5 and S6 in Multimedia Appendix 1). Overall, the team reflected that ChatGPT-generated differentials were considered a helpful supplement to team-generated ones for confirming clinical reasoning and completeness but did not meaningfully change leading diagnoses or care plans.

**Challenging Medical Axioms**

An axiom is defined as a rule, principle, or truth that is often widely accepted on its merit without proof or basis for further analysis [16]. In this case, ChatGPT was used to query axiomatic practice on rounds and either justify or challenge them (Figures S7–S9 in Multimedia Appendix 1). Common practice habits (eg, timing of medication doses) or practices used widely across patients (eg, electrolyte repletion) were queried most often—generally in response to a team member’s question (Why do we do this?)—resulting in deeper knowledge for the team regarding their daily work. The use of ChatGPT to confirm or challenge medical axioms generated a thoughtful discussion among the team members regarding the basis of medical knowledge and the transmission of learning in medical education.

**Cognitive Aids in Acute Care Scenarios**

Acute care scenarios and medical emergencies are stressful, complex events that require rapid response and coordination of care that is often time-critical. The use of cognitive aids and checklists has been widely studied and proven effective in optimizing care and reducing errors in emergency response scenarios [17]. During the pilot study, ChatGPT was deployed as part of a rapid response team (RRT) debrief, in which the tool was used after the team had performed an RRT to replicate and review the clinical scenario (Figures S10 and S11 in Multimedia Appendix 1). On review, the overall structure of ChatGPT’s RRT management response (eg, initial scene assessment and triage process) was considered poorly organized and insufficiently specific to guide real-time RRT management. Conversely, outputs did include thorough reasoning for recommended RRT procedures (eg, laboratory testing and imaging), which the team perceived as helpful to recall during high-stress scenarios (Figure S11 in Multimedia Appendix 1). Overall, the team felt the level of prompting and interaction required to get appropriate and well-structured information to guide an RRT was overly burdensome and inefficient compared to existing processes.

**Facilitating Conversations With Subspecialties**

Another use for ChatGPT was identified during a complex patient case involving multiple specialties and ongoing goals of care conversation. During rounds, the team expressed concern regarding their ability to effectively care for the patient and their communications with specialists as a result. In response, ChatGPT was queried regarding specific best practice guidelines for medical versus surgical management of the condition (Figures S12 and S13 in Multimedia Appendix 1). Outputs provided by ChatGPT were supplemented with web-based inquiry to validate the content and identify the most up-to-date information; through this combined internet and ChatGPT process, a previously unknown set of guidelines (Figure S14 in Multimedia Appendix 1) [18] were identified, which provided the team with specialty-specific references to guide further high-level conversations with consultants. Overall, the team felt that ChatGPT had given them a better understanding of management options, which empowered them to advocate for their patient and work collaboratively with specialists on the case.

**Other Topics**

Upon completing the 7-day pilot study, the team met to debrief on the experience and identify additional potential use cases. These included engaging in difficult conversations with patients, exploring ethical challenges and general medical ethics teaching, personal continuing medical education resources, developing teaching tools (eg, “teaching on the wards” aids), supporting and automating clinical documentation, and supporting productivity and task management. Of note, when prompted throughout the investigation to cite specific references, ChatGPT provided outputs but noted it could not cite the specific location of the information provided; further review of the citations by the team revealed none referred to actual papers. Reference provisions were therefore considered unreliable use case by the team. The team also discussed ongoing ethical issues in ChatGPT uses for health care, including its well-documented biases and potential to result in health inequities, medico-legal implications, data privacy and security, and potential nefarious uses. This discussion resulted in a prompt inquiry to ChatGPT on the future of medical education as generative AI technologies advance (Figure S15 in Multimedia Appendix 1).

**Discussion**

**Principal Findings**

This exploratory case study examined various use cases of the commercially available LLM, ChatGPT, as an educational tool for attending teaching rounds of an inpatient resident medicine service at a large urban academic medical center. We identified several key areas for which ChatGPT was used, including addressing team or individual knowledge gaps and validating funds of knowledge through discrete medical knowledge inquiries; expanding differential diagnoses and engaging dual systems process thinking to validate and expand clinical reasoning; challenging axioms through active investigation of default medical knowledge and practice heuristics; supporting triage, diagnostic, and care decision-making during acute care emergencies; and facilitating patient advocacy and complex care management through improved specialty consultations. Other topics not directly explored but identified as potential use cases included challenging patient scenarios and conflicts, medical ethics inquiries, general continuing medical education, and team productivity and efficiency. Overall, the tool was considered a promising addition to the learning environment.
while it was also noted to be limited in accuracy, reliability, and usability in its current state. In particular, using ChatGPT as a real-time educational aid during attending rounds enhanced team learning by fostering a discussion of responses and generating further areas of exploration and inquiry.

Contributions to the Literature and Limitations

While exploratory, this case study adds to the rapidly growing literature exploring the various uses and limitations of generative AI technologies, such as LLMs, in medical education. As previously stated, ChatGPT has been successfully tested in a growing number of medical training contexts, including writing medical notes and academic abstracts and completing the United States Medical Licensing Examination (USMLE); a surfeit of opportunities to explore the impact of generative AI tools in medical education has also been identified at the undergraduate, graduate, and professional levels [5,6,19-22]. Significant work remains, however, to better understand the roles of these technologies in medical education, including both the extent to which these technologies have already been integrated into educational programs (either formally or through casual use) as well as the true appetite for their future use [23,24]. At the same time, there is an ongoing need to better characterize and actively mitigate the risks of these tools’ use in care delivery, particularly among developing trainees. This study identified numerous limitations of the technology, many of which have been described elsewhere. These included difficulties validating information sources and specific references, inconsistent responses to similar prompts, and incomplete access to major databases and up-to-date material [19]. Although we were unable to confirm examples of misinformation or bias—such as the generated missing references—that may have occurred during our pilot study, our experience reflects larger grievances with the current publicly available LLM tools, in particular around issues of output “trustworthiness” and response fidelity over time [25].

There is also the potential of ChatGPT and other tools to perpetuate cognitive and sociocultural biases [26,27], impacting both trainee development and overall care delivery; processes to better center equity in LLMs and mitigate bias-related AI harms are needed to address this [1,28,29].

In addition to the larger issues of LLMs identified earlier, this study has several important limitations. Using ChatGPT on a single team over the 7 days of a shared clinical rotation restricted the depth and range of analysis to a small-scale pilot study with a short duration; future investigations should extend this period as well as evaluate the impact on different clinical care teams (eg, nurses and pharmacists) and other team factors. Significantly, our team had limited knowledge of prompt engineering and optimization in generating desirable ChatGPT outputs, which likely limited our interaction potential with technology and may have introduced specific interaction biases; conversely, our interactions with the ChatGPT user interface also likely represent an “average” user experience for a health care provider at the time, which will likely evolve as clinicians gain familiarity and skill with the tool. Future work should emphasize the role of prompt engineering and various other approaches to priming and interacting with LLMs (eg, “zero-shot” vs few-shot learning) [30]. Additionally, ChatGPT output quality was assessed without the aid of existing validated tools to measure performance or in comparison to other LLMs; there is a need for both general objective measures as well as comparative metrics across products to more rigorously benchmark and compare these tools.

Conclusions

This case study explored ChatGPT as an educational tool in an inpatient academic medical service. Several noteworthy use cases of LLM were identified, including addressing knowledge gaps, expanding differential diagnoses, challenging medical axioms, supporting acute care decision-making, and facilitating complex care management through improved specialty consultations. Overall, ChatGPT enhanced team learning by prompting engaged discussion and further areas of exploration and inquiry. LLMs continue to demonstrate promise and peril in health care, with particular opportunities and risks in educational spaces. Concerns related to biases, misinformation, and ethical implications in health care emphasize the need for further consideration and regulatory guidelines for LLM application in medical education and clinical practice. Ultimately, technical progress (or stagnation), regulatory oversight, and social appetite will likely decide their future. Researchers should continue to study impacts by identifying further use cases for investigation, conducting meta-analyses on the myriad of case studies currently being conducted, better defining study designs and evaluation tools, and advocating for the safe, ethical, and equitable use of these technologies.

Acknowledgments

The authors received no financial support for writing this paper. The authors express their gratitude to the inpatient general internal medicine team, including the medical interns, medical students, and physician assistant students who actively participated in this study. ChatGPT was not used in the creation of the text of this paper.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample ChatGPT queries and outputs.

References


Abbreviations

AI: artificial intelligence
LLM: large language model
RRT: rapid response team

©Anthony Skryd, Katharine Lawrence. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Acceptance of Medical Artificial Intelligence in Skin Cancer Screening: Choice-Based Conjoint Survey

Inga Jagemann, MSc; Ole Wensing, BSc; Manuel Stegemann, Prof Dr; Gerrit Hirschfeld, Prof Dr

School of Business, University of Applied Sciences and Arts Bielefeld, Bielefeld, Germany

Corresponding Author:
Inga Jagemann, MSc
School of Business
University of Applied Sciences and Arts Bielefeld
Interaktion 1
Bielefeld, 33619
Germany
Phone: 49 521106 ext 70508
Email: inga.jagemann@hsbi.de

Abstract

Background: There is great interest in using artificial intelligence (AI) to screen for skin cancer. This is fueled by a rising incidence of skin cancer and an increasing scarcity of trained dermatologists. AI systems capable of identifying melanoma could save lives, enable immediate access to screenings, and reduce unnecessary care and health care costs. While such AI-based systems are useful from a public health perspective, past research has shown that individual patients are very hesitant about being examined by an AI system.

Objective: The aim of this study was two-fold: (1) to determine the relative importance of the provider (in-person physician, physician via teledermatology, AI, personalized AI), costs of screening (free, 10€, 25€, 40€; 1€=US $1.09), and waiting time (immediate, 1 day, 1 week, 4 weeks) as attributes contributing to patients’ choices of a particular mode of skin cancer screening; and (2) to investigate whether sociodemographic characteristics, especially age, were systematically related to participants’ individual choices.

Methods: A choice-based conjoint analysis was used to examine the acceptance of medical AI for a skin cancer screening from the patient’s perspective. Participants responded to 12 choice sets, each containing three screening variants, where each variant was described through the attributes of provider, costs, and waiting time. Furthermore, the impacts of sociodemographic characteristics (age, gender, income, job status, and educational background) on the choices were assessed.

Results: Among the 383 clicks on the survey link, a total of 126 (32.9%) respondents completed the online survey. The conjoint analysis showed that the three attributes had more or less equal importance in contributing to the participants’ choices, with provider being the most important attribute. Inspecting the individual part-worths of conjoint attributes showed that treatment by a physician was the most preferred modality, followed by electronic consultation with a physician and personalized AI; the lowest scores were found for the three AI levels. Concerning the relationship between sociodemographic characteristics and relative importance, only age showed a significant positive association to the importance of the attribute provider ($r=0.21$, $P=0.02$), in which younger participants put less importance on the provider than older participants. All other correlations were not significant.

Conclusions: This study adds to the growing body of research using choice-based experiments to investigate the acceptance of AI in health contexts. Future studies are needed to explore the reasons why AI is accepted or rejected and whether sociodemographic characteristics are associated with this decision.

(JMIR Form Res 2024;8:e46402) doi: 10.2196/46402

KEYWORDS

artificial intelligence; skin cancer screening; choice experiment; melanoma; conjoint analysis, technology acceptance; adoption; technology use; dermatology; skin cancer; oncology; screening; choice based; trust
Introduction

Skin cancers are the most common groups of cancers diagnosed worldwide, with more than 1.5 million new cases estimated in 2020 [1]. Melanoma is the deadliest form of skin cancer. Based on demographic changes, it is estimated that more than 500,000 new cases of melanoma and almost 100,000 deaths from melanoma should be expected worldwide by 2040 [1]. As melanoma case numbers are expected to increase in the future, high-cost treatments will continue to put a strain on the already overburdened health care budgets. To combat the rising mortality rate of melanoma, early detection is critical. Currently, the German national treatment guidelines [2] recommend skin cancer screening as a standardized full-body skin examination performed by dermatologists who have completed specialized training in the early detection of skin cancer. In addition, dermatologists should use dermoscopy to diagnose suspected skin cancer. Given the rising number of cases as well as increasing scarcity of trained dermatologists [3-5], there has been substantial research into the feasibility of artificial intelligence (AI) to augment or replace traditional skin cancer screening regimens [6].

AI describes machines (or computers) that mimic the cognitive functions associated with human thought, such as learning and problem-solving. These systems observe their surroundings and adopt action to reach their targets directly [7]. Further, AI has the ability to learn from images and subsequently provide an image-based diagnosis. Dermatology, as an image-based field of medicine, retains a dominant position in the AI evolution with the ability to classify skin lesions [8].

Research into the technical quality of AI-based skin cancer screening technologies has shown that these systems achieve detection rates that are on par or better than those of highly trained clinicians [9-13]. This highlights the great potential of AI for future skin cancer screening in the general population. As part of apps, AI systems offer immediate access to dermatological screening for all patients with mobile digital devices, enabling health care and treatment to be provided regardless of time and place and close to everyday life [6]. Thus, AI systems capable of detecting melanoma and non-melanoma skin cancer could avoid unnecessary care, reduce health care costs, offer solutions to the increasing scarcity of clinicians, and reduce the waiting times for an appointment and for a diagnosis [3-5]. However, there is a risk that some melanomas will be missed and treatment delayed if the apps incorrectly reassure the user that their lesion is of low risk [14].

Although the technical quality has improved, there is also a growing awareness that patients do not generally accept the use of AI-based systems in health care settings. There is still no consistent definition of technology acceptance in the literature. Terms such as “acceptability,” “acceptance,” and “adoption” are often employed in this context, sometimes interchangeably. Dillon and Morris [15] defined user acceptance “as the demonstrable willingness within a user group to employ IT [information technology] for the tasks it is designed to support.”

Khullar et al [16] conducted an online survey to examine patients’ perspectives about applications of AI in health care, showing that 31% of respondents reported being very uncomfortable and 40.5% were somewhat uncomfortable with receiving a diagnosis from an AI algorithm that was accurate 90% of the time but incapable of explaining its rationale. Longoni et al [17] demonstrated that consumers are very hesitant about being examined by an AI system and consumers’ willingness to pay decreases when an equivalent service is performed by an AI system. Additionally, they concluded that patients’ perceived neglect of uniqueness leads to more resistance to medical AI [17].

Past research has also identified several factors that might impact patients’ preferences to use AI-based health care services. The European Commission [18] interviewed citizens of the 28 member states of the European Union (N=27,900) and concluded that younger participants with a high educational level are more likely to use online health care services. This finding was also replicated in oncology patients, where younger patients indicated higher acceptance of and a greater intention to use digital tools and apps to manage their cancer [19]. The European Commission [18] also found that the opinion on AI strongly depends on exposure to related information and knowledge. This relationship is also supported by a series of experiments showing that resistance to the utilization of medical AI is driven by the subjective difficulty of understanding algorithms [13].

Concerning skin cancer, previous research has shown that patients were generally reluctant to use AI-based systems in the field of dermatology. Snoswell et al [6] examined the consumer preference and willingness to pay for mobile teledermoscopy services in Australia using a discrete-choice experiment (N=199). They found that patients prefer a trained medical professional to be involved in their skin cancer screening and that patients are less willing to pay money for teledermatology [6]. However, Snoswell et al [6] did not take into account sociodemographic factors that may have had an impact on the patients’ decisions. In a multicenter clinical study assessing the performance of automated diagnosis of melanoma with a self-completion questionnaire (N=65), Fink et al [20] found that most patients agreed that computer-assisted diagnoses are trustworthy and may generally improve the diagnostic performance of physicians. However, participants rejected the idea of AI-based systems completely replacing physicians and instead strongly favored hybrid solutions in which diagnoses by a physician are supported by automated systems [20].

To date, only three studies have directly addressed the question of which factors are associated with patients’ preferences regarding AI-based skin cancer screening [21-23]. Ghanit et al [22] studied public interest in teledermatology, which was found to be positively associated with a younger age, higher educational attainment, and higher household income. Chang et al [21] examined sociodemographic differences in teledermatology acceptance with a cross-sectional survey (N=13,996), showing that respondents who were interested in teledermatology were more frequently 18–39 years of age, men, college graduates, and tablet or smartphone users. Similarly, young age, male gender, a previous history of melanoma, and higher educational level were significantly associated with a more positive attitude toward skin cancer–related apps [23]. However, it is unclear whether these results from questionnaires...
can be replicated in choice-based experiments that rely to a lesser degree on introspection and are thus one step closer to actual behavior [24].

As described above, provider and costs for a skin cancer screening have high relevance for the user [6,16,17,20]. For this choice-based conjoint analysis, we further added the attribute waiting time for a diagnosis, because studies have shown a strong negative correlation between patient satisfaction and waiting time [25,26]. AI provides the opportunity to get a skin cancer screening immediately, without any waiting time [3]. Due to the shortage of medical professionals, waiting time for a skin cancer diagnosis is also an important attribute for the user [5,7].

The aim of this study was two-fold based on the following two research questions: (1) How important are the attributes provider, costs for screening, and waiting time for diagnosis for participants’ preference for skin cancer screening? (2) Are sociodemographic characteristics, especially age, systematically related to the relative importance scores of participants to the various attributes?

**Methods**

**Study Design**

This cross-sectional study used a choice-based conjoint analysis to examine the acceptance of medical AI for a skin cancer screening from the user perspective. Conjoint analysis is a quantitative marketing research method that quantifies the value consumers place on the attributes of a product [27]. Respondents are asked to make a choice between 2 or more different choice sets, where each set is described in terms of several predefined attributes, each with different levels. Given a sufficient number of choices per respondent, it is then possible to statistically estimate the importance of each attribute and level for the choice in terms of part-worth utilities. This method offers a behavioral approach and is less susceptible to social desirability and other biases [28].

This study systematically manipulated three attributes (provider, cost, and waiting time) for a hypothetical skin cancer screening. Participants were presented with 12 different choice sets one after another, each consisting of three different modes of skin cancer screenings that were generated by combining different levels of the three attributes (see Figure 1 for an example). The choice sets were generated by the conjointly algorithm using default settings [29].

**Survey**

Before the survey was conducted, it was tested with the “think-aloud” method by three volunteers to find out if there were any comprehensibility problems. For this purpose, the pretest participants had to speak their thoughts aloud while completing the survey [30].

The questionnaire started with informed consent, where participants were informed about the nature and scope of the survey and about the protection of their data. Before starting the questionnaire, participants completed the consent form and agreed to participate in the anonymous study. The participants then moved on to the choice-based conjoint task, which consisted of 12 different choice sets. The participant’s task for each choice set was to indicate the skin cancer screening that they most prefer (ie, they selected one of the three options as their preferred choice). After responding to the choice sets, participants were asked whether they had undergone a skin cancer screening in the last year and at which type of provider.

Finally, the sociodemographic characteristics (age, gender, education, status, income) were assessed. Finally, the survey asked again whether the data could be used for analysis in anonymized form in case respondents changed their minds during the course of the survey and to filter out people who just wanted to “click through” without seriously answering the questions.

**Participants**

Recruitment was based on a convenience sample through the social environment; individuals were asked to participate in the open voluntary survey shared with contacts via WhatsApp and Instagram. Standard procedures for conducting and reporting online surveys [31] were followed. Furthermore, conjointly’s default methods were used to identify and bar potential duplicate entries from the same user. Data were collected during the time period of September 29, 2022, through October 20, 2022. The link to the survey was clicked 383 times by unique site visitors. Of these potential respondents, 126 (32.9%) people...
filled out the conjoint survey completely and gave their agreement for processing their data. In total, 220 (57.4%) respondents opened the link but did not complete the survey and another 33 (8.6%) people were disqualified from the study because they answered the survey several times. Three people (0.8%) did not give their agreement to process their data and a single respondent (0.3%) was excluded because the survey was answered too quickly. Respondents took an average of 4.7 minutes to complete the survey. Table 1 provides an overview of the respondents’ sociodemographic characteristics. There was a relatively equal proportion of participants identifying as male and female. The average age of the participants was 37.6 years and the median age was 29 years.

Table 1. Sociodemographic characteristics of participants sampled from September 29, 2022, to October 20, 2022 in Germany (N=126).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>58 (46.0)</td>
</tr>
<tr>
<td>Female</td>
<td>67 (53.2)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Still a student</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>School-leaving qualification</td>
<td>25 (19.8)</td>
</tr>
<tr>
<td>Vocational qualification</td>
<td>34 (27)</td>
</tr>
<tr>
<td>University degree</td>
<td>58 (46)</td>
</tr>
<tr>
<td>Doctorate</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Other degree</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Not specified</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
</tr>
<tr>
<td>Elementary/high school student</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>University student</td>
<td>29 (23)</td>
</tr>
<tr>
<td>Apprentice</td>
<td>11 (8.7)</td>
</tr>
<tr>
<td>Employee</td>
<td>58 (46)</td>
</tr>
<tr>
<td>Civil servant</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>8 (6.3)</td>
</tr>
<tr>
<td>Not employed</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Retired without income</td>
<td>9 (7.1)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Not specified</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td><strong>Monthly income (Euro; 1€=US$1.09)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;250</td>
<td>5 (4.0)</td>
</tr>
<tr>
<td>250–499</td>
<td>8 (6.3)</td>
</tr>
<tr>
<td>500–999</td>
<td>17 (13.5)</td>
</tr>
<tr>
<td>1000–1499</td>
<td>12 (9.5)</td>
</tr>
<tr>
<td>1500–1999</td>
<td>13 (10.3)</td>
</tr>
<tr>
<td>2000–2999</td>
<td>26 (20.6)</td>
</tr>
<tr>
<td>3000–3999</td>
<td>16 (12.7)</td>
</tr>
<tr>
<td>4000–4999</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>&gt;5000</td>
<td>13 (10.3)</td>
</tr>
<tr>
<td>Not specified</td>
<td>10 (7.9)</td>
</tr>
</tbody>
</table>
Ethical Considerations

Our online study was conducted in accordance with the American Psychological Association’s Ethical Principles of Psychologists and Code of Conduct. In particular, data collection was anonymous; harmless to participants; and did not involve deception, injure, or place participants under high levels of physical or emotional stress. In line with 2023 guidelines of the German Research Foundation, formal ethical approval was not required because our study did not include aspects that would necessitate a statement, per subsection two of the “Information on proposals in the field of psychology” [32]. Informed consent was obtained from all participants after the purpose of the study and the data collection were outlined in the survey introduction. Participants indicated their consent by clicking a button. Study data and identifiers were anonymized during the data collection and data analysis to maintain confidentiality. No compensation was awarded to participants.

Results

Table 2 provides an overview of the relative importance values of the attributes and part-worth values of each level for each attribute as determined by conjointly to answer the first research question [29]. A treatment by a physician that is completely compensated by insurance and has no waiting time emerged as the most preferred mode of treatment. Overall, provider was the most important attribute, followed by costs and waiting time. For all attributes, we found two levels with part worths around zero and one positive and negative level. For provider, the physician had a positive part worth and the AI system had a negative part worth, while both the personalized AI and teledermatology had near-zero part worths. For waiting time, immediate results had a positive part worth and a 4-week wait had a negative part worth, while a 1-day and 1-week wait had similar near-zero part worths.

Table 2. Part-worth and relative importance values of the attributes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Part worth (95% CI)</th>
<th>Relative importance, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider</td>
<td></td>
<td>38.6 (35.3 to 41.6)</td>
</tr>
<tr>
<td>AI</td>
<td>−0.15 (−0.17 to −0.13)</td>
<td></td>
</tr>
<tr>
<td>Personalized AI</td>
<td>−0.06 (−0.08 to −0.04)</td>
<td></td>
</tr>
<tr>
<td>Physician treatment</td>
<td>0.21 (0.19 to 0.24)</td>
<td></td>
</tr>
<tr>
<td>Electronic consultation with physician (teledermatology)</td>
<td>0.005 (−0.01 to 0.02)</td>
<td></td>
</tr>
<tr>
<td>Costs for screening</td>
<td></td>
<td>31.6 (29.0 to 34.0)</td>
</tr>
<tr>
<td>0€ (completely covered by health insurance)</td>
<td>0.15 (0.13 to 0.16)</td>
<td></td>
</tr>
<tr>
<td>10€ copayment</td>
<td>0.06 (0.06 to 0.07)</td>
<td></td>
</tr>
<tr>
<td>25€ copayment</td>
<td>−0.03 (−0.04 to −0.03)</td>
<td></td>
</tr>
<tr>
<td>40€ copayment</td>
<td>−0.18 (−0.19 to −0.16)</td>
<td></td>
</tr>
<tr>
<td>Waiting time for diagnosis</td>
<td></td>
<td>29.8 (27.2 to 32.3)</td>
</tr>
<tr>
<td>Immediate</td>
<td>0.10 (0.09 to 0.11)</td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td>0.082 (0.07 to 0.09)</td>
<td></td>
</tr>
<tr>
<td>1 week</td>
<td>−0.004 (−0.01 to 0.003)</td>
<td></td>
</tr>
<tr>
<td>4 weeks</td>
<td>−0.18 (−0.20 to −0.17)</td>
<td></td>
</tr>
</tbody>
</table>

aAI: artificial intelligence.
b1€=US $1.09.

Figure 2 shows an overview of the relationships between sociodemographic characteristics and the relative importances to answer the second research question. We found a medium-sized positive relationship between age and provider. In addition, there were two nonsignificant trends. The first indicated an inverse relationship between age and the importance of costs and the second indicated an inverse relationship between income and the importance for costs. All other importances were not systematically related to sociodemographic variables (Table 3).
Figure 2. Relationship between relative importances and sociodemographic characteristics: (A) age, (B) gender, (C) education, (D) income, (E) employment status.
Table 3. Correlation coefficients (Spearman ρ) for the importance values.

<table>
<thead>
<tr>
<th>Sociodemographic characteristics</th>
<th>Relative importance</th>
<th>Provider</th>
<th>Costs for screening</th>
<th>Waiting time for diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Coefficient</td>
<td>0.21a</td>
<td>0.17</td>
<td>0.11a</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.02</td>
<td>.05</td>
<td>.25</td>
</tr>
<tr>
<td>Gender</td>
<td>Coefficient</td>
<td>–0.003</td>
<td>–0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.97</td>
<td>.71</td>
<td>.60</td>
</tr>
<tr>
<td>Education</td>
<td>Coefficient</td>
<td>–0.04</td>
<td>–0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.64</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td>Employment status</td>
<td>Coefficient</td>
<td>0.05</td>
<td>–0.11</td>
<td>–0.004</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.53</td>
<td>.22</td>
<td>.96</td>
</tr>
<tr>
<td>Income</td>
<td>Coefficient</td>
<td>0.02</td>
<td>–0.17</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.81</td>
<td>.07</td>
<td>.24</td>
</tr>
</tbody>
</table>

aPearson correlation coefficient.

Discussion

The aim of this study was to determine how important the attributes provider, costs, and waiting time are for users’ preference for skin cancer screening and to investigate whether sociodemographic characteristics, especially age, are systematically related to participants’ individual importances. We found that provider was as equally important a factor for participants’ decisions as cost and waiting time. While a physician was the most preferred level of this attribute, AI-based treatment was disliked and a personalized AI had the same value for participants as teledermatology. Concerning the relationship between sociodemographic characteristics and relative importances, we found that only age showed a reliable positive association to provider, in which younger participants place less importance on the provider than older participants. In the following, we discuss these findings in turn before discussing the limitations of the study and providing a general outlook.

Regarding the role of the provider in users’ decisions, other studies underline our results that patients exhibit hesitant behavior toward medical AI. Patients would rather not have a treatment than be examined by an AI system, even if the AI system shows the same or better accuracy as a physician [17]. However, the same study also found that patients prefer personalized AI over nonpersonalized AI. Similarly, earlier discrete-choice experiments [6] as well as surveys [33] found that patients prefer a trained medical professional to be involved in their skin cancer screening; 41% of respondents were open to using AI as a standalone system for skin cancer screening and 94% were open to using it as a support system for physicians [33]. Together, existing studies indicate that personalized AI and teledermatology are generally more accepted than nonpersonalized AI for skin cancer screening, while the physician remains the most preferred option.

Concerning the impact of age differences on the acceptance of AI in dermatology, our findings also support some earlier results [21-23]. Higher interest in using teledermatology [21,22] and in using skin cancer–related apps [23] was associated with younger age. The results of cross-sectional studies back up our findings from the choice-based conjoint analysis. Based on these trends, it is possible to imagine that the acceptance of AI in skin cancer screening will rise in the future due to the aging of digital natives and their increased acceptance of AI.

Regarding income and educational factors, our findings do not align with those of previous studies. Ghani et al [22] concluded that higher education attainment and a higher household income increased the interest in using teledermatology. Chang et al [21] came to similar conclusions, indicating that college graduates showed the greatest interest in teledermatology. In addition, Steeb et al [23] showed that a high educational level was associated with a positive attitude toward skin cancer–related apps. While we were not able to show significant relationships to income and educational background, the smaller sample size in this study compared to those of earlier studies might explain this inconsistency.

Previous studies also identified gender differences in the acceptance of AI in skin cancer screening. Chang et al [21] came to the conclusion that men are more likely to use teledermatology than women. Steeb at al [23] found similar results in which male gender was significantly associated with a positive attitude toward skin cancer–related apps. However, the gender difference that was reported in earlier studies was not visible in our data. Again, this might be a factor of sample size.
Several aspects must be considered in interpretation of our findings. First, the sample was not randomly selected but was based on a convenience sample. While a wide range of recruitment means were used, the results are likely not generalizable to the general public but rather more specific to highly educated young adults. Further research is needed with the target group. Although the sample size may not seem particularly large, sensitivity analysis showed that this sample size was in fact sufficient to detect a medium-sized correlation \( r = 0.28 \) with a power of 90% and error rate of 5%. Second, some participants contacted us about the meaning of the attribute waiting time because they were unsure whether this pertained to the waiting time for a diagnosis or the waiting time for an appointment. Future studies should make this distinction more explicit to study possible differential effects of these two types of waiting times.

Taken together, we believe that this study adds to the growing body of research using choice-based experiments to investigate the acceptance of AI in health contexts. This approach offers additional insights and is less susceptible to social desirability and other biases [35]. However, the choice-based conjoint analysis only allows studying a small number of potential attributes at a time [24]. Because we included personalized AI as a level for the attribute provider, our study adopted the findings of Longoni et al [17] that personalized AI increases patient acceptance. In addition, we examined factors that may have an impact on patients’ decision-making following the study of Snoswell et al [6].

For the future, it could be interesting to add “AI as a physician support system” to the choice set [33]. It might also be interesting to find out whether patients who perceived themselves as more individualized are less accepting of AI [17]. Additionally, it could be interesting to explore whether specialized knowledge about AI systems would increase patient acceptance [13] and which other factors might have an influence on patients’ acceptance. Ideally, this would not only rely on correlational evidence as used here but also on experimental evidence that shows how preferences and importances may be altered. The variables such as income and educational background cannot be manipulated easily. Nevertheless, we believe that the magnitude of these effects provides some benchmarks for future studies that aim to use experimental methods to alter preferences.

In summary, while there have been technological advances in the effectiveness of AI for supporting skin cancer screening and health care more generally, we believe that the true potential of AI systems can only be realized if patients’ needs and demands are taken into account.

Acknowledgments
This research was conducted within the framework of the project “SAIL: SustAInable Lifecycle of Intelligent Socio-Technical Systems,” funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia under grant NW21-059B. The sole responsibility for the content of this publication lies with the authors. Publication was funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under grant 490988677 and the University of Applied Sciences Bielefeld. Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability
The data sets generated and analyzed during this study are available in the zenodo repository [36].

Authors’ Contributions
IJ: study conceptualization, analysis, writing first draft, approval of final manuscript; MS: study conceptualization, analysis, approval of final manuscript; OW: study conceptualization, data collection, approval of final manuscript; GH: study conceptualization, data collection, analysis, approval of final manuscript.

Conflicts of Interest
None declared.

References


28. Orme B. Interpreting the results of conjoint analysis 2010. Sawtooth Software. 2010. URL: https://content.sawtoothsoftware.com/assets/3501e79f-3a9f-4d99-a0f5-1522525e7f3f [accessed 2023-08-08]


31. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]


Abbreviations

AI: artificial intelligence

©Inga Jagemann, Ole Wensing, Manuel Stegemann, Gerrit Hirschfeld. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Digital Health Needs and Preferences During Pregnancy and the Postpartum Period: Mixed Methods Study

Natalie Henrich1, MPH, PhD; Alison Brinson1,2, MSPH; Alyssa Arnold1, MPH; Hannah R Jahnke1, PhD

1Maven Clinic, New York, NY, United States
2Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Natalie Henrich, MPH, PhD
Maven Clinic
160 Varick St
New York, NY, 10013
United States
Phone: 1 212 457 1790
Email: natalie.henrich@mavenclinic.com

Abstract

Background: Digital health is increasingly used to meet the needs of perinatal people, with estimates of pregnancy-related internet use ranging from 90% to 97% of pregnant people. As digital health takes on greater importance during the perinatal period, it is essential that providers and developers of digital health content understand why perinatal people use these resources and the features that enhance their experience. However, gaps remain in understanding the content that is most helpful and how the platforms are navigated. Learning directly from perinatal people about their needs will help ensure alignment between perinatal needs and available content.

Objective: This formative study aims to identify the reasons why perinatal people use digital health resources; the features of the digital health platforms that are of greatest importance to them; and how these differ by perinatal stage (pregnancy vs postpartum), mental health conditions, parity, and demographics (race and ethnicity).

Methods: This mixed methods study used interviews; surveys; and secondary data on demographic, health, and pregnancy characteristics to identify the digital health needs and preferences of pregnant and postpartum people who used the Maven digital health platform in the United States during their pregnancy or postpartum period. The interviews informed the content of the surveys and provided additional insights and examples for interpreting the survey results. The surveys were used to collect data from a sample of Maven users, and the results were linked to the secondary data set. The interviews were thematically analyzed, and survey data were analyzed using descriptive statistics and stratified by parity, race, and mental health status.

Results: Overall, 13 people were interviewed (including n=4, 31% pilot interviews), and 147 pregnancy and 110 postpartum survey respondents completed the surveys and had linkable secondary data. Top reasons for using digital health resources during pregnancy were to (1) know what is normal or typical during pregnancy, (2) have access to a health care provider when needed, and (3) know how the baby is developing. Top reasons for postpartum use were to (1) help with breastfeeding, (2) know what normal baby development is, and (3) help with the baby’s health issues. Top platform features during pregnancy and the postpartum period were (1) credible and trustworthy information and providers, (2) nonjudgmental information and support, and (3) no cost to the user. In general, more reasons for using digital resources were identified as extremely important during pregnancy compared with post partum. The results showed minor variations across strata.

Conclusions: This formative research found minor differences in digital resource needs and preferences across user characteristics among perinatal people in the United States. Future work should examine whether there are variations in interests within topics by user characteristics, which may provide additional opportunities to better meet user needs.

*(JMIR Form Res 2024;8:e48960) doi:10.2196/48960*

**KEYWORDS**
digital health; perinatal; pregnancy; postpartum; interviews; survey; user needs; patient centered; mixed methods
Introduction

Background

Digital health, which refers to digital health applications, ecosystems, and platforms [1-4], is increasingly used to meet the needs of perinatal people [5-8], with estimates of pregnancy-related internet use ranging from 90% to 97% of pregnant people [9-13]. Digital health is expected to have growing importance because of its potential to provide health prevention, consultation, treatment, and management while simultaneously providing an opportunity to reduce costs and improve access to care and patient satisfaction, which together may lead to better informed and more engaged patients [1,4,14,15].

To date, research on perinatal digital health needs falls into 2 categories: self-report via surveys or qualitative data collection (interviews and focus groups) and content analysis of perinatal digital forums. Previous research has shown that access to perinatal digital health information provides reassurance and support [5,16,17]. Both types of studies have found that most people use digital resources during pregnancy to access information on maternal health (eg, pregnancy ailments, pain and complications, and health and nutrition during pregnancy) and during the postpartum period to access information on baby-related topics such as infant sleep and feeding (especially breastfeeding) [18,19]. Self-report studies have also indicated interest in childbirth, fetal and infant development, and infant illnesses, and studies of web-based forums have revealed interest in people and relationships [5,10,11,18-21]. The relative interest in these topics and the specific content within them vary throughout the perinatal period [17,19,20]. Research has also found that a positive experience using digital health resources is characterized by being easy to access and navigate, providing trustworthy and unbiased information, and being customized (eg, trackers and localized information) [4,22,23]. Perinatal digital health users specifically value an experience that provides information that is proactively delivered; offers immediate access; and provides information that is practical and concise, entertaining, and reassuring [5,24,25].

As digital health takes on greater importance during the perinatal period, it is essential that providers and developers of digital health content understand why perinatal people use these resources and what contributes to a positive user experience [26]. The information sought by perinatal people differs from the information that health care providers prioritize and offer. For example, postpartum people seek information on sleep, emotional changes, and breastfeeding, whereas providers focus on medical issues such as bleeding and infection [27,28]. Although there is a growing body of literature, there remain gaps in understanding of the content that is most helpful and how the resources are navigated [5]. There is a demand for more detailed analyses, including which digital resources are most helpful [16], with a particular need to understand what is valued during the postpartum period [18]. Research is needed on digital health needs and experience among perinatal people with mental health conditions, especially pre- and postpartum anxiety [20], and more insights are needed on barriers to and facilitators of using digital health, especially with newer technologies such as apps [4,20]. Learning directly from perinatal people about their needs will help ensure alignment between perinatal needs and available content [18,29]. The design of digital health content should be directly informed by individuals and their experiences. However, the research on patients’ user experience of digital health is limited, especially during the perinatal period [4].

Objectives

This mixed methods study, which used formative interviews and surveys, investigated the reasons why pregnant and postpartum people use digital health resources and what contributes to a positive user experience. Using a national sample of pregnant and postpartum users of Maven [30], a comprehensive reproductive and family digital health platform that includes patient education, care coordination, and provider services, this formative study identified how needs and preferences differ by perinatal stage (ie, pregnancy vs postpartum), mental health conditions, parity, and demographics (race and ethnicity). The results from this study can be used to inform a user-centered approach to developing the content and design of digital health resources and improve the alignment between resources and users. In recognition of the fact that patients have diverse gender identities, we have used gender-inclusive language throughout this paper. If a participant described themselves using gendered language (ie, “as a working mother...”), we used their terminology. When describing the results of previous research, we used the terminology used by the study investigators.

Methods

Study Design

In this mixed methods study, we identified the digital health needs and preferences of pregnant and postpartum people who used the Maven digital health platform during their pregnancy or postpartum period. The study used formative interviews to inform the content of surveys and provide additional insights and examples for interpreting the survey results. Surveys were used to collect data from a sample of Maven users that could be linked to a secondary data set of demographic, pregnancy, and health characteristics in the Maven platform (Figure 1).
Digital Platform Background

Maven was developed in 2014 as a digital platform to support perinatal people and their families. Users have a primary point of contact called a care advocate, an allied health professional (e.g., social worker or nurse) who works with the user to navigate the services and resources on the platform. Services and resources include web-based classes; educational materials; and internet-based appointments with a diverse team of health care providers, including obstetrician-gynecologists, mental health professionals, midwives, and doulas. Access to Maven is a sponsored benefit through an employer or the health plan of the user or their partner. The study investigators are Maven employees of the Maven clinical research team. When activating their Maven accounts, users consent to the use of their contact information by Maven for various reasons, including research. The investigators only have direct contact with Maven users or access to identifiable data within the scope of specific research projects, in accordance with institutional review board (IRB) approvals, and in concordance with Health Insurance Portability and Accountability Act regulations. Access to user data is minimized by requesting recruitment lists and platform-based data from the Maven product analytics team, which deidentifies the information before sharing it with the research team. To recruit Maven users for interviews and the survey, the recruitment parameters were sent to the product analytics team, which generated a list of deidentified IDs of users who met the parameters. This list of deidentified IDs was sent to the Maven team responsible for direct communication with users, which sent out recruitment emails (using email addresses linked to the deidentified IDs).

Formative Interviews

In September 2022 and October 2022, semistructured interviews with postpartum people were conducted virtually via Zoom (Zoom Video Communications) to identify the reasons why perinatal people use any digital health resources, not just Maven. To help encourage participants to think about digital use broadly, at the start of the interview, they were asked to list all the digital resources used during their pregnancy and the postpartum period, and throughout the interview, they were reminded to think about digital resources in general. The first part of the interview was open-ended, and participants were prompted not only to describe the reasons why they used digital resources but also to identify additional reasons for which they would like to use them. In the second part of the interview, participants were shown (via screen sharing on Zoom videoconferencing) a list of reasons why digital resources could be helpful during the perinatal period. Participants were asked whether they had used a digital resource for each individual reason or whether they would like to use a digital resource for each individual reason in the future. Finally, participants were asked about the importance of digital health resources being inclusive of their entire identity, which was explained as including anything that the participant considered part of their identity, such as race, gender, sexual identity, geography, social or economic status, and mental and physical health conditions, and to describe their experience with (lack of) inclusion.

Pilot interviews were conducted with 4 postpartum Maven staff members who had used Maven during pregnancy and the postpartum period and were post partum at the time of the interview. The interview guide was modified based on their feedback and the interviewer’s learnings from conducting the pilot interviews. Following piloting, interviews were conducted with 9 Maven users, all of whom were post partum so that they could speak to their digital needs and preferences throughout the perinatal period. Participants were recruited with an emphasis on including people from a range of ethnicities and races, health conditions, and parity. Participants were recruited via email. Interviews lasted approximately 1 hour, were conducted via Zoom, and were recorded. An interviewer and a notetaker were present, and transcripts were autogenerated on Zoom. Participants (Maven staff and nonstaff) consented to the interview, had the option of having the video on or off, and could decline to have the interview recorded (one participant...
declined, and we relied on detailed notes for that interview. The consent form was included in the recruitment email and reviewed at the start of the interview, and consent was provided verbally. Participants were given a US $50 Amazon gift card as a thank you for their participation.

The prevalence of each reason for using digital resources was calculated by counting the number of interviews (Maven staff and nonstaff) in which the reason was mentioned as important. The reasons that were mentioned as important in the most interviews were included in the survey.

Survey Development
In addition to the semistructured interviews, separate quantitative surveys were developed to inquire about digital resource use during pregnancy and the postpartum period. Through the qualitative analysis, it emerged that participants were describing 2 aspects of digital resource use: the information or support they were seeking and features of the digital health platforms. Consequently, the survey contained 2 sections, each addressing one of these aspects. To focus on common needs and remove outliers that were mentioned by only 1 or 2 participants, reasons for using digital resources and features affecting experience were excluded from the surveys if they were cited by less than one-third of the interview participants. In each section of the survey, participants indicated the importance of each item on a Likert scale that included a description of each option (a little important or doesn’t matter to me—I don’t care much about this, important—I would like to have this but I’m OK without it, very important—I care about this a lot but it’s not absolutely necessary, and extremely important—a “must have” ) and subsequently picked the 5 most important items (not ranked). We opted to allow participants to select up to 5 items as a means to balance the tension between the high number of reasons why pregnant and postpartum people are turning to digital resources and the need to prioritize to inform actionable decisions about what content to provide in these resources. To reduce the potential bias of basing responses on their use of Maven, the instructions in the survey reminded respondents to think holistically about digital health resources. The following is an example of the instructions from a section of the pregnancy version of the survey: “Thinking about the reasons you might use any digital resource during pregnancy, rate how much each of these reasons matter to you. Digital resources include websites, apps, videos, blogs, search engines, etc.”

The surveys were piloted with 4 Maven staff members who had used Maven during pregnancy and the postpartum period and were post partum at the time of piloting. The wording and format were modified based on their feedback. The revised surveys were piloted with 4 Maven nonstaff users. During piloting, we identified that nearly all items were selected as very important. To minimize ceiling effects, the response options were revised to provide more response options that reflected higher levels of importance and add definitions (eg, extremely important: a must have). We also opted to reverse the order of the options (with a little important or not important to me as the first option rather than the last). In addition, we included a section for selecting the 5 most important reasons for use and platform features, which would provide information on the relative importance of the items if we continued to have a ceiling effect.

On the basis of feedback from the staff and nonstaff member pilots, we added 2 items that had not emerged in the interviews: using digital resources to help with formula feeding and feeding solid foods. These were added as participants felt that the survey was not inclusive or relevant to the entire postpartum period if it only asked about breastfeeding.

The survey was built in SurveyMonkey (SurveyMonkey Inc) using a Health Insurance Portability and Accountability Act–compliant account.

Survey Administration
Survey respondents were recruited from current Maven users. To be eligible for the pregnancy survey, users had to be between 0 and 6 months post partum and have had a Maven account during pregnancy. To be eligible for the postpartum survey, users had to be 3 to 12 months post partum and have a Maven account during the postpartum period. Recruitment was via email, which included a link to the survey. Each survey link was associated with the user’s unique Maven identifier to link the survey responses to user demographic characteristics previously recorded on the Maven platform (described in the Secondary Data section). A reminder email was sent 1 week after the initial email to anyone who had not clicked through to the survey. The survey remained open for 2 weeks from the time of the initial email. A consent form appeared before the start of the survey, and consent had to be provided digitally to access the survey. Participants were entered into a draw for a chance to win 1 of 5 US $100 Amazon gift cards per survey (5 for the pregnancy survey and 5 for the postpartum survey). Data were collected from January 18 to 29, 2023.

Secondary Data
Maven users are encouraged but not required to fill out a questionnaire when they activate their account. The Maven pregnancy questionnaire collects data on physical and mental health, pregnancy conditions, demographics, and parity. The Maven postpartum questionnaire collects information on demographics, mode of delivery, and previous pregnancy conditions. For survey participants who completed the Maven questionnaires, we linked their survey responses with their Maven questionnaire data using their unique Maven identifier. Maven users consent to the use of their Maven questionnaires for research when they activate their accounts, and the research team does not have access to any identifying information.

Analyses
Survey Analysis
Respondents who completed the survey and a Maven questionnaire were included in the survey analysis. Figure 2 shows the pathway for inclusion in the analysis, from eligibility to be recruited to starting and completing the survey and completing a Maven questionnaire. The quantitative data from both the pregnancy and postpartum surveys were assessed primarily through the evaluation of descriptive statistics. Chi-square and Fisher exact tests were used to assess differences among pregnancy items, postpartum items, and user experience.
features by parity, presence of any mental health conditions, and race and ethnicity. A user was considered to have a mental health condition if they reported any of the following: a history of anxiety, depression, perinatal mood disorder, or high pregnancy-related anxiety. A history of anxiety or depression was assessed in the Maven questionnaires using the following question—“Do any of these conditions apply to you or did they in the past?”—with the selection of Anxiety or depression from a list of conditions. Experience of perinatal mood disorder was assessed using the following question—“Have you experienced any of the following during this pregnancy?”—with a selection of Perinatal mood disorder. Pregnancy-related anxiety was assessed on a 5-item Likert scale in response to the following question—“On a scale of 1-5, how anxious are you feeling about your pregnancy?”—with responses of 4 (very) or 5 (extremely) indicating the presence of pregnancy-related anxiety. For analyses stratified by race and ethnicity, we recategorized individuals who identified as Hispanic, Black, Asian or Pacific Islander, or multiracial into 1 category labeled non-White because of our small sample of people of racial and ethnic minorities. Consistent with the study’s aim of identifying differences in needs and preferences between respondents with these characteristics, P values were calculated to determine whether the differences were statistically significant (with results considered to be statistically significant if the P value was <.05). Statistical analyses were conducted using RStudio (Posit Software, PBC).

Figure 2. Flowchart of the pregnancy survey respondents (A) and postpartum survey respondents (B) included in the survey analysis.

Interview Analysis

The interviews were analyzed thematically using a framework analysis [31-36]. Themes were identified deductively based on the interview guide and inductively based on the messages that emerged from the interviews. Themes and their definitions were reviewed by all team members who participated in the interviews (as interviewers or notetakers) and were modified based on group discussion and consensus. The analysis table was populated based on the notes taken during the interviews and subsequent analysis discussions. Audio recordings and transcripts were referenced if the notes were unclear or lacked sufficient details to populate the analysis table. All entries in the analysis table were reviewed by NH, who attended all but one of the interviews. Any information that was missing or potentially miscategorized in the table was flagged for review by the person who entered the information and, if necessary, discussed to reach a consensus. To better understand the experiences and perspectives associated with the topic of each of the items in the quantitative surveys, the corresponding interview themes were summarized, and representative anecdotes or quotes were selected. All quotes were verified using the audio recordings. Quotes and anecdotes are cited using the interview participant number (eg, P1 refers to participant 1), and it is specified whether the participant was from the pilot interviews (eg, Pilot P1 refers to the first interview among the pilot participants).

Ethics Approval

The study protocol was approved by the WCG IRB (study 1338443). The use of secondary data was deemed exempt by the WCG IRB under 45 Code of Federal Regulations § 46.104(d)(4).

Results

Survey Sample Characteristics

In total, we had 147 pregnancy survey respondents and 110 postpartum survey respondents with complete Maven questionnaires (Table 1 and Figure 2). All survey responses

https://formative.jmir.org/2024/1/e48960 Jmir Form Res 2024 | vol. 8 | e48960 | p.2803 (page number not for citation purposes)
were collected within 48 hours of sending the initial recruitment email. The average age of the respondents was 32.8 (SD 4.6) years and 32.7 (SD 3.5) years for pregnancy and postpartum survey respondents, respectively. On the basis of those who reported their race and ethnicity, our sample was comparable with the overall US population in terms of the percentage of people identifying as non-Hispanic White (134/257, 52.1%) and multiracial (5/257, 1.9%), with an overrepresentation of people identifying as Asian (31/257, 12.1%) and an underrepresentation of people identifying as Black (4/257, 1.6%) or Hispanic (16/257, 6.2%) [37]. A total of 44.9% (66/147; pregnancy survey) and 51.8% (57/110; postpartum survey) of the respondents had household incomes of >US $100,000. Approximately 28% (72/257) of the sample reported a history of a mental health condition (Table 1).

### Table 1. Characteristics of respondents who took the pregnancy survey and the postpartum survey.

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Pregnancy survey respondents (n=147)</th>
<th>Postpartum survey respondents (n=110)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>32.8 (4.6)</td>
<td>32.7 (3.5)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8 (5.4)</td>
<td>8 (7.3)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>21 (14.3)</td>
<td>10 (9.1)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>3 (2)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Non-Hispanic multiracial or American Indian</td>
<td>3 (2)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>77 (52.4)</td>
<td>57 (51.8)</td>
</tr>
<tr>
<td>I prefer not to say</td>
<td>35 (23.8)</td>
<td>17 (15.5)</td>
</tr>
<tr>
<td><strong>BMI (kg/m²), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤18.5</td>
<td>4 (2.7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>18.5 to 24.9</td>
<td>76 (51.7)</td>
<td>49 (44.5)</td>
</tr>
<tr>
<td>25 to 29.9</td>
<td>30 (20.4)</td>
<td>15 (13.6)</td>
</tr>
<tr>
<td>≥30</td>
<td>29 (19.7)</td>
<td>15 (13.6)</td>
</tr>
<tr>
<td>Social Vulnerability Index (high), n (%)</td>
<td>10 (6.8)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Household income (US $), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20,000-34,999</td>
<td>1 (0.7)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>35,000-49,999</td>
<td>3 (2)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>7 (4.8)</td>
<td>3 (2.7)</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>18 (12.2)</td>
<td>9 (8.2)</td>
</tr>
<tr>
<td>≥100,000</td>
<td>66 (44.9)</td>
<td>57 (51.8)</td>
</tr>
<tr>
<td>I prefer not to say</td>
<td>27 (18.4)</td>
<td>10 (9.1)</td>
</tr>
<tr>
<td>Nulliparous (yes), n (%)</td>
<td>97 (66)</td>
<td>58 (80.6)</td>
</tr>
</tbody>
</table>

|Mental health history, n (%) |                                      |                                       |
| Anxiety                    | 33 (22.4)                            | 18 (25)                               |
| Depression                 | 14 (9.5)                             | 4 (5.6)                               |
| Perinatal mood disorder    | 1 (0.7)                              | 2 (2.8)                               |

**a** Variables with <15% of missing data: age, race, social vulnerability, BMI (pregnancy survey respondents), mental health history (pregnancy survey respondents), and parity (pregnancy survey respondents). Variables with 15% to 30% of missing data: household income and BMI (postpartum survey respondents). Variables with 30% to 40% of missing data: parity (postpartum survey respondents) and mental health history (postpartum survey respondents).

**b** Measured using the Centers for Disease Control and Prevention Social Vulnerability Index.

**c** A total of 65.5% (72/110) of postpartum survey respondents provided parity or mental health data. Therefore, the sample size for those variables is 72.

**Reasons for Use of Digital Resources During Pregnancy**

On the basis of the survey responses, the most important reasons for using digital resources during pregnancy were to know what is normal or typical during pregnancy, have access to a health care provider when needed, and know how the baby is changing and developing during pregnancy (Table 2).
### Table 2. Reasons identified as extremely important when using digital health resources during pregnancy by parity, mental health status, and race and ethnicity (N=147).

<table>
<thead>
<tr>
<th>Reasons for using digital health resources during pregnancy</th>
<th>Overall, n (%)</th>
<th>Parity</th>
<th>Mental health</th>
<th>Race and ethnicity&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>To know what is normal or typical during pregnancy</td>
<td>95 (64.6)</td>
<td>24 (48)</td>
<td>68 (64.8)</td>
<td>18 (51.4)</td>
</tr>
<tr>
<td>To have access to a health care provider when I need them</td>
<td>84 (57.1)</td>
<td>26 (52)</td>
<td>57 (54.3)</td>
<td>22 (62.9)</td>
</tr>
<tr>
<td>To know how my baby is changing and developing during pregnancy</td>
<td>76 (51.7)</td>
<td>20 (40)</td>
<td>58 (55.2)</td>
<td>18 (51.4)</td>
</tr>
<tr>
<td>To help me manage discomfort during pregnancy</td>
<td>61 (41.5)</td>
<td>19 (38)</td>
<td>44 (41.9)</td>
<td>18 (51.4)</td>
</tr>
<tr>
<td>To become more empowered to speak up for myself during pregnancy and labor</td>
<td>56 (38.1)</td>
<td>16 (32)</td>
<td>41 (39)</td>
<td>14 (40)</td>
</tr>
<tr>
<td>To explore my options for labor and delivery</td>
<td>55 (37.4)</td>
<td>14 (28)</td>
<td>42 (40)</td>
<td>14 (40)</td>
</tr>
<tr>
<td>To help me manage anxiety or depression</td>
<td>55 (37.4)</td>
<td>19 (38)</td>
<td>37 (35.2)</td>
<td>21 (60)</td>
</tr>
<tr>
<td>To help me manage my health problems (chronic, ongoing, or related to pregnancy)</td>
<td>45 (30.6)</td>
<td>17 (34)</td>
<td>32 (30.5)</td>
<td>18 (51.4)</td>
</tr>
<tr>
<td>To get additional information or clarification on things my doctor tells me</td>
<td>44 (29.9)</td>
<td>15 (30)</td>
<td>29 (29.9)</td>
<td>13 (37.1)</td>
</tr>
<tr>
<td>To know the size of my baby throughout pregnancy</td>
<td>35 (23.8)</td>
<td>14 (28)</td>
<td>21 (21.6)</td>
<td>13 (37.1)</td>
</tr>
<tr>
<td>To help me with healthy eating during pregnancy</td>
<td>32 (21.8)</td>
<td>7 (14)</td>
<td>25 (25.8)</td>
<td>11 (31.4)</td>
</tr>
<tr>
<td>To help me with being physically active during pregnancy</td>
<td>30 (20.4)</td>
<td>9 (18)</td>
<td>21 (21.6)</td>
<td>9 (25.7)</td>
</tr>
<tr>
<td>To get recommendations of things to buy to care for my baby or myself</td>
<td>25 (17)</td>
<td>7 (14)</td>
<td>18 (18.6)</td>
<td>9 (25.7)</td>
</tr>
<tr>
<td>To connect with a community of people who are in the same stage of pregnancy as me</td>
<td>17 (11.6)</td>
<td>6 (12)</td>
<td>11 (11.3)</td>
<td>10 (13)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Respondents who selected *I prefer not to say* for race and ethnicity were not included in this comparison.

In interviews, participants described the role that digital resources play in meeting these needs (Table 3), and they emphasized the importance of these resources during their first pregnancy (Pilot P1 and P2). Among the information they sought about what is normal, participants wanted to know what they would encounter during each phase of labor and delivery and were looking for information such as how long to expect to push. A participant described learning about the phases during a digital birth-planning appointment:

*The digital appointment with the birth planning specialist was very helpful to know what to expect at the hospital during birth, and actually I feel like they were more helpful than the in person doula that I hired...I remember them stepping through the birth process. You know, active labor and transition, the pushing and talking about the ring of fire.* [P5]
### Table 3. Illustrative quotes for using digital resources (pregnancy and postpartum interviews).

<table>
<thead>
<tr>
<th>Reasons for using digital health resources</th>
<th>Illustrative quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>During pregnancy</strong></td>
<td></td>
</tr>
<tr>
<td>To know the size of my baby throughout pregnancy</td>
<td>What is baby now today? Let’s see. I will also open the app up and see. So that was a good thing as well. So this is like ‘What vegetable is your baby right now?’ “ (P2)</td>
</tr>
<tr>
<td>To help me manage discomfort during pregnancy or postpartum</td>
<td>[The digital resource helped with] cracked and bleeding nipples” (P9).</td>
</tr>
<tr>
<td>To become more empowered to speak up for myself during pregnancy and labor</td>
<td>I read a ton of articles [online] because I felt like otherwise it was like the blind leading the blind. My husband had never done that before. I’m an only child like my mom was going to be there, but she hadn’t done it in thirty years. So it was kind of like, Okay, Who in this room knows what they’re doing? The answer was going to be nobody...So I read a lot just to know these are all the possible outcomes. This is how it might go. These are the things that I’m going to be asked, and I wanted to know, while I’m in a calm and sound state of mind, like is my answer Yes? Is my answer No? What are the implications if I say yes or no. And then I can communicate that to my partner and my mother. So if they needed to advocate for me [they could]” (P4).</td>
</tr>
<tr>
<td>To explore my options for labor and delivery</td>
<td>[Exploring my labor and delivery options] I would say [getting support on Maven] helped a little but that was more so because of me, and I didn’t look into it as much as I should have...Because, I thought I would have a natural birth. But I ended up having a C section, but just not knowing what to expect, and more so it was on the process of like what an induction is, and like how your body would look like afterwards, because I didn’t get to in my entire pregnancy. But all the drugs they pumped into you to induce you caused me to blow up after, and I just didn’t expect that, no one tells you, right? They just tell you to consent, and you say yes, so like all that, I wish I knew. But I don’t think that I looked into it as much, because I just didn’t know” (P8).</td>
</tr>
<tr>
<td><strong>During the postpartum period</strong></td>
<td></td>
</tr>
<tr>
<td>To help me with feeding my baby breast milk</td>
<td>I didn’t see lactation consulting in person until after birth, which was at the pediatricians office. So it’s nice to have that, to be able to talk to one [online] beforehand, to know what to prepare for. And then also I talked to one after as well, after giving birth. That was super helpful” (P5).</td>
</tr>
<tr>
<td>To help me with my baby’s health issues</td>
<td>[My baby] was very colicky, I took an [online] appointment and showed the doctor what was happening, so they also gave me a hint she might have a milk allergy, [and said to] check with your doctor and check the stools and all” (P2).</td>
</tr>
<tr>
<td><strong>During pregnancy or the postpartum period</strong></td>
<td></td>
</tr>
<tr>
<td>To know what is normal or typical during pregnancy or postpartum</td>
<td>[Getting digital help on my baby’s sleep] helped a lot, because I had a lot of questions about what’s normal, what’s not normal and they answered everything. It made me feel like I wasn’t doing something bad” (P8).</td>
</tr>
<tr>
<td>To know how my baby is changing and developing during or after pregnancy</td>
<td>I had like three apps on my phone to track [how my baby was changing] like every week” (P4).</td>
</tr>
<tr>
<td>To have access to a health care provider when I need them</td>
<td>It wasn’t just 8 to 5 that I could access someone...The sleep consultant was like you can just message me any time, and I’ll respond and help you real time with what you need. That’s something that you can’t get with any in person care provider” (P3).</td>
</tr>
<tr>
<td>To help me with being physically active during pregnancy or the postpartum period</td>
<td>I was trying to be physically active during pregnancy but then there were times where I stopped because of the spotting. They didn’t know what was going on. Then I found out it was okay. But then the last trimester it got tougher since I was tired a lot and having pain...It would have been nice to see what I could do, like modifications and exercises I could do around the pain I was having” (P5).</td>
</tr>
<tr>
<td>To help me with healthy eating during pregnancy or the postpartum period</td>
<td>Just offering ideas of foods, more nutritious foods to eat, healthier options, because you can look around on the Internet but just talking to a person who is, you know, that’s their specialty, and you can give them personal information about yourself so they can help you” (P9).</td>
</tr>
<tr>
<td>To help me manage anxiety or depression</td>
<td>I was in so much pain and I thought maybe [the baby will] be also in pain, or somehow it will affect the baby, because I’m not happy. So then I talked to [an online] doctor, and one of the doctors was very good and she talked to me more than the time of appointment, and she was very kind, and she said, I know how you are feeling, but she gave me so much support, emotional support, and she said, ‘You can come to talk to me anytime.’...So she then she assured me, ‘Everything is fine, and you don’t need to worry’” (P2).</td>
</tr>
<tr>
<td>To help me manage my health problems (chronic, ongoing, or related to pregnancy or postpartum)</td>
<td>I have arthritis and Hashimoto. So two autoimmune diseases. So when I got pregnant, I had a lot of concerns about that, and how that would affect my pregnancy, or vice versa” (P8).</td>
</tr>
<tr>
<td>To get recommendations of things to buy to care for my baby or myself</td>
<td>“I’m a working mother. At the time, you know, working while pregnant, and so I didn’t have time to spend hours and hours going down the rabbit hole of like, this stroller versus that stroller. Like just like, tell me what you recommend, and we’re moving on. So it felt like the research had been done for you” (P4).</td>
</tr>
</tbody>
</table>
Reasons for using digital health resources | Illustrative quote
--- | ---
To connect with a community of people who are in the same stage of pregnancy as me or who have a baby the same age as mine | “I used Reddit honestly before I was telling people I was pregnant...I really kind of used Reddit to ask people different questions early on in my pregnancy like ‘Oh, this happened to me. Is this normal?’ Like I said, it was in the very early stages of before we were talking to anybody, so I didn’t really have anyone to talk to’” (P7).
To get additional information or clarification on things my physician or pediatrician tells me | “The pediatrician’s really good for, like, is the baby’s body healthy, but they’re not necessarily specifically trained in infant nutrition and infant sleep. They kind of know enough to be dangerous. You know, I would ask some questions like, ‘How do I get my baby sleeping through the night longer?’ And they would say something really old school like ‘Put rice cereal in their bottle’; and I’m like well, one I’m not even giving a bottle, I’m breastfeeding and two, that’s like not really even something people say anymore. So it’s just like our pediatrician wasn’t super helpful. I totally trust him for everything else. But just not those services” (P3).

In interviews, participants talked about being able “to count on” accessing health care providers on the internet for urgent situations, such as a participant (P1) who used a digital health care provider when she did not feel her baby kick for a while and wanted to know whether she should go to the emergency room.

Almost all the participants talked about the importance of knowing how their baby was growing and developing during pregnancy. They described wanting to know about the development week by week and found that digital resources could easily provide this information.

In the survey, nulliparous respondents were more likely than parous respondents to find knowing what was normal (P=.004) and how their baby was developing (P=.06) to be extremely important (Table 2). Compared with non-Hispanic White respondents, respondents identifying as Hispanic or people of racial and ethnic minorities were more likely to report that digital health resources were extremely important for the management of anxiety or depression (P=.003) and chronic health problems (P=.002). Compared with respondents without a mental health history, respondents with a mental health history were more likely to find it important to receive web-based help with healthy eating during pregnancy (P=.04; Table 2).

Reasons for Use of Digital Resources During the Postpartum Period

On the basis of survey responses, during the postpartum period, the most important reasons for using digital resources were to help with feeding the baby breast milk, know what is normal baby development (eg, developmental milestones), help with the baby’s health issues, and help with feeding the baby solid foods (Table 4).

The reasons for using digital resources during the postpartum period were described in interviews (Table 3). When asked specifically about the use of digital health resources for help with feeding, interview participants primarily spoke about support for breastfeeding. In particular, participants found it valuable to connect with a lactation consultant during pregnancy to prepare for breastfeeding and during the postpartum period for tips that facilitated feeding:

> I didn’t see lactation consulting in person until after birth, which was at the pediatrician’s office. So it’s nice to have that, to be able to talk to one [online] beforehand, to know what to prepare for. And then also I talked to one after as well, after giving birth. That was super helpful. Just things I could do for milk supply. And then the one major helpful thing that I learned is about the pump parts where I can just store them in the fridge, and I don’t have to wash it after every time, because it was a hassle, too. [P5]

In interviews, participants described using digital resources to know what to expect about life with a baby and to find out whether the things they encountered were normal, such as struggling with the baby’s sleep schedule, and what it is really like to breastfeed. Participants voiced frustration that this type of information was not provided by in-person providers, so they looked for answers on the internet:

> They don’t tell you anything [at the hospital]. A big one for me was breastfeeding, like just how complex it is. There’s just so many things in the hospital they don’t tell you anything, either. They just ask you like, “Are you going to breastfeed, yes or no,” and you have to decide on the spot, and, like you don’t know anything about this. After that, coming home, I definitely did research on [a digital platform] and the resources that they had for help. [P8]

Participants also wanted digital resources to provide information on what is normal for babies as they grow, including what developmental milestones to expect at each age. A participant described using an app to know what to expect as her baby changed:

> [There] is an app that when you put in the baby’s due date, this helps you to see your baby’s development, like milestone-wise. So week by week, you know some of the things to look for. And mainly I use it to see like, when are one of the cranky times coming up because they’re going through a milestone leap where they’re learning, you know lots of new things. [P9]

In interviews, participants provided anecdotes about the value of digital resources for helping with the baby’s health issues, especially the ability to access support when it is needed. A new parent described meeting with a digital sleep coach in the middle of the night when her baby would not sleep, a service she said she could not obtain from her in-person providers as they only handled emergencies after hours:

> It wasn’t just 8 to 5 that I could access someone... The sleep consultant was like you can just message me any time, and I’ll respond and help you real time with
what you need. That’s something that you can’t get with any in person care provider. If you call their after hours number it’s because it’s an emergency right? Not because you are just, like, wondering why your baby is fussy, or awake window, or, you know whatever. [P3]

In the survey, nulliparous respondents were more likely than parous respondents to find it extremely important that digital resources help with feeding their baby solid food ($P=.01$) and with their own health problems ($P=.06$; Table 4). Respondents identifying as Hispanic or people of racial and ethnic minorities were more likely than non-Hispanic White respondents to report that it is extremely important that digital resources can be used to connect them with a community of people who have a baby of the same age ($P=.02$) and to obtain recommendations of things to buy to care for their baby or themselves ($P=.08$). Respondents with a mental health history were more likely than those without a mental health history to find it important to obtain web-based help with breastfeeding their baby ($P=.01$; Table 4).

Table 4. Reasons identified as extremely important when using digital health resources during the postpartum period by parity, mental health status, and race and ethnicity (N=110)$^a$.

<table>
<thead>
<tr>
<th>Reasons for using digital health resources during the postpartum period</th>
<th>Overall, n (%)</th>
<th>Parity</th>
<th>Mental health</th>
<th>Race and ethnicity$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Parous (n=14), n (%)</td>
<td>Nulliparous (n=58), n (%)</td>
<td>$P$ value</td>
</tr>
<tr>
<td>To help me with feeding my baby breast milk</td>
<td>64 (58.2)</td>
<td>6 (42.9)</td>
<td>37 (63.8)</td>
<td>.12</td>
</tr>
<tr>
<td>To know what is normal baby development (milestones)</td>
<td>54 (49.1)</td>
<td>7 (50)</td>
<td>30 (51.7)</td>
<td>.89</td>
</tr>
<tr>
<td>To help me with my baby’s health issues</td>
<td>54 (49.1)</td>
<td>6 (42.9)</td>
<td>32 (55.2)</td>
<td>.38</td>
</tr>
<tr>
<td>To help me with feeding my baby solid foods</td>
<td>53 (48.2)</td>
<td>4 (28.6)</td>
<td>36 (62.1)</td>
<td>.01</td>
</tr>
<tr>
<td>To help me manage anxiety or depression</td>
<td>45 (40.9)</td>
<td>5 (35.7)</td>
<td>22 (37.9)</td>
<td>.77</td>
</tr>
<tr>
<td>To help me with feeding my baby formula</td>
<td>30 (27.3)</td>
<td>2 (14.3)</td>
<td>18 (31)</td>
<td>.20</td>
</tr>
<tr>
<td>To get additional information or clarification on things my physician or pediatrician tells me</td>
<td>29 (26.4)</td>
<td>4 (28.6)</td>
<td>14 (24.1)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>To help me with my health problems (chronic, ongoing, or related to pregnancy or postpartum)</td>
<td>22 (20)</td>
<td>0 (0)</td>
<td>12 (20.7)</td>
<td>.06</td>
</tr>
<tr>
<td>To help me with being physically active after my baby is born (postpartum exercise)</td>
<td>18 (16.4)</td>
<td>4 (28.6)</td>
<td>10 (17.2)</td>
<td>.47</td>
</tr>
<tr>
<td>To help me with healthy eating after my baby is born</td>
<td>16 (14.5)</td>
<td>2 (14.3)</td>
<td>10 (17.2)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>To help me with losing weight after my baby is born</td>
<td>12 (10.9)</td>
<td>1 (7.1)</td>
<td>7 (12.1)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>To obtain recommendations of things to buy to care for my baby or myself</td>
<td>11 (10)</td>
<td>1 (7.1)</td>
<td>9 (15.5)</td>
<td>.67</td>
</tr>
<tr>
<td>To connect with a community of people who have a baby the same age as mine</td>
<td>9 (8.2)</td>
<td>0 (0)</td>
<td>6 (10.3)</td>
<td>.33</td>
</tr>
</tbody>
</table>

$^a$A total of 65.5% (72/110) of postpartum survey respondents provided parity or mental health data.

$^b$Respondents who selected “I prefer not to say” for race and ethnicity were not included in this comparison.
The reasons for using digital resources that the fewest survey respondents selected as extremely important during the perinatal period were uses related to general wellness (ie, healthy eating, exercise, and losing weight), connecting with a community in the same perinatal stage, and obtaining recommendations of things to buy to care for their baby or themselves. Notably, these lowest-ranked reasons were still selected by 8.2% (9/110) to 22% (32/147) of the respondents (Tables 2 and 4).

In contrast to the surveys, help with healthy eating was one of the topics that the most interview participants identified as a reason for using digital resources, although it was usually in response to specifically being asked about it in the structured part of the interview. When discussing the use of digital resources for healthy eating, participants tended to refer to access to personalized information such as food recommendations that fit with their dietary restrictions. A participant described why tailored information would be helpful to her on a digital health platform:

Because if I had dietary restrictions I’m going to have to go searching more for that information on my own. [P9]

Importance of Digital Health Platform Features During Pregnancy and the Postpartum Period

The 3 most important features of digital health platforms identified through both the pregnancy and postpartum surveys were credible and trustworthy information and providers, digital resources that are free to the user, and nonjudgmental information and support. Although access to appointments (ie, fast access and access at convenient times) was not among the top 3 most important features, approximately half of the respondents felt that it was extremely important (Table 5). See Multimedia Appendices 1 and 2 for digital health platform features identified as extremely important by parity, mental health status, and race and ethnicity during pregnancy and the postpartum period, respectively.

<table>
<thead>
<tr>
<th>Digital health platform features</th>
<th>Pregnancy survey respondents (n=147), n (%)</th>
<th>Postpartum survey respondents (n=110), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credible and trustworthy information and providers</td>
<td>117 (79.6)</td>
<td>82 (74.5)</td>
</tr>
<tr>
<td>Nonjudgmental information and support</td>
<td>97 (66)</td>
<td>58 (52.7)</td>
</tr>
<tr>
<td>Digital resources that are free to me</td>
<td>89 (60.5)</td>
<td>60 (54.5)</td>
</tr>
<tr>
<td>Fast access to appointments</td>
<td>75 (51)</td>
<td>52 (47.3)</td>
</tr>
<tr>
<td>Easy to find information; easy to navigate</td>
<td>74 (50.3)</td>
<td>53 (48.2)</td>
</tr>
<tr>
<td>Access to appointments at convenient times</td>
<td>72 (49)</td>
<td>50 (45.5)</td>
</tr>
<tr>
<td>Information that is actionable (specific recommendations for what to do)</td>
<td>64 (43.5)</td>
<td>57 (51.8)</td>
</tr>
<tr>
<td>Receive fast responses to my digital messages</td>
<td>59 (40.1)</td>
<td>40 (36.4)</td>
</tr>
<tr>
<td>Resources that are specific to my needs (personalized)</td>
<td>49 (33.3)</td>
<td>33 (30)</td>
</tr>
<tr>
<td>Access to a lot of information on each topic (depth of information)</td>
<td>36 (24.5)</td>
<td>22 (20)</td>
</tr>
<tr>
<td>Access to information on a lot of topics (breadth of topics)</td>
<td>35 (23.8)</td>
<td>22 (20)</td>
</tr>
<tr>
<td>Consistent care or support from the same people over time on the digital platform</td>
<td>31 (21.1)</td>
<td>33 (30)</td>
</tr>
<tr>
<td>Proactive outreach from the digital resource (pushes content to me; provider reaches out to me)</td>
<td>26 (17.7)</td>
<td>12 (10.9)</td>
</tr>
<tr>
<td>Care or content that fits with my culture and identity</td>
<td>26 (17.7)</td>
<td>14 (12.7)</td>
</tr>
</tbody>
</table>

Interview participants described why the features of digital health resources were important (Textbox 1). When looking to interact with content that was trustworthy, participants mentioned a range of ways in which they determined whether digital content was credible. A couple of participants explained that they trusted the content if it was consistent with the information they received from their in-person providers. This was done by taking the information from the digital content to the in-person provider for verification or by seeing that the content was consistent with what they had already been told:

Whatever I heard from the [web-based] doctor, it was matching with my experience and my other doctor as well...it builds trust since my doctor is saying the same thing. [P2]

So it was like things that the [web-based] practitioner might have suggested or recommended that I would then mention at my doctor’s appointment...kind of to get a check box from my regular doctor, but it was always like, she agreed with what had been communicated. [P4]

https://formative.jmir.org/2024/1/e48960

JMIR Form Res 2024 | vol. 8 | e48960 | p.2809

XSL-FO RenderX
**Box 1.** Illustrative quotes regarding features of digital health resources (pregnancy and postpartum interviews).

<table>
<thead>
<tr>
<th>Feature Description</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Credible and trustworthy information and providers</strong></td>
<td>“Sleep schedules, again, with so much information online, you get really caught up in, like, ‘Am I doing the right thing?’ especially in the beginning. I remember one thing, I was struggling with my baby’s sleep schedule. And so I went to this sleep expert from [my digital platform], and they said that’s normal, like, a baby shouldn’t really have a sleep schedule until six months. They’ll never have one, and it’s like everything online, says the opposite! So just like having information that is accurate and like coming from experts [really helped]” (P8).</td>
</tr>
<tr>
<td><strong>Digital resources that are free to me</strong></td>
<td>“But because this service was free to me I felt like ‘well, it’s there, and I might as well use it,’ so I’ve taken advantage of it” (P3).</td>
</tr>
<tr>
<td><strong>Nonjudgmental information and support</strong></td>
<td>“And these platforms really help because, one, you’re able to speak to someone different that may not even have your medical history. They don’t have your medical history in front of you. So maybe it’s the medical system, health care system thinking or judging you based on your history, right, rather than just listening to this specific one time incident. So maybe they’re less biased ‘cause they just have the one thing they’re looking at” (P8).</td>
</tr>
<tr>
<td><strong>Fast access to appointments</strong></td>
<td>“If I’m struggling with a situation that I think is super important with the baby like the baby has a cold and has a runny nose, fever, I’m gonna want whoever is available next” (P9).</td>
</tr>
<tr>
<td><strong>Easy to find information; easy to navigate</strong></td>
<td>“[My app] did not help [for finding information on baby’s development] because the information, I guess I didn’t seek it out, but I also didn’t see it readily available” (P9).</td>
</tr>
<tr>
<td><strong>Access to appointments at convenient times</strong></td>
<td>“Especially for a new mom, if you’re trying to work around is my baby napping or not, and that’s not always predictable, just to be able to quickly touch base with someone. I used it during pregnancy as well, so I feel it kind of rounds out the care you are receiving from your in person doctor” (P4).</td>
</tr>
<tr>
<td><strong>Information that is actionable (specific recommendations for what to do)</strong></td>
<td>“In my state it is legal to end the pregnancy, right? So [the online provider] gave me the support that if you want to go for this, this will happen, or if you go, these are the steps you need to follow, and all...She was telling me all the steps, what I need to do, and what I not to do, and she gave me a support for keeping this pregnancy” (P2).</td>
</tr>
<tr>
<td><strong>Receive fast responses to my digital messages</strong></td>
<td>“When I messaged, it wasn’t always [the same person] that responded, it was someone else. But it’s nice to have that, too. Like, someone quickly getting back to me. So I didn’t mind that” (P5).</td>
</tr>
<tr>
<td><strong>Resources that are specific to my needs (personalized)</strong></td>
<td>“Because if I had dietary restrictions I’m going to have to go searching more for that information on my own. [And it would be helpful to] come up with a list of foods that would be specific to me, because we didn’t do that. I realized that may be asking a lot of someone. But that’s just what I see would be super helpful” (P9).</td>
</tr>
<tr>
<td><strong>Access to a lot of information on each topic (depth of information)</strong></td>
<td>“I like the options when I went in, and I would read through an article. It’s like, Oh, here’s some other related stuff to this. If you’re still interested, I really like that because there was a couple of times with a few articles where, like I want to know more about this, and so I would keep clicking and read a little bit more” (P6).</td>
</tr>
<tr>
<td><strong>Access to information on a lot of topics (breadth of topics)</strong></td>
<td>“[It would help if the platform had], like, a more robust portfolio of different like classes and articles” (P4).</td>
</tr>
<tr>
<td><strong>Consistent care or support from the same people over time on the digital platform</strong></td>
<td>“The sleep coach that I am talking to now. She’s the same one I’ve talked to every single time...[It’s important to have continuity], especially just like to not have to answer the same questions over and over again, like they just kind of know who I am, looking for me, my style, and just more get to the care rather than have to do the basic information each session” (P3).</td>
</tr>
<tr>
<td><strong>Proactive outreach from the digital resource (pushes content to me; provider reaches out to me)</strong></td>
<td>“One thing I like about the Baby Center or a lot of the articles I get from other providers is that they’re just in the email, and I don’t have to go anywhere” (P7).</td>
</tr>
</tbody>
</table>
Others said that they trusted digital content that was reviewed by medical professionals (Pilot P3) or had a “provider stamp of approval” (P9). A couple of participants were looking for digital resources that provided information from the people who conducted the research (P8) and that included “the latest research and peer reviewed evaluation” (P3).

In interviews, participants explained that access to free digital health services enables users to consume more health resources and spend more time interacting with health care providers than during in-person appointments:

> [When there are] no limit on appointments...suppose I forgot something and I want another appointment and I can do it today. I don’t have to wait for another appointment. [P2]

They also accessed types of resources that they valued but would not use if there was an out-of-pocket cost:

> So I think I just wouldn’t have gotten what I needed [in-person] because I have two other children, and I didn’t [get what I needed] with them. Just, like you know, Google things or Pinterest, or talk to friends, and that’s it. But because this service was free to me I felt like ‘well, it’s there, and I might as well use it,’” so I’ve taken advantage of it. And so things like a sleep consultant I probably wouldn’t have paid for. But it was just so nice to have a sounding board...So yeah, I think just having services there that I wouldn’t have made the effort to use before, but because they’re just sitting there on my phone, I might as well, and it’s free to me. [P3]

Participants highlighted that free digital resources are particularly important as the cost of in-person services is very high and they were looking for a cheaper alternative:

> Free resources—that’s big, since in person visits are so expensive. [P5]

Interview participants also wanted to interact with digital health content and people who provided information and support without judgment. A participant explained that it is easier to receive nonjudgmental care from a digital health care provider as the person only knows the information that is being shared in the moment and cannot be biased by previous interactions or conditions. Another participant explained that she did not want to discuss her baby’s sleep issues with her in-person pediatrician as she was worried that the physician would judge her as a bad mother. It felt safer to explore the problem with the web-based provider, with whom she did not have an ongoing relationship (Pilot P4).

Interview participants shared anecdotes illustrating the accessibility of web-based providers and described connecting with these providers at night and on weekends or around their work or family schedules:

> Whereas in person, they’re always busy. So there’s some waiting time...If I had concerns, I want to consult with somebody...[it’s] very helpful where I can just find someone to talk to within the week, sometimes even the same day...Being able to get in touch with someone quickly during pregnancy and even postpartum has been super helpful. [P5]

> There was one time where I had a concern and it was 2AM and I reached out to a practitioner over the phone, it might have been with Maven or it might have been with Telehealth, and [I had] flexibility with getting help at abnormal times of day. [P8]

In general, the importance of digital health platform features was consistent by race, parity, and mental health. Differences emerged for free resources, which were more important to respondents without a mental health history compared with those with a mental health history (P = .02), and consistent care from the same people over time on the digital platform (P = .01) and access to information on a lot of topics (P = .02) were more important for Hispanic and people of racial and ethnic minorities respondents than for non-Hispanic White respondents.

For both pregnant and postpartum survey respondents, the 2 least important features were care or content that fit with their culture and identity and proactive outreach from the digital resources. These features were selected as extremely important by less than 20% of respondents. In interviews, participants described experiences on digital health platforms that were enhanced by resources and support that were respectful of and relevant to their culture or identity. When specifically asked about the importance of digital health resources that “take their whole identity into account,” participants consistently interpreted this to be about race, ethnicity, and gender or sexual identity, and participants from non–historically marginalized populations tended to report that inclusion was not relevant for them. This emerged in comments such as “Personally, I’m a woman, heterosexual, so I fit in fine, more than fine” (Pilot P1) and “[Health care is] not hard for me to navigate as a straight, White woman” (Pilot P3). However, as they shared anecdotes about their use of digital health platforms, they described experiences that were meaningfully inclusive (or noninclusive) related to other characteristics, such as their lifestyle (eg, family structure, work choices, and how they cared for their child) or religious practices. A participant appreciated that a nutritionist took the time to accommodate the fact that she kept kosher and “went above and beyond to research kosher supplements and make sure they were appropriate” for her to take (Pilot P2), and a working mother described her appreciation for resources that recognized her more limited availability to be with her children.
I’m probably in the majority, for most of the like, you know, bubbles you would fill in of race, gender, like I’m not a minority in any of those, but I think that everyone was just very respectful...maybe because they got to just like learn a little bit about me. [P3]

People of racial and ethnic minorities participants appreciated when there were racially diverse providers on the platform. Even though the participants did not necessarily choose to interact with a member of their ethnicity or race, they valued the presence of diversity:

I felt included throughout the process. I think even seeing the diversity of the people you can choose for a specific topic also helps just, like, feel more included, like just her background, I am Mexican American. So I was born in the US but Mexican background. And so just seeing the diversity on the team was important to me. [P8]

**Discussion**

**Principal Findings in Relation to Prior Research**

This mixed methods study found that the reasons why perinatal people use digital resources are consistent with previous research. During pregnancy, our survey and interview participants were highly interested in fetal development [10,11,16,18,20,21], reassurance about what is normal [5,16,17,38], and childbirth [10,11,20,21,39], although in our study, the latter topic was not among the top reasons for use. Unlike other studies, we found less interest in health and nutrition during pregnancy. During the postpartum period, our results were concordant with those of previous research with regard to users having a high interest in feeding, especially breastfeeding; developmental milestones; and the baby’s health [18,19,21]. Although interest in infant sleep is also common in the literature [18,19], we did not explicitly ask about this. However, this need is consistent with our overall finding that respondents were looking for information about caring for their baby and with anecdotes in our interviews that centered on infant sleep when participants discussed wanting to know what is normal, reaching out to digital providers at night, and receiving nonjudgmental support. We found an especially high interest in using digital resources for breastfeeding help among respondents with a mental health history (17/20, 85%), which is consistent with studies showing an association between maternal anxiety and breastfeeding outcomes [40-42]. Aligning with findings from content analyses of perinatal forums, top interests included baby-related topics and pain during pregnancy [17,19], but we found less interest in using digital resources for the health of the birthing person, and we did not ask about relationships (a common topic in web-based perinatal forums [17,19]) as this did not emerge as a key interest during the formative interviews. In addition to finding many of the same known reasons for using digital resources, we found that using these resources to help manage anxiety or depression was commonly selected as extremely important (55/147, 37.4% during pregnancy and 45/110, 40.9% during the postpartum period), especially for people of racial and ethnic minorities during pregnancy (21/35, 60%).

Our survey identified that, during the postpartum period, reasons for using digital resources related to the health of the birthing person were ranked lower than reasons related to obtaining information or help about the baby. This aligns with findings that mothers tend to focus on their babies and deprioritize their own needs [43]. However, there are physical and emotional challenges during the fourth trimester that often go unaddressed and can negatively affect maternal and child outcomes [44,45]. Contributing to underuse is a misalignment between the issues being addressed by providers and the birthing person’s needs, discomfort with discussing “embarrassing” issues with providers and feeling judged by them, and not knowing where to access information [43,46,47]. Digital health resources provide an opportunity for birthing people to be more in control of when, how, and for which needs they access self-care resources. Even though self-care may not be the primary priority for using digital health resources, making these supports available is valuable for those times when the birthing person is ready and interested in using them.

Emerging from both our surveys and interviews was the preference for using digital health resources to access providers. The characteristics that make digitally based providers helpful are differentiators from in-person care, with providers on platforms being reported as spending more time, focusing on issues and questions that matter to the user, and providing support quickly and conveniently [48-50]. Although study participants were explicitly asked about what is important to them when using digital health resources in general, they all had access to Maven, which offers access to a wide range of provider services. Study participants may have been more likely to express the importance of digitally based providers as they had the opportunity to use them during their perinatal stage. Users of digital health services who do not have access to digitally based providers may not identify them as important as they have not had the opportunity to experience their value. Although our results may not replicate those among digital health users who have not had access to these providers, this does not mean that digitally based providers are not an important part of what should be offered in digital health. Rather, it is an opportunity to meet a need that people may not realize they have.

Generally, the needs and preferences for perinatal digital resource use did not show much variation by perinatal phase (pregnancy vs postpartum), parity, demographics, or health characteristics. The least important reasons for use were the same during pregnancy and the postpartum period: help with healthy eating and being physically active (and, for the postpartum period, help with losing weight), connecting with a community of people in the same perinatal stage, and obtaining recommendations of things to buy to care for the birthing person or baby. Platform features were ranked in generally the same order during pregnancy and the postpartum period. However, there seems to be a difference in the overall importance of information during pregnancy versus the postpartum period, with reasons for using digital resources being selected as extremely important by a greater percentage of respondents during pregnancy than postpartum. During pregnancy, the only reasons for using digital resources that differed by race were
non-pregnancy-related (help with managing mental health and physical health problems), which may reflect gaps in in-person care for people of racial and ethnic minorities [51,52]. In interviews, we found a difference in participants who said that care or content that fits with their cultural, racial, or ethnic identity was important to them, with people of racial and ethnic minorities describing the value of being able to access providers with a shared identity or who provided culturally relevant support, whereas non-Hispanic White participants tended to say that this was not relevant to them. We were surprised that we did not find this difference in the surveys. A possible reason for this discrepancy is that our survey analysis focused on “extremely important” responses, and it may be that culturally relevant support is important but not extremely important. This would align with comments in the interviews about liking that providers of the same race were available even if they did not choose to use these providers.

This study included a national sample from the United States. Although our sample underrepresented people who identify as Black and Hispanic and lower-income people, it includes a significant portion of sustained users of digital resources (based on their completion of questionnaires on the Maven platform at multiple time points throughout pregnancy and the postpartum period), which is valuable as they represent people who will often use digital resources and have the potential to be significantly helped and supported by them. However, future research should also explore these topics among lower-frequency users to identify and understand barriers or deficiencies in the resources that may be contributing to reduced use.

This study adds to existing research by linking survey responses to other pregnancy, health, and demographic data about the respondents. Psychological states and mental health, especially perinatal anxiety, in relation to digital use are understudied [20], as are race and ethnicity. This formative research found only minor differences in digital resource needs and preferences across several user characteristics, including perinatal phase, parity, demographics, and mental and chronic health conditions. If this finding holds up in additional research, it has implications for the design of digital resources as it suggests that it may be appropriate to cover the same high-level perinatal topics for users across these groups. However, additional research is needed to better understand whether the content within these topics needs to be tailored to different populations. For example, future research could find that, although it is appropriate to include resources on childbirth for all users, there may be a need for more information about birthing centers in resources targeting Black pregnant people. Additional insights that are needed to improve the relevance and impact of these resources include understanding the use of the internet by nonbirthing partners [20] and the impact of use on clinical outcomes and user satisfaction [51]. There is also an opportunity to develop a standardized set of methods and metrics for assessing perinatal digital resources [4].

Limitations
This study has some limitations. First, the generalizability of these results is limited as all survey and interview participants had access to private insurance, the sample had a relatively high income, and Hispanic and Black respondents were underrepresented. Second, there was a high amount of missingness within our race and ethnicity data from respondents selecting that they prefer not to provide this information (17/110, 15.5% and 35/147, 23.8% for postpartum and pregnancy survey respondents, respectively). This missingness may limit the reliability of our results regarding differences in digital needs by race. Finally, not all survey respondents completed the Maven questionnaires, and those who did not complete at least one Maven questionnaire were excluded from the analyses of the survey. This potentially biased the sample toward heavier users of digital resources, although this creates an opportunity to understand use needs among this segment of users.

Data Availability
Owing to the sensitive nature of the questions asked in this study, respondents were assured that raw data would remain confidential but may be shared upon request.

Conflicts of Interest
NH, AB, AA, and HRJ are employed by Maven Clinic. NH, AA, and HRJ have equity in Maven Clinic.

Multimedia Appendix 1
Digital health platform features identified as extremely important during pregnancy by parity, mental health status, race, and ethnicity (N=147).

[PDF File (Adobe PDF File), 50 KB - formative_v8i1e48960_app1.pdf ]

Multimedia Appendix 2
Digital health platform features identified as extremely important during the postpartum period by parity, mental health status, race, and ethnicity (N=110).

[PDF File (Adobe PDF File), 49 KB - formative_v8i1e48960_app2.pdf ]

References


Abbreviations

IRB: institutional review board

©Natalie Henrich, Alison Brinson, Alyssa Arnold, Hannah R Jahnke. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Acceptability and Feasibility of a Smartphone-Based Real-Time Assessment of Suicide Among Black Men: Mixed Methods Pilot Study

Leslie B Adams, MPH, PhD; Thomasina Watts, MSPH; Aubrey Devinney, BA; Emily E Haroz, PhD; Johannes Thrul, PhD; Jasmin Brooks Stephens, MA, LPA; Mia N Campbell, MHS; Denis Antoine, MD; Benjamin Lê Cook, PhD; Sean Joe, PhD; Roland J Thorpe Jr, PhD

1Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
2Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia
3Department of Psychology, University of Houston, Houston, TX, United States
4Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, United States
5Department of Psychiatry, Cambridge Health Alliance, Cambridge, MA, United States
6George Warren Brown School of Social Work, Washington University at St. Louis, St. Louis, MO, United States
7Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Leslie B Adams, MPH, PhD
Department of Mental Health
Johns Hopkins Bloomberg School of Public Health
624 N. Broadway
Baltimore, MD, 21205
United States
Phone: 1 410 955 1906
Email: ladams36@jhu.edu

Abstract

Background: Suicide rates in the United States have increased recently among Black men. To address this public health crisis, smartphone-based ecological momentary assessment (EMA) platforms are a promising way to collect dynamic, real-time data that can help improve suicide prevention efforts. Despite the promise of this methodology, little is known about its suitability in detecting experiences related to suicidal thoughts and behavior (STB) among Black men.

Objective: This study aims to clarify the acceptability and feasibility of using smartphone-based EMA through a pilot study that assesses the user experience among Black men.

Methods: We recruited Black men aged 18 years and older using the MyChart patient portal messaging (the patient-facing side of the Epic electronic medical record system) or outpatient provider referrals. Eligible participants self-identified as Black men with a previous history of STB and ownership of an Android or iOS smartphone. Eligible participants completed a 7-day smartphone-based EMA study. They received a prompt 4 times per day to complete a brief survey detailing their STB, as well as proximal risk factors, such as depression, social isolation, and feeling like a burden to others. At the conclusion of each day, participants also received a daily diary survey detailing their sleep quality and their daily experiences of everyday discrimination. Participants completed a semistructured exit interview of 60-90 minutes at the study’s conclusion.

Results: In total, 10 participants completed 166 EMA surveys and 39 daily diary entries. A total of 4 of the 10 participants completed 75% (21/28) or more of the EMA surveys, while 9 (90%) out of 10 completed 25% (7/28) or more. The average completion rate of all surveys was 58% (20.3/35), with a minimum of 17% (6/35) and maximum of 100% (35/35). A total of 4 (40%) out of 10 participants completed daily diary entries for the full pilot study. No safety-related incidents were reported. On average, participants took 2.08 minutes to complete EMA prompts and 2.72 minutes for daily diary surveys. Our qualitative results generally affirm the acceptability and feasibility of the study procedures, but the participants noted difficulties with the technology and the redundancy of the survey questions. Emerging themes also addressed issues such as reduced EMA survey compliance and diminished mood related to deficit-framed questions related to suicide.
Conclusions: Findings from this study will be used to clarify the suitability of EMA for Black men. Overall, our EMA pilot study demonstrated mixed feasibility and acceptability when delivered through smartphone-based apps to Black men. Specific recommendations are provided for managing safety within these study designs and for refinements in future intervention and implementation science research.

International Registered Report Identifier (IRRID): RR2-10.2196/31241

(JMIR Form Res 2024;8:e48992) doi:10.2196/48992

KEYWORDS
Black men; suicide; ecological momentary assessment; feasibility; acceptability; mixed methods; smartphone; real-time assessment; suicide prevention; user experience; behavior; implementation; intervention; mobile phone

Introduction

Suicide among Black Americans is a critical public health priority that requires immediate attention. Studies have shown that suicide rates among this population are increasing and that Black men, in particular, are at a higher risk of suicide-related mortality [1,2]. It is crucial for public health efforts to prioritize suicide prevention in the Black community and address the systemic issues that contribute to mental health disparities [1,2]. Epidemiological data reveal that Black men die by suicide at rates 4 to 6 times greater than Black women and that suicide is the third leading cause of death for young Black men [2,3]. While nationwide suicide rates provide a broader perspective on this alarming public health issue, delving into state-specific data, such as in Maryland between 2016 and 2020, reveal a localized increase in suicide death for Black Americans, compared to their White counterparts and highlights the urgent need for tailored research priorities [4]. The COVID-19 pandemic further exacerbated these alarming trends, with Black Marylanders experiencing an uptick in suicide deaths during the initial lockdown periods in the United States [5]. These national and state-specific trends highlight the pressing need to address more robust approaches to suicide prevention, emphasizing the importance of ensuring equitable access to support and resources for all communities.

Mental health disorders, such as depression and substance use, are known risk factors for suicide among Black Americans [6,7], but recent research also highlights racism and associated contextual stressors as important, but underresearched, explanatory risk factors for suicidal thoughts and behaviors (STB) among Black adults [8-10]. Previous studies have demonstrated the significant impact of racism and associated daily stressors on mental health outcomes, including depression, anxiety, sleep disturbances, and posttraumatic stress disorder (PTSD), among Black individuals [11-15]. However, the emerging research highlights the need to examine the specific ways in which racism contributes to STB among this population. Further research in this area is crucial to identify effective prevention strategies that can reduce the impact of racism and mitigate the rising trend of suicide completion among Black adults.

Black Americans also face significant barriers in accessing mental health care due to stigma, systemic lack of access to services, and cultural mistrust [16-19]. Although Black men are approximately 30% more likely to report having a mental health condition compared to non-Hispanic White individuals [20], they also have significantly lower percentages of mental health-related visits prior to a suicide attempt than men of other racial and ethnic backgrounds and Black women [21,22]. In addition, traditional clinical records used to assess suicide risk and prevention often rely on distal and static measures of risk, such as prior diagnoses and family history, which may overlook the dynamic interplay of proximal factors that contribute to STB and patient-provider synergy. Consequentially, while Black Americans face considerable obstacles in accessing mental health care, these disparities are particularly pronounced among Black men, highlighting the urgency of addressing these disparities and adopting approaches to assess more proximal aspects of suicide risk and prevention.

Smartphone-based ecological momentary assessment (EMA) platforms are one such approach to collect dynamic, real-time data and encompass a range of different active and passive information, including, but not limited to, spatial trajectories (via GPS), physical mobility patterns (via accelerometer), social networks and social dynamics (via call and text logs and Bluetooth), and EMA surveys [23-25]. Understanding dynamic risk in suicide prevention efforts is important since decades of previous research have shown that single, static risk factors often add little to our understanding of who is at elevated risk and when [26]. More real-time data on STB may also provide key information about when to intervene with potential just-in-time interventions.

Although participant burden, noncompliance, and reactivity to the protocol measures have been cited as potential limitations to EMA approaches, studies using this methodology to assess STB demonstrate a favorable median response rate of 70%, suggesting feasibility [25,27-30]. These studies highlight the potential of EMA as a tool for capturing the complexity and variability of STB, ultimately aiding efforts to improve mental health outcomes. Specifically, one systematic review of EMA studies found suicidality fluctuates considerably over short periods of time and that those with higher levels of overall suicidality also have more fluctuations [27]. This review also noted risk factors such as negative affect, hopelessness, burdensomeness, and sleep characteristics that impacted suicidality [27]. In a sample of psychiatric inpatients, a separate study discovered that the use of real-time data collections significantly enhanced the accuracy of predictions for suicide attempts post-discharge [31]. Despite the promise of this methodology for the study and prevention of suicide [25,29,32], there has been no study, to date, that uses this approach to assess experiences among Black men at critical periods for early
intervention [26]. Black men face daily societal and cultural stressors that may contribute to their heightened risk of suicide, including but not limited to, systemic and interpersonal racism, economic disparities, and limited access to mental health resources [8,33,34]. To address this gap, additional research is needed to clarify the efficacy of EMA monitoring as a suicide prevention tool for this at-risk population.

Methods

Eligibility and Recruitment Procedures

We recruited adult Black or African American men (18 years of age or older) with a lifetime history of suicidal ideation or attempt residing in Maryland counties where mobile crisis support was available. To be eligible, participants had to (1) own a smartphone and (2) not be present with active psychosis or cognitive deficits. We used two recruitment approaches: (1) providing recruitment information to eligible, active patients via the Johns Hopkins Health MyChart, the web-based patient portal of the Epic electronic medical record system that allows for health communication between patients and health care providers and (2) clinician referral. Participants who met the eligibility criteria through either of these approaches were referred to the study coordinator and were required to complete a screening survey to confirm their interest and eligibility.

Ethical Considerations

The study was approved by the institutional review boards of Johns Hopkins Bloomberg School of Public Health (IRB 00013672). All participants were provided detailed information about the study procedures and expectations, risks, and benefits as part of the informed consent process conducted by a research coordinator.

Baseline Survey

We asked enrolled participants to complete a brief baseline assessment including demographics and psychosocial measures associated with affective, gender, and race-specific factors associated with STB, such as anger, sadness, attributional style, and racial identity. A complete list of baseline measures is described elsewhere. During the informed consent process, participants received an overview of the MetricWire smartphone app from the study coordinator. The study coordinator also determined the county in which the participant resided, in order to align our safety protocol with mobile health crisis units in their area. After completion of the baseline survey, we asked participants to download the MetricWire app onto their personal smartphones for the duration of the study. The MetricWire app is available for both iOS and Android smartphone platforms at no cost in the Apple App Store (Apple Inc) or Google Play Store (Google), respectively. Examples of the user interface of the MetricWire app are presented in Figure 1.

Figure 1. Screenshot of suicidal ideation and intensity questions on the ecological momentary assessment user interface. CSSR: Columbia Suicide Severity Rating Scale.

EMA and Daily Diary Data Collection Procedures

This study used the MetricWire app to deliver daily EMA surveys to participants at 4 time points between 10 AM and 6 PM. Each of the prompts was designed to occur at least 2 hours after the previous prompt and included 3 push notification reminders at 20, 40, and 55 minutes after the initial prompt. If the participant did not complete the EMA survey within 60 minutes, it was marked as incomplete. The EMA surveys were designed to be brief and take no more than 3 minutes to complete to reduce respondent burden. To capture instances of racism-related stressors occurring outside of the random EMA survey prompts, participants were allowed to record event-driven entries detailing their daily experiences (see Figure 2).
We administered a brief daily diary survey via the MetricWire app once per day to assess participants’ everyday experiences, including sleep-related impairment and quality, as well as their daily experiences with racism-related stress. The daily diary was prompted each day at 8 PM. This survey was not designed to capture momentary instances but rather to provide a more comprehensive picture of participants’ daily experiences.

Exit Interview
After the 7-day data collection period, each participant underwent a qualitative semistructured exit interview conducted by the study team. The interviews probed participants on issues such as question difficulty and clarity, potential revisions to question prompts, and overall satisfaction with the study protocol and EMA surveys. Verbatim transcripts of the recorded interviews were produced and analyzed to evaluate the feasibility and effectiveness of the study protocol.

Participant Incentives
To encourage higher EMA survey completion rates, participants were eligible to receive up to US $110 throughout the study duration. The incentive was incrementally phased, with participants receiving US $10 after completing the baseline survey, US $20 for completing 20% (7/35) of all surveys, US $50 for completing 50% (17.5/35) of all surveys, and US $80 for completing 80% (28/35) of all surveys. We offered an additional US $20 for the completion of the exit interview.

Safety Protocols
Upon enrollment, we provided participants with a document containing information about local and national mental health resources, including suicide crisis hotlines, to support their well-being throughout the study. To ensure participant safety during the data collection period, we implemented a 3-tiered safety protocol. Moderate risk, which is defined as any suicidal ideation (“Have you had thoughts of killing yourself?”), since the last assessment, but without any plan or intent, resulted in a notification to the participant guiding them to the online support groups and community-based mental health services and urging them to seek support. Elevated risk, defined as suicidal ideation with intent or a plan within the last 24 hours (“Have you planned out how you would do it?” or “When you thought about making yourself not alive anymore, did you think that this was something you might actually do?”), resulted in the same notification that was given to participants with moderate risk and participants will be asked if they would like to handle the matter themselves or if they would like us to contact a mobile crisis response unit in their area on their behalf. Acute risk, defined as suicidal ideation with an action since the last assessment (“Did you do anything to make yourself not alive anymore or kill yourself?”), resulted in a call to the participant’s closest mobile crisis response unit made by a member of the study team on the participant’s behalf. To protect participant privacy, data were not stored on their smartphones. Instead, survey response data were automatically synced to the MetricWire servers when participants were connected to the

Figure 2. Screenshot of the respondent-driven racial stress recording option on the ecological momentary assessment user interface.
internet, and encrypted response data were stored until the next connection. These servers were continuously backed up.

To provide additional support, all research personnel interacting with participants received training in psychological first aid to assist with identifying any mental health needs. EMA responses were closely monitored by the study coordinator and regular updates were provided during biweekly research team meetings between January 2022 and May 2023.

**Quantitative and Qualitative Data Analysis**

To measure the acceptability and feasibility of the EMA method, we tracked four key metrics using Excel (Microsoft): (1) the proportion of eligible patients who enrolled in the study, (2) the percentage of completed EMA sessions out of the total number of scheduled sessions, (3) the number of safety-related incidents reported to mobile health crisis support teams, and (4) the average time taken to complete the EMA surveys.

Semistructured exit interviews were audio recorded and transcribed verbatim by the research team. The transcripts were then imported into Dedoose (version 8; Sociocultural Research Consultants, LLC) qualitative data analysis software for analysis. A deductive coding framework was used based on previous conceptual definitions related to implementation science. The research team independently reviewed the transcripts and generated a preliminary codebook based on a priori concepts. These concepts focused on prevailing definitions of acceptability, feasibility, adoption, and fidelity to the EMA methodology.

**Results**

**Demographics**

Participants ranged from 18 to 34 years of age, with an average age of 27 (SD 5.31) years. The majority (4/10, 40%) of our sample had completed a high school degree, followed by an associate or bachelor’s degree (n=2, 20%), while some completed high school (n=1, 10%), college (n=2, 20%), and a master’s degree (n=1, 10%). Participants lived across 3 counties in Maryland, including Baltimore, Anne Arundel, and Howard. Furthermore, participants identified as a spectrum of sexual orientations, including heterosexual (n=3, 30%), gay (n=2, 20%), bisexual (n=1, 10%), pansexual (n=2, 20%), and questioning (n=1, 10%). One participant chose not to disclose his sexual orientation.

**Feasibility of EMA and Daily Diary**

The MyChart recruitment service identified and sent study information to 744 active patients residing in Maryland. A total of 58 individuals completed our interest survey. Of those, 53 (91%) were attributed to the MyChart recruitment service. Three submissions were from clinician referrals and 2 were from community outreach events. Out of the 58 individuals completing the interest survey, 10 (17%) participants were enrolled in the EMA study. The primary reasons for exclusion after completing the interest survey were active psychosis (n=19) and no history of suicidal thoughts or behavior (n=10). A total of 7 individuals were excluded because they reported having no mental health care provider.

The 10 enrolled participants completed 166 EMA surveys and 39 daily diary entries. A total of 4 (40%) out of 10 participants completed 75% (21/28) or more of the EMA surveys. The average completion rate of all surveys was 58% (20.3/35), with a minimum of 17% (6/35), and maximum of 100% (35/35). A total of 4 (40%) out of 10 participants completed at least 5 out of 7 daily diary entries. The response rate for each day of the week ranged from 64% (3.2 surveys submitted on average; day 1) to 54% (2.7 surveys submitted on average; days 6 and 7). All 10 participants completed qualitative exit interviews at the conclusion of the study. There were no safety-related incidents that required mobile health crisis service response teams. On average, participants took 2 minutes and 5 seconds to complete EMA surveys and 2 minutes and 43 seconds to complete daily diary surveys.

**Qualitative Findings From Exit Interviews**

After coding, several key themes arose, including the study implementation, the iatrogenic effects of repeated suicidality assessments, and strategies to improve usability and effectiveness in suicide-related smartphone app development for Black men. We present key details and findings on these themes as follows.

**Study Implementation**

Table 1 shows representative quotes that illustrate participant perspectives on adoption, acceptability, and feasibility to the study protocol.
### Table 1. Qualitative codes and selected quotations related to implementation of smartphone-based ecological momentary assessment study.

<table>
<thead>
<tr>
<th>Code</th>
<th>Key quotes</th>
<th>Key critiques or recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptability</td>
<td><strong>“They’re good questions. It makes sense. It’s exactly the information that you’d be looking for.”</strong> [Participant 1]</td>
<td>“The only thing that I can really think of as far as that is the hours...Maybe they’re asleep, maybe they’re working etc., etc.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td><strong>“I like how it did call my therapist, because I finally set up a meeting with her. But she, you know, was like, ‘Hey, um, you know, come in contact with me, because I know you’re not doing well.’ And I was like, ‘Yeah, that’s fair.’”</strong> [Participant 2]</td>
<td>“I thought it shouldn’t have been the same questions. It got boring and repetitive and seemed like more of a hassle than something that was going to help me, because it’s annoying now, because it’s the same questions.” [Participant 5]</td>
</tr>
<tr>
<td></td>
<td><strong>“I thought the questions were pretty straightforward.”</strong> [Participant 4]</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>“I think for the most part, it was good, it was self-explanatory. It was easy to get to.”</strong> [Participant 5]</td>
<td></td>
</tr>
<tr>
<td>Feasibility</td>
<td><strong>“I normally don’t go through racial stuff, like...”</strong> [Participant 6]</td>
<td>“The wording was, yeah, as I was saying before, it was very, extremely clinical, which I mean, you’re not really supposed to be beating around the bush as far as the stuff is concerned.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td><strong>“Um, that was fine. I was. I’m not sure if you remember, but I was a bit wary beforehand. I was worried that it might be a little bit too many, but no, it was fine. The amount was perfectly okay. The timing was fine too...”</strong> [Participant 1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>“The only thing that I can really think of as far as that is the hours but as I had mentioned, before we even started this, anyone who's downloading this resource is probably doing it with the intention on actually utilizing it. So I can't really see someone actively like disregarding the surveys, unless they either don’t notice them or, well, yeah, unless they don’t notice them or somehow it slips past their schedule. Maybe they're asleep, maybe they're working etc., etc.”</strong> [Participant 1]</td>
<td></td>
</tr>
<tr>
<td>Fidelity</td>
<td>Prompts to complete surveys: “It was good, because some days I will remember, I have to take the survey so such and such time but other days, I wouldn’t to remember. I’m glad they, the prompts you know, were there.” [Participant 6]</td>
<td>“One of the things that was a little annoying was like oh, it would crash on me a lot. So I would have to re-open it, but when I would re-open it, it would take me right back to the question that it crashed on me on. So, I was like, ‘Oh, okay, that was convenient.’” [Participant 4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“it seemed like every time I tried to open the app, it wouldn’t even open, like sometimes it...ugh like it was so frustrating, so I just wouldn’t try to do it for a while.” [Participant 6]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lack of use due to app issues: “I didn’t always have like issues as far as that’s concerned, I didn’t really use it very often, explicitly, because around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
</tr>
<tr>
<td>Appropriateness</td>
<td><strong>Timing of prompts: “I’m not sure if you remember, but I was a bit wary beforehand. I was worried that it might be a little bit too many, but no, it was fine. The amount was perfectly okay. The timing was fine too...”</strong> [Participant 1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Completing racial stress record in a predominantly Black neighborhood: “I normally don’t go through racial stuff, like I'm in a neighborhood is predominately Black. I see one White person a day, maybe.”</strong> [Participant 5]</td>
<td></td>
</tr>
<tr>
<td>Adoption</td>
<td><strong>“I found myself being very, at least the first or second time, I felt myself being very, how am I supposed to answer this? It’s, it just felt strange that, well, yeah.”</strong> [Participant 1]</td>
<td>Intention influenced by number of prompts: “It’s not like, something like a text or the person will later call me and be like, ‘Hey, did you see my text?’ I’ll probably completely forget it exists, which I think is what I did for like a lot of the surveys.” [Participant 2]</td>
</tr>
<tr>
<td></td>
<td><strong>“I think that the first couple of questions are pretty intense ones. So, what was interesting for me at first—so I was having like a very good week, so the questions were a little bit jarring almost, if you're like, in the middle of a good day, and then you get those questions.”</strong> [Participant 8]</td>
<td></td>
</tr>
</tbody>
</table>

\[\text{N/A: not applicable.}\]
Overall, participants found the frequency and timing of the EMA survey prompts to be appropriate. One participant commented that the questions provided “exactly the information needed,” while another appreciated the reminder to contact their therapist. Participants generally found the questions straightforward and easy to understand, and the survey itself was considered self-explanatory and easy to access. Participant 1 noted that they were initially concerned about the number of prompts, but ultimately found them to be manageable:

I'm not sure if you remember, but I was a bit wary beforehand. I was worried that it might be a little bit too many, but no, it was fine. The amount was perfectly okay. The timing was fine too.

However, some participants reported feeling that the racial stress record and daily diary questions regarding discrimination may not have been relevant to their daily experiences.

I normally don't go through racial stuff, like I'm in a neighborhood that is predominately Black. I see one White person a day, maybe. [Participant 5]

Our study identified some challenges related to technology issues that impacted the fidelity of our study implementation and, ultimately, participants’ ability to complete the surveys in a timely manner. Participants identified app crashes as a significant issue, which led to frustration and difficulty in completing the surveys.

One of the things that was a little annoying was like oh, it would crash on me a lot. So I would have to re-open it, but when I would re-open it, it would take me right back to the question that it crashed on me on. So, I was like, “Oh, okay, that was convenient.” [Participant 4]

This issue not only caused inconvenience but also resulted in participants potentially losing progress on their survey responses. Consequently, this problem may have discouraged some participants from continuing with the study.

Influence of Mood on EMA Compliance

The narratives shared by the participants highlighted the possibility of diminished effect resulting from frequent evaluations of suicidality, which could potentially exacerbate negative mental health outcomes. Participants’ accounts of their study experience suggest that depressed mood prior to a notification to complete a survey may impact compliance with suicide-related assessments. Overall, as stated by Participant 1, “during moments where I already wasn’t feeling too great, it exacerbated things a bit.” Further, another participant in the study described how their survey compliance can be influenced by their mood:

It’s sort of like a reflex for like, whenever something buzzes on my phone, and I’m like, not in a good mood, I immediately, like, swipe it [get rid of the notification]

Participant 2 noted that repeated assessments of suicidality could increase their awareness of passive suicidal thoughts, stating, “Not really trigger me, but be like, oh, yeah, I am, like, passively suicidal 99% of the time.” Another participant noted in his interview that he did not complete the last day of the trial due to an anniversary of a loved one’s death:

It was an emotional day and I didn’t want to continue because my responses would have changed drastically. I didn’t want nobody to look at me differently...I didn’t want to put myself in an uncomfortable position.

This response further cements the idea that response rates may be influenced by low mood. Several participants provided feedback on the survey questions, suggesting that shorter surveys may be more effective in assessing suicidality. They noted that if they responded with a positive mood score, assessed by the Patient Health Questionnaire-2 item measure (PHQ-2), it likely indicated they were doing well and not feeling suicidal. Multiple participants suggested that gauging mood before asking directly about suicide may improve the effectiveness of the survey.

Those questions are pretty good. But I would probably, like I’d probably try and, like keep the survey shorter...So maybe like gauge the mood before asking like, “Hey, are you going to kill yourself?”

Because if I’m at like 3, then the answer is like, it’s no, it’s, it’s gonna be no, you know, because I’m in like, a pretty good place. [Participant 2]

Given our study’s focus on suicide risk assessment, a significant portion of the survey questions pertained to assessing proximal or momentary risk for self-harm. Unfortunately, these questions were predominantly framed from a negative or deficit-oriented perspective, such as “Have you thought about killing yourself since the last prompt?” Several participants pointed out that this focus on negative experiences was contributing to their mood states and recommended that the survey be balanced with more positive questions.

It seems like everything was about if I have a terrible day...there should have been some questions that show record of your good days, too. [Participant 5]

Recommendations to Improve Usability and Effectiveness

As part of the qualitative exit interview, participants were asked to provide feedback and suggestions for the future use of the app. In relation to the app’s accessibility, several participants suggested incorporating features that would promote ease of use. Specifically, participants mentioned that a dark mode would be beneficial, as the bright white background of the app was uncomfortable for some. In addition, participants suggested incorporating a text-to-speech feature that would read the survey questions aloud to make it easier for individuals with visual impairments or reading difficulties.

To encourage better participation and compliance, some participants suggested that a more flexible schedule would better align with their daily routines and lifestyles. Some participants pointed out that they are less inclined to fill out surveys when they are occupied with other social activities. As a result, participants suggested a schedule that can be customized according to their personal needs and circumstances.
I guess, maybe if I could like, sort of choose my own schedule...like put it at times where I know that I'm going to be like, on my phone, or like, in my bed...I'm probably going to be in my bed not doing anything else, like probably playing a game on my phone. So I'll tell myself, “Hey, you have no excuse, fill this out so that they know you haven't, like died.” But, if it's like, in the middle of like, me going to the mall with my friends, I'm most definitely not going to stop and be like, “Hold on everybody, I have to fill this out” I'm just gonna forget about it. [Participant 2]

Participants provided various suggestions for enhancing the interactivity of the smartphone-based app, including incorporating a journaling component to allow them to document their thoughts in written form. Participant 2 described the potential benefit of such a feature, stating that it could provide a private space for them to express their emotions and thoughts more freely, without the fear of judgment or negative consequences.

I was thinking, maybe something where, you know, if you're not feeling that bad, you can just write down something...And knowing that it's going to medical professionals, I probably will be as kind of open as I am in my journal.

Additionally, other participants expressed interest in in-app features that would allow for personalized coping strategies, such as guided breathing exercises or links to mental health resources. These in-app features were seen as potential tools to promote engagement with the app and ultimately improve mental health outcomes. Finally, participants recommended that the survey include questions that allow for a more comprehensive and nuanced picture of their mood and well-being, and that these questions should be presented before questions pertaining to suicide risk to “ease into the heavier stuff,” as suggested by participant 1.

Discussion

Principal Findings

Our study aimed to assess the feasibility of using smartphone-based EMA to evaluate suicidality in real time among individuals in a high-risk sample of Black men. During the interviews, most participants generally reported that the frequency and timing of the prompts and reminders were suitable for the project and their daily lives. The results indicate that this approach shows promise for future research in this population. The results also suggested that while suicide-related smartphone apps have the potential to be effective tools for suicide prevention, they need to be developed and implemented with care. Even though the completion rate for EMA surveys was not optimal, we are encouraged by the high level of compliance with the daily diary entries. The study also did not pose any safety concerns, and the completion time for the surveys was found to be relatively short. Participants reported that the questions were generally relevant to their daily experiences and easy to understand. Additionally, the user feature that allowed for open-ended audio recordings was found to be useful in capturing daily experiences of racism, which is an important factor that could contribute to suicidality in this population. Finally, the qualitative findings from exit interviews suggest that the study procedures were perceived by participants to be acceptable and feasible.

Despite the promising results, the study revealed that the compliance rate was lower than in other studies not focused on Black men that leveraged EMA approaches to assess suicide risk [25,28]. One possible reason for this finding is that participants encountered technological issues and cited them as reasons for disengagement from the study. Participants noted redundancy in survey questions, which can lead to survey fatigue and reduced engagement. This redundancy might have contributed to the lower completion rates observed in the study. Moreover, the frequency of prompts may have contributed to the overall low compliance rate. The participants recommended fewer prompts and more flexibility in timing to enhance compliance.

Our findings also highlight the need for careful consideration of the potential harms of repeated assessments of suicidality, which focuses on a negative framing of mental well-being, and the need for the development of personalized interventions to mitigate any negative effects. Specifically, participant narratives from our qualitative exit interviews suggest that individuals experiencing negative moods may be less likely to engage with suicide-related assessments delivered via smartphone app notifications. Additionally, respondents felt that the EMA survey questionnaire could be more balanced by including both questions that assessed suicide risk and positive emotions as protective factors. One such measure, the Reasons for Living Inventory (RLI), was developed to identify factors that serve as deterrents against suicidal behavior. This inventory assesses positive emotions that support an individual’s decision to avoid death by suicide, should such thoughts emerge. Including measures, such as the RLI, that ground Black men to life is vital for addressing mental health challenges and fostering well-being for this demographic. By understanding how to balance the benefits of suicide-related assessments with their potentially harmful impact, scholars will be more equipped to identify protective factors and develop tailored interventions that are more likely to be used by Black men.

Our findings have important implications for the use of EMA among Black men. Smartphone-based methods, such as EMA and daily diaries, offer a unique opportunity to assess and evaluate suicide risk in real time. This approach represents a significant advancement in capturing momentary phenomena that could aid in suicide prevention, such as (1) characterizing dynamic mood changes over a short time span, (2) untangling stressful and racialized daily experiences, and (3) transmitting critical health information to health care providers to support care management. By using this approach, researchers can gain a more nuanced understanding of the experiences of high-risk populations and provide timely culturally relevant interventions to prevent suicide [23,24,32,33].

Moreover, researchers should be aware of certain populations that may be cautious to participate in research, such as Black men, particularly when it involves the explicit tracking and reporting of their daily experiences. This hesitancy can be
rational, given the history of systemic exploitation of Black communities in biomedical research [36]. Therefore, it is important for researchers to acknowledge and address these historical injustices to build trust and facilitate participation among these populations. Only by taking these steps can researchers ensure that their work is ethical, respectful, and ultimately beneficial to the communities they seek to serve [19,37,38].

Smartphone-based suicide prevention is also increasingly becoming a viable approach for combating rising suicide rates, especially among young adults. The prevalence of health-related internet searches via smartphones, as evidenced by the high percentage of people who are 18-29 and 30-49 years of age and who use their phones for medical information, suggests that smartphones can be effective tools for suicide prevention. Additionally, the fact that Black and Hispanic Americans use their phones for medical information more frequently than White non-Hispanic Americans suggests that smartphone-based interventions may be especially important in addressing suicide risk in communities that experience higher levels of stigma and barriers to accessing traditional health care resources [39]. By leveraging the widespread use of smartphones, suicide prevention efforts can reach a larger and more diverse population, potentially reducing the burden of suicide on individuals, families, and communities.

Limitations
There are several potential limitations to our study that should be noted. First, our primary recruitment method was through MyChart messaging, which may have limited the pool of potential participants to those who are more likely to have additional physical comorbidities. The recruitment approach may have influenced the response rate, since participants may not have had the same motivation to respond to an interest survey, compared to a direct referral from a medical professional. Second, all participants were required to own a smartphone, which may have excluded individuals who did not have access to this technology. Additionally, we excluded Black men with active psychosis, which may have further limited the diversity of our sample. Given the potential for overdiagnosis of psychosis in underserved communities [40], it is possible that we also excluded potentially eligible individuals. Our study was also geographically restricted to Black men receiving health care in Baltimore, Maryland, and its surrounding counties. Future studies should consider additional venues and settings to recruit Black men who are not engaged in psychiatric care, including but not limited to social media, advocacy groups, and peer-led and community-based organizations. Finally, our study involved a relatively small number of participants, with 9 eligible participants completing the EMA surveys. This limited sample size may not fully represent the diversity and complexity of experiences among Black men and can affect the generalizability of findings to a larger population. Future studies should expand from our preliminary findings with a sufficient sample size to statistically investigate predictors of low compliance and user uptake.

Conclusions
This study is the first to our knowledge to address critical gaps in suicide research by incorporating EMA to improve the care for Black men who are at risk of suicide. Using EMA may be an important tool to help stem challenges to the timely assessment of suicide among Black men, who comprise the largest percentage of deaths by suicide (81%) within the Black community [4]. Our findings highlight how acceptable and feasible this method is for this high-priority population, as well as potential approaches to improve its fit.

Our study provides critical insights into the use of smartphones to capture real-time data for assessing the mental and emotional health of individuals in high-risk clinical samples of Black men. Our results highlight the suitability of EMA using smartphone-based approaches for studying sensitive topics related to suicide in vulnerable populations. Furthermore, the findings shed light on the next steps for creating more equitable suicide prevention approaches, including identifying areas of missing data and the cultural acceptability of smartphone-based tools for health promotion. However, further research is necessary to expand these tools to assess structural racism and other racialized factors that influence Black men’s daily lives.

Acknowledgments
LBA was supported by a grant from the Bloomberg American Health Initiative at Johns Hopkins Bloomberg School of Public Health, from the National Institutes on Mental Health (K01MH127310) and the American Foundation on Suicide Prevention (YIG-0-001-19). RJTJ was supported by grants National Institutes of Health (K02AG059140) and (U54MD000214). EEH was supported through the National Institutes on Mental Health (K01MH226335). The MyChart Recruitment Service is a service supported by the Johns Hopkins Institute for Clinical and Translation Research. The Johns Hopkins Institute for Clinical and Translation Research is funded by the National Center for Advancing Translational Sciences through the Clinical & Translational Science Awards Program.

Data Availability
The data set generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.
References


Abbreviations

- **EMA**: Ecological momentary assessment
- **PHQ-2**: Patient Health Questionnaire-2
- **PTSD**: posttraumatic stress disorder
- **RLI**: Reasons for Living Inventory
- **STB**: suicidal thoughts and behavior
An Initial Validation of Community-Based Air-Conduction Audiometry in Adults With Simulated Hearing Impairment Using a New Web App, DigiBel: Validation Study

Anna Sienko¹, BA; Arun James Thirunavukarasu², BA, MBBCHIR; Tanya Kuzmich¹, BSc; Louise Allen³,⁴, MBBS, MA, MD

¹School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom
²Oxford University Clinical Academic Graduate School, University of Oxford, Oxford, United Kingdom
³Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
⁴Department of Paediatrics, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Anna Sienko, BA
School of Clinical Medicine
University of Cambridge
Hills Road
Cambridge, CB2 0SP
United Kingdom
Phone: 44 1223 336700
Email: as2866@cam.ac.uk

Abstract

Background: Approximately 80% of primary school children in the United States and Europe experience glue ear, which may impair hearing at a critical time for speech acquisition and social development. A web-based app, DigiBel, has been developed primarily to identify individuals with conductive hearing impairment who may benefit from the temporary use of bone-conduction assistive technology in the community.

Objective: This preliminary study aims to determine the screening accuracy and usability of DigiBel self-assessed air-conduction (AC) pure tone audiometry in adult volunteers with simulated hearing impairment prior to formal clinical validation.

Methods: Healthy adults, each with 1 ear plugged, underwent automated AC pure tone audiometry (reference test) and DigiBel audiometry in quiet community settings. Threshold measurements were compared across 6 tone frequencies and DigiBel test-retest reliability was calculated. The accuracy of DigiBel for detecting more than 20 dB of hearing impairment was assessed. A total of 30 adults (30 unplugged ears and 30 plugged ears) completed both audiometry tests.

Results: DigiBel had 100% sensitivity (95% CI 87.23-100) and 72.73% (95% CI 54.48-86.70) specificity in detecting hearing impairment. Threshold mean bias was insignificant except at 4000 and 8000 Hz where a small but significant overestimation of threshold measurement was identified. All 24 participants completing feedback rated the DigiBel test as good or excellent and 21 (88%) participants agreed or strongly agreed that they would be able to do the test at home without help.

Conclusions: This study supports the potential use of DigiBel as a screening tool for hearing impairment. The findings will be used to improve the software further prior to undertaking a formal clinical trial of AC and bone-conduction audiometry in individuals with suspected conductive hearing impairment.

(JMIR Form Res 2024;8:e51770) doi:10.2196/51770

KEYWORDS

audiology; audiometry; hearing test; eHealth; mobile application; automated audiometry; hearing loss; hearing impairment; web-app; web-apps; web-application; digital health; hearing; adult; adults; mobile health; mhealth; community-based; home-based; assistive technology; screening; usability; ears; ear
Introduction

Over 5% of the world’s population, approximately 430 million people worldwide, have disabling hearing loss; 34 million of these are children [1]. The main causes of hearing loss in adulthood are age-related hearing loss, noise-related hearing loss, and hearing loss due to chronic otitis media. In children, the most common cause of hearing impairment is otitis media with effusion, also known as “glue ear”. Approximately 80% of primary school children in the United States and Europe experience glue ear [2]. In contrast to most forms of adult hearing loss, hearing impairment in children with glue ear fluctuates. Serial testing is often required to detect and manage the condition to mitigate its adverse impact on social development. Children in low- and middle-income countries, families experiencing socioeconomic deprivation, and disadvantaged populations are disproportionally affected by conductive hearing loss caused by glue ear and its complications, or from damage to the eardrum as seen in chronic tympanic perforations or chronic serous otitis media [3,4]. Since delayed recognition and management of childhood hearing impairment have long-term consequences for socialization and educational attainment, screening audiometry is recommended in primary school [5]. However, population screening programs are hindered by cost, standardization, requirement for staff training, false-positive referrals, and poor data capture [6,7]. Even where available, school screening may miss children with fluctuating hearing loss due to glue ears. Additionally, several year groups have missed screening during the COVID pandemic [8]. These children may face months of impaired hearing before diagnosis and management due to backlogs in audiology and specialist services.

Hearing thresholds are assessed using pure tone audiometry (PTA) with air-conduction (AC) headphones and bone-conduction (BC) transducers. This usually requires specialist equipment and trained clinicians. Automated audiometry and, more recently, validated self-testing hearing software apps, may improve the accessibility of screening and threshold audiometry testing, particularly in rural areas. The DigiBel web app is a recently developed Class 1 CE marked medical device that enables self-testing of AC and BC hearing levels. Like some other audiometry apps, it is suitable for community use in adults and children without clinical support. DigiBel has the novel facility to undertake BC audiometry with the same transducer used in a BC hearing assistance kit (BC headphones with Bluetooth-connected microphone, Raspberry Pi, Cambridge, United Kingdom). This could identify children who may benefit from this assistive technology while waiting for diagnosis, spontaneous resolution, or definitive management of their glue ear [9].

The purpose of this study of DigiBel audiometry is to determine the app’s sensitivity and specificity for detecting simulated conductive hearing impairment of more than 20 dB and to identify software modifications required prior to formal trials in a clinical population. Given the long protocol of testing or retesting and the requirement for usability feedback to inform improvement in the app design, this preliminary study involved healthy adult volunteers rather than children.

Methods

Overview

Healthy adult volunteers from the community without a previous history of hearing impairment were invited to participate in this comparative study of automated PTA and DigiBel audiometry. After receiving an explanation of the study, participants provided verbal consent to proceed to audiometry testing. Each participant was assigned a unique study identification number; no personal identifiable information was recorded.

Testing was undertaken in community settings such as participants’ homes and classrooms by nonaudiologist technicians. Prior to testing, each volunteer was instructed to place a foam earplug firmly into their left auditory canal and requested not to adjust it until testing was complete. This simulated a conductive hearing loss in the plugged ear, making each ear an independent entity for the purpose of statistical analysis and providing a range of hearing levels.

The reference automated AC PTA and the index DigiBel audiometry test were undertaken sequentially in random order. DigiBel audiometry for 4 frequencies was repeated immediately after the initial test, to assess within-session test-retest (TRT) reliability. After completing both audiometry tests, each volunteer was asked to complete a feedback questionnaire covering test preference and usability (Multimedia Appendix 1).

Automated Pure Tone Audiometry—The Gold-Standard Reference Test

Automated PTA was undertaken using an Oscilla (Oscilla A/S Aarhus Denmark) USB300 audiometer with TDH-39 headphones. The modified Hughson-Westlake algorithm was used for determining the reference AC audiometric threshold [10]. After an initial explanation by the technician and a conditioning test at 1000 Hz, thresholds were recorded at 2000, 4000, 8000, 1000, 500, and 250 Hz in accordance with British Society of Audiology recommendations [11]. The hearing threshold criterion for each frequency was determined as the lowest intensity at which participants accurately signaled 2 confirmations out of 3 presentations. The number of false positive responses was manually recorded.

The DigiBel Index Test

DigiBel has been laboratory and biologically calibrated (Institute of Sound and Vibration Research, University of Southampton, United Kingdom; and Chears-audiology, Royston, United Kingdom) to run specifically on any model iPad tablet (Apple) with Sennheiser HD 400S AC headphones (Wedemark, Germany).

A pretest embedded video provides instructions for use and a checklist ensures that the iPad volume is on maximum and the headphones are fitted correctly. Noise sampling, using the inbuilt Sennheiser headphone microphone, ensures that the ambient noise level is less than 40 dB prior to testing. Testing can be paused and restarted at any stage if the ambient noise level changes unexpectedly or interruptions occur. Configurable...
settings include a choice of warble or pure tone, test frequencies, ambient noise setting, and child or adult version of the test.

The user taps a central animated button on the iPad display when they hear the tone (Figure 1A). A conditioning step requires the user to accurately tap the button on hearing a random onset suprathreshold tone at 1000 Hz before testing can begin. During testing, the onset of the tone is randomized from 0 to 3 seconds after the appearance of the response button to avoid a predictable response pattern. The tone stops in response to the tap and a psychophysical staircase algorithm starting from 60 dB (10 dB down, 5 dB up) is followed in a 2 down, 1 up rule, requiring 5 reversals dependent on the user’s input. The final threshold is calculated as the mean of the final 3 reversal thresholds. Once completed, a standard audiometry graph and the number of false positive responses are displayed (Figure 1B). The user can choose to repeat testing or undertake BC testing to determine the functional effect on hearing levels.

Figure 1. (A) DigiBel test interface and (B) DigiBel audiometry graph.

For this study, DigiBel AC PTA was undertaken in the sequence 2000, 4000, 8000, 1000, 500, and 250 Hz and retested in the sequence 2000, 4000, 8000, and 500 Hz using the adult test version.

Statistics
Data were analyzed in R (version 4.1.2; R Foundation for Statistical Computing) [12,13]. Accuracy and TRT reliability were assessed through Bland-Altman analysis, a statistical approach enabling analysis of the agreement between 2 measurement methods by assessing their mean differences (mean bias) and upper and lower limits of agreement (SD 1.96). Qualitative appraisal of mixed effects model 2-way intraclass correlation coefficients (ICCs) between the threshold measurements from each device was based on conventional standards with moderate, good, and excellent agreement indicated by an ICC of ≥0.50, ≥0.75, and ≥0.90, respectively [14]. The correlation between calculated bias and mean threshold values was analyzed using Pearson correlation coefficients (PCCs). The percentages of DigiBel threshold measurements lying within 10 dB of both the reference test and the repeated DigiBel test were calculated [15,16]. Statistical significance was calculated for the mean bias where confidence intervals did not cross zero, and for ICC and PCCs where $P<.05$. To assess diagnostic efficacy, sensitivity and specificity for detection of hearing thresholds above 20 Hz were calculated, using automated PTA as the reference. The Student t test was used to compare the number of false positive responses for each test. Throughout, magnitudes were reported as the mean (SD) unless otherwise stated.

Ethical Considerations
The study adhered to the tenets of the Declaration of Helsinki and the protocol was approved by the Quality and Safety Committee of Cambridge University Hospitals National Health Service Foundation Trust as part of a service improvement project. The Quality and Safety Committee at Cambridge University Hospitals considered that formal ethics committee approval was not required for this no-risk service improvement study in a healthy, adult population and opined that verbal consent was sufficient in these circumstances to prevent the collection of patient identifiable information (approval for
Service Evaluation Project – PRN9288; dated April 11, 2023). No other demographic information other than age was recorded to maintain anonymity for this service improvement study. No compensation was offered for participating in the study.

**Results**

**Participant Characteristics**

A total of 32 healthy participants agreed to take part, but 2 participants were excluded due to malfunction of the reference automated PTA test. The 30 participants who completed both the reference and index tests were 21-66 (mean 27.9, SD 10.3) years. TRT data were collected from 29 participants (one volunteer left early due to time constraints). Feedback forms were completed by 24 participants (Figure 2).

**Figure 2.** Participant flow diagram.

---

**Accuracy and Reliability of DigiBel**

Across the 6 tested frequencies, threshold hearing levels of DigiBel compared with automated PTA gave an ICC of above 0.75 (good or excellent agreement) and \( P < .001 \) in every trial (Table 1; Multimedia Appendices 2 and 3). The mean lower limit of agreement (LOA) was –17.04 dB (SD 2.12 dB) and the mean upper LOA was 20.39 dB (SD 3.41 dB). No significant bias was apparent except at 4000 and 8000 Hz where a small but statistically significant bias was apparent (2.62 and 4.60 dB, respectively), with DigiBel providing systematically higher threshold results at those frequencies (Figure 3). An agreeable level of threshold difference in audiology assessments has previously been defined as 10 dB; 263 out of 360 (73%) of DigiBel threshold measurements were within this standard [16]. There was a significant positive PCC between the measurement bias and mean at 500 (\( P < .001 \)), 1000 (\( P = .005 \)), 2000 (\( P < .001 \)), and 4000 Hz (\( P < .001 \)) test frequencies.
<table>
<thead>
<tr>
<th>Comparison and frequency (Hz)</th>
<th>Number of ears tested</th>
<th>Bias (dB), 95% CI</th>
<th>LLoA&lt;sup&gt;a&lt;/sup&gt; (dB), 95% CI</th>
<th>ULoA&lt;sup&gt;b&lt;/sup&gt; (dB), 95% CI</th>
<th>ICC&lt;sup&gt;c&lt;/sup&gt;, 95% CI</th>
<th>PCC&lt;sup&gt;d&lt;/sup&gt; (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DigiBel compared with standard automated audiometry&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>60</td>
<td>1.72 (–0.32 to 3.76)</td>
<td>–13.75 (–17.25 to –10.25)</td>
<td>17.18 (13.68 to 20.69)</td>
<td>0.85 (0.77 to 0.91)</td>
<td>0.16</td>
</tr>
<tr>
<td>500</td>
<td>60</td>
<td>–1.55 (–3.93 to 0.83)</td>
<td>–19.62 (–23.71 to –15.52)</td>
<td>16.52 (12.42 to 20.61)</td>
<td>0.87 (0.80 to 0.92)</td>
<td>0.48</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.23 (–2.30 to 2.77)</td>
<td>–19.01 (–23.36 to –14.65)</td>
<td>19.47 (15.11 to 23.83)</td>
<td>0.88 (0.80 to 0.92)</td>
<td>0.36</td>
</tr>
<tr>
<td>2000</td>
<td>60</td>
<td>2.43 (–0.15 to 5.02)</td>
<td>–17.18 (–21.62 to –12.74)</td>
<td>22.05 (17.61 to 26.49)</td>
<td>0.88 (0.81 to 0.93)</td>
<td>0.64</td>
</tr>
<tr>
<td>4000</td>
<td>60</td>
<td>2.62 (0.14 to 5.10)</td>
<td>–16.19 (–20.45 to –11.93)</td>
<td>21.42 (17.16 to 25.68)</td>
<td>0.89 (0.83 to 0.94)</td>
<td>0.48</td>
</tr>
<tr>
<td>8000</td>
<td>60</td>
<td>4.60 (1.82 to 7.38)</td>
<td>–16.51 (–21.29 to –11.73)</td>
<td>25.71 (20.93 to 30.49)</td>
<td>0.91 (0.82 to 0.95)</td>
<td>0.13</td>
</tr>
<tr>
<td>DigiBel test-retest comparison&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>58</td>
<td>0.26 (–1.91 to 2.43)</td>
<td>–15.90 (–19.62 to –12.17)</td>
<td>16.42 (12.69 to 20.14)</td>
<td>0.92 (0.87 to 0.95)</td>
<td>N/A&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>2000</td>
<td>58</td>
<td>1.40 (–0.65 to 3.44)</td>
<td>–13.84 (–17.36 to –10.33)</td>
<td>16.64 (13.12 to 20.15)</td>
<td>0.95 (0.91 to 0.97)</td>
<td>N/A</td>
</tr>
<tr>
<td>4000</td>
<td>58</td>
<td>0.90 (–0.66 to 2.46)</td>
<td>–10.74 (–13.42 to –8.06)</td>
<td>12.53 (9.85 to 15.22)</td>
<td>0.97 (0.94 to 0.98)</td>
<td>N/A</td>
</tr>
<tr>
<td>8000</td>
<td>58</td>
<td>0.98 (–1.23 to 3.19)</td>
<td>–15.48 (–19.28 to –11.68)</td>
<td>17.44 (13.65 to 21.24)</td>
<td>0.95 (0.92 to 0.97)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup> LLoA: lower limit of agreement.
<sup>b</sup> ULoA: upper limit of agreement.
<sup>c</sup> ICC: intraclass correlation coefficient.
<sup>d</sup> PCC: Pearson correlation coefficient (calculated for the relationship between measurement bias and mean for each frequency).
<sup>e</sup> Comparison of threshold values (in dB) using Bland-Altman statistics and ICCs between DigiBel and standard automated PTA and DigiBel test and retest.
<sup>f</sup> N/A: not applicable.

Figure 3. Comparison of DigiBel and automated pure tone audiometry using Bland-Altman plots at 6 frequencies. Bland-Altman plots comparing mean and difference in threshold measurements (in dB) for DigiBel and standard automated pure tone audiometry at 6 frequencies (in Hz). The 95% CIs are shaded for the bias (red) and 95% limits of agreement (blue).

In TRT comparisons, ICC was above 0.90 (excellent) at all tested frequencies: 500, 2000, 4000, and 8000 Hz (Table 1). No statistically significant mean bias was exhibited at any frequency. The mean lower LOA was –13.99 dB (SD 2.34 dB); the mean upper LOA was 15.76 dB (SD 2.20 dB). Overall, 85% of TRT thresholds were within 10 dB of each other.

A mean of 3.57 (SD 4.68) false positive responses were recorded during reference testing and 11.43 (SD 7.12) during the first DigiBel test ($P<.001$).
Sensitivity and Specificity

The sensitivity and specificity of DigiBel for detecting 20 dB hearing loss at each frequency are shown in Table 2. When applied to the 4 frequencies used for screening (250, 1000, 2000, and 4000 Hz), DigiBel had 100% sensitivity (95% CI 87.23-100) and 72.73% (95% CI 54.48-86.70) specificity for detecting 20 dB hearing loss in adults in a quiet setting.

Table 2. Screening accuracy of DigiBel for hearing threshold >20 dB identified by automated pure tone audiometry.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Sensitivity (%)</th>
<th>95% CI</th>
<th>Specificity (%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>95.00 (75.13 to 99.87)</td>
<td></td>
<td>85.00 (70.16 to 94.29)</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>91.30 (71.96 to 98.93)</td>
<td></td>
<td>89.19 (74.58 to 96.97)</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>100.00 (86.28 to 100.00)</td>
<td></td>
<td>85.71 (69.74 to 95.19)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>100.00 (86.77 to 100.00)</td>
<td></td>
<td>79.41 (62.10 to 91.30)</td>
<td></td>
</tr>
<tr>
<td>4000</td>
<td>100.00 (86.28 to 100.00)</td>
<td></td>
<td>88.57 (73.26 to 96.80)</td>
<td></td>
</tr>
<tr>
<td>8000</td>
<td>100.00 (86.28 to 100.00)</td>
<td></td>
<td>88.57 (73.26 to 96.80)</td>
<td></td>
</tr>
<tr>
<td>Screening frequencies (250, 1000, 2000, and 4000)</td>
<td>100.00 (87.23 to 100.00)</td>
<td></td>
<td>72.73 (54.48 to 86.70)</td>
<td></td>
</tr>
</tbody>
</table>

Usability

Out of the 24 participants, 21 (88%) participants, while completing the questionnaire, did not regularly use digital health apps. All 24 participants rated DigiBel either good (15/24, 63% of participants) or excellent (9/24, 38% of participants); (7/24, 29% participants preferred the DigiBel test; 6/24, 25% participants preferred the standard test; and 11/24, 46% participants gave no test preference). A total of 21 (88%) participants agreed or strongly agreed that they would be confident to use DigiBel at home without help. The most common qualitative feedback given to the question “what is the best thing about the app?” was that it was easy or intuitive to use (17/24, 71% of participants). Answering “what is the worst thing about the app?” the commonest complaint was that the test was too long or boring (10/24, 42% of participants). One participant commented on environmental noise leaks through the headphones (Multimedia Appendix 1).

Discussion

Principal Findings

In 30 healthy adults with simulated unilateral conductive hearing impairment, DigiBel’s screening sensitivity and specificity were 100% (95% CI 87.2-100) and 72.73% (95% CI 54.45-86.7), respectively. The hearing threshold measurement mean bias between DigiBel and automated AC PTA was not significant except at 4000 and 8000 Hz, where it reached statistical significance (2.62 and 4.60 dB higher than the reference, respectively).

At least 5 validated downloadable apps enable automated AC audiometry, several of which support self-testing without clinician involvement [17]. Two apps, uHear (Unitron Ltd) and ShoeBOX (SHOEBOX Ltd) for iOS, include a BC audiometry facility [18,19]. uHear has been calibrated for use with commercial in-ear Apple headphones and ShoeBOX uses purpose-built audiometry headphones. DigiBel is not yet commercialized but its potential advantage is its calibration to purpose-built audiometry headphones. DigiBel is the only app to apply a staircase-reversal machine learning techniques have been developed [24]. To our knowledge, DigiBel is the only app to apply a staircase-reversal technique to audiometry, although it is commonly used for visual threshold testing [25]. In this study, the number of false positive participant responses was substantially higher with headsets (retail price US $28, equivalent to €26) for BC audiometry which may quantify the potential benefit from their use (with a paired microphone) as an assistive technology.

DigiBel’s screening sensitivity and specificity for hearing impairment (more than 20 dB) is comparable to previous studies of both uHear (98.2%-100% sensitivity and 60.0-82.1 specificity) and ShoeBOX (91.2%-93.3% and specificity of 57.8%-94.5%) [18-22]. In this study comparing the threshold measurements of DigiBel to automated AC PTA in 30 individuals, there was a significant positive correlation between the mean bias and the mean threshold measurement at 500, 1000, 2000, and 4000 Hz frequencies, resulting in an overestimation of hearing ability at normal hearing levels and an underestimation of hearing ability in ears with subnormal hearing. This was particularly evident at 2000 Hz testing and may have resulted in the lower specificity demonstrated at this frequency. This is likely to reflect the sound output characteristic of the AC headphones and will require software corrections prior to future clinical studies. Overall, 73% of threshold measurements with DigiBel were within 10 dB of standard PTA; previous studies of ShoeBOX have found over 90% of measurements were within this range [23]. The comparatively poor performance of DigiBel for this metric may be due to environmental noise leakage through the Sennheiser headphones, a disadvantage of their comfort.

Masking of the unplugged ear was not used for either automated PTA or DigiBel because this facility is unlikely to be used by nontrained observers in community settings. A minority of participants in this study had a simulated intra-aural threshold difference exceeding 40 dB. It is possible that intra-aural transmission may have affected threshold values in these individuals, but this would be expected to affect both tests similarly.

The Hughson-Westlake algorithm and other adaptive methods are widely used to assess audiometric threshold, more recently, machine learning techniques have been developed [24]. To our knowledge, DigiBel is the only app to apply a staircase-reversal technique to audiometry, although it is commonly used for visual threshold testing [25]. In this study, the number of false positive participant responses was substantially higher with...
DigiBel than standard testing. This may be due to the higher number of stimulus presentations compared with the ascending method used in standard automated PTA. The sensitivity of the iPad screen to a tap compared with the standard audiometer’s hand-held responder may be an additional factor. There was no evidence in this study that the higher false positive rate translated into a systematic overestimation of hearing ability.

All participants rated DigiBel as good or excellent, but 42% (n=10) of participants complained that the test took too long or was boring. Test duration was not measured during this study, but threshold testing for 6 frequencies is expected to take approximately 13 minutes with DigiBel, several minutes longer than standard automated PTA, primarily due to its staircase-reversal algorithm. Study participants had performed retesting which may have contributed to the perceived length of testing. The children’s version of the app has cartoons designed to increase interest but, even so, the test duration of threshold audiometry may limit its usability in young children. The DigiBel screening test of 4 frequencies takes approximately 3 minutes and may prove more feasible in this, its target population.

To simulate a range of hearing thresholds in the participant cohort, an earplug was used. This is a major limitation of the study because the effectiveness of the earplug may have altered during testing and earplugs may not accurately mimic genuine hearing impairment. Although this study indicates that DigiBel has acceptable accuracy for detecting simulated hearing impairment in healthy adult volunteers, these results are not generalizable and software corrections are required prior to clinical use.

This preliminary study confirms that the DigiBel app is an acceptable and easy-to-use self-testing web-based tool that accurately detects more than 20 dB of simulated hearing impairment in adults. Minor modifications to the 1000, 2000, and 4000 Hz frequency-specific normalization factors used in the software algorithm which converts sound pressure levels to hearing level are required to ensure uniformity of sound output and accuracy across the range of hearing abilities. A study, conducted in primary school children attending a hospital audiology clinic, is underway to assess the accuracy of DigiBel in identifying conductive hearing impairment. Additionally, the innovative concept behind DigiBel will be tested: its ability to identify those children who could benefit from the temporary use of a BC hearing assistance kit for use at school and home, and the impact this has on quality of life (using a parent- or patient-reported outcome measure questionnaire) while waiting for specialist care.

Conclusions

The World Health Organization identifies hearing loss as a major global health issue, with two-thirds of people with severe hearing loss living in low and middle countries with poor access both to hearing testing (audiometry) or conventional hearing aids. It can affect many aspects of life such as education, employment, and communication, and result in social isolation. Several software apps like DigiBel, studied here, have been developed to enable individuals to test their own hearing in the community. Uniquely, DigiBel has the additional potential to identify individuals with hearing loss who could derive immediate hearing support from an affordable and rechargeable bone-conduction hearing assistance kit while waiting for specialist care.

This initial study of DigiBel provides confirmation that the app is easy to use and accurate at detecting simulated hearing impairment. It identifies some software corrections that may improve its accuracy and lays the groundwork for future clinical studies to assess DigiBel’s performance in children and adults with hearing impairment.

Acknowledgments

The authors thank Josephine Marriage and Martha Mann at Chears-audiology, Royston, Cambridge for their support with the biologic calibration of the DigiBel app and Tamsin Holland-Brown for improving the manuscript. This research was supported by the National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre (NIHR203312). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

AS, AJT, and TK have declared that no competing interests exist. LA declares nonfinancial competing interests as the inventor of DigiBel, which was developed with funding from a Medical Research Council Confidence in Concepts Grant.
Audiometry dataset.

[XLSX File (Microsoft Excel File), 18 KB - formative_v8i1e51770_app3.xlsx ]

References


17. Sienko et alJMIR FORMATIVE RESEARCH


Abbreviations

AC: air-conduction
BC: bone-conduction
ICC: Intraclass correlation coefficient
LOA: limit of agreement
PCC: Pearson correlation coefficient
PTA: pure tone audiometry
TRT: test-retest

©Anna Sienko, Arun James Thirunavukarasu, Tanya Kuzmich, Louise Allen. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original work on https://formative.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: Poor sleep hygiene persists in college students today, despite its heavy implications on adolescent development and academic performance. Although sleep patterns in undergraduates have been broadly investigated, no study has exclusively assessed the sleep patterns of premedical undergraduate students. A gap also exists in the knowledge of how students perceive their sleep patterns compared to their actual sleep patterns.

Objective: This study aims to address 2 research questions: What are the sleep patterns of premedical undergraduate students? Would the proposed study protocol be feasible to examine the perception of sleep quality and promote sleep behavioral changes in premedical undergraduate students?

Methods: An anonymous survey was conducted with premedical students in the Medical Science Baccalaureate program at an R1: doctoral university in the Midwest United States to investigate their sleep habits and understand their demographics. The survey consisted of both Pittsburgh Sleep Quality Index (PSQI) questionnaire items (1-9) and participant demographic questions. To examine the proposed protocol feasibility, we recruited 5 students from the survey pool for addressing the perception of sleep quality and changes. These participants followed a 2-week protocol wearing Fitbit Inspire 2 watches and underwent preassessments, midassessments, and postassessments. Participants completed daily reflections and semistructured interviews along with PSQI questionnaires during assessments.

Results: According to 103 survey responses, premedical students slept an average of 7.1 hours per night. Only a quarter (26/103) of the participants experienced good sleep quality (PSQI<5), although there was no significant difference (P=.11) in the proportions of good (PSQI<5) versus poor sleepers (PSQI>5) across cohorts. When students perceived no problem at all in their sleep quality, 50% (14/28) of them actually had poor sleep quality. Among the larger proportion of students who perceived sleep quality as only a slight problem, 26% (11/43) of them presented poor sleep quality. High stress levels were associated with poor sleep quality. This study reveals Fitbit as a beneficial tool in raising sleep awareness. Participants highlighted Fitbit elements that aid in comprehension such as being able to visualize their sleep stage breakdown and receive an overview of their sleep pattern by simply looking at their Fitbit sleep scores. In terms of protocol evaluation, participants believed that assessments were conducted within the expected duration, and they did not have a strong opinion about the frequency of survey administration. However, Fitbit was found to provide notable variation daily, leading to missing data. Moreover, the Fitbit app’s feature description was vague and could lead to confusion.
Conclusions: Poor sleep quality experienced by unaware premedical students points to a need for raising sleep awareness and developing effective interventions. Future work should refine our study protocol based on lessons learned and health behavior theories and use Fitbit as an informatics solution to promote healthy sleep behaviors.

(JMIR Form Res 2024;8:e45910) doi: 10.2196/45910

KEYWORDS
patient-generated health data; Fitbit wearables; sleep quality; premedical college students; sleep; sleep hygiene; student; colleges; university; postsecondary; higher education; survey; sleep pattern; medical student; adolescence; behavior change

Introduction
Sleep health has been recognized as an unresolved public health concern since 2006 [1]. Although sleep plays a vital role in adolescent health and development, poor sleep patterns continue to be associated with college students. A 2015 study showed that 70%-96% of college students sleep less than the recommended 8 hours on weeknights [2,3]. With the heavy prevalence of sleep deprivation, it is important to recognize the detrimental effects it can have. Starting with cognitive processing, poor sleep has been shown to have negative effects on working memory, attention span, and speed of processing [4]. With poor sleep, our bodies can also experience impaired immune function, more susceptibility to illnesses, increased risk of stress, and decreased physical functions [5]. In addition, emotional dysregulation and lower levels of subjective well-being can be observed [6]. As a result, students experience a consequential effect on their academic performance [7]. It is therefore imperative to understand and discuss the sleep patterns of adolescents, especially college students, in order to aid them with tools that can help avoid these consequences.

Additionally, it is possible that students who may appear to receive an adequate amount of sleep can experience poor sleep due to bad sleep quality. Sleep quality refers to both quantitative aspects of sleep such as duration, sleep latency, and number of arousals, and qualitative aspects such as depth and restfulness of sleep according to a 1998 Psychiatry and Research publication [8]. Poor sleep quality refers to the struggle of falling asleep or maintaining sleep. A recent 2022 study clarifies this definition by adding the component of individuals’ self-satisfaction with their sleep as another part of sleep quality [9]. A previous survey on college undergraduates found that only 34.1% of students displayed good sleep quality based on Pittsburg Sleep Quality Index (PSQI) scores [10]. The PSQI is a self-rated questionnaire that serves as a standard tool for measuring sleep quality [8]. The responses to questionnaire items allow researchers to determine a global PSQI score, which helps distinguish between good-quality and poor-quality sleepers. A PSQI score smaller than 5 (<5) signifies good sleep quality, whereas a PSQI score equal to or larger than 5 (≥5) indicates poor sleep quality. However, to diagnostically study sleep, polysomnography (PSG) can be conducted, which involves recording brainwaves, oxygen level in the blood, respiration rate, and physical movements to diagnose sleep disorders [11]. PSG is used to understand the sleep patterns of patients and why there might be disruptions [12].

A recent study on the sleep quality of college students explored the determinants of sleep quality [13], interventions for improving sleep quality, overall demographics of sleep patterns, and associations of poor sleep quality with health problems. Some studies [7,14] have approached the understanding of sleep through a comparison of different college majors. One such major or a group of students that is of interest to us are students who are premedicine (premedical) majors or undergraduate students on the prehealth track. However, there is limited exploration looking specifically at the sleep patterns of premedical undergraduate students (premedical students hereafter). Previous research in this area has largely been centered on the sleep quality of students who are in a medical school or graduate medical education program and focused on sleep-related disorders among them [15-19]. Numerous studies have measured the sleep patterns of medical students but have not focused on understanding the patterns [20-33]. To date, only 1 study has tested medical students on their sleep knowledge and has found a disconnect between their understanding of sleep facts and actual sleep quality [34]. This led to the focus of our study on the sleep quality perception of premedical students.

A student’s perception of his or her sleep habits can give us insights into whether perception influences sleep behaviors and quality. Many prior studies [35-38] have looked at sleep state misperception. This concept discusses how there is a disconnect between the amount of sleep people think they get versus their actual duration of sleep. Research on sleep state misperception has primarily been conducted on people dealing with varying mental health conditions or sleep disorders [35-38]. In this manner, our research focused on the effect of sleep quality perception on sleep patterns, which considers overall sleep quality rather than just duration, in a relatively healthy population comprised of premedical students. Literature shows that in medical students, perception of stress can have an impact on their well-being [39], while another study has confirmed that there is a positive effect of education on sleep in undergraduates [40]. However, there is a gap in the literature in terms of understanding how premedical students perceive their sleep versus their actual sleep patterns. Although one study comes close to answering this question by examining the beliefs of sleep hygiene in medical students compared to their actual sleep practices [41], the behavior of premedical undergraduates is yet to be explored and understood.

Current validation studies suggest that while PSG remains the gold standard, Fitbit is capable of monitoring various sleep parameters that do not significantly deviate from PSG measurements [42-44]. The overall objective of our research was to understand the sleep patterns of premedical students and create interventions to promote positive behavioral changes. Our pilot study focuses on answering the following 2 research questions
questions (RQs). What are the sleep patterns of premedical undergraduate students (RQ1)? Would the study protocol be feasible to examine the perception of sleep quality and promote sleep behavioral changes in premedical undergraduate students (RQ2)?

**Methods**

**Setting**

This study was conducted with premedical students from a Medical Sciences Baccalaureate Program (MSBP) at the University of Cincinnati, an R1: doctoral university and an undergraduate College of Medicine in the Midwest United States. The MSBP is a program designed to help premedical students prepare for professional graduate schools in the health care field. The 4-year program consists of 376 students as of October 2022.

**Study Design**

To address RQ1, a “Well-being and Sleep Survey” (the well-being survey hereafter) was distributed to collect subjective sleep patterns. To address RQ2, a small group of participants (MSBP premedical students) were recruited to evaluate whether the study protocol (Figure 1) was feasible and could generate valid data for analysis. Specifically, the participants put on the Fitbit Inspire 2 wearable watch to monitor their sleep patterns. These participants were then invited to a semistructured interview to understand their Fitbit app experience. As shown in Figure 1, during the 2-week pilot study period, participants progressed through preassessment (week 0), midassessment (week 1), and postassessment (week 2) periods. The well-being survey was filled out by participants at all 3 fixed time points. During week 0, a semistructured interview was conducted assessing participants about their typical sleep habits and their previous experience with wearable devices. Participants at this point were also prompted to fill out a form daily, reflecting their thoughts regarding the Fitbit sleep score they received the night before. On week 1, a midassessment was performed in which the participants were asked to speak about their previous week’s experience with the aid of the daily reflections they had previously filled for days 1-7. In order to prompt the participants, researchers displayed a line chart of the 7 responses gathered from participants’ daily forms to project their sleep score trends over the week. Researchers also listed all participants’ free-text responses so that they could elaborate specifically about their sleep corresponding to certain dates. Lastly, in week 2, participants reviewed the daily reflections for days 8-14, which was followed by a semistructured interview.
Figure 1. Flowchart of the study protocol tested on 5 premedical participants during the 2-week pilot study with Fitbit Inspire 2. Week 0 started with preassessment, conducted in person at the University of Cincinnati Medical Sciences building. Week 1 midassessment and Week 2 postassessment were conducted virtually through Microsoft Teams meeting. The data collection was conducted in 2 cycles. Participants 1, 2, and 3 began the 2-week protocol in the last week of July 2022 and finished in the second week of August 2022. Participants 4 and 5 began the 2-week protocol in the first week of September 2022 and finished in the third week of September 2022. All of the participants had varying start dates during this time frame, but each participant was available for 10 weekdays and 4 weekend days during the duration of the 2-week data collection. PGHD: patient-generated health data; PSQI: Pittsburg Sleep Quality Index.

Ethics Approval

This study was reviewed and approved by the University of Cincinnati Institutional Review Board (approval 2021-1119) and determined as not human subject research due to its collection on de-identified data and the focus on quality improvement. This institutional review board determination resulted in the avoidance of collecting detailed student demographics such as sex and age and the deidentification of all participant information in the data.
Participant Recruitment

The survey participants were recruited on a voluntary basis via email and GroupMe messages through the research team’s professional and personal network. The small group of participants for the protocol evaluation was selected from the survey participants as a convenience sample. This study was expected to target 100 survey participants and 5 protocol evaluation participants.

Data Collection

The well-being survey was distributed among all cohorts (years 1, 2, 3, and 4 students) of the MSBP program (N=376). The survey consisted of PSQI items (1-9) adjusted to items (1-11) to accommodate the web-based survey format due to some combined line items in the original paper format, requiring different lines in the web-based survey. The response components combined to calculate a global PSQI score (with a Cronbach \( \alpha \) of .83) [8], which suggests a high degree of internal consistency. These items were followed by a few questions assessing students’ demographics and their perception of their sleep quality. The well-being survey was created in Microsoft forms with survey questions as listed in Table 1. For the protocol evaluation, 5 students were selected from the well-being survey pool. In addition to the sleep patterns collected from the self-reported PSQI survey and the objective Fitbit wearables, these 5 participants’ feedback on the protocol was collected through semistructured interviews. The interview was recorded and conducted on a virtual platform and further analyzed thematically [45]. For the Fitbit data collection, 5 Fitbit Inspire 2 watches were deployed—one for each participant. Fitbit data were collected by linking the watch to the participant’s Fitbit web account. During the semistructured interview in the midassessment, the participants were instructed to log into their account on Fitbit.com. Next, they were asked to navigate to “Settings,” select “Data Export,” and then click on “Request Data.” This generated an email request to the participant to “Confirm Export Request.” Finally, after confirmation, the participant received a downloaded data zip file with JSON files containing complete archive data. The participants were then prompted to email the zip file to the researcher. This process was again repeated during the postassessment. Table 1 identifies each question sequentially, as it was presented in the web-based Microsoft form, as well as categorizes it as either a PSQI question or a demographic question. PSQI questions are directly derived from PSQI questionnaire items (1-9), which are adjusted to items 1-11 in our survey to accommodate the web-based format as compared to the original PSQI questionnaire paper format. The description lists exactly how the questions were asked to premedical MSBP students. The response type specifies whether survey participants responded to the question by inputting free text or selecting their answers on the Likert scale or multiple choice.
Table 1. Well-being survey questions that were administered to 376 students in the Medical Sciences Baccalaureate program at the University of Cincinnati.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Category</th>
<th>Description</th>
<th>Response type/scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PSQI</td>
<td>During Week 01, when have you usually gone to bed at night?</td>
<td>Free text</td>
</tr>
<tr>
<td>2</td>
<td>PSQI</td>
<td>During Week 01, how long (in minutes) has it usually taken you to fall asleep each night?</td>
<td>Free text</td>
</tr>
<tr>
<td>3</td>
<td>PSQI</td>
<td>During Week 01, when have you usually gotten up in the morning?</td>
<td>Free text</td>
</tr>
<tr>
<td>4</td>
<td>PSQI</td>
<td>During Week 01, how many hours of actual sleep did you get at night? (This may be different from the number of hours you spend in bed)</td>
<td>Free text</td>
</tr>
<tr>
<td>5</td>
<td>PSQI</td>
<td>During Week 01, how often have you had trouble sleeping because you...</td>
<td>Likert</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a...cannot get to sleep within 30 min</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>b...wake up in the middle of the night or early morning</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c...have to get up to use the bathroom</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d...cannot breathe comfortably</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>e...cough or snore loudly</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>f...feel too cold</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>g...feel too hot</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>h...had bad dreams</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>i...have pain</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PSQI</td>
<td>If you have other reasons for having trouble sleeping, please describe...</td>
<td>Free text</td>
</tr>
<tr>
<td>7</td>
<td>PSQI</td>
<td>How often during Week 01 have you had trouble sleeping because of Q6?</td>
<td>Likert</td>
</tr>
<tr>
<td>8</td>
<td>PSQI</td>
<td>How often during Week 01 have you had trouble sleeping because of this?</td>
<td>Likert</td>
</tr>
<tr>
<td>9</td>
<td>PSQI</td>
<td>During Week 01, how often have you taken medicine (prescribed or “over the counter”) to help you sleep?</td>
<td>Likert</td>
</tr>
<tr>
<td>10</td>
<td>PSQI</td>
<td>During Week 01, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?</td>
<td>Likert</td>
</tr>
<tr>
<td>11</td>
<td>PSQI</td>
<td>During Week 01, how much of a problem has it been for you to keep up enough enthusiasm to get things done?</td>
<td>Likert</td>
</tr>
<tr>
<td>12</td>
<td>Demographics</td>
<td>Year in MSBP(^b) program?</td>
<td>Multiple choice</td>
</tr>
<tr>
<td>13</td>
<td>Demographics</td>
<td>Sleep quality:</td>
<td>Likert</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Do you have concerns about your sleep quality?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>How much role does stress play in your sleep quality?</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Demographics</td>
<td>Have you ever seen a sleep specialist for sleep problems?</td>
<td>Likert</td>
</tr>
<tr>
<td>15</td>
<td>Demographics</td>
<td>Have you ever worn a wearable device to track your sleep?</td>
<td>Multiple choice</td>
</tr>
<tr>
<td>16</td>
<td>Demographics</td>
<td>Are you interested in joining a Fitbit usability study?</td>
<td>Multiple choice</td>
</tr>
<tr>
<td>17</td>
<td>Demographics</td>
<td>If yes to Q16, then enter your email.</td>
<td>Free text</td>
</tr>
</tbody>
</table>

\(^a\)PSQI: Pittsburg Sleep Quality Index.

\(^b\)MSBP: Medical Sciences Baccalaureate Program.

Survey Data Analysis

After MSBP premedical students completed the well-being survey, the responses were analyzed as follows. First, the PSQI items (questions 1-11) were scored based on PSQI scoring guidelines. Then, a global PSQI score was assigned to each survey participant. Next, 3 analyses were made: (1) comparing student distributions with good PSQI scores (≤5) versus poor PSQI scores (≥5) categorized by each cohort, (2) detecting the consistency between the perception of sleep quality and the PSQI score levels, and (3) finding the PSQI score differences on the perception of stress effect levels. Statistical analyses were performed to test the distribution homogeneity of each cohort using the chi-square test. The consistency between the perception of sleep quality and PSQI score will be tested using Cohen \(\kappa\) by grouping PSQI scores into corresponding sleep quality levels. For comparing the global PSQI score differences between stress effect levels, global PSQI scores were first treated as a continuous variable and the equality of medians on each group was tested using the Kruskal-Wallis test (nonparametric equivalent test of 1-way analysis of variance) to show any statistical difference. Furthermore, the post hoc Dwass-Steel-Critchlow-Fligner multiple comparisons test was applied to show the pairwise comparisons if at least one group median was significantly different from others.
Protocol Evaluation

The protocol evaluation was 2-fold. First, each assessment had an expected duration to avoid overwhelming the participants. The expected durations of the preassessments, midassessments, and postassessments were 30 minutes, 30 minutes, and 60 minutes, respectively. The postassessment was longer due to the semistructured interview. Second, the semistructured interview collected the participants’ feedback on the study protocol in 4 areas: (1) duration of the 3 (pre, mid, and post) assessments, (2) frequency of administering the well-being surveys, (3) completeness of Fitbit data, and (4) other suggestions. An additional analysis was conducted to assess the degree to which Fitbit data aligns with the self-reported data obtained through surveys by using Pearson correlation coefficients where $>0.7$ is strong correlation, $[0.3]-[0.7]$ is moderate correlation, and $<0.7$ is poor correlation (of note, the vertical lines present the absolute value sign).

Results

Survey Responses and Analysis Results

The survey responses (n=103) demonstrated that premedical MSBP students sleep an average of 7.1 hours each night with 81.3% habitual sleep efficiency (average hours slept versus hours spent in bed, as defined in PSQI). Those who experienced trouble sleeping commonly expressed reasons such as not being able to sleep within 30 minutes, waking up in the middle of the night or early morning, anxiety, stress, and a restless mind. Among the cohorts (Figure 2), no significant difference was found in the median PSQI scores as indicated by the Kruskal-Wallis test at a .05 significant level ($P=.11$). Similarly, the proportion of students who had good versus poor sleep was not significantly different as indicated by the chi-square test ($P=.48$). This concludes that the distributions have no significant differences in the good or bad PSQI scores among the student cohorts. It is worth noting that the response rates of the survey for the first-, second-, third-, and fourth-year cohorts were 38% (33/88), 23.4% (25/107), 23% (20/86), and 26% (25/95), respectively. Further, the same analysis was applied to the categorical variables collected in the survey data (Multimedia Appendix 1), including 2 non-PSQI variables (perceived sleep quality and perceived stress). The only significant difference was that the first-year students tended to take medicines to help themselves to sleep unlike the other cohorts ($P=.03$). However, there was no significant difference between the low frequency group (less than once a week and no medicine) and the high frequency group (once a week or more). To demonstrate the consistency between the perceived sleep quality and the PSQI scores, we assumed that students who perceived “no problem at all” will receive PSQI global scores $<5$, those who perceived “only a very slight problem” will receive PSQI global scores of $5-7$, those who perceived “somewhat of a problem” will receive PSQI global scores of $8-10$, and those who perceived “a very big problem” will receive PSQI global scores $\geq 11$. The Cohen $\kappa$ (with 1 being a perfect consistency) showed that participants’ perceived sleep quality only had a slight consistency with the PSQI scores ($\kappa=0.19$). Table 2 shows the contingency of these 2 variables with an overpositive tendency from participants, that is, the bigger the problem the participants perceived in their sleep quality, the poorer the PSQI score they would receive. However, the table shows some inconsistencies. When participants perceived no problem at all in their sleep quality, 50% (14/28) of them actually had a nonoptimal PSQI score. When the participants thought there was only a very slight problem in their sleep, about a quarter of them (11/43, 26%) had a poor PSQI score. When the participants appeared to recognize that they had somewhat of a problem (or a big one) in their sleep, the majority of them had nonoptimal PSQI scores. Finally, the comparisons between perceived stress impact and median global PSQI scores in each perceived stress level were analyzed (Figure 3). The Kruskal-Wallis test shows there is at least one median that is significantly different ($P<.001$) from others at the .05 significant level. The post hoc Dwass-Steel-Critchlow-Fligner multiple comparisons was applied and showed that the students who perceived “a very big problem” on the stress effect had significantly greater PSQI scores than the students who perceived “no problem at all” ($P=.048$) and “only a very slight problem” ($P<.001$). Similarly, for the students who perceived “somewhat of a problem,” the stress effect was significantly greater than that in students who perceived “only a very slight problem” ($P=.001$) but not significantly different from the “no problem at all” stress level (Table 3). This indicates that participants’ perceived stress effects have high correlation with their PSQI scores, but it may not be the only effect at play.
Figure 2. Stacked bar chart display of premedical student responses based on good sleep (global PSQI<5) and poor sleep (global PSQI≥5) with percentages within different medical sciences baccalaureate program cohorts. PSQI: Pittsburg Sleep Quality Index.

Table 2. The contingency table of Medical Sciences Baccalaureate program premedical students’ perceived sleep quality by their actual global Pittsburg Sleep Quality Index score categories.

<table>
<thead>
<tr>
<th>Pittsburg Sleep Quality Index score (perceived sleep quality)</th>
<th>Optimal (&lt;5)</th>
<th>Slightly poor (5-7)</th>
<th>Borderline poor (8-10)</th>
<th>Poor (11+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No problem at all (total score=28)</td>
<td>14 (13.6)</td>
<td>9 (8.7)b</td>
<td>4 (3.9)b</td>
<td>1 (1)b</td>
</tr>
<tr>
<td>Only a very slight problem (total score=43)</td>
<td>11 (10.7)b</td>
<td>21 (20.4)</td>
<td>9 (8.7)b</td>
<td>2 (2)b</td>
</tr>
<tr>
<td>Somewhat of a problem (total score=28)</td>
<td>1 (1)b</td>
<td>13 (12.6)b</td>
<td>7 (6.8)</td>
<td>7 (7)b</td>
</tr>
<tr>
<td>A very big problem (total score=4)</td>
<td>0 (0)</td>
<td>0 (0)b</td>
<td>1 (1)b</td>
<td>3 (3)</td>
</tr>
</tbody>
</table>

a0-4 (<5)=optimal sleep; 5-7=slightly poor; 8-10=borderline poor; ≥11=poor with overall percentages.
bInconsistency between the perceived sleep quality and Pittsburg Sleep Quality Index scores.

Figure 3. Boxplot of perceived stress effect on sleep by students compared to their global Pittsburg Sleep Quality Index scores. PSQI: Pittsburg Sleep Quality Index.
Process Evaluation and Participant Feedback

Overall, the participants reported the Fitbit app as a helpful tool in understanding their sleep. They were especially excited to visualize their sleep stage breakdown and obtain an overview of their sleep by simply looking at their Fitbit sleep score. The participants’ feedback on the study protocol was summarized in the following 4 areas.

First, all assessments were conducted in the expected duration, that is, 30 minutes for the preassessment and midassessment and 60 minutes for the postassessment. The web-based mode of midassessments and postassessments did not create any barriers to the data collection. The feedback also revealed that the interview component in all 3 assessments was effective and served different purposes. Specifically, the interview in the preassessment made the participants think about their previous experiences regarding sleep and patient-generated health data devices. The midassessment interview, which reviewed daily reflections at a glance, invoked awareness among participants. The interview in the postassessment was the most comprehensive as it gathered thoughts on daily reflections and assessed Fitbit data comprehension. The administered questionnaires in preassessments, midassessments, and postassessments, as shown in Figure 1, had a 100% response rate because participants were required to complete it during the semistructured interview. The global PSQI scores for 5 participants across the 2 weeks are shown in Figure S1 in Multimedia Appendix 2. The daily reflections also had a 100% response rate, as several reminder emails were sent out to ensure data collection.

Second, the participants did not have a strong opinion about the frequency of the well-being survey administration (3 times over a 2-week period in this study). However, the survey scores did not display much variation or clear trends. Although this can be due to the small group of participants (n=5), it may be more ideal to conduct the survey with more spaced durations.

Third, the Fitbit data provided notable variation daily and was able to grasp all 14 days well for the most part. Some participants experienced cases during which Fitbit would sense bed and wake time but would not be able to provide detailed analytics on the sleep stages or on the sleep score. For these participants, some of the issues included wearing the Fitbit too snug or loose or wearing the Fitbit only right before going to bed. In other cases, participants simply appeared to forget to put on their Fitbit watch after removing it for charging, leading to a void of data points. The participants were fairly adherent to completing daily reflections on time and found that a daily email reminder from the research team helped greatly with this task.

Lastly, the participants identified that the Fitbit app tends to provide vague descriptions. For example, the biggest concern raised was not being able to understand how the sleep score was calculated. This made participants question the score’s reliability and accuracy. Similarly, some participants highlighted redundancy of information (sleep stage condensed cycling), whereas others suggested presenting information in a more visually appealing manner (pie chart to see sleep stage breakdown), individualized manner (suggesting sleep recommendations based on score trends), and sleep score explainability (being specific as to why a score may have dropped/raised on a day-by-day basis).

Finally, the degree to which Fitbit data aligns with the PSQI self-reported data obtained through surveys was assessed. Overall, across the 2-week period, there was a moderately negative correlation ($r=-0.598$) calculated from the data shown in Table S1 of Multimedia Appendix 2. In context, this means high Fitbit scores, indicating good sleep, were correlated with low global PSQI scores, also indicating good quality of sleep. Week 1 had a strong negative correlation ($r=-0.87$) compared to Week 2, which had a moderate negative correlation ($r=-0.648$).

**Discussion**

**Principal Findings**

This study presents the sleep patterns of premedical students through a sleep and well-being survey and evaluates a study protocol in preparation for a large-scale study. The survey showed that although the participants did have an average of 7 hours of sleep in general, only a quarter of them had good sleep quality. There was no difference in sleep quality among the student cohorts. Furthermore, there was an inconsistency between the perceived sleep quality by the students versus what their PSQI scores reflected. The poor sleep quality experienced by unaware students points to the need for self-monitoring through wearables such as Fitbit devices to increase sleep quality awareness. Lastly, a comparison of stress effect on sleep with PSQI scores establishes that sleep is not the only dominant factor in PSQI. The participants who perceived stress to play no role in their sleep still displayed a wide range of PSQI scores. Our findings show that the study protocol is highly feasible in terms of assessment durations and interview effectiveness. The usage of Fitbit devices did increase participants’ awareness...
regarding their sleep based on the qualitative feedback, along with their overall understanding of sleep factors. Participants keeping track of their sleep via daily reflections also aided in their cognizance. Based on the feedback, the study protocol can be improved by extending the duration, educating individuals about the Fitbit-wearing behaviors, and increasing the sleep score interpretability.

In assessing how the Fitbit sleep score data align with the global PSQI scores, we found a moderate negative correlation. This makes sense because higher Fitbit scores, suggesting good sleep, aligned with lower global PSQI scores, which indicates good sleep quality. Therefore, achieving moderate to high correlation displays a good potential for Fitbit sleep scores to serve as a proxy for PSQI global scores. Although week 1 gathered a stronger correlation than week 2, we believe that this result may be due to greater missing data in week 2, either from fatigue or nonadherence.

**Implications**

Our study shows that only 25.2% (26/103) of the participating premedical students had good sleep quality, as indicated by the PSQI scores. This proportion is lower than that (34.1%) previously reported among college students in general [10]. This leads us to a possible inference that premedical students may have sleep patterns different from those of college students overall. Our study also introduces a new direction, in which sleep patterns are not only measured but further analyzed to determine discrepancies between perception versus actual sleep quality. Prior studies [17,46] looking at sleep patterns in medical school students or undergraduate college students in general could also benefit from this approach and take their sleep analysis to the next step. For example, a 2016 study conducted on undergraduates at the University of California explored relations between sleep and mental health in individuals with healthy sleep habits [17]. Hence, further understanding the sleep perceptions of these healthy students could be used to create a benchmark to which other types of sleepers can be measured relatively. Another application could include a longitudinal study, similar to the one proposed by the University of South Wales in Australia [46]. This 5-year study provides insights into the mental health of medical students, with sleep being one of the factors measured via Fitbit. By incorporating the concept of sleep and well-being perceptions, the longitudinal approach may provide a better understanding of whether perceptions stay consistent or change over time.

The increase in awareness regarding sleep due to Fitbit usage mentioned by the participants also opens an area for possible interventions. As previous literature has reported, when individuals are educated in a personally relevant manner, their awareness leads to behavior change [47-49]. An important component of behavior change is self-efficacy, which is the perception of one’s ability to achieve certain outcomes [50-52]. Hence, interventions can begin with Fitbit usage to increase participant awareness regarding why they need to make a change, establish how certain health behaviors will lead to expected outcomes, and then provide ways for participants to reach those outcomes [53,54]. Conversely, I study has tested a habit loop model, which is based on cue, routine, and reward [55]. The Fitbit usage behavior in terms of the habit loop model can be as follows: trust in Fitbit accuracy of physical and sleep data (cue), intensity of Fitbit use (routine), and adjustment of physical and sleep behaviors (reward). Although this study shows that trust in Fitbit produced little change in sleep and health behaviors, future studies should be built upon this foundation to examine the potential behavioral change through the supplementation of educational information to participants.

In the evaluation of the protocol itself, there were inconsistencies in obtaining sleep scores from Fitbit at times. A pre-existing measure during this study was placing instructions in the daily reflection reminder email to contact the research team immediately if they received no score in order for the research team to troubleshoot. Some other troubleshooting measures could help ensure participants wear their Fitbit at all times. For example, the study team can propose a standardized charging time that prompts participants to remove and charge their Fitbits during our week 1 midassessment session and wear the Fitbit again at the conclusion of the meeting. This can prevent participants from forgetting to wear their Fitbits during the study period. Not all participants explored all the features of Fitbit’s sleep interface. This points to the need to give participants a tutorial about the various types of analytical information available that participants can browse through to understand their sleep patterns. Finally, for the Fitbit data retrieval process, researchers asked the participants to download their data from the Fitbit.com website in week 1 and week 2 assessments. This proved to be a time-consuming process; therefore, the protocol was updated to begin data download at the beginning of the assessment meeting, which allowed the download to be finished by the end of the meeting. These lessons learned help to refine our study protocol and will lead to higher data quality in our large-scale study.

**Limitations**

This study is limited in several ways. First, the assessment was conducted with premedical students from a single program in a single institution. Hence, it is possible that the findings were limited to the student demographics in our institution. Additionally, the well-being survey administered in this study did not ask for their sex, race, or age due to the constraints in the institutional review board protocol. Thus, this analysis does not demonstrate whether demographic factors are linked with certain sleep patterns. Moreover, this study did not consider the academic performance of the participants and its relationship with sleep patterns. Demographic information and academic performance will be collected in our future studies with an updated institutional review board protocol. Next, we only recruited 5 students via convenience sampling for protocol evaluation. Small sample sizes introduce the possibility of skewed opinions. However, it should be noted that the goal of this study was to test the protocol rather than any hypothesis. Additionally, the Fitbit usage habits demonstrated by the pilot study group point to the need to improve the Fitbit usage protocol. For example, students were advised to always wear their Fitbits except while charging. However, some students after charging their Fitbit forgot to wear it overnight, thereby leading to nonadherence. Another instance involved a participant using the “sleep now” function offered by the watch to track
sleep, even though the participants were not instructed on that feature’s use. Using that function did dysregulate the visualization data, which returned to normal once the participant was asked to terminate the function’s usage.

**Future Work**

With the findings gathered from this pilot study, we hope to expand to a larger experiment within the MSBP and expand to premedical programs at other institutions. We hope to address sleep quality discrepancies and determine a feasible solution for sleep hygiene improvement in premedical students. Another future work involves the calculation of Fitbit sleep scores especially when the sleep data are collected but the sleep score is missing. In this study, we tried to explore how the Fitbit score was calculated and attempted to assess trends. Fitbit informs everyone that their sleep score is devised from 3 pieces of information: time asleep and awake, deep and rapid eye movement sleep, and sleeping heart rate and restlessness. However, it is unclear how exactly the sleep score values are formed. Therefore, it was difficult to generate actionable information. Hence, this topic should be further explored to determine whether sleep scores are indeed valid measures of sleep.

**Conclusions**

We conducted a pilot study to understand the sleep patterns of premedical undergraduate students and evaluated the study protocol. The survey results showed a generally lower sleep quality for this student population and posited a gap between the perceived sleep quality versus PSQI-measured sleep quality. This gap may be solved by the awareness raised by Fitbit usage. We will revise the study protocol based on the lessons learned and health behavior theories and conduct a large-scale experiment to use Fitbit as an informatics solution to promote healthy sleep behaviors.

**Acknowledgments**

This research did not receive any funding. We thank Dr Anil Menon, Director of the Medical Sciences Baccalaureate Program at the University of Cincinnati College of Medicine, for supporting the implementation of this study along with the student tribunal for the recruitment efforts. In addition, we thank Dr Claudia Rebola, Co-Principal Investigator of the WorkWell project, for providing Fitbit watches for data collection.

**Data Availability**

All data generated or analyzed during this study are included in this paper and multimedia appendices.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Survey data.
[XLSX File (Microsoft Excel File), 32 KB - formative_v8i1e45910_app1.xlsx]

Multimedia Appendix 2

Supplementary data.
[DOCX File, 196 KB - formative_v8i1e45910_app2.docx]

**References**


Abbreviations

MSBP: Medical Sciences Baccalaureate Program
PSG: polysomnography
PSQI: Pittsburg Sleep Quality Index
RQ: research question
Acceptability and Utility of a Digital Group Intervention to Prevent Perinatal Depression in Youths via Interactive Maternal Group for Information and Emotional Support (IMAGINE): Pilot Cohort Study

Keshet Ronen¹, MPH, PhD; Anupa Gewali¹, MPH, MSW; Kristin Dachelet¹, MSW; Erica White¹, MPH; Marimirca Jean-Baptiste¹; Yolanda N Evans², MPH, MD; Jennifer A Unger³, MPH, MD; S Darius Tandon⁴,⁵, PhD; Amritha Bhat⁶, MPH, MBBS, MD

¹Department of Global Health, University of Washington, Seattle, WA, United States
²Division of Adolescent Medicine, Seattle Children’s Hospital, Seattle, WA, United States
³Department of Obstetrics and Gynecology, Warren Alpert Medical School at Brown University, Providence, RI, United States
⁴Department of Medical Social Sciences, Northwestern Feinberg School of Medicine, Chicago, IL, United States
⁵Center for Community Health, Northwestern Feinberg School of Medicine, Chicago, IL, United States
⁶Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

Corresponding Author:
Keshet Ronen, MPH, PhD
Department of Global Health
University of Washington
3980 15th Avenue North East
Seattle, WA, 98195
United States
Phone: 1 2066854363
Email: keshet@uw.edu

Abstract

Background: Perinatal depression (depression during pregnancy or the first year postpartum) affects 10%-25% of perinatal individuals, with a higher risk among youths aged <25 years. The Mothers and Babies Course (MB) is an evidence-based intervention for the prevention of perinatal depression, grounded in cognitive behavioral therapy, attachment theory, and psychoeducation.

Objective: We developed a digital adaptation of MB (Interactive Maternal Group for Information and Emotional Support [IMAGINE]) and evaluated it in a pre-post mixed methods pilot among young perinatal people in the United States.

Methods: IMAGINE was a structured digital group of up to 7 participants, with scheduled MB content and open discussion for 12 weeks, facilitated by a social worker. Scheduled content included asynchronous SMS text messages, graphics, prerecorded videos, mood polls, and optional weekly synchronous video calls. Eligible participants were pregnant or ≤80 days postpartum, aged 16 to 24 years, had access to a smartphone, spoke English, and had a Patient Health Questionnaire score <10. Participants were recruited throughout the United States from August 2020 to January 2021 through paid social media ads, in-person outreach at clinics, and respondent-driven sampling. Participants completed quantitative questionnaires at enrollment and 3 months, and qualitative interviews at 3 months. We determined uptake, acceptability (by Acceptability of Intervention Measure score), and utility (by use of cognitive behavioral therapy skills). We compared depression symptoms (by Patient Health Questionnaire score), social support (by abbreviated Social Support Behavior score), and perceived stress (by Perceived Stress Score) between enrollment and follow-up by paired 2-tailed t test.

Results: Among 68 individuals who contacted this study, 22 were screened, 13 were eligible, and 10 enrolled, for an uptake of 76.9%. Furthermore, 4 (40%) participants were pregnant at enrollment. Participants had a median age of 17.9 (IQR 17.4-21.7) years, 6 (67%) identified as Black, 5 (56%) Latinx, and 6 (67%) using Medicaid health insurance. Further, 9 (90%) participants completed follow-up. Among these, the mean acceptability score was 4.3 out of 5 (SD 0.6) and all participants said they would recommend IMAGINE to a friend. Participants reported using a median of 7 of 11 skills (IQR 5-7 skills) at least half the days. We found no significant changes in depression symptoms, perceived stress, or social support. Qualitatively, participants reported one-to-one support from the facilitator, connection with other parents, and regular mood reflection were especially helpful aspects.
of the intervention. Additionally, participants reported that the intervention normalized their mental health challenges, improved their ability to manage their mood, and increased their openness to mental health care.

**Conclusions:** This pilot study provides promising evidence of the acceptability and utility of IMAGINE among perinatal youths. Our study’s small sample size did not detect changes in clinical outcomes; our findings suggest IMAGINE warrants larger-scale evaluation.

*(JMIR Form Res 2024;8:e51066) doi: 10.2196/51066*

**KEYWORDS**

perinatal depression; youth; mHealth; digital health; acceptability; utility; depression; pilot study; pregnancy; postpartum; prevention; cognitive behavioral therapy; psychoeducation; mixed methods; manage; mood; mobile phone

**Introduction**

Perinatal depression, defined as depression during pregnancy or up to 1 year after childbirth, affects an estimated 10%-25% of birthing people in the United States [1,2]. Untreated, perinatal depression can have long-term negative impacts on birthing parents and newborns, including elevated risk of suicide [3], preterm birth [4], low birth weight [5], and impaired infant attachment [6]. Young parents (aged 15 to 24 years), those with low income, and people of color experience elevated risks of perinatal depression [7,8].

Despite many programs aimed at reducing the incidence of adolescent pregnancy in the United States, nearly 23% of all pregnancies occurred in people aged 15 to 24 years in 2020 [9]. Young parents are more likely to have risk factors that place them at greater risk for perinatal depression, including unplanned pregnancy, social isolation, and intersecting social determinants of health [10,11]. Many barriers to mental health care exist for perinatal youths, including lack of financial resources to pay for treatment, lack of access to or money for transportation, and difficulty finding childcare and missing work to attend appointments [8]. Young parents are therefore a key group needing innovative and tailored perinatal mental health support.

Evidence-based interventions have been developed to prevent and treat perinatal depression. The Mothers and Babies Course (MB) is a cognitive behavioral therapy (CBT) program developed for low-income racial minority women in the United States [12]. MB can be delivered to individuals or groups and is focused on building participants’ skills in modifying their thoughts, social contact, and pleasant activities. Further, 4 randomized controlled trials of the MB program found it reduced depressive symptoms in the perinatal period, and it has been rolled out at scale in the United States through home visiting programs [13-18]. MB has been recognized by the US Preventive Services Task Force as an evidence-based intervention that should be recommended for individuals at high risk of perinatal depression [19]. Despite these advances, there are barriers to accessing interventions such as MB, due to regional provider shortages, lack of transportation and childcare, clients’ difficulty committing to a prespecified time to attend sessions, and experienced or internalized stigma of seeking mental health care [20]. These barriers may be higher among those at the highest risk of perinatal depression, and barriers were further exacerbated by the COVID-19 pandemic and consequent physical distancing protocols [21-24].

Mobile phones can be used to offer more accessible mental health support to young parents (known as mobile health [mHealth]). mHealth interventions have shown promising results in improving perinatal mental health [25-30], and young people in the United States are an ideal audience for such programs: nearly all Americans aged 18 to 29 years own a phone, and 96% own a smartphone [31]. mHealth can eliminate the need to travel to appointments, and asynchronous mHealth programs allow patients to access care within their schedules, which can be especially helpful to parents [32]. Group mHealth interventions allow new parents to connect with and share ideas with peers, which can lower feelings of social isolation and increase social support and mental well-being [33].

We developed a digital adaptation of MB for the prevention of perinatal depression in youths, named Interactive Maternal Group for Information and Emotional Support (IMAGINE) [34]. In this paper, we present a mixed methods evaluation of a pilot study of the IMAGINE intervention, assessing the uptake, acceptability, and utility of IMAGINE.

**Methods**

**Study Design**

We conducted a single-arm pilot study with pre-post mixed methods evaluation.

**Study Population and Recruitment**

Participants were eligible to participate in the IMAGINE pilot if they were: pregnant or ≤180 days postpartum, aged 16 to 24 years during pregnancy, had daily access to a smartphone, and were comfortable conducting study visits and reading and responding to social media messages in English. Individuals who exhibited elevated depression symptoms at screening (Patient Health Questionnaire [PHQ-9] score ≥10 [35]) were excluded and referred to individual clinical care.

Participants were recruited between August 2020 and January 2021, through 3 main methods. First, study information was shared through paid, targeted advertisements on Instagram and Facebook throughout the United States. Specific parameters used to target participants were: female sex, aged 16 to 25 years, located anywhere in the United States. Second, this study’s team identified health care providers and community-based organizations in several cities in the United States (Seattle, WA; Olympia, WA; Philadelphia, PA; and Temple, TX) and provided materials for staff at these organizations to promote this study by distributing flyers. Finally, we used respondent-driven...
Potential participants who learned about this study through any method contacted this study by phone call, text message, email, Instagram message, or by sending a message through this study’s website. Study staff then contacted potential participants to conduct eligibility screening by phone or video call. If participants were ineligible due to elevated depression symptoms at screening, study staff shared the National Crisis Line phone number and offered support to find treatment near their location. This study’s team psychiatrist was available to support linkage with resources.

Informed Consent
Informed consent was obtained by study staff for screening and enrollment. Participants provided verbal consent to participate in eligibility screening. Eligible participants provided written consent for study enrollment, using a web-based consent form on REDCap (Research Electronic Data Capture; Vanderbilt University) [37].

Study Visit Schedule
Participants enrolling in the IMAGINE pilot attended 2 study visits by video call: enrollment and 3-month follow-up after completion of the intervention. Quantitative data were collected at enrollment and follow-up using a REDCap web-based questionnaire. Qualitative interviews were additionally conducted at follow-up.

Intervention
The IMAGINE intervention was a facilitated digital group adaptation of the evidence-based MB program [12], developed through a human-centered design process described elsewhere [34]. MB content focuses on engagement in pleasant activities, healthy thinking, and social support, and is grounded in CBT, attachment theory, and psychoeducation. IMAGINE was delivered using the messaging platform, Slack (Slack Technologies, Salesforce Inc), to groups of up to 10 participants, facilitated by a member of this study’s team with Master’s-level training as a social worker. Groups ran for approximately 12 weeks. Participants were grouped based on the timing of enrollment: groups were filled sequentially as participants enrolled. Guided by our formative work [34], the intervention consisted of multimedia adaptation of MB content, delivered through 5 components. First, MB session content was delivered asynchronously, through short SMS text messages, summary graphics, and prerecorded videos sent approximately 4 times per week. Messages were designed to promote group discussion or personal reflection. Participants were encouraged, but not required, to participate in group discussions by sending messages and reacting to other participants’ messages, at a time that was convenient for them. Second, an automated “mood poll” was sent to each participant individually 3 times per week, prompting the participant to reflect on their mood, activities, thoughts, and social contacts. Third, the facilitator was available for individual messaging through Slack. This was used for the facilitator to answer questions and send messages when a participant showed low engagement in other parts of the intervention. Fourth, participants could send messages on topics beyond the MB curriculum, through separate Slack “channels” (parallel conversations all members had access to): “ask an expert,” where participants could send questions for members of this study’s team with expertise in obstetrics and psychiatry; “random,” where participants could share any content; and “references,” where the facilitator posted graphics summarizing intervention content and links to resources. Fifth, in addition to asynchronous messaging content, the facilitator held a weekly 1-hour synchronous group video call, using the videoconferencing platform, Zoom (Zoom Video Communications, Qumu Corporation). Participation in the call was optional, to reduce barriers to participation due to scheduling and attendance challenges. No new content was delivered on the call, but participants could ask questions, share experiences, and receive support from the facilitator and other group members. Intervention content was developed in advance of the intervention and manually sent by the facilitator, except for mood polls, which were automatically scheduled within Slack. The facilitator could exercise discretion in message pacing based on participant feedback and questions during the intervention period. All 5 elements of the intervention were considered to be active components in engaging MB’s mechanism of action.

Quantitative Data Collection
Recruitment Log
A spreadsheet was used by study staff to record participants who contacted the IMAGINE study and their completion of eligibility screening.

Screening Questionnaire
Screening was completed verbally by phone or video call and responses were entered by study staff into an electronic questionnaire using REDCap, hosted at the University of Washington [37]. The screening questionnaire ascertained pregnancy status, age, access to a smartphone, comfort in English, and depression symptoms by PHQ9.

Enrollment Questionnaire
Enrollment of eligible participants was conducted either immediately following consent or at a separate scheduled visit, based on participant preference. The enrollment questionnaire was administered using REDCap by study staff through Zoom or phone calls. If conducted by video call, study staff screen-shared the REDCap questionnaire and read each question aloud so the participant could see and hear the questions and responses; data were entered by study staff. The enrollment questionnaire ascertained demographic characteristics and technology access. It also included an abbreviated 12-item version of the Social Support Behavior (SSB) instrument [38] to ascertain social support (score range 5-60). We used an abbreviated version of SSB to reduce the burden of data collection on participants. This version of the SSB has not been psychometrically validated. We used the Perceived Stress Score (PSS-4) instrument [39] to ascertain perceived stress (score range of 0-16). This instrument has previously been used in MB studies and is recommended for comparability [40,41].
Instrument reliability was previously reported as Cronbach $\alpha$.72 [39].

**Follow-Up Questionnaire**
A follow-up visit was conducted at 3 months. The week after the completion of the IMAGINE intervention, a member of this study’s team other than the group facilitator contacted participants and arranged a follow-up study visit, conducted via Zoom video conference. The follow-up electronic questionnaire was administered using REDCap and assessed pregnancy status, PHQ9, abbreviated 12-item SSB, PSS-4, and acceptability of IMAGINE via the Acceptability of Intervention Measure (AIM) [42]. The AIM scale is 4 Likert scale questions with possible responses scored 1-5: completely disagree, disagree, neither agree nor disagree, agree, and strongly agree. The score is calculated as the mean numerical value of the responses across the 4 items. Instrument internal consistency and validity were previously reported as Cronbach $\alpha$.85 and confirmatory factor analysis loadings 0.75-0.89 [42]. Participants were also asked how often they had used key CBT skills over the past month, using a Likert scale with possible responses: not at all, a few times, half the days, most of the days, and every day. The following skills were asked about mood tracking, engaging in pleasant activities, overcoming obstacles to engage in pleasant activities, thought interruption to reduce harmful thoughts, designated worry time to reduce harmful thoughts, time projection to imagine a better time in the future, self-instruction, positive contact with others, soliciting positive support from others, and using assertive communication. These questions were modeled on the core CBT components of the MB program [17]. If a participant reported the use of a skill, they were asked how helpful the skill was, using a Likert scale with possible responses: not helpful at all, somewhat helpful, or very helpful. Study staff screen-shared the REDCap questionnaire, as in the enrollment visit.

**Engagement data**
We quantitatively assessed engagement in the different components of IMAGINE. Messaging data from all Slack channels was exported as an HTML file. Message counts for each group member were determined by searching for each member’s username and counting the number of occurrences in the file. Completion of automated mood polls was determined based on reports from the Polly tool within Slack that was used to send polls. Attendance of each group member in Zoom calls was recorded by the facilitator.

**Quantitative Data Analysis**
We calculated descriptive measures of uptake, acceptability, and utility. Uptake was defined as the percentage of screened, eligible participants who enrolled in the intervention. Acceptability was determined based on self-report responses to the AIM. Utility was defined as the percentage of participants who reported using each CBT skill discussed in MB at least half the time. Additionally, among those who used each skill at least half the time, the percentage of participants who found it to be helpful was calculated.

We determined pre-post change in indicators of mental wellness. Depression symptoms were determined by PHQ9 score, calculated according to instrument guidelines, with a possible range of 0-27. Social support was calculated using the abbreviated SSB, as the sum score over all questions referring to family support and separately those referring to friend support. The mean and SD for each score were summarized for each time point. Scores at the 2 time-points were compared by paired 2-tailed $t$ test. RStudio (version 2023.06.0+421; Posit) was used for all data analysis.

**Qualitative Data Collection**
Qualitative In-Depth Interviews (IDIs) were conducted with all study participants at the 3-month follow-up visit. IDIs were conducted virtually over Zoom using a semistructured interview guide designed by this study’s team. The guide explored participants’ experience in the intervention, utility and potential improvements to each component of the intervention, level of engagement with intervention content, barriers to participation, and recommended improvements for future intervention implementations.

**Qualitative Data Analysis**
We conducted a thematic analysis of IDIs using a mixture of inductive coding driven by themes emerging from the transcripts and deductive coding based on themes from the interview guide. Qualitative analysis focused on the perceived acceptability and mental health impact of the IMAGINE intervention as well as recommendations for future iterations. First, 3 members of this study’s team (KR, EW, and AG) read all transcripts and separately developed initial codebooks based on themes that emerged from the transcripts and were explored in the interview guide. Initial codebooks were compared and discussed to create an agreed-upon combined codebook. Transcripts were then coded by 2 analysts (EW and AG) using Dedoose (Sociocultural Research Consultants, LLC) software [43]. Disagreements between analysts were resolved through discussion.

**Ethical Considerations**
The IMAGINE study was approved by the University of Washington Institutional Review Board (STUDY00008278). All participants provided informed consent for eligibility screening, exposure to the intervention, and data collection. Waivers were obtained for written documentation of informed consent and parental consent for adolescents younger than 18 years.

**Results**

**Participant Flow and Intervention Uptake**
Figure 1 summarizes participant flow from contacting this study to completing screening, enrolling in this study, and completing follow-up. In total, 68 individuals contacted this study between October 16, 2020, and January 29, 2021. Of these, 22 were assessed for eligibility, while 46 individuals did not complete screening due to challenges scheduling screening calls or participants declining to complete screening. Of the 22 who were assessed, 13 were eligible and 9 were ineligible, most (n=7, 77.8%) due to elevated depression scores warranting referral to individual care. In total, 3 eligible participants declined participation and 10 eligible participants were enrolled.
in the pilot, for an uptake of 76.9% (10/13 eligible participants). Participants were divided into 2 intervention groups: 7 in group 1 (active December 2020 to February 2021) and 3 in group 2 (active February to May 2021). While groups could be up to 10 participants, we elected to initiate the first group when 7 participants had been enrolled to minimize participants’ wait, while recruitment of the remaining participants continued. Of the 10 study participants, 9 completed the 3-month follow-up questionnaire and IDI.

**Figure 1.** Participant flow. PHQ-9: Patient Health Questionnaire.

**Participant Demographic Characteristics**

Demographic characteristics of enrolled participants are summarized in Table 1. All participants identified as female, and the median age was 17.9 (IQR 17.4-21.7) years. Of 9 participants who provided their race or ethnicity, 6 (67%) identified as Black, 5 (56%) identified as Latinx, 1 identified as an unlisted category, and 2 (22%) identified as White. Further, 4 (40%) participants were bilingual and 5 (50%) had completed at least a high school diploma or general education development. In total, 4 (40%) participants were pregnant at the time of this study and 6 (67%) used Medicaid health insurance. All but one of the participants reported that they were stably housed.
Table 1. Participant baseline characteristics.

<table>
<thead>
<tr>
<th>Participant characteristic</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y; N=10), median (IQR)</td>
<td>17.9 (17.4-21.7)</td>
</tr>
<tr>
<td>Sex (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10 (100)</td>
</tr>
<tr>
<td>Race or ethnicity a (n=9), n (%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>6 (66.7)</td>
</tr>
<tr>
<td>Latinx</td>
<td>5 (55.6)</td>
</tr>
<tr>
<td>Not listed</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td>White</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td>Pregnancy status (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>Pregnant at the time of enrollment</td>
<td>4 (40)</td>
</tr>
<tr>
<td>English proficiency (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>Fluent</td>
<td>10 (100)</td>
</tr>
<tr>
<td>Bilingual (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Education level (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>9-12th grade</td>
<td>5 (50)</td>
</tr>
<tr>
<td>High school diploma or GED b</td>
<td>1 (10)</td>
</tr>
<tr>
<td>&gt;High school</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Employment (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>8 (80)</td>
</tr>
<tr>
<td>Part-time (&lt;40h/wk)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Full-time (40h/wk)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Health insurance status (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>Medicaid</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Employer-provided private</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Housing status (N=10), n (%)</td>
<td></td>
</tr>
<tr>
<td>Stably</td>
<td>9 (90)</td>
</tr>
<tr>
<td>Unstably</td>
<td>1 (10)</td>
</tr>
</tbody>
</table>

aRace or ethnicity categories are not mutually exclusive.
bGED: general education development.

Intervention Acceptability and Utility

Quantitative Assessment

Among the 9 participants who completed follow-up, we found a mean acceptability score of 4.3 out of 5 (SD 0.6) on the AIM questionnaire. All participants reported that they would recommend IMAGINE to a friend. When asked about the use of core CBT skills covered in MB, all participants reported engaging in “playing with baby,” “contact with others,” and “talking to/contacting someone who has been a positive support for self and baby” skills during at least half the days in the prior month (Table 2). The majority reported using the following skills at least half the days: “mood tracking” (n=6, 66.7%), “engaging in pleasant activities” (n=6, 66.7%), “overcoming obstacles to doing pleasant activities” (n=6, 66.7%), “thought interruption to reduce harmful thoughts” (n=5, 55.6%), “using time projection to imagine a better time in the future” (n=6, 66.7%), and “using self-instruction to give oneself helpful directions” (n=7, 77.8%). When asked about the helpfulness of each skill, 100% of participants who used them (n=9) reported they were helpful (Table 2). Of the 11 skills we asked about, participants reported using a median of 7 (IQR 5-7) skills at least half the days.

We also analyzed participant engagement in the intervention. Figure 2 summarizes 3 measures of engagement: the number of SMS text messages sent on all Slack channels over the course of the intervention, the proportion of video calls attended, and the proportion of mood polls completed. Participants sent a median of 12 (IQR 11-15.8) messages during the 12-week intervention period, attended a median of 9.7% (IQR 0%-45.8%)
of the weekly video calls, and responded to a median of 32% (IQR 14.5%-47%) of the mood polls. Levels of engagement, particularly messaging, generally decreased over time but increased at intervention close (Multimedia Appendix 1).

### Table 2. Frequency and helpfulness of MB<sup>a</sup> skill use.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Participants who used skill for half of the time or more (n=9), n (%)</th>
<th>Participants who found skill helpful&lt;sup&gt;b&lt;/sup&gt;, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kept track of mood</td>
<td>6 (66.7)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Engaged in pleasant activities</td>
<td>6 (66.7)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Overcame obstacles to engage in pleasant activities</td>
<td>6 (66.7)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Used thought interruption to reduce harmful thoughts</td>
<td>5 (55.6)</td>
<td>5 (100)</td>
</tr>
<tr>
<td>Used worry time to reduce harmful thoughts</td>
<td>1 (11.1)</td>
<td>1 (100)</td>
</tr>
<tr>
<td>Used time projection to imagine a better time in the future</td>
<td>6 (66.7)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Use self-instruction to give yourself helpful directions</td>
<td>7 (77.8)</td>
<td>7 (100)</td>
</tr>
<tr>
<td>Played with baby</td>
<td>9 (100)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Had positive contact with others</td>
<td>9 (100)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Talked to or contacted someone who has been a positive support to yourself or baby</td>
<td>9 (100)</td>
<td>4 (100)</td>
</tr>
<tr>
<td>Made a request using assertive communication</td>
<td>3 (33.3)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Met a new person who can provide support for you and your baby&lt;sup&gt;c&lt;/sup&gt;</td>
<td>4 (44.4)</td>
<td>4 (100)</td>
</tr>
</tbody>
</table>

<sup>a</sup>MB: Mothers and Babies Course.

<sup>b</sup>Among those who reported using the skill; includes “somewhat helpful” and “very helpful” responses.

<sup>c</sup>Responses to this question asked for the number of new people who met who can provide support. This n represents the number of those who met at least 1 new support person.

### Figure 2. Participant engagement in the intervention.

**Qualitative Assessment**

IDIs explored the acceptability and perceived utility of IMAGINE. Participants highlighted several aspects of the intervention that were especially well-received (Textbox 1). Support from both the facilitator and other participants was viewed as beneficial. Most participants (n=8) reported valuing the connection with the facilitator and feeling that they could go to them for guidance and support. Participants (n=7) highlighted that support from other young parents in the program normalized and validated their experiences; several participants spoke about the value of connecting with others in similar situations to them. Mood polls were mentioned by all 9 participants as an impactful component of the intervention, with participants valuing the opportunity to reflect on their emotional state. In total, 2 participants highlighted the value of the intervention’s flexibility, both that the asynchronous design allowed them to access content at a time and pace that worked best for them, and that the facilitator made adaptations to respond to participant needs during intervention delivery.
In-Depth Interview themes related to intervention acceptability.

**Support from the study facilitator**
- When I had something come up or just anything, like, I would just talk to her, or if I had a question, I would talk to her. I asked her the question, and she would like answer the best way that she could or stuff. And I was just like, that's really helpful.
  [Participant 5, aged 21 y, postpartum]
- I had a kind of Facetime with her one time, just us, because the other girls weren’t able to make it. But that was very - I really appreciated that. It was kind of something that I didn't know that I needed, but she willing to like, hear me out and ask the right questions for me. So I think that was pretty awesome of her.
  [Participant 7, aged 23 y, pregnant]

**Connection with other participants**
- Seeing other women, not necessarily the facilitator, but being around the women that are going through the same things that you're going through, like that are also pregnant and just had babies, that makes a difference. That allows the connections that you form are a lot stronger and a lot tighter
  [Participant 9, aged 23 y, postpartum]
- We all went through this together and I felt like we all kind of grew together and, you know, they could, we could all relate to each other. And I didn't really have many mom friends I guess before this.
  [Participant 17, aged 17 y, pregnant]

**Mood polls**
- Normally you don't really reflect on your day unless you have a bad day. So I just like the way how it got me thinking, 'Oh, well, I did have a good day but I didn't notice it because it wasn't a bad day' if that makes sense.
  [Participant 12, aged 23 y, postpartum]
- I think they were helpful just because it was like a second to self-reflect, maybe look in the real world. Sometimes they ask you how you doing but it's kind of difficult to explain to someone, but when you're doing it for yourself, I think it can be a little bit more honest.
  [Participant 7, aged 23 y, pregnant]

**Intervention flexibility**
- At the beginning I think it started off slow and then she was sending a lot of messages, every day, so I wouldn't get time to read them. I'd have to go back a lot, and then that's something that I told her, and so she slowed down on the number of messages, so she sent them every other day, instead of every single day.
  [Participant 17, aged 17 y, pregnant]
- Having lots of interactions in written form gave everyone an opportunity to share their perspective and their experiences and what works for them, and we didn't all have to be right then and there, like I could look at it at two o'clock in the morning and another girl could look at it at two o'clock in the afternoon.
  [Participant 9, aged 23 y, postpartum]

**Mental Wellness**

**Quantitative Assessment**
Summary statistics for mental wellness outcomes are presented in Table 3 and individual participant trajectories are displayed in Multimedia Appendix 2. We compared depression symptoms (by PHQ-9), perceived stress (by PSS-4), and social support (by abbreviated SSB) between enrollment and follow-up. No significant changes in scores were found. At enrollment, the median PHQ-9 score was 4.0 out of 27 (IQR 2.0-5.0), compared with 2.0 (IQR 2.0-3.0) at follow-up ($P=.25$); 6 of 9 participants demonstrated reductions in PHQ-9 scores from baseline to follow-up. Median PSS-4 at enrollment was 8.0 out of 16 (IQR 8.0-10.0) and 9.0 out of 16 (IQR 8.0-10.0) at follow-up ($P=.46$). The median social support score related to participants’ family was 50.0 out of 100 (IQR 42.0-53.0) at enrollment, and 48.5 out of 100 (IQR 47.5-50.5) at follow-up ($P=.11$). The median social support score for friends was 51.0 out of 100 (IQR 47.0-54.0) at enrollment and 50.0 out of 100 (IQR 46.0-53.0) at follow up ($P=.88$).
Table 3. Change in participant mental wellness from baseline to follow-up.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Enrollment, median (IQR)</th>
<th>Follow-up, median (IQR)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression (PHQ-9&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>4.0 (2.0-5.0)</td>
<td>2.0 (2.0-3.0)</td>
<td>.25</td>
</tr>
<tr>
<td>Perceived Stress Score (PSS-4)</td>
<td>8.0 (8.0-10.0)</td>
<td>9.0 (8.0-10.0)</td>
<td>.46</td>
</tr>
<tr>
<td>Social support</td>
<td>Family: 50.0 (42.0-53.0); friends: 51.0 (47.0-54.0)</td>
<td>Family: 48.5 (47.5-50.5); friends: 50.0 (46.0-53.0)</td>
<td>Family: .11; friends: .88</td>
</tr>
</tbody>
</table>

<sup>a</sup>PHQ-9: Patient Health Questionnaire.

**Qualitative Assessment**

In IDIs, participants highlighted 3 ways they perceived participating in IMAGINE benefited their mental health (Textbox 2). First, participants (n=5) expressed that being in a group with other young parents normalized and validated their experiences, which helped them feel more connected with others and strengthened their belief that they could make changes to improve their mood. Second, several participants (n=5) noted that they had gained skills in regulating their emotions. Some stated that they still used some of the mood monitoring and management techniques discussed in the program after it had ended. Participants highlighted that they had few opportunities in their day-to-day lives to reflect on their mood and that there was value and ease in reflecting “for yourself” through the digital platform, rather than to another person who may not understand their feelings. Participants also commented that they learned to pay closer attention to emotions in the middle of the spectrum (ie, not crisis or elation), and that helped them monitor when and why their mood changed. Third, several participants (n=6) stated that they would be more open to mental health services in the future after they completed IMAGINE. Some common feedback from participants related to this was that it helped them realize they needed to return to psychotherapy if they had sought it in the past, or for those without prior experience, IMAGINE helped normalize some of the unknown or off-putting aspects of mental health care. It became more familiar to discuss moods and feelings with others, which increased their openness to seeking care.
Textbox 2. In-Depth Interview themes related to mental health impact.

Normalized experiences

- It helped me, like, open up a little more, and realize that I’m not the only person suffering, and that there’s more people like me that I can talk about it with. I don’t have to, you know, suffer by myself.
  [Participant 17, aged 17 y, pregnant]

- [IMAGINE helped me] understand… my emotions right now were normal… Before, I felt like I was broken or I needed to be fixed or there’s something wrong with me and now… there’s nothing wrong with me. [I’m] normal.
  [Participant 9, aged 23 y, postpartum]

Gained emotional regulation tools

- I had a pregnancy before that and it was also a preterm birth. And so, that baby passed away. And so it was really hard for me. I was really depressed, and so I know that if I would have had like the Imagine group it like it would help it would have helped me to manage my, my emotions and stuff.
  [Participant 5, aged 21 y, postpartum]

- I definitely added a handful more tools to my arsenal to be able to calm myself down, stay calm, or not become overwhelmed with everything that’s going on… I’ve been using these - to me they’re like lower tier ways of coping. So it’s like when I’m not at a ten, I’m at like a five, I can use these.
  [Participant 9, aged 23 y, postpartum]

- Reading the mood tree1… identifying the stressors… how to regulate it, different forms of communication, just all these different things for your mood and focusing on that. It just helped me to realize taking myself is a priority. And not just my baby and making sure she’s okay but making sure I’m okay as well.
  [Participant 17, aged 17 y, pregnant]

Opened to mental health services in the future

- I like to bottle up my anger or my problems and they explode on the wrong person at the wrong time so that being in a group kind of made me realize like I need to go back to counseling. I was already in counseling but I don’t feel like I was taking it as serious but when I sat back and realized… No, I do need to get it together, that’s when I was like, I’mma just go back to counseling.
  [Participant 14, aged 18 y, pregnant]

- When you talk to [my family] about… feeling depressed and you’re not mentally okay they’re, like, ”oh that’s not real, it’s fake, it’s all in your head”… And so, [IMAGINE] just helped me to think that there is help, and it is real, and that it’s not just in your head.
  [Participant 5, aged 20 y, postpartum]

Recommended Improvements

IDIs explored participants’ barriers to intervention participation and recommended improvements to IMAGINE. Frequently identified barriers to participation included being too busy, introversion, and low engagement from other members (Textbox 3). Participants (n=7) mentioned being unable to read all the messages or join Zoom meetings because they were working, in school, or busy with parenting responsibilities. Introversion was mentioned by a few participants (n=3), and 1 participant stated that a reason she did not participate in more Zoom calls was the thought that she was expected to be on camera, which she found uncomfortable. Limited engagement from other group members was discussed as a deterrent to engagement: participants (n=6) felt they would have engaged more and benefited more if synchronous and asynchronous conversations had included more voices. In total, 2 participants also reported technology challenges, both of which were resolved. Participants shared recommendations to address their barriers to engagement (Textbox 3). Several participants (n=8) recommended changes to the frequency and timing of the synchronous Zoom calls to facilitate their attendance. Calls were scheduled each week based on polls with the group to identify the optimal time. Participants suggested that offering calls in the evenings, during the weekends, or announcing the schedule for all calls at the start of the program rather than scheduling them week-by-week could improve attendance. A few participants suggested offering multiple call times each week. Some participants (n=3) also requested more onboarding at the start of the group, including more orientation to the Slack platform and more icebreakers and rapport-building activities with other group members to encourage greater comfort and collective engagement in the group. While IMAGINE was designed to be virtual and COVID-19 restrictions meant in-person contact was not possible at the time, a common piece of feedback (n=7 participants) was to have at least one in-person meeting or offer a hybrid version of IMAGINE to deepen relationship-building with other group members.
Textbox 3. In-Depth Interview themes related to barriers to participation and recommended improvements.

**Barriers**

- Being busy
  - I couldn't read or join in on the calls, because I had started back working. [Participant 24, aged 16 y, pregnant]
  - I started working like right after I got into [the study], and I didn't know when I was actually going to be available… Some of the polls or some of the Zooms I wasn’t even able to join, because I work. [Participant 18, aged 19 y, pregnant]

- Introversion
  - I’m a shy person. So that’s what, I just want to push myself more to be able to, you know, do more things like this because this is like out of my comfort zone. [Participant 17, aged 17 y, pregnant]
  - Being on video is just not for me, I’m just not gonna keep with this. [Participant 12, aged 23 y, postpartum]

- Low engagement from other study participants
  - It would have been better if everybody, you know, responded in and reacted. But it was a good thing to still be able to have somebody there talking to us, even though we weren’t always responding. [Participant 20, aged 18 y, postpartum]
  - I just wish more people participated, because even in the zoom calls like sometimes I would be the only person that would join, or it would just be one or two other moms. [Participant 17, aged 17 y, pregnant]

- Issues with technology
  - My phone like we started and it kind of deleted the app so I was just like, whoa. Okay. [Participant 5, aged 21 y, postpartum]
  - I had to switch [cell phones], so I had turned in a phone that I was leasing for Sprint and then I had an old phone, but I couldn’t do certain things on it. Like I couldn’t download a lot of apps because I didn’t have the latest iOS and then, when I got the new phone, I had to wait to switch carriers. It was a lot going on, when I was trying to get a new phone. [Participant 12, aged 23 y, postpartum]

**Recommended improvements**

- More flexibility for Zoom calls
  - I didn’t really get to join any of the zoom meetings because I was usually busy… I wish I could have. [Participant 5, aged 21 y, postpartum]
  - If we had more zoom calls like twice a week, that would be really great for me. [Participant 7, aged 23 y, pregnant]

- Hybrid or in-person meetings
  - I think that if we have more in person, it would have been better because it's more something that you can engage in more than just being online. [Participant 20, aged 18 y, postpartum]
  - I feel if you were to go in person and meet everyone, I think that'd be great. You get to interact. [Participant 5, aged 21 y, postpartum]

- More onboarding
  - At the very end, it was really nice because we were kind of more comfortable with each other, but at the beginning it was, understandably, a little bit awkward. We don’t know each other, but I think we could have benefited with some more icebreakers or some more like getting to know each other, rather than just kind of jumping in and expecting [us] to be open to each other. Because not everybody is willing to, you know, put themselves in that situation where they kind of have to be vulnerable. [Participant 7, aged 23 y, pregnant]
  - Me personally I didn't like the Slack platform just because I'm not really tech savvy I guess you could say, and it was a lot of different components to the app, so it's kinda like I'm still trying to learn how to use it. [Participant 14, aged 18 y, postpartum]

- Low engagement from other study participants
  - It would have been better if everybody, you know, responded and reacted. But it was a good thing to still be able to have somebody there talking to us, even though we weren’t always responding. [Participant 20, aged 18 y, postpartum]
  - I just wish more people participated, because even in the zoom calls like sometimes I would be the only person that would join, or it would just be one or two other moms. [Participant 17, aged 17 y, pregnant]

**Discussion**

**Principal Results**

In this mixed methods pilot study, we found that IMAGINE, a novel digital adaptation of the evidence-based MB course, had high acceptability and perceived utility. Uptake of IMAGINE, defined as the proportion of eligible participants who enrolled, was high. However, many individuals who initially contacted this study did not complete the screening process, suggesting these screening procedures presented barriers. These barriers may include the need to arrange a call with study staff to...
complete screening; it is possible the reach would be higher in a lower-barrier delivery model where clients could self-register in the group without arranging a screening call. We also found that a substantial proportion of screened participants were ineligible, most commonly due to elevated PHQ-9 scores. While our study’s focus was on the prevention of perinatal depression, this observation suggests that IMAGINE may also be appealing to individuals who are already experiencing elevated symptoms of depression; this observation is important in considering potential future applications of IMAGINE.

Acceptability scores were high and all participants stated they would recommend IMAGINE to a friend. Quantitative measures of engagement in the intervention, such as frequency of messaging and attendance of video calls, demonstrated low-moderate engagement. Nevertheless, most CBT approaches discussed in the IMAGINE intervention were used frequently and reported to be useful by the majority of participants. Qualitatively, participants reported one-to-one support from the facilitator, connection with other parents, and regular opportunities to reflect on their mood through mood polls were especially helpful aspects of the intervention. Additionally, participants reported that the intervention normalized their mental health challenges, improved their ability to manage their mood, and increased their openness to mental health care. We found no significant changes in depression scores, perceived stress, or perceived social support, although our study’s small sample size was not powered to detect changes in these outcomes.

Participants also made recommendations for improvements to the IMAGINE intervention in future iterations, including more opportunities for synchronous connection with other group members, additional rapport-building activities, and further simplification of message content.

Comparison With Prior Work

Our findings add to a small but growing body of literature on the use of digital interventions to prevent perinatal depression. While several studies have found that internet-delivered CBT is effective in the treatment of depression, including in the perinatal period [28,44-49], digital interventions for the prevention of perinatal depression have been less well studied [30,50].

Our findings are consistent with data from in-person MB. The levels of CBT skill use that IMAGINE participants reported were similar to that reported in a cluster randomized trial of in-person MB [51]: 66% (n=6) of IMAGINE participants reported that they engaged in pleasant activities and 100% (N=9) talked to or contacted someone who has been a positive support, compared with 78% and 80% respectively in Tandon et al’s [51] study previously adapted MB to a self-guided web-based format with informational pages, audio and video clips, and worksheets that follow MB modules in English and Spanish. A pilot randomized trial of the intervention, named Mothers and Babies Online Course, found nonsignificant improvement in depression symptoms; this study’s power was limited by sample size and the authors commented that facilitator guidance may improve uptake and efficacy [25,52]. The IMAGINE intervention differs from the Mothers and Babies Online Course in its inclusion of a facilitator and group delivery format, both of which were viewed as helpful components by participants in our pilot. Our findings of high acceptability, low-moderate engagement in intervention content, and high perceived utility are consistent with those of Barrera et al [25], suggesting digital delivery of MB with and without facilitation is a promising strategy that warrants further evaluation.

Strengths and Limitations

Our study has several strengths. The IMAGINE intervention was developed through systematic adaptation of the evidence-based MB course that prioritized fidelity to core MB components while adapting to participants’ design recommendations [34]. Our recruitment strategy feasibly reached potential participants across the country. Our evaluation is strengthened by including measures of CBT skill use that align with the mechanism of MB and were used in previous MB studies, allowing comparison of our findings with other MB studies. Furthermore, our study explored the experiences of young perinatal people, whose often marginalized perspectives are critical to developing responsive interventions. Guided by the principles of human-centered design [53], we collected in-depth insights from users before and after our pilot to drive future iteration and improvement of the intervention.

The primary limitations of our study are its small sample size and nonrandomized design. This study was not powered to evaluate the intervention’s effect on mental health outcomes, and pre-post comparisons are susceptible to confounding by changes over time that are not attributable to the intervention. Additionally, recruitment was primarily through Facebook and Instagram, which were selected for participants who were already using social media platforms and may be more open to a digital intervention. Due to resource constraints for this pilot study, our eligibility criteria included the ability to read and write in English, which systematically excluded non-English speakers, whose needs may differ. Future evaluations should address the limitations of this study by achieving a larger sample size, employing a randomized design, recruiting from nonsocial media sources, and developing translations in additional languages, particularly Spanish, in which MB materials already exist.

Conclusions

This pilot study provides promising evidence of the acceptability and utility of a digital group adaptation of MB among perinatal youths. These findings support further development and evaluation of the IMAGINE intervention to increase access to evidence-based interventions for the prevention of perinatal depression. Future iterations of IMAGINE will incorporate user recommendations from this study and use randomized powered evaluations to test clinical impact.
Acknowledgments
The authors gratefully acknowledge IMAGINE study participants and recruitment partners. This study was funded by a grant from the Technology and Adolescent Mental Wellness program at the University of Wisconsin-Madison. This study was also supported by the National Institutes of Health (grant K18MH122978) and the University of Washington Behavioral Research Center for HIV (BIRCH), a National Institute of Mental Health–funded program (P30 MH123248). Research support was provided by the University of Washington Global Center for Integrated Health of Women, Adolescents, and Children (Global WACH). REDCap access was provided by the University of Washington Institute of Translational Health Science (grants UL1TR002319 and KL2TR002317) and the National Center for Advancing Translational Sciences of the National Institutes of Health (TL1TR002318). The sponsor was not involved in the review and approval of this paper for publication.

Data Availability
The data sets analyzed during this study are available on GitHub [54].

Authors’ Contributions
KR designed this study, conceptualized this paper, guided the analysis, supported this paper’s writing, and obtained funding. AG, EW, and KR conducted the data analyses. AG, KD, MJ-B, YNE, AB, and KR contributed to participant recruitment, intervention delivery, and data collection. All authors contributed to the development of the intervention and reviewed and approved this paper for submission.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Participant interaction over time (group 1 and 2 combined).

Multimedia Appendix 2
Participant trajectories of mental wellness outcomes.

References


41. Mothers and Babies Program. URL: https://www.mothersandbabiesprogram.org/ [accessed 2023-01-31]


46. Mothers and Babies Program. URL: https://www.mothersandbabiesprogram.org/ [accessed 2023-01-31]


54. IMAGINE_public. GitHub. URL: https://github.com/keshetronen/IMAGINE_public/ [accessed 2023-12-29]

Abbreviations
AIM: Acceptability of Intervention Measure
CBT: cognitive behavioral therapy
IDI: In-Depth Interview
IMAGINE: Interactive Maternal Group for Information and Emotional Support
MB: Mothers and Babies Course
mHealth: mobile health
PHQ9: Patient Health Questionnaire
PSS-4: Perceived Stress Score
REDCap: Research Electronic Data Capture
SSB: Social Support Behavior

Please cite as:
Ronen K, Gewali A, Dachelet K, White E, Jean-Baptiste M, Evans YN, Unger JA, Tandon SD, Bhat A
Acceptability and Utility of a Digital Group Intervention to Prevent Perinatal Depression in Youths via Interactive Maternal Group for Information and Emotional Support (IMAGINE): Pilot Cohort Study
JMIR Form Res 2024;8:e51066
URL: https://formative.jmir.org/2024/1/e51066
doi:10.2196/51066
PMID: 38306159

©Keshet Ronen, Anupa Gewali, Kristin Dachelet, Erica White, Marimirca Jean-Baptiste, Yolanda N Evans, Jennifer A Unger, S Darius Tandon, Amritha Bhat. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Benefits, Recruitment, Dropout, and Acceptability of the Strength Back Digital Health Intervention for Patients Undergoing Spinal Surgery: Nonrandomized, Qualitative, and Quantitative Pilot Feasibility Study

Annemieke van der Horst1,2, PhD; Laura Meijer1, MSc; Harmieke van Os - Medendorp1, PhD; Jan S Jukema1, PhD; Ernst Bohlmeijer2, Prof Dr; Karlein MG Schreurs2,3, Prof Dr; Saskia Kelders2,4, PhD

1Research Group Smart Health, Saxion University of Applied Sciences, Deventer, Netherlands
2Department of Psychology, Health and Technology, Centre for eHealth & Well-being Research - Behavioural, Management and Social Sciences, University of Twente, Enschede, Netherlands
3Roessingh Research & Development, Enschede, Netherlands
4Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa

Corresponding Author:
Annemieke van der Horst, PhD
Department of Psychology, Health and Technology
Centre for eHealth & Well-being Research - Behavioural, Management and Social Sciences
University of Twente
PO BOX 217
Enschede, 7500 AE
Netherlands
Phone: 31 (0)53 489 4470
Email: a.vanderhorst-1@utwente.nl

Abstract

Background: Patients undergoing spinal surgery report high levels of insecurity, pain, stress, and anxiety before and after surgery. Unfortunately, there is no guarantee that surgery will resolve all issues; postsurgical recovery often entails moderate to severe postoperative pain, and some patients undergoing spinal surgery do not experience (long-term) pain relief after surgery. Therefore, focusing on sustainable coping skills and resilience is crucial for these patients. A digital health intervention based on acceptance and commitment therapy (ACT) and positive psychology (PP) was developed to enhance psychological flexibility and well-being and reduce postsurgical pain.

Objective: The objective of this study was 3-fold: to explore the potential benefits for patients undergoing spinal surgery of the digital ACT and PP intervention Strength Back (research question [RQ] 1), explore the feasibility of a future randomized controlled trial in terms of recruitment and dropout (RQ 2), and assess the acceptability of Strength Back by patients undergoing spinal surgery (RQ 3).

Methods: We used a nonrandomized experimental design with an intervention group (n=17) and a control group (n=20). To explore the potential benefits of the intervention, participants in both groups filled out questionnaires before and after surgery. These questionnaires included measurements of pain intensity (Numeric Pain Rating Scale), pain interference (Multidimensional Pain Inventory), anxiety and depression (Hospital Anxiety and Depression Scale), valued living (Engaged Living Scale), psychological flexibility (Psychological Inflexibility in Pain Scale), and mental well-being (Mental Health Continuum–Short Form). Semistructured interviews combined with log data and scores on the Twente Engagement With eHealth Technologies Scale were used to assess the acceptability of the intervention.

Results: A significant improvement over time in emotional (V=99; P=.03) and overall (V=55; P=.004) well-being (Mental Health Continuum–Short Form) was observed only in the intervention group. In addition, the intervention group showed a significantly larger decline in pain intensity (Numeric Pain Rating Scale) than did the control group (U=75; P=.003). Of the available weekly modules on average 80% (12/15) was completed by patients undergoing spinal fusion and 67% (6/9) was completed by patients undergoing decompression surgery. A total of 68% (17/25) of the participants used the intervention until the final interview. Most participants (15/17, 88%) in the intervention group would recommend the intervention to future patients.
Conclusions: This pilot feasibility study showed that combining ACT and PP in a digital health intervention is promising for patients undergoing spinal surgery as the content was accepted by most of the participants and (larger) improvements in pain intensity and well-being were observed in the intervention group. A digital intervention for patients undergoing (spinal) surgery can use teachable moments, when patients are open to learning more about the surgery and rehabilitation afterward. A larger randomized controlled trial is now warranted.

(JMIR Form Res 2024;8:e54600) doi: 10.2196/54600

KEYWORDS
pilot feasibility study; spinal surgery; digital health intervention; positive psychology; acceptance and commitment therapy; mobile phone

Introduction

Background

Patients undergoing spinal surgery report high levels of physical complaints as well as insecurity, pain, low well-being, stress, and anxiety before and after surgery [1-3]. Postsurgical recovery often entails moderate to severe postoperative pain, and approximately 20% to 30% of patients undergoing spinal surgery do not experience (long-term) pain relief after surgery [3,4]. This results in a longer hospital stay, higher health care costs, longer physical and mental recovery, delayed return to work, and the potential development of chronic pain. The potential transition from postoperative pain to chronic pain is a major issue as chronic pain affects many aspects of a patient’s life, such as work; physical, emotional, and social well-being; and quality of life [5,6].

Mental factors such as cognition, emotions, and expectations play a significant role in the experience of pain [7]. The fear-avoidance model explains the trajectory from acute to chronic pain through fear and catastrophizing, which is the tendency to enlarge the threat of pain and a feeling of helplessness, leading to an increase in pain avoidance as a dominant coping strategy [8]. Pain avoidance then leads to a less active lifestyle, thereby worsening rather than relieving pain. High levels of catastrophizing and fear have been found to predict higher levels of (postoperative) pain and pain chronicity and a lower quality of life in patients who undergo surgery [4,9-11]. In addition, realistic expectation management is key as unrealistic or unfulfilled expectations about surgery and preoperative stress may lead to the experience of higher levels of postoperative pain [12-18]. This link between mental factors and expectation management with postoperative outcomes shows that psychological preparation before surgery is essential to improve recovery after surgery.

Although the potential benefits of psychological preparation before surgery have long been known [19], a more recent meta-analysis of psychological preparation techniques before surgery could not find strong evidence from high-quality research to verify these claims [20]. Potentially, interventions that focus on promoting adaptive coping skills and reducing maladaptive emotion regulation skills are more promising. Smith and Zautra [21] and Sturgeon and Zautra [22] found, for example, evidence that coping strategies such as (pain) acceptance, engaging in beneficial social interactions, and experiencing a value-based purpose in life have the potential to improve mental well-being and promote resilience in the face of (chronic) pain. Sustainable resilience to (chronic) pain requires skills promoting adaptation and mental health in the long term [22]. In other words, adaptation to different contexts and circumstances requires (psychological) flexibility and a long-term focus. In the context of pain, psychological flexibility implies that painful sensations, feelings, and thoughts are accepted as opposed to avoided and that attention shifts toward personally valued goals [23]. For patients undergoing spinal surgery, psychological flexibility enables them to cope with the fluctuating circumstances surrounding surgery (eg, insecurity, fear, and pain) in a flexible rather than rigid manner. Being able to accept negative emotions or sensations such as pain and insecurity in the face of surgery might prevent catastrophizing, fear, and avoidance behavior.

Psychological flexibility is the aim of acceptance and commitment therapy (ACT) [24]. ACT is based on the relational frame theory and focuses on performing value-based activities in life even in the face of insecurity and adversity [24]. In a similar way, positive psychology (PP) is the scientific study of well-being and optimal functioning, focusing on human flourishing instead of reducing the risk factors for psychopathology and malfunctioning. PP involves topics such as strengths, virtues, meaning, happiness, gratitude, compassion, resilience, and flourishing [25]. ACT and PP can help experience a value-based purpose in life, promote resilience, and improve mental well-being. Mental well-being comprises positive emotional, psychological, and social functioning [26]. The presence of higher levels of these 3 dimensions of well-being is an indicator of flourishing [26-28]. Promoting positive resources and skills that contribute to successful adaptation and (mental) health are the aims of PP interventions (PPIs) [25,29]. PPIs and ACT have been found to be effective in the treatment of chronic pain [30-35], improving affect and functional ability after knee surgery [21], and quicker cessation of pain and opioid use in veterans after orthopedic surgery [36]. In summary, PP and ACT could potentially benefit patients undergoing spinal surgery in terms of physical (eg, pain) and psychological (eg, well-being) factors.

In the context of surgery, digital health interventions can be used to aid patients during their recovery process [37]. These interventions can be used to improve postoperative outcomes by supporting healthy lifestyle behavior change before and after surgery [38,39] and to improve medication adherence [40,41]. Digital health interventions can also better prepare patients for surgery or shorten postoperative recovery through behavioral modification, patient monitoring, or protocol adherence [42-45].
Behavioral modification can also be applied to promote a healthy lifestyle. This is important as healthy lifestyle behavior changes before and after orthopedic surgery, such as increased preoperative physical activity or smoking cessation, have been associated with improved postoperative bone healing [46] and wound healing [47], quicker recovery times, and reduced pain scores [48]. Moreover, patients undergoing surgery who engage with digital health interventions show better medication adherence, better adherence to discharge instructions, greater patient satisfaction, improved clinic attendance, lower readmission, and less emergency department visits after surgery [49]. van der Meij et al [50] found in their review that, in most studies, perioperative digital health interventions improved clinical patient-related outcomes compared with face-to-face perioperative care alone for patients who had undergone various forms of surgery. This shows the potential of digital health interventions to improve perioperative care in addition to face-to-face meetings with professionals.

Despite the clear potential of ACT and PPIs for mental well-being and long-term resilience, they have not yet been used to develop interventions targeting psychological flexibility in patients undergoing spinal surgery. In addition, seeing the potential of digital health interventions in the context of surgery, this form also seems promising for patients undergoing spinal surgery. For this reason, a digital health intervention called Strength Back [51] was developed. Strength Back aims to increase psychological flexibility and well-being and improve postoperative recovery in patients undergoing spinal surgery. This intervention, developed through cocreation with different stakeholders, contains procedural information, pain education, and PPI and ACT exercises.

**Objectives**

Although the intervention is based on a scientific, theoretical framework, the effects and impact of the intervention need to be proven to be able to implement it in an evidence-based health care setting. This requires a collection of evidence, such as through a randomized controlled trial (RCT). As an RCT is complex and expensive, preparation and insights in advance into the required parameters, such as measures, recruitment, and dropout numbers, are essential to perform a thorough RCT. In light of the aforementioned research on PP and ACT in other target populations, several outcome measures covering physical and psychological factors need to be explored as potential benefits of this intervention. In addition, in an RCT, the best possible version of the intervention needs to be tested. This is in line with the recommendation of the Medical Research Council of first testing and refining an intervention to ensure it is acceptable to successfully evaluate whether it is effective [52]. In addition, to reduce psychological and practical barriers and maximize acceptability, it is important to understand and accommodate patients’ views [53]. As of yet, it is unknown how patients undergoing spinal surgery value the intervention Strength Back in terms of acceptability, potential benefits, and feasibility in a real-life setting and what the required parameters for a future RCT would be. Therefore, the objective of this study was 3-fold: to explore the potential benefits for patients undergoing spinal surgery of the digital ACT and PPI Strength Back (research question [RQ] 1), explore the feasibility of a future RCT in terms of recruitment and dropout (RQ 2), and assess the acceptability of Strength Back by patients undergoing spinal surgery (RQ 3).

**Methods**

**Study Design**

This was a nonrandomized pilot study, a subset of feasibility studies as described by Eldridge et al [54]. Our study design included an intervention group and a historic control group (Figure 1). We used a historic control group as the development of the intervention was still ongoing and the target population was small. The participants in the control group received care as usual. In addition to receiving care as usual, the participants in the intervention group were given access to the digital health intervention Strength Back. To test the acceptability of the intervention, individual interviews were conducted with the participants in the intervention group. Both groups filled out a questionnaire before (pretest time point) and after (posttest time point) surgery.

**Figure 1.** Study design and timeline for individual participants (patients undergoing spinal surgery) in the control and intervention groups of this pilot feasibility study.
As suggested by Orsmond and Cohn [55], we used a combination of qualitative and quantitative measures in this feasibility study. We aimed to include 20 participants in our control group and 20 participants in our intervention group, which is in line with the study by Billingham et al [56], who found a median sample size of 36 in their review of feasibility studies.

**Participants**

Patients undergoing decompression or spinal fusion surgery were eligible for inclusion in either the control or intervention group of this study. Patients undergoing spinal surgery other than decompression or spinal fusion surgery (eg, patients with hernias or patients of oncology) were excluded. Other inclusion criteria were age of >18 years, proficiency in Dutch, and an email address. Participants in the intervention group needed an Android or iOS smartphone or tablet at home and had to be willing to use a digital health intervention both before and after surgery.

All participants in this study were recruited through purposive sampling at an orthopedic center in the Netherlands. The participants in the historic control group were recruited between February 2020 and July 2020. The participants in the intervention group were recruited between September 2020 and February 2021.

**Conditions**

**Care as Usual**

The historic control group received care as usual. They were provided with brochures containing information about their diagnosis, surgical procedure, surgery preparation, physical guidelines during recovery, contact information, and possible complications. In addition, usual care consisted of several appointments besides the surgery itself: an intake and a preoperative screening before surgery, a consultation by phone 10 days after surgery, and a physical checkup at the orthopedic center 6 weeks (both surgery types) and 12 weeks (spinal fusion surgery) after surgery.

**Strength Back Digital Health Intervention**

In addition to care as usual, participants in the intervention group were given access to the Strength Back digital health intervention (see Figure 2 for screenshots). The content of this app was developed through a process of cocreation. The stakeholders involved in this process were patients who had undergone spinal surgery, orthopedic surgeons, a physical therapist, a nurse practitioner, a research coordinator, and several nurses. The content of the intervention was based on the results of interviews and focus group sessions with the aforementioned stakeholders (patients and health care professionals) combined with ACT and PP exercises. Multimedia Appendices 1 to 3 of this publication and our previous publication on the cocreation of Strength Back [51] provide more details on the developmental process.
The intervention consisted of 13 information modules, which were continuously available, and 6 to 12 weekly modules. Participants did not receive personal messages from a health care professional. The contact information of the orthopedic clinic was provided in the intervention, making sure participants could contact a health care professional in case of any questions.

The information modules of the intervention were based on existing brochures and leaflets of the orthopedic center (Multimedia Appendix 1).

In addition, to support them during recovery, intervention participants received 6 to 12 weekly modules. These modules were timed: participants received 3 modules in the 4 weeks before surgery and 6 or 12 weekly modules after surgery depending on the type of surgery they had undergone. In total, 6 weekly modules were provided for patients undergoing decompression surgery, and 12 weekly modules were provided for patients undergoing spinal fusion, matching the expected recovery time after surgery as suggested by the orthopedic center (Multimedia Appendices 2 and 3).

During the intervention, participants received automated reminders every time a new weekly module was available. This was voluntary and could be turned on or off by the participants.

The digital health intervention was downloaded by the participants themselves using a step-by-step guide provided by the researcher.

**Procedure of Data Collection**

For both conditions, patients received a leaflet from their personal health care professional (ie, orthopedic surgeon, specialized physical therapist, or advanced nurse practitioner) explaining this study during their visit to the orthopedic center and were subsequently contacted by phone by the researcher. When patients agreed to take part in the study, their email address was collected, after which they received an information email with a link to the first web-based questionnaire (through Qualtrics [Qualtrics International Inc]; pretest assessment). We chose this specific moment for the pretest assessment to clarify the current and most up-to-date picture of the preoperative situation. In the questionnaire, the participants first had to

---

**Figure 2.** Screenshots of the Strength Back intervention (in Dutch, Kracht TeRUG) for patients undergoing spinal surgery. (A) The information on the spinal condition, the animation on pain education, and the video of the nursing ward. (B) The “wish question,” text about acceptance of pain, and “mindful breathing” exercise.
provide active informed consent to continue. Participants in both the control and intervention groups filled out this questionnaire (pretest assessment; Figure 1).

A total of 3 months after surgery, participants in both groups received an invitation by email to fill in the posttest questionnaire. At this time, it was expected that (nearly) all participants were in the final phase of their recovery process. We chose this specific moment to collect the data, when the recovery situation had stabilized. At this point, the participants in the intervention group were contacted by the principal investigator (AH) to make an appointment for an interview to assess how they valued the intervention. These semistructured interviews were conducted between January 2021 and June 2021 by 2 researchers (AH and LM). Owing to the COVID-19 pandemic, interviews were conducted via telephone. The interviews were audio recorded and lasted between 30 and 60 minutes.

The recruitment and dropout numbers in the intervention group were explored in this study to provide insights into what is needed for the recruitment process of a future RCT to reach the desired power.

Materials

Regarding RQ 1 (potential benefits), the questionnaires contained several outcome measures (ie, pain intensity and interference, anxiety, depression, psychological inflexibility, valued living, and mental well-being). These potential benefits were measured at the pre- and posttest time points for both the control and intervention groups.

Pain intensity was measured using the Numeric Pain Rating Scale (NRS). For the NRS, participants were asked to circle the number (0-10) that best described their experience of pain in the last week in general and in the last week at worst. The questions asked were as follows: “Please indicate the average intensity of your pain in the last week” and “Please indicate the intensity of your pain at the worst moments in the last week.” A higher score indicates a higher intensity of pain. The NRS is considered a valid, reliable, and appropriate pain rating scale for use in clinical practice, with good sensitivity [57].

Pain interference was measured using the pain interference subscale of the Multidimensional Pain Inventory (MPI) [58]. This subscale consists of 11 items measuring to what degree pain influences the daily life of a respondent, with total mean scores ranging from 0 to 6. A higher score indicates a higher degree of pain interference. The subscale showed good reliability in our study population, with a Cronbach α of .874.

Anxiety and depression were measured using the Hospital Anxiety and Depression Scale (HADS) [59]. Both the anxiety and depression subscales consist of 7 items with scores ranging from 0 to 21. A higher score indicates a higher degree of anxiety or depression. Both subscales showed good reliability, with a Cronbach α of .874 for the anxiety subscale and .789 for the depression subscale.

Psychological inflexibility was measured using the Psychological Inflexibility in Pain Scale (PIPS) [60]. The 2 subscales measure avoidance and cognitive fusion, concepts derived from the theoretical framework of ACT [24]. As the cognitive fusion subscale showed low reliability in our study population with a Cronbach α of .524, only the total score of the PIPS was used in this study. This total score had good reliability, with a Cronbach α of .893. A higher score indicates more inflexibility with pain (ie, lower acceptance and higher cognitive fusion).

Value-based living was measured using the “Valued living” subscale of the Engaged Living Scale (ELS; Trompetter et al [61]), which covers the concept of “valued living”—recognizing personal values and undertaking committed actions that are congruent with these values. The “Valued living” subscale consists of 10 items and uses a 5-point scale, with scores ranging from 10 to 50. A higher score indicates a higher degree of valued living. This subscale of the ELS showed high reliability, with a Cronbach α of .917.

Mental well-being was measured using the Mental Health Continuum–Short Form [62] on three dimensions: (1) emotional well-being, defined as positive feelings (3 items); (2) psychological well-being, described as individual functioning (eg, environmental mastery and purpose in life; 6 items); and (3) social well-being, measuring functioning in community life (eg, social contribution and social acceptance; 5 items). The mean scores per subscale range from 0 to 5, with a higher score indicating a higher degree of emotional, psychological, or social well-being. The total score (Cronbach α=.930) as well as that of the different subscales (emotional: Cronbach α=.930; psychological: Cronbach α=.874; social: Cronbach α=.799) showed good reliability, with a high Cronbach α in our study population.

Regarding RQ 2 (recruitment and dropout), we kept an overview in Microsoft Excel (Microsoft Corp) of the recruitment and dropout numbers in the intervention group in this study.

Regarding RQ 3 (acceptability), the scheme used for the interviews consisted of 2 parts. The first part of the interview started with questions on the intervention in general (eg, topics on layout and user-friendliness based on the Mobile App Rating Scale [63]). The second part continued by discussing the content of the modules (Multimedia Appendix 1) followed by the exercises in the weekly modules (Multimedia Appendix 2) as topics. Engagement with the intervention was measured using the Twente Engagement With eHealth Technologies Scale (TWEETS) [64]. This scale of 9 items comprises 3 subscales: behavioral, cognitive, and affective engagement. In this study, the total score, with a Cronbach α of .883, was used. In addition, the automated logs from the system were analyzed to see when participants opened which weekly module. The information modules were continuously available (before and after surgery), and activity in these modules was not recorded in the automated logs.

Data Analysis

All statistical analyses were conducted using SPSS (version 26.0; IBM Corp). For the exploratory aim of this study, full cases provide the most valuable information. Therefore, only full-case analyses were conducted. In addition, as the quantitative data were used to generate a general picture of
between-group differences, no imputation techniques were used. When participants filled out the pre- or posttest assessment more than once, the first completed test was used for the analysis.

For RQ 1 (potential benefits), between-group differences were assessed using nonparametric Mann-Whitney U tests because of the small sample size and nonnormal distribution of the data. First, between-group differences were analyzed at baseline using the Mann-Whitney U test. Second, within-group differences were analyzed for both groups between the pre- and posttest assessment using a Wilcoxon signed rank test. The effect sizes were calculated using the Cohen d: (mean at the posttest assessment – mean at the pretest assessment) / (SD). When the SD of both moments differed, the mean SD of the pre- and posttest assessments was calculated and used. Values for the Cohen d of 0.2 were considered a small effect size, 0.5 was considered a medium effect size, and values of ≥0.8 were considered a large effect size.

Third, between-group differences were analyzed by comparing the pre- and posttest difference scores of both groups (Mann-Whitney U test). All statistical tests were 2 sided. P values of <.05 were considered statistically significant. As this was a feasibility pilot study, the results were considered exploratory, and no correction for multiple testing was performed.

For RQ 2 (recruitment and dropout), we analyzed the recruitment and dropout numbers of the intervention group in this study. For the analysis, we looked at the number of approached patients, the number of patients who consented to participate in the study, and the time span in which all these patients were approached. Dropout was defined as agreeing to participate in the study but not downloading the intervention, not filling out the pretest questionnaire, or not filling out the posttest questionnaire.

For RQ 3 (acceptability), interview data were collected, and the audio recordings were transcribed verbatim. For the analysis, the transcripts of all interviews were read and reread by 2 researchers (AH and LM) to familiarize themselves with the data. Subsequently, the transcripts were open coded by both researchers independently, followed by a discussion to reach a consensus regarding the coding. As a next step, both researchers sorted the codes into several groups. A deductive coding approach was used based on the topics on the interview scheme combined with an inductive approach for important content within these topics. For the coding and analysis process, the ATLAS.ti software (version 9.0; ATLAS.ti Scientific Software Development GmbH) was used.

**Ethical Considerations**

Ethics approval was granted by the Medical Ethical Committee of the Radboud University Medical Centre in Nijmegen (2019-5608) and the Ethics Committee of the Faculty of Behavioural, Management, and Social Sciences of the University of Twente (191080). The study is registered at the repository of the University of Twente for processing of personal data (AVG Register; WBP18ME0048).

Participants in both the control group and the intervention group filled out a questionnaire before and after surgery. When patients agreed to take part in the study, they received a link to the first questionnaire, in which they had to provide active informed consent to continue.

All human participant data were deidentified before analysis. Study participants received no compensation for their inclusion in the study.

**Results**

**Sample**

Figure 3 shows an overview of the inclusion of participants in the intervention and control groups. A total of 25 patients agreed to participate in the intervention group, of whom 3 (12%) participants did not fill out the pretest questionnaire; the reason for this is unknown. Therefore, a total of 22 participants filled out the pretest questionnaire. After filling out the pretest questionnaire, 9% (2/22) of the participants failed to download the intervention to their phones. Therefore, 20 participants started using the intervention (20/35, 57% of the approached patients). Before the posttest questionnaire, 15% (3/20) of the patients discontinued their participation. This resulted in a total of 17 participants filling out the pre- and posttest questionnaires. All these participants (17/17, 100%) also took part in a telephone interview.
A total of 31 patients were approached by phone to participate in the control group. This resulted in 71% (22/31) of these participants completing the pretest questionnaire. A total of 9% (2/22) of these participants did not complete the posttest questionnaire without giving a reason. In total, 9% (2/22) of the participants started the posttest questionnaire but only partially completed it. This resulted in a control group of 20 participants, of whom 18 (90%) completed the entire pre- and posttest questionnaires.

The participant characteristics at baseline are presented in Table 1. The Mann-Whitney U test revealed no substantial differences between the intervention and control groups in terms of these characteristics.
Table 1. Participant characteristics at baseline by group.

<table>
<thead>
<tr>
<th></th>
<th>Intervention group (n=17)</th>
<th>Control group (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>6 (35)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Male</td>
<td>11 (65)</td>
<td>8 (40)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>62 (12)</td>
<td>66 (14)</td>
</tr>
<tr>
<td><strong>Surgery type, n, (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decompression</td>
<td>11 (65)</td>
<td>10 (50)</td>
</tr>
<tr>
<td>Spinal fusion</td>
<td>6 (35)</td>
<td>10 (50)</td>
</tr>
<tr>
<td><strong>Educational level(^a), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>4 (24)</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Middle</td>
<td>8 (47)</td>
<td>4 (20)</td>
</tr>
<tr>
<td>High</td>
<td>5 (29)</td>
<td>7 (35)</td>
</tr>
<tr>
<td><strong>Occupational status, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paid work</td>
<td>6 (35)</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Retired</td>
<td>8 (47)</td>
<td>10 (50)</td>
</tr>
<tr>
<td>Housewife or househusband</td>
<td>1 (6)</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Incapacitated</td>
<td>1 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>1 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married or living together</td>
<td>15 (88)</td>
<td>18 (90)</td>
</tr>
<tr>
<td>In a relationship; living apart</td>
<td>1 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Widowed</td>
<td>1 (6)</td>
<td>2 (10)</td>
</tr>
</tbody>
</table>

\(^a\)Low: primary and lower secondary education; middle: upper secondary education; high: higher vocational training and university.

RQ 1: What Are the Potential Benefits for Patients Undergoing Spinal Surgery of a Digital ACT and PPI?

First, between-group differences were analyzed at baseline (Mann-Whitney U test). The total score on the PIPS was significantly higher in the control group ($U=262.5; P=.004$). Other measures did not differ significantly between both groups at baseline.

Second, within-group differences were analyzed for both groups between the pre- and posttest assessments (Wilcoxon signed rank test; see Table 2 for the results). When looking at the differences between the pre- and posttest assessment, it should be taken into account that both groups had undergone surgery, which is an intervention by itself. Both groups showed significant improvement in the NRS (pain intensity week average and week worst), HADS anxiety subscale, HADS depression subscale, pain interference subscale of the MPI, and total score on the PIPS between the pre- and posttest time points. The score on the valued living scale of the ELS did not significantly change over time in either of the 2 groups. In the control group, the subscores as well as the total score on the Mental Health Continuum–Short Form did not differ significantly between the pre- and posttest time points. In the intervention group, the total score ($V=99; P=.03$) as well as the score on emotional well-being ($V=55; P=.004$) increased significantly over time.
Table 2. Potential benefits for patients undergoing spinal surgery in the control and intervention groups.

<table>
<thead>
<tr>
<th></th>
<th>Pretest assessment, mean (SD)</th>
<th>Posttest assessment, mean (SD)</th>
<th>Pretest-posttest difference, V</th>
<th>Pretest-posttest difference, P value</th>
<th>Pretest-posttest effect size, Cohen d</th>
<th>Pretest-posttest mean difference—full sample (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PIPS&lt;sup&gt;a&lt;/sup&gt;—total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−13.34 (−19.19 to −7.50)</td>
</tr>
<tr>
<td>Control group (n=18)</td>
<td>62.2 (9.8)</td>
<td>49.7 (11.2)</td>
<td>19</td>
<td>.004</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>52.5 (11.6)</td>
<td>43.8 (11.7)</td>
<td>24.5</td>
<td>.01</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td><strong>NRS&lt;sup&gt;b&lt;/sup&gt;—week average</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−3.80 (−4.57 to −3.03)</td>
</tr>
<tr>
<td>Control group (n=20)</td>
<td>6.7 (2.3)</td>
<td>3.8 (2.6)</td>
<td>0</td>
<td>&lt;.001</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>7.1 (1.1)</td>
<td>2.0 (1.6)</td>
<td>0</td>
<td>&lt;.001</td>
<td>3.71</td>
<td></td>
</tr>
<tr>
<td><strong>NRS—week at worst</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−4.09 (−4.94 to −3.23)</td>
</tr>
<tr>
<td>Control group (n=20)</td>
<td>8.5 (1.5)</td>
<td>5.0 (3.1)</td>
<td>0</td>
<td>&lt;.001</td>
<td>1.44</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>8.3 (1.2)</td>
<td>3.3 (2.5)</td>
<td>0</td>
<td>&lt;.001</td>
<td>2.55</td>
<td></td>
</tr>
<tr>
<td><strong>HADS&lt;sup&gt;c&lt;/sup&gt;—anxiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−1.89 (−2.92 to −0.86)</td>
</tr>
<tr>
<td>Control group (n=20)</td>
<td>7.3 (3.8)</td>
<td>5.2 (3.8)</td>
<td>14</td>
<td>.008</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>4.6 (3.2)</td>
<td>2.8 (2.2)</td>
<td>13.5</td>
<td>.02</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td><strong>HADS—depression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−3.46 (−4.62 to −2.29)</td>
</tr>
<tr>
<td>Control group (n=20)</td>
<td>7.0 (3.8)</td>
<td>4.3 (3.4)</td>
<td>26.5</td>
<td>.006</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>5.8 (3.3)</td>
<td>1.9 (1.8)</td>
<td>4.5</td>
<td>&lt;.001</td>
<td>1.47</td>
<td></td>
</tr>
<tr>
<td><strong>ELS&lt;sup&gt;d&lt;/sup&gt;—valued living</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.60 (−2.28 to 1.08)</td>
</tr>
<tr>
<td>Control group (n=18)</td>
<td>37.8 (4.9)</td>
<td>36.5 (5.9)</td>
<td>50</td>
<td>.57</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>40.4 (5.5)</td>
<td>40.2 (5.8)</td>
<td>57.5</td>
<td>.89</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td><strong>MHC-SF&lt;sup&gt;e&lt;/sup&gt;—emotional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.46 (0.12 to 0.80)</td>
</tr>
<tr>
<td>Control group (n=18)</td>
<td>3.4 (1.2)</td>
<td>3.8 (0.7)</td>
<td>73.5</td>
<td>.44</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>3.7 (1.2)</td>
<td>4.2 (0.7)</td>
<td>55</td>
<td>.004</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td><strong>MHC-SF—social</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.12 (−0.14 to 0.38)</td>
</tr>
<tr>
<td>Control group (n=18)</td>
<td>2.7 (1.1)</td>
<td>2.7 (1.1)</td>
<td>89</td>
<td>.55</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>2.8 (1.0)</td>
<td>2.9 (1.2)</td>
<td>63</td>
<td>.21</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td><strong>MHC-SF—psychological</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.21 (−0.14 to 0.57)</td>
</tr>
<tr>
<td>Control group (n=18)</td>
<td>3.1 (1.1)</td>
<td>3.3 (1.0)</td>
<td>86</td>
<td>.35</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>3.6 (1.0)</td>
<td>3.7 (1.0)</td>
<td>80</td>
<td>.08</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td><strong>MHC-SF total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.23 (−0.04 to 0.50)</td>
</tr>
<tr>
<td>Control group (n=18)</td>
<td>3.0 (1.1)</td>
<td>3.2 (0.9)</td>
<td>106.5</td>
<td>.36</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>3.3 (0.9)</td>
<td>3.5 (0.9)</td>
<td>99</td>
<td>.03</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td><strong>MPI&lt;sup&gt;f&lt;/sup&gt;—pain interference scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−2.07 (−2.59 to −1.55)</td>
</tr>
<tr>
<td>Control group (n=19)</td>
<td>4.4 (0.9)</td>
<td>2.6 (1.4)</td>
<td>9.5</td>
<td>&lt;.001</td>
<td>1.53</td>
<td></td>
</tr>
<tr>
<td>Intervention group (n=17)</td>
<td>4.1 (0.9)</td>
<td>1.6 (1.1)</td>
<td>0</td>
<td>&lt;.001</td>
<td>2.49</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>PIPS: Psychological Inflexibility in Pain Scale.

<sup>b</sup>NRS: Numeric Pain Rating Scale.

<sup>c</sup>HADS: Hospital Anxiety and Depression Scale.

<sup>d</sup>ELS: Engaged Living Scale.

<sup>e</sup>MHC-SF: Mental Health Continuum–Short Form.

<sup>f</sup>MPI: Multidimensional Pain Inventory.
Third, between-group differences were analyzed by comparing the pre- and posttest difference scores of both groups (Mann-Whitney U test). A significant difference was found in the weekly average NRS score between the intervention and control groups ($U=75; P=.003$), showing a significantly larger decline in pain intensity in the intervention group.

**RQ 2: What Is the Feasibility of a Future RCT in Terms of Recruitment and Dropout Rates?**

To determine the feasibility of a future RCT, recruitment, dropout rates, and dropout characteristics were examined. Over a period of 5 months, a total of 35 patients were approached by phone to participate in the intervention group (Figure 3). In total, 51% (18/35) of the approached patients did not want to participate and did not start with or dropped out of the intervention. Of these 18 patients, 14 (78%) were female and 14 (78%) underwent decompression surgery. Most of the dropouts did so before the intervention started as they did not want to participate in the study (10/18, 56%, of whom 9/10, 90% were female; age range 19-79 years). Most of the additional 44% (8/18) of dropouts did so at the pretest (3/8, 38%) or posttest (3/8, 38%) time points, and only 25% (2/8) of these participants dropped out because of nonadherence as they did fill out the pretest questionnaire but did not download the intervention. Of these 2 nonadherers, 1 (50%) was male and 1 (50%) was female; both underwent decompression surgery, and both were aged >65 years.

**RQ 3: What Is the Acceptability of a Digital ACT and PPI Called “Strength Back” by Patients Undergoing Spinal Surgery?**

**Overview**

Individual semistructured interviews were conducted with all 17 participants in the intervention group. The results of these interviews are described in 3 sections: general impression of the intervention, content of the intervention, and suggestions for improvement. Subsequently, engagement with the intervention is discussed by combining input from the interviews, log data, and the scores on the TWEETS [64].

**General Impression of the Intervention**

The general impression of the intervention can be described in terms of added value, recommendation, ease of use, repeated answers, feedback, and notifications.

Most of the participants (15/17, 88%) felt that the intervention had at least some added value in terms of support and as a source of information. PT3 stated the following:

> I always liked it when there was a new module that you had to read, I always found it very interesting. Because each time I read the app, it might contain something that really helps me. That helps me recover faster perhaps. [PT3]

Even when some participants at first stated that they did not really think the intervention was useful, they corrected themselves later in the interviews. For instance, PT8 stated the following:

> It did make me realize that the surgery was not a thing to take too lightly, that I could experience a relapse, so it did comfort me to know that in advance. [PT8]

Of the 17 participants who were interviewed, 2 (12%; PT4 and PT5) stated that the intervention did not have any added value for them. For instance, PT5 stated the following:

> The information from the physician was enough for me, but it might be helpful for other patients. [PT5]

A total of 12% (2/17) of the participants (PT4 and PT8) stated that they found the intervention to be too fluffy or vague and not suitable for them as they did not experience any pain or ups and downs.

When comparing the intervention with the paper brochure that patients normally receive from the hospital, some participants stated that either one would have been fine (eg, PT1 and PT9), whereas others stated that they preferred the intervention as the information was always nearby and, therefore, more accessible than a paper brochure that might get lost (eg, PT2, PT5, and PT6). Some participants stated that they had looked on the internet for information as well but preferred the information in the intervention as this was tailored to their own hospital and, therefore, also more reliable.

Similarly, 88% (15/17) of the participants stated that they would recommend the intervention; one participant (PT4) would not recommend it as it did not offer him anything additional to a paper brochure, and with another participant (PT12), this topic was not discussed. Participants mainly stated that they would recommend the intervention as it helped them prepare for surgery and enabled them to read and reread important information.

Participants indicated that the intervention was easy to use. A total of 12% (2/17) of the participants (PT5 and PT18) mentioned some minor layout suggestions, but the main consensus was that the intervention was easy to download and use, had a nice layout, and contained text in clear language. The functionality of previous answers (eg, preoperative answers on valuable activities to do after surgery) being repeated in later modules was appreciated by the participants:

> Nice to read what I had filled in previously and to see, on a later date that I was doing better when I answered a similar question. [PT3]

As the intervention only provided general information and no personal feedback from a professional, participants were asked whether they had missed this. Several participants (5/17, 29%) indicated that they did miss this personal feedback during the intervention or in general during recovery. This preferred feedback ranged from the ability to click on a help button (PT15) or the possibility of contacting someone for more information (PT1) to reaching out to someone guiding the intervention to reduce the amount of psychological content (PT11 and PT16) or feeling insecure and missing a contact person to consult (PT7).

Owing to an error, participants did not correctly receive a notification every time a new weekly module was available. Participants (6/17, 35%) indicated that receiving a notification...
made them use the intervention more. They would have liked to receive the reminder every week, stating that they would have used the intervention more if they had.

Content of the Intervention
During the interview, all separate components of the intervention (Multimedia Appendices 1 and 2) were discussed with the participants. These elements were clustered as preparation and peer support, hospital information, and positive psychology and ACT content.

Preparation and Peer Support
Almost all participants (14/17, 82%) appreciated the videos of the nursing ward and the surgery room. These participants stated that they found watching the videos informative and comforting. One of the participants who did not appreciate the videos (PT11) stated the following:

It might be comforting for little children, but as an adult you know what a hospital looks like and what will happen. [PT11]

The practical tips from previous patients were valued by most participants (11/17, 65%). In total, 24% (4/17) of the participants did not see or remember this element, and 12% (2/17) of the participants would have liked to see more tips. The module describing the fluctuating recovery process was not remembered by several participants (8/17, 47%) but was valued greatly by those who did remember it:

I could really recognize myself in the text, having had pain for so long myself. [PT14]
It really resembled the reality. [PT18]

Participants reacted with mixed feelings to the quotes of previous patients. In total, 12% (2/17) of the participants (PT1 and PT16) stated that it might benefit other people but it was not for them. Several participants (7/17, 41%) could not remember the module (PT2, PT8, and PT15) or did not find it interesting (PT4, PT9, PT11, and PT13). These participants stated that they wanted to do things their own way and were not interested in the stories of other patients. Other participants (8/17, 47%) found the quotes valuable, recognizable, and comforting.

Hospital Information
The information on the spinal condition and on the surgical procedure was appreciated by the participants. Some stated that it was clearer than what the physician had told them or that it was nice to be able to read and reread it in the intervention, whereas other participants stated that it had no added value to them as the information resembled that of the paper brochure. The information on physical therapy did not match reality for some participants (3/17, 18%), but the discharge criteria were useful to see before surgery. In total, 18% (3/17) of the participants (PT1, PT4, and PT9) thought that the physical guidelines in the intervention had no added value as they were also available in the paper brochure. Another 18% (3/17) of the participants (PT11, PT15, and PT18) liked the fact that the physical guidelines were in the intervention but would have preferred them to be more specific or elaborate. The other participants (11/17, 65%) found the guidelines pleasant, useful, and supportive. Most participants (11/17, 65%) appreciated the pain medication module. The other participants (6/17, 35%) liked the fact that the physical guidelines were in the intervention but would have preferred them to be more specific or elaborate. The other participants (11/17, 65%) found the guidelines pleasant, useful, and supportive. Most participants (11/17, 65%) appreciated the pain medication module. The other participants (6/17, 35%) preferred the paper brochure, did not see the module, or did not use it. The contact details of the hospital in the intervention were considered clear and useful by all participants.

PP and ACT Content
The PP and ACT content in the intervention evoked the strongest opinions in the participants (Table 3). Several participants (5/17, 29%) stated that the amount of psychological content in the intervention was too high and made them less engaged with the intervention (like or use it less):

The amount of psychological content was too much for me. I wasn’t raised that way and it doesn’t appeal to me. It almost made me stop using the intervention all together. [PT17]
Table 3. Participants’ opinions on the positive psychology and acceptance and commitment therapy content of the intervention ranked in order of most to least appreciated exercise (n=17).

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Appreciated, n (%)</th>
<th>Not appreciated, n (%)</th>
<th>Not remembered, n (%)</th>
<th>Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotion quadrant</td>
<td>13 (76)</td>
<td>4 (24)</td>
<td>0 (0)</td>
<td>“It was good to be made aware of how I was feeling at that moment.” [PT5]</td>
</tr>
<tr>
<td>What makes the surgery worthwhile?</td>
<td>13 (76)</td>
<td>4 (24)</td>
<td>0 (0)</td>
<td>“A really good question. A question I would have liked to have asked myself two years ago, because this is what it is really about, in the end. Why to choose for surgery or not.” [PT12]</td>
</tr>
<tr>
<td>Formulating positive statements</td>
<td>12 (71)</td>
<td>5 (29)</td>
<td>0 (0)</td>
<td>“It made me start a conversation with my spouse about my recovery process.” [PT6]</td>
</tr>
<tr>
<td>Mindful enjoying</td>
<td>11 (65)</td>
<td>3 (18)</td>
<td>3 (18)</td>
<td>“Walking is something you do without thinking about it. And before surgery it was terrible, walking. So becoming aware of the fact that walking is possible without pain...It’s good to become aware of these things.” [PT6]</td>
</tr>
<tr>
<td>What do you desire to do (again) after surgery?</td>
<td>11 (65)</td>
<td>4 (24)</td>
<td>2 (12)</td>
<td>“It made me melancholic or sad, because it were things that I had not been able to do for a very long time.” [PT7]</td>
</tr>
<tr>
<td>3 positive things</td>
<td>9 (53)</td>
<td>5 (29)</td>
<td>3 (18)</td>
<td>“Yes, I absolutely loved it. I was busy in the kitchen, cooking. I was able to do that all by myself again. It was really a great day.” [PT11]</td>
</tr>
<tr>
<td>Wish question</td>
<td>8 (47)</td>
<td>7 (41)</td>
<td>2 (12)</td>
<td>“It got me out of the moment when I was in pain and helped me to see what made it worthwhile.” [PT7]</td>
</tr>
<tr>
<td>Mindfulness exercises</td>
<td>6 (35)</td>
<td>5 (29)</td>
<td>0 (0)</td>
<td>“It has really helped me, especially in the beginning. Further along during recovery I didn’t need it anymore.” [PT7]</td>
</tr>
<tr>
<td>Uploading a valuable picture</td>
<td>6 (35)</td>
<td>5 (29)</td>
<td>6 (35)</td>
<td>“I could upload a picture of my grandchildren, but I don’t see the use or need for such an exercise.” [PT5]</td>
</tr>
<tr>
<td>Video of how pain works</td>
<td>4 (24)</td>
<td>5 (29)</td>
<td>8 (47)</td>
<td>“That really was an eye-opener for me. It all fell into place.” [PT12]</td>
</tr>
<tr>
<td>Write a letter to yourself</td>
<td>2 (12)</td>
<td>15 (88)</td>
<td>0 (0)</td>
<td>“I really did not see the use of this exercise.” [PT9]</td>
</tr>
</tbody>
</table>

The potential relaxing effect of the mindfulness exercises was not experienced by 12% (2/17) of the participants (PT3 and PT9) even though they did do the exercises. Other participants (4/17, 24%) stated that the exercises were not for them as they already knew the techniques (eg, from yoga) but suggested that they might benefit other patients. The least appreciated exercise was writing a letter to themselves to support themselves in difficult times during recovery. Exercises focusing on values and positive statements were the most appreciated by the participants. Table 3 provides a more detailed overview.

**Suggestions for Improvement**

**Overview**

The participants also gave some suggestions for improvement. For instance, PT1, PT2, PT6, and PT15 stated that the information in the intervention was sometimes too elaborate and could have been more concise. Some suggestions about the usability of the intervention were mentioned (eg, on where to position the menu of the intervention or to enlarge the font size for older participants). Several participants would have liked the intervention to have less psychological content (eg, PT4, PT12, and PT15) or at least have the option to influence the amount of these exercises or content while using the intervention (eg, PT17). Other points of interest were the need for more specific physical guidelines (eg, PT11), more physical (therapy) exercises (PT18), and information about returning to work (PT17).

**Engagement With the Intervention**

Engagement with the intervention is described through the topics usage, reasons for not using, and number of modules, as discussed in the interviews, together with scores on the TWEETS and log data.

In the interviews, some participants (3/17, 18%) mentioned that they mainly used the intervention before surgery, and PT1 even used it almost daily before surgery. Other participants primarily used the intervention after surgery (eg, “almost daily to check the physical guidelines” [PT2 and PT11]), especially the first few weeks after surgery, when they were more restricted in movements and were in bed for most of the day (eg, PT9, PT12, PT14, and PT17). These participants stated that, as their recovery—and, therefore, their mobility—progressed, their use of the intervention declined.
I noticed that the more I recovered, the less I used the app. [PT9]

Other participants used the intervention both before and after surgery (eg, PT18 and PT3):

I must have read the entire app at least 10 times, I only missed one weekly module because I got COVID. [PT3]

Interestingly, all participants stated that the fact that they received a new module every week helped them use the intervention more often, as did viewing the other content of the intervention while filling out the weekly module.

Reasons for not using the intervention (more) were the lack of automatic notifications or reminders (eg, PT16), feeling no pain postoperatively, feeling that pain management was the main focus of the intervention (PT11), or lack of interest in the (large amount of) psychological content (eg, PT9, PT11, PT16, and PT17).

Most participants (12/17, 71%) felt that the number of modules was correct. The other 29% (5/17) of the participants stated that fewer modules would have sufficed, especially in the later weeks after surgery.

The mean total score of all participants in the intervention group on the TWEETS was 2.6 out of a possible 4 (SD 0.7), which is more than the average answering option of the scale (65% of the maximum score). The highest-scoring items on the TWEETS were “Strength Back is easy to use” and “Strength Back helps me to get more insight into my preparation before surgery and recovery after surgery” with an average score of 2.9 and 3.1, respectively. The lowest-scoring items on the TWEETS were “Strength Back is part of my daily routine” with an average score of 1.9 and “Strength Back fits me as a person” with an average score of 2.1.

A total of 15 weekly modules were offered to the patients undergoing spinal fusion (6/17, 35%). Log data of these patients showed that the average number of weekly modules completed was 10.2 out of 15 (SD 3.9; Table 4). The least completed weekly modules were POST-6, POST-7, and POST-12 (2/6, 33% in all cases). All patients undergoing spinal fusion completed the weekly POST-2 and POST-3 modules.

### Table 4. Overview of completed weekly modules per participant—patients undergoing spinal fusion (N=6).

<table>
<thead>
<tr>
<th>Participant number</th>
<th>PRE</th>
<th>POST</th>
<th>Total, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>PT3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PT5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PT8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PT9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PT12</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PT17</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Total, n (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5 (83)</td>
<td>4 (67)</td>
<td>4 (67)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Total number of completed modules; maximum of 15 modules.

<sup>b</sup>Total number of participant completing a specific module; maximum of 6 participants.

<sup>c</sup>N/A: not applicable.

Several participants kept using the intervention until the last weekly module. The least completed weekly modules for patients undergoing decompression surgery were the PRE-module 3 (5/11, 45%) and POST-module 5 (4/11, 36%), whereas the PRE-module 1 was completed by almost all participants (10/11, 91%; Table 5).
Table 5. Overview of completed weekly modules per participant—patients undergoing decompression surgery (N=11).

<table>
<thead>
<tr>
<th>Participant number</th>
<th>PRE</th>
<th>POST</th>
<th>Total completed, n (%)^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT1</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT2</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT4</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT6</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT7</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT10</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT11</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT13</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT14</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT15</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>PT16</td>
<td>✓</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
</tbody>
</table>

Total respondents, n (%)^b | 10 (91) | 9 (82) | 5 (45) | 8 (73) | 7 (64) | 6 (55) | 7 (64) | 4 (36) | 7 (64) | N/A^c |

^aTotal number of completed modules; maximum of 9 modules.
^bTotal number of participant completing a specific module; maximum of 11 participants.
^cN/A: not applicable.

Discussion

Principal Findings

The objective of this study was 3-fold: to explore the potential benefits for patients undergoing spinal surgery of the digital ACT and PPI Strength Back (RQ 1), explore the feasibility of a future RCT in terms of recruitment and dropout (RQ 2), and assess the acceptability of Strength Back by patients undergoing spinal surgery (RQ 3).

The focus of this study was to explore the potential benefits of the intervention and not to determine its effectiveness. The latter should be the focus of a future RCT including more participants and allowing for more robust statements. Nonetheless, this study does show that Strength Back seems promising as pain intensity decreased more in the intervention group than in the control group and emotional well-being as well as overall well-being improved in the intervention group but not in the control group (RQ 1). This is in line with research in patients with chronic pain showing beneficial effects of ACT and PPIs compared with a control group on pain intensity and emotional functioning [65,66], in the treatment of chronic pain [30-35], in improving affect and functional ability after knee surgery [21], and in quicker cessation of pain and opioid use in veterans after orthopedic surgery [36].

Contrary to our expectations, no differences between the groups were observed in improvement in pain interference and psychological flexibility. Possibly, the effect of the ACT and PPI content shows through in the significantly larger decline in pain intensity in the intervention group. It is reasonable to suppose that a decline in pain intensity also helps prevent pain interference with daily activities and improve psychological flexibility. However, the sample size may have been too small, or the effect of surgery may have been so massive that these effects did not appear in this study. Moreover, in the intervention group as well as the control group, valued living and psychological and social well-being were almost the same at the pre- and posttest time points. The amount of ACT and PPI content may have been too small to influence valued living or psychological and social well-being. It is also possible that these effects only occur in the longer term. Living more accordingly to one’s values, resulting in enhanced psychological and social well-being, may only start later in the recovery process and not immediately after surgery.

In terms of recruitment and dropout rates, a future RCT seems feasible (RQ 2). A large proportion of the approached patients wanted to participate in the study, and once they started with the intervention, almost all participants used it until the posttest assessment and final interview. The willingness to use a digital health intervention seemed quite high in our study compared with in previous studies [67,68]. It seems that, once participants started the intervention in this study, they saw its value and kept using it over a longer period. In line with other research [69], patients stated that receiving notifications improved their use of the intervention. The dropout numbers found in this study were low and correspond to those of a recent RCT on PP exercises for patients with chronic pain [66]. However, to recruit a larger number of participants for an RCT, multiple hospitals or orthopedic centers should be included as there are a limited number of potential participants in a single facility. Interestingly, almost all the patients who were approached and did not want to participate in our study were female. Perhaps this group needs more attention during recruitment to explore their reasons for not wanting to participate.

The vast majority of participants in the intervention group were positive about the digital health intervention and would recommend it to future patients (RQ 3). The scores on the
TWEETS indicated a moderate to above-moderate engagement with the intervention, which was in line with log data showing that participants completed on average approximately 75% of the available weekly modules. The information modules containing videos of the surgery room and nursing ward, practical tips from previous patients, physical activity guidelines, and pain medication were the most appreciated by participants. The PP and ACT content evoked the strongest opinions in the participants, with a small minority indicating that they preferred no psychological content at all, whereas most of the participants saw added value in a number of specific exercises. Although the development of the intervention was based on a participatory design process and most participants appreciated the content, others felt that the focus was too much on mental health, whereas they experienced their issues as physical. At the same time, to increase the effectiveness of the intervention on the ACT- and PP-related outcomes, even more psychoeducation and exercises might be needed. In addition, the current version of Strength Back focuses on certain elements of ACT, whereas it may be necessary to address all ACT processes to achieve optimal benefits for patients undergoing spinal surgery. Indeed, Carr et al [29] found in their meta-analysis of PPIs that interventions were more effective when they contained multiple PPIs, were of longer duration, and contained more sessions. This poses a dilemma for the further development of this intervention. Information on the psychological content of the intervention before inclusion and tailoring the content and dose of the intervention to individual patients might further improve acceptance of the PP exercises, as suggested by previous research [70-72]. In a future version of Strength Back, patients could be introduced with a few of the most appreciated ACT and PP exercises and from there on be given the possibility to determine the amount of ACT and PP content themselves. This might reduce the potential effect on some patients who opt for a lower amount of ACT and PP content but might increase overall adherence to and acceptance of the intervention.

Our results question what the most appropriate primary outcome is for a subsequent RCT. As the intervention is based on ACT and PP, it makes sense to choose a measure corresponding to this content, such as well-being or pain interference. However, our feasibility study shows that the intervention and control groups particularly differed in improvement in pain intensity, which is an outcome that is important to patients but not directly targeted within the intervention. Perhaps targeting both pain intensity and increasing well-being might be the best option for sustainable resilience in patients undergoing spinal surgery. This is in line with previous research proposing a balanced, complaint and strength–oriented approach to reach sustainable mental health [73].

Second, attention should be paid to what the appropriate process measurements in a subsequent RCT would have to be. On the basis of the content and assumed working mechanism of this intervention, we propose psychological flexibility as a process measure in a future RCT. Nonetheless, we would not advise using the PIPS to measure this concept as we found low reliability in our study. Other studies have found similar issues with this scale and have recommended only interpreting it as a whole [74] or only using the avoidance subscale [75]. In addition, the PIPS measures inflexibility and not flexibility, which does not fit as well with the focus of ACT and PP. Perhaps the Personalized Psychological Flexibility Index [76] is a suitable alternative measure of psychological flexibility for a future RCT. In addition, we propose pain interference as a process measure in a future RCT, measured using the MPI subscale of pain interference. In addition, the ELS might be used in a future RCT to measure engaged living, including valued living.

The final discussion point relates to the length and timing of the intervention. The current intervention starts before surgery and lasts 6 to 12 weeks after surgery (for decompression and spinal fusion surgery, respectively). This was regarded positively by most participants, and from both the interviews and log data, we learned that most participants were already active in the intervention before surgery. This focus on pre- and postoperative timing is supported by a recent systematic review on perioperative psychological interventions that found that interventions (also) delivered after surgery tended to be more effective for postsurgical pain and disability than interventions delivered (only) before surgery [77]. It seems that, while waiting for surgery, patients are quite eager to learn more about the surgery and rehabilitation afterward. An intervention such as Strength Back can take advantage of these “teachable moments” [39] by providing content that participants may not have found or chosen themselves but that is known to help them in the recovery process. The duration of our intervention is in line with a recent review on perioperative psychological interventions for patients undergoing spinal fusion showing a reduction in pain and disability starting from immediately after surgery up to 3 months [78].

Strengths of This Study

The high recruitment and relatively low dropout rates were a strength of this study. In addition, qualitative data were collected through individual interviews with almost all the users of the intervention. This yielded crucial information on the value of the intervention and the reasons behind participants’ use of the intervention. Using multiple methods combining qualitative data with log data and quantitative data retrieved through the questionnaires provided a complete picture of the value of the intervention. Therefore, this research design is close to a convergent parallel mixed methods design [79]. Another strength of this study is the underlying theoretical framework of the intervention. As this is lacking for many digital health interventions, there are calls for further research to enhance the scope and use of this technology [80,81]. The design of digital interventions should be anchored in behavior change theories to optimally engage patients in the intervention and behavior change [82,83]. In a review of 85 studies using the Behavior Change Support Systems by Oinas-Kukkonen and Harjumaa [82], less than half referred to theories of behavior change, but those that did were uniformly successful [84]. Clearly, a sound theoretical base for digital health interventions is warranted for optimal behavior change and effectiveness.

Limitations of This Study

As all participants underwent surgery, which is an intervention by itself, we wanted to see in this study what the digital health
intervention might add to the effect of the surgery. An exploratory design regarding outcome measures was used to see which potential benefits could be achieved for patients undergoing spinal surgery. Owing to this exploratory design and the small sample size, more and larger effects of the intervention were not expected to be found in this study. In addition, no correction for multiple testing was performed in this study, and thus, conclusions can only be drawn with caution. Therefore, the preliminary findings of this study need to be replicated in a future larger study.

This study used a historic control group. A future RCT should use a design in which the intervention group and the control group run parallel in time and participants are assigned to them randomly.

In this study, log data were gathered on modules in which participants answered questions. For a future study, we would recommend registering every log-in and activity of users instead of only activities in which participants provide answers. We now see certain patterns in the use of participants, but more data are needed to further investigate use of and engagement with the intervention.

Previous hospital experience, preexisting comorbidities, patient technology readiness, health literacy, and ward factors such as staffing were not measured in this study. As these factors may influence how prepared patients are and what their capability is to engage with the intervention, they should be considered in a future RCT.

Conclusions
This study shows that combining ACT and PP in a digital health intervention is promising for patients undergoing spinal surgery as the content was accepted by most of the participants and (larger) improvements in pain intensity and well-being were found in the intervention group. A digital intervention for patients undergoing (spinal) surgery can use teachable moments, when patients are open to learning more about the surgery and rehabilitation afterward, to also provide patients with content that they may not have found or chosen themselves but that is known to help them in the recovery process.

Acknowledgments
The authors would like to thank all the participating patients as well as Leonie Olde Keizer-Spanjer, Leid de Vries-Gerritsen, Mark Havinga, and their colleagues at (Orthopedic Centre Oost Nederland) OCON for their help and contribution during this study. This work was supported by an educational grant from the Dutch Ministry of Education, Culture, and Science provided by the Saxion University of Applied Sciences. The funding body had no role in the study design; collection, analysis, or interpretation of the data; writing of the report; or the decision to submit the paper for publication.

Data Availability
The data sets analyzed during this study are not publicly available because of the exploratory and mixed methods nature of the study, but they are available from the corresponding author upon reasonable request.

Authors’ Contributions
AVDH was responsible for conceptualization, methodology, validation, formal analysis (qualitative and quantitative data), investigation, data curation, visualization, writing—original draft, project administration, and funding acquisition. LM was responsible for validation, formal analysis (qualitative data), investigation, and writing—review and editing. HVOM was responsible for validation, formal analysis (quantitative data), and writing—review and editing. JSJ was responsible for conceptualization, methodology, writing—review and editing, and supervision. EB was responsible for conceptualization, methodology, validation, writing—review and editing, supervision, and funding acquisition. KMGS was responsible for conceptualization, methodology, validation, writing—review and editing, supervision, and funding acquisition. SK was responsible for conceptualization, methodology, validation, formal analysis, writing—original draft, writing—review and editing, and supervision.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Content of the Strength Back digital health intervention for patients undergoing spinal surgery.
[DOCX File, 21 KB - formative_v8i1e54600_app1.docx ]

Multimedia Appendix 2
Overview of positive psychology and acceptance and commitment therapy exercises in the weekly modules of the Strength Back digital health intervention for patients undergoing spinal surgery.
[DOCX File, 22 KB - formative_v8i1e54600_app2.docx ]

Multimedia Appendix 3
Overview of the timing of the positive psychological and acceptance and commitment therapy exercises in the weekly modules of the Strength Back digital health intervention for patients undergoing spinal surgery.

References


75. Barke A, Riecke J, Rief W, Glombiewski JA. The Psychological Inflexibility in Pain Scale (PIPS) - validation, factor structure and comparison to the Chronic Pain Acceptance Questionnaire (CPAQ) and other validated measures in German chronic back pain patients. BMC Musculoskelet Disord 2015 Jul 28;16:171 [FREE Full text] [doi: 10.1186/s12891-015-0641-z] [Medline: 26215038]


Abbreviations

ACT: acceptance and commitment therapy
ELS: Engaged Living Scale
HADS: Hospital Anxiety and Depression Scale
MPI: Multidimensional Pain Inventory
NRS: Numeric Pain Rating Scale
PIPS: Psychological Inflexibility in Pain Scale
Edited by A Mavragani; submitted 16.11.23; peer-reviewed by J McDonall, L Yu; comments to author 15.12.23; revised version received 21.12.23; accepted 22.12.23; published 07.02.24.

Please cite as:
van der Horst A, Meijer L, van Os - Medendorp H, Jukema JS, Bohlmeijer E, Schreurs KMG, Kelders S
Benefits, Recruitment, Dropout, and Acceptability of the Strength Back Digital Health Intervention for Patients Undergoing Spinal Surgery: Nonrandomized, Qualitative, and Quantitative Pilot Feasibility Study
JMIR Form Res 2024;8:e54600
URL: https://formative.jmir.org/2024/1/e54600
doi: 10.2196/54600
PMID: 38324374

©Annemiek van der Horst, Laura Meijer, Harmieke van Os - Medendorp, Jan S Jukema, Ernst Bohlmeijer, Karlein MG Schreurs, Saskia Kelders. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Web-Based Screening, Brief Intervention, and Referral to Treatment for Traumatic Stress and Alcohol Misuse Among Survivors of Sexual Assault and Intimate Partner Violence: Usability and Acceptability Study

Christine Hahn¹, PhD; Emily Tilstra-Ferrell¹, PhD; Selime Salim¹, PhD; Nada Goodrum², PhD; Alyssa Rheingold¹, PhD; Amanda K Gilmore³, PhD; Sara Barber⁴, MA; Angela Moreland¹, PhD

¹Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, SC, United States
²Department of Psychology, University of South Carolina, Charleston, SC, United States
³Department of Health Policy & Behavioral Sciences, Georgia State University, Atlanta, GA, United States
⁴South Carolina Coalition Against Domestic Violence and Sexual Assault, Columbia, SC, United States

Corresponding Author:
Christine Hahn, PhD
Department of Psychiatry and Behavioral Sciences
Medical University of South Carolina
67 President St. MSC 861
Charleston, SC, 29425
United States
Phone: 1 8437928209
Email: hahnc@musc.edu

Abstract

Background: Recent survivors of intimate partner violence (IPV) and sexual assault (SA) are at a high risk for traumatic stress and alcohol misuse. IPV and SA survivors face barriers to services for traumatic stress and alcohol misuse and have low service utilization rates. One way to increase access to services for this population is the use of web-based screening, brief intervention, and referral to treatment (SBIRT), an evidence-informed approach for early identification of traumatic stress and alcohol and drug misuse and connecting individuals to treatment.

Objective: This study aims to assess the usability and acceptability of a web-based SBIRT called CHAT (Choices For Your Health After Trauma) tailored to address traumatic stress and alcohol misuse following past-year IPV, SA, or both.

Methods: Phase 1 involved gathering feedback about usability and acceptability from focus groups with victim service professionals (22/52, 42%) and interviews with past-year survivors of IPV, SA, or both (13/52, 25%). Phase 2 involved gathering feedback about the acceptability of an adapted version of CHAT in an additional sample of recent survivors (17/52, 33%). Survey data on history of IPV and SA, posttraumatic stress disorder symptoms, alcohol and drug use, and service use were collected from survivors in both phases to characterize the samples. Qualitative content and thematic analyses of the interviews and focus group data were conducted using a coding template analysis comprising 6 a priori themes (usability, visual design, user engagement, content, therapeutic persuasiveness, and therapeutic alliance).

Results: Six themes emerged during the focus groups and interviews related to CHAT: usability, visual design, user engagement, content, therapeutic persuasiveness, and therapeutic alliance. Phase 1 providers and survivors viewed CHAT as acceptable, easy to understand, and helpful. Participants reported that the intervention could facilitate higher engagement in this population as the web-based modality is anonymous, easily accessible, and brief. Participants offered helpful suggestions for improving CHAT by updating images, increasing content personalization, reducing text, and making users aware that the intervention is confidential. The recommendations of phase 1 participants were incorporated into CHAT. Phase 2 survivors viewed the revised intervention and found it highly acceptable (mean 4.1 out of 5, SD 1.29). A total of 4 themes encapsulated participant’s favorite aspects of CHAT: (1) content and features, (2) accessible and easy to use, (3) education, and (4) personalization. Six survivors denied disliking any aspect. The themes on recommended changes included content and features, brevity, personalization, and language access. Participants provided dissemination recommendations.
Conclusions: Overall, CHAT was acceptable among victim service professionals and survivors. Positive reactions to CHAT show promise for future research investigating the efficacy and potential benefit of CHAT when integrated into services for people with traumatic stress and alcohol misuse after recent IPV and SA.

(JMIR Form Res 2024;8:e49557) doi: 10.2196/49557

KEYWORDS
screening, brief intervention, and referral to treatment; brief intervention; intimate partner violence; sexual assault; substance use; alcohol use; mobile phone

Introduction

Background
Sexual assault (SA) and intimate partner violence (IPV) remain major public health concerns for both women and men. In the United States, an estimated 43.6% of women and 24.8% of men experience some form of contact sexual violence in their lifetime, and 21.3% of women and 2.6% of men report being a survivor of attempted or completed rape [1]. Furthermore, more than one-third of women (36.4%) and men (33.6%) in the United States have experienced contact sexual violence, physical violence, or stalking by an intimate partner [1]. Individuals whose gender identity is transgender, gender queer, or nonbinary experience higher rates of SA and IPV compared with cisgender individuals [2,3]. Survivors of IPV and SA are at high risk for traumatic stress symptoms and alcohol misuse (ie, >4 drinks per day or 14 drinks per week for men and >3 drinks per day or 7 drinks per week for women) [4-6]. Although evidence-based treatments for traumatic stress and alcohol misuse exist, survivors of recent IPV and SA receive related services at low rates [7-9] and face many barriers to engaging in treatment [10]. Screening, brief intervention, and referral to treatment (SBIRT) is an evidence-based approach for identifying and reducing alcohol misuse, and web-based delivery shows promise in decreasing alcohol use, addressing barriers, and increasing treatment engagement among survivors of IPV and SA [11,12]. This study examined the usability of a web-based SBIRT intervention for traumatic stress and alcohol misuse tailored to recent survivors of IPV and SA.

Traumatic Stress and Related Alcohol Misuse Among Survivors of IPV and SA
Survivors of IPV and SA are at a high risk for traumatic stress and alcohol misuse [4-6]. A nationally representative study revealed that, when accounting for sociodemographic factors such as age, ethnicity, marital status, income, and education, people who experienced IPV in the past year were more likely to have problematic use of alcohol in the past year [13]. The effect of recent IPV remained for alcohol use even when accounting for past-year mood and anxiety disorders, lifetime personality disorders, and IPV perpetration. Among a national sample of lifetime female survivors of SA, the current prevalence of alcohol use disorder (AUD) ranged from 5% to 20% depending on the type of rape (ie, forced rape, incapacitated rape, or combined type) [6], and more than half of the people who receive an SA medical forensic examination report alcohol misuse [14]. Symptoms of traumatic stress are also high after recent IPV and SA, with 57% of the lifetime survivors of IPV [15] and 74% of the past-month survivors of SA [5] reporting traumatic stress symptoms.

Self-medication or using alcohol to cope with trauma-related distress is theorized to account for high rates of alcohol misuse among survivors of IPV and SA [16-18]. Existing research supports the self-medication hypothesis among survivors of IPV and SA. The negative sequelae of exposure to IPV and SA, including posttraumatic stress disorder (PTSD), depression, and other mental health difficulties, serve as mediating factors linking IPV and SA to increased alcohol use, consistent with the self-medication hypothesis [16,17]. Qualitative data also point to alcohol use as means to cope with the emotional distress of IPV [19]. Similarly, longitudinal evidence suggests that traumatic stress symptoms predict subsequent increase in alcohol use [16].

Aligned with the self-medication hypothesis, the motivational model of alcohol use suggests that people make decisions about drinking based on the expected positive consequences (eg, avoid negative affect) and negative consequences (eg, approach positive affect) [20]. Among people with SA or IPV histories, negative affect from trauma-related distress increases the likelihood of using alcohol to cope [21,22]. Thus, increasing motivation to reduce alcohol use and teaching alternative coping strategies to use during periods of negative affect, particularly in response to trauma-related distress, recently after exposure to IPV and SA could have a substantial impact on decreasing the development and sustainment of alcohol misuse. In addition, the empowerment model, which is often applied to interventions for survivors of SA, highlights the importance of aligning a survivor’s long-term goals with their behaviors in trauma recovery [23]. Therefore, aligned with the empowerment process model, empowering survivors by giving power and control over choices about their health [24] and helping survivors to identify how their values align with their current alcohol use could enhance the likelihood of reducing use after IPV and SA.

Mental Health Service Use for Traumatic Stress and Alcohol Misuse Among Survivors of IPV and SA
Interventions that are effective at addressing traumatic stress and co-occurring alcohol misuse are available. For example, COPE (Concurrent Treatment of PTSD and Substance Use Disorders Using Prolonged Exposure) [25] and Seeking Safety [26] are effective integrated treatments for addressing traumatic stress and alcohol misuse simultaneously. In addition, traumatic stress treatment alone (ie, Cognitive Processing Therapy) has been shown to reduce alcohol and other substance use among survivors of SA [27]. Finally, a psychoeducational video shown to survivors of recent SA during a forensic examination was...
found to be an effective early intervention to reduce the risk of alcohol misuse among recent survivors with a previous history of exposure to SA [28]. Although numerous treatment options for traumatic stress and alcohol misuse are available, survivors of IPV and SA face barriers to accessing treatment for these conditions. Despite high rates of alcohol misuse among survivors, it is estimated that only 20%-35% of survivors seek medical, psychiatric, or mental health services [29,30], with an even smaller subset seeking treatment for substance use disorders. People in abusive partner relationships may be hindered by their partner from seeking treatment, may not have independent financial resources, or may be subject to exacerbated violence or retribution for seeking treatment [31]. Furthermore, survivors of IPV and SA may not disclose their alcohol misuse to service providers as active substance use can be an exclusion criterion for accessing safe housing in zero-tolerance IPV shelters [19].

Additional barriers to treatment engagement include the lack of available mental health and substance use resources in the community, limited transportation particularly for rural residents, stigma associated with behavioral health service utilization, immigration status for undocumented survivors, and a lack of culturally sensitive services [19,31]. Stigma and discrimination associated with intersecting racial, ethnic, gender, and sexual identities can create additional obstacles to receiving care [32,33]. Furthermore, research highlights the potential for further harm when survivors seek services owing to limited training and awareness of IPV and SA-related issues among staff, which may increase fears that survivors will not be believed or will be blamed for the IPV or SA [10]. Staff in substance use disorder treatment settings report feeling ill-equipped to identify and address IPV [30]. In addition, even for survivors who access victim-related services, treatment services for alcohol misuse are typically not integrated within these settings [34], leaving a critical care gap for survivors of IPV and SA with alcohol misuse.

Use of SBIRT for Alcohol and Drug Misuse

SBIRT is a public health–based approach to screening for alcohol misuse, assessing the level of risk, and providing an appropriate intervention that can include no intervention, brief motivational interviewing, or referral to AUD treatment [35]. SBIRT has been effectively extended to address the screening, intervention, and referral to treatment needs of survivors of trauma with alcohol misuse, known as T-SBIRT. This approach has shown promise in elevating the rates of referrals for survivors of trauma experiencing alcohol misuse to specialized mental health services [36]. SBIRT has shown efficacy in reducing alcohol use and has been applied across a range of clinical and community settings, including primary care clinics, emergency departments, outpatient medical settings, and employee assistance programs [37].

In recent years, there has been an increase in efforts to adapt SBIRT to electronic health technologies (eg, computer, web, and phone based) because it may increase disclosure of alcohol misuse owing to increased comfort and decrease provider barriers to screening and providing brief interventions [38]. However, limited research has examined web-based SBIRT as an early intervention among survivors of IPV and SA. A randomized controlled trial of a web-based SBIRT intervention for identifying and addressing IPV among women who used substances, Women Initiating New Goals of Safety, was developed and increased survivors’ likelihood of seeking follow-up care for IPV and reduced drug use at a 3-month follow-up [12]. Similarly, Brief Spousal Assault Form for the Evaluation of Risk [11], a web-based intervention for co-occurring substance use and IPV was feasible and acceptable among a sample of women presenting to the emergency department. Safe and Healthy Experiences is another computerized SBIRT intervention specific for alcohol misuse delivered on an iPad tailored for female Veterans who have lifetime experiences of SA seeking services in primary care [39]. Results from these studies support that SBIRT delivered via eHealth is feasible and acceptable; however, preliminary efficacy results on substance use outcomes were mixed.

These existing interventions are limited because they are not specific to recent IPV, SA, alcohol misuse, or traumatic stress. SBIRT may be advantageous in the months after IPV and SA because this is a period when there is a risk for patterns of alcohol misuse to intensify because of elevated trauma-related stress [5]. SBIRT was developed for use in medical settings, such as emergency care centers, clinics, and primary care, with the intention of reaching people who are at a high risk for alcohol misuse. However, most recent survivors, who would be appropriate for SBIRT given the high-risk period for alcohol misuse, do not receive related medical care [14,40,41]. Although some limited studies have identified early interventions for survivors of recent IPV and SA, there is no consensus on the best approach to early intervention following recent SA and IPV [42]. To address gaps in service provision and research, we developed a web-based SBIRT called CHAT (Choices For Your Health After Trauma), which is compatible for use on a smartphone and, therefore, has the potential to be disseminated in a variety of community settings and on social media, increasing reach. CHAT is tailored for recent survivors of IPV and SA and provides SBIRT primarily for alcohol misuse, while also screening and offering psychoeducation about traumatic stress and drug use, which commonly co-occur with alcohol misuse among survivors of SA [14,41]. The SBIRT intervention is based on the motivational model of alcohol use [20], self-medication theory [16], and the empowerment model [23] and applies principles of motivational interviewing to reduce motives to drink alcohol and increase valued living, particularly when experiencing negative affect, which are theoretical and empirically supported intervention targets for alcohol misuse. CHAT follows the core SBIRT model components. First, the intervention provides psychoeducation about alcohol use specific to IPV and SA. Next, screening and assessment of alcohol misuse, drug use, and traumatic stress symptoms are completed by administering brief validated, standardized measures (ie, Alcohol Use Disorder Identification Test–Concise, AUDIT-C [43], item 2 from the National Institute of Drug Abuse-Modified Assist [44], and the Primary Care PTSD Screen for DSM-5 [45]). Next, participants receive personalized feedback about trauma symptoms, coping with traumatic stress symptoms (eg, self-blame and nightmares), drinking quantity, recommended drinking limits, money spent on alcohol and drugs, and the...
impact of drinking and drugs on recovery from IPV and SA adapted from the National Institute on Alcohol Abuse and Alcoholism Rethinking Drinking [46]. In addition, brief exercises aimed at increasing motivation and healthy alternatives to alcohol use, including value identification, readiness to change ruler, goal setting, identification of coping skills, and social support adapted from Brief Spousal Assault Form for the Evaluation of Risk [11]. Finally, a personalized printable plan with all psychoeducational information and personalized goals created when using the intervention is provided with numerous referrals in the local community for traumatic stress or alcohol or drug-related services. Throughout the intervention, users are provided feedback on their responses and provided tailored recommendations for care (eg, “Your reactions are common and natural responses to violence. If these reactions are bothering you, it might be time to try therapy or other services. We will provide you with a list of providers at the end.”).

Empowerment is an essential component of recovery after IPV and SA as increased agency and control over one’s choices is associated with improved mental health outcomes [47]. Therefore, throughout the intervention, empowering imagery and themes are integrated. For example, aligned with the empowerment process model, the intervention emphasizes the importance of personal choice and autonomy and places survivors in control by providing options (eg, choices for personal goals that range from abstinence to a smaller reduction amount or choosing to not set a goal). Furthermore, by providing the intervention in a web-based format, the intervention is survivor led and self-paced, which is intended to enhance feelings of agency and choice throughout the intervention.

This Study

IPV and SA remain highly prevalent [1-3]. Unfortunately, there is a strong bidirectional association of IPV and SA with traumatic stress and alcohol misuse [4-6]. However, many survivors of IPV and SA do not engage in the services needed for traumatic stress and alcohol misuse because of several access barriers. One avenue for addressing barriers and increasing access for this population is the use of web-based SBIRT for traumatic stress and alcohol misuse, which incorporates tailored IPV and SA content. The purpose of this study was to examine the usability and acceptability of a web-based SBIRT intervention (CHAT) designed for recent survivors of IPV and SA, adapted from previous web-based SBIRT applications [11], among a sample of victim service professionals (VSPs) within IPV and SA advocacy centers and survivors of IPV and SA. Usability testing for web-based interventions refers to formal evaluation for use within the population of interest to identify methods to receive feedback that improves the design and addresses errors in the application [48]. Testing usability of eHealth interventions using a combination of quantitative and qualitative methods is crucial for obtaining adequate feedback to improve and tailor web-based interventions for use in specific populations [48].

This study had 2 phases. The first phase involved refining CHAT by making iterative adaptations based on feedback about usability and acceptability gathered from VSPs in focus groups and people who have experienced recent IPV and SA in individual interviews. After making iterative adaptations to the intervention, the second phase was to gather feedback about the acceptability of the adapted version of CHAT in an additional sample of people who have experienced recent IPV and SA.

Methods

Ethical Considerations

The institutional review board (IRB) at the Medical University of South Carolina (Pro00080368) approved this study. Informed consent was obtained from all survivors before their participation in the study. A Waiver of Consent was issued by the IRB for VSPs, and all VSPs were made aware of the risk to loss of confidentiality before participating in focus groups. The provider and survivor participants in both phases were assigned a random ID to protect their anonymity. Providers were compensated with US $25 Amazon gift cards for participation in focus groups. Phase 1 survivors received US $75 Amazon gift cards as compensation. Phase 2 survivors received US $50 Amazon gift cards as compensation. Transcriptions of focus groups were deidentified and stored in a secure location.

SBIRT Development and Design

CHAT was created on REDCap (Research Electronic Data Capture; Vanderbilt University) [49], a secure application for creating and managing surveys, that is hosted on the Medical University of South Carolina server. We chose to create CHAT on REDCap for several reasons, including low cost, accessibility among several institutions, ability to limit privileges to potential future providers and agencies interested in the intervention to help maintain the confidentiality of users, interventions that allow for personalizing intervention content such as branching and piping logic, ability to embed videos and photos into content, and available distribution methods (eg, links, URL codes, and email). We also selected REDCap because it is compatible with use on smartphones, which is a promising approach for reaching survivors of IPV and SA given the numerous barriers to accessing formal services [50].

CHAT was based on previous SBIRT interventions for survivors of traumatic events, including an SBIRT intervention for IPV and alcohol use (Brief Spousal Assault Form for the Evaluation of Risk [11] and National Institute on Alcohol Abuse and Alcoholism Rethinking Drinking [46]). The SBIRT intervention is based on the motivational model of alcohol use [20], self-medication theory [16], and the empowerment model [23] and applies principles of motivational interviewing to reduce motives to drink alcohol and increase valued living, particularly when experiencing negative affect, which are theoretical and empirically supported intervention targets for alcohol misuse. The SBIRT intervention, “CHAT” involves (1) psychoeducation about alcohol and drug use specific to IPV and SA; (2) screening and assessment for alcohol misuse, AUD, drug use, and traumatic stress symptoms using standardized measures; (3) personalized feedback about drinking quantity, recommended drinking limits, impact of drinking and drugs on recovery from SA, and money spent on alcohol and drugs; and (4) brief exercises aimed at increasing motivation and healthy alternatives to alcohol use, including value identification, readiness to change ruler, goal setting, identification of coping skills, and...

https://formative.jmir.org/2024/1/e49557

JMIR Form Res 2024 | vol. 8 | e49557 | p.2893

(page number not for citation purposes)
social support. The intervention concludes with a personalized plan and lists national and local referrals to IPV, SA, and substance use agencies. It takes approximately 20 minutes to complete. At the beginning of CHAT, users select whether they would like to view female, male, or gender nonbinary content, and branching logic provides images and drinking recommendations tailored to their preference. Piping logic is also used to personalize the intervention in several ways, including providing feedback about the quantity of the user’s alcohol use compared with recommended limits, relating the user’s reasons for drinking to potential long-term consequences (eg, “Making healthy choices about alcohol use helps me to take good care of my children”; response options range from strongly disagree to strongly agree), integrating the top values the user initially selects into motivational exercises at the end of the intervention, and providing them with a plan at the end that summarizes their selections throughout the intervention.

Phase 1: Refinement

Participants

Phase 1 included 2 samples. The first sample comprised adult VSPs (22/52, 42%) who were employed at 1 of 2 local nonprofit agencies that serve people who have experienced IPV and SA and completed semistructured focus groups. The VSPs included master-level clinical providers (9/22, 41%) and victim advocates who did not have a degree in counseling (12/22, 55%); educational level of 1 participant was unknown. The average experience of working with survivors of IPV and SA was 6.95 (SD 8.4) years and ranged from 4 months to 30 years. The second sample included adult survivors of past-year SA or physical IPV who drank alcohol (13/52, 25%).

Alcohol use was selected as an inclusion criterion because it is the most common substance used among survivors of IPV and SA and is the primary focus of intervention; however, participants could also use drugs as CHAT includes content on drug use after IPV and SA. Participants who used drugs were also permitted to be included in the sample; however, this was not required. The survivors were asked to complete a web-based screener to determine study eligibility. There were no exclusion criteria other than inability or unwillingness of the participant to provide informed consent.

Procedure

For phase 1, a total of 3 focus groups with 7 to 8 VSPs in each group took place at the agencies. Study flyers and email invitations to participate in the study were sent to all staff at an SA advocacy program and a nonprofit for IPV. VSPs completed a paper-based survey before the interview about demographics and experiences of treating survivors as well as an additional survey completed after the interview focused on assessing perceptions of CHAT. Survey data were entered into REDCap by 2 research assistants and checked for accuracy. Overall, 69.2% (9/13) of the survivors were recruited from the community using social media advertisements and flyers to participate. Four participants were recruited from an outpatient mental health clinic that served survivors of crime. After obtaining informed consent, interviews with survivors (13/52, 25%) were conducted in person in an outpatient clinic that served survivors of crime or over a secure web-based video platform. Baseline surveys administered on REDCap were completed before conducting the interview, and a brief second survey was administered after the interview to assess perceptions of CHAT. The interviewers and focus group moderators included 1 clinical psychologist and 2 master-level service providers.

Quantitative Measures

Demographics and Background Information

Self-report surveys were used to collect demographic and background variables from survivors and VSPs. VSPs were asked to indicate the percentage of time per week they provided services related to alcohol misuse and the type of evidence-based treatment they provided (ie, “Do you provide any of the following evidence-based treatments for PTSD and/or alcohol use in your agency?”).

Mental Health Self-Report Measures

Phase 1 survivors were administered measures of exposure to IPV and SA, traumatic stress symptoms, alcohol and drug use, and treatment utilization.

SA Victimization and IPV

SA exposure among phase 1 survivors was assessed using 2 items adapted from the Trauma History Questionnaire [50]: “Has anyone ever made you have intercourse or anal sex against your will?” and “Has anyone ever touched private parts of your body, or made you touch theirs, under force or threat?”. Physical IPV exposure among phase 1 and 2 survivors was assessed using an item from the HITS (Hurt, Insult, Threaten, Scream Measure; eg, Has a partner ever physically hurt you?) [51]. Survivors were asked to indicate if the IPV and SA events occurred in the lifetime or the past year.

1. PTSD Checklist for DSM-5 (PCL-5) [52]: The traumatic stress symptoms associated with the most distressing incident of IPV or SA among phase 1 survivors were assessed using the 20-item PCL-5. Items are rated on a 5-point scale ranging from 0 (not at all) to 4 (extremely), and a total score is created with higher scores indicating greater traumatic stress symptoms. PCL-5 scores >31 indicated clinically relevant traumatic stress symptoms [43]. Internal consistency was excellent for the PCL-5 for the phase 1 sample (Cronbach α=.92).

2. AUDIT-C [43]: The 3-item AUDIT-C was used to identify alcohol misuse. Survivors responded to 3 items about alcohol use (eg, How often do you have a drink containing alcohol?), with response options ranging from 0 (ie, never/no) to 4 (ie, ≥4 times a week/daily or almost daily). Responses were summed, and scores of ≥3 for women and ≥4 for men were used to indicate alcohol misuse. Internal consistency was good for the AUDIT-C for the phase 1 sample (Cronbach α=.91).

Qualitative Measures

To assess factors related to perceptions of the intervention created to address substance misuse, a semistructured focus group discussion and qualitative interview was developed. The term interpersonal violence was defined (ie, “Interpersonal
violence means physical or sexual violence such as sexual assault or domestic violence.”) for survivors and VSP participants at the start of the interview. Next, the web-based SBIRT intervention was described to survivors and VSP participants as a self-help intervention for substance use after IPV and SA. The interview consisted of asking survivors and VSP participants up to 4 questions as they looked at each content area of the intervention, including “What are you thinking about as you look at this page?,” “What do you like about this page?,” “What don’t you like about this page?,” and “How can we make this more interesting?” Issues related to errors in branching, options, or presentation of content were noted by the interviewer. Follow-up probes were used to clarify information provided whenever necessary. After viewing CHAT, survivors and VSP participants were asked about usefulness (eg, “Do you think this would be useful, why or why not?”; “Do you think other people with use it, why or why not?”; and “Is this something that you think people should be told about shortly after they speak to a provider about experiencing interpersonal violence, why or why not?”).

**Data Analysis**

Descriptive statistics were computed (ie, mean and SD) for participant age, PCL-5 scores, and AUDIT-C scores. Rates of past-year and lifetime exposure to IPV and SA, race, gender, and previous service use were aggregated. Clinical psychologists with expertise in qualitative methods conducted qualitative analyses.

The interviews and focus group discussions were audio-recorded and transcribed verbatim by an IRB-approved third party. Data from qualitative interviews were organized using coding template analysis [53], applying thematic and content analysis approaches using 6 themes outlined by Baumel et al [54] to examine the quality and usability of mobile apps, including usability, visual design, user engagement, content, therapeutic persuasiveness, and therapeutic alliance. Using a deductive coding strategy not only allowed for examination of themes proposed in existing usability literature but also allowed for the development of inductive categories that emerged through coding [55].

A clinical psychologist with expertise in substance use intervention development and qualitative analysis examined each line of the transcripts and mapped participant’s responses to the coding template [55,56]. More than 1 code could be applied. A second coder, who was also a clinical psychologist with expertise in interpersonal violence and substance use, reviewed responses against the coding template. Interrater discrepancies were discussed and resolved by 2 independent coders. NVivo software (version 11.1; QSR International) was used for data management and analysis. Demographics and background variables were computed using SPSS (version 27; IBM Corp [57]).

Results from formative usability trials have shown that 80% of usability issues can be identified with a sample of at least 5 people involved in usability testing [58,59]. Thus, this study was suitably powered to assess usability.

**Phase 2: Acceptability**

**Participants**

The phase 2 sample included adult survivors of past-year SA, physical IPV, or both who drank alcohol (17/52, 33%). The inclusion criteria for the survivor sample of phase 2 mirrored the inclusion criteria for survivors in phase 1.

**Procedures**

For phase 2, survivors were recruited through Facebook advertisements (14/17, 82%), through Craigslist advertisements (1/17, 6%), and from an outpatient clinic that served survivors of crime (2/17, 12%). Phase 2 survivors completed the study remotely using their own devices to complete study surveys and CHAT. They completed a baseline survey on REDCap comprising questions about demographics, IPV and SA exposure, and alcohol use. Next, the survivors completed CHAT and a survey on acceptability of the intervention.

**Measures**

Demographics, descriptive characteristics, exposure to IPV and SA, alcohol misuse, and traumatic stress were gathered from phase 2 survivors using the same validated measures as in phase 1 (PCL-5, AUDIT-C, HITS, and Trauma History Questionnaire items). Internal consistency was fair for the AUDIT-C (Cronbach α=.70) and PTSD Checklist (Cronbach α=.80) in the phase 2 sample. Phase 2 survivors were also asked whether they needed care for past-year SA, IPV, alcohol use, or drug use (eg, In the past year, did you ever want or need help with any alcohol use concerns?). The following additional measures were collected from phase 2 participants:

1. **Daily Drinking Questionnaire** [60]: The Daily Drinking Questionnaire was used to assess the number of standard drinks consumed per week by phase 2 survivors (eg, “On a typical Monday, I had _ drinks.”). The mean weekly drinks were calculated.

2. **Acceptability of Intervention Measure** [61]: After viewing the web-based SBIRT intervention, phase 2 survivors completed the 4-item Acceptability of Intervention Measure. Items were rated on a 5-point scale (1=completely disagree to 5=completely agree) and averaged. Phase 2 survivors were also asked open-ended survey questions about most-liked aspects of the intervention, least-liked aspects of the intervention, and dissemination recommendations (ie, What would be the best way to inform people about the tool?).

**Analyses**

Descriptive statistics were computed in the same way as in phase 1. Participant’s brief responses to open-ended questions were coded into representative themes and subthemes. Previous research indicates that a sample size of at least 10 should suffice for acceptability testing [62]. Therefore, the phase 2 sample size had sufficient power to assess acceptability.
Results

Phase 1: Refinement Qualitative Results

Participant Demographics

VSPs' (22/52, 42%) experience working with survivors ranged from 4 months to 30 years, and the average experience of working with survivors was 6.95 (SD 8.4) years. Two-thirds (13/22, 59%) of the participants had experience providing services to individuals engaging in alcohol misuse. Table 1 shows the demographics of the VSPs. Overall, 85% (11/13) of the phase 1 survivors reported lifetime exposure to both SA and physical IPV, and 31% (4/13) of the survivors reported both types of exposure in the past year. Phase 1 survivors reported a high prevalence of alcohol misuse, traumatic stress symptoms, and previous substance use and mental health service use (Table 1).

Table 1. Demographics, trauma history, and mental health information for phase 1 victim service professionals (VSPs), phase 1 survivors, and phase 2 survivors (N=52).

<table>
<thead>
<tr>
<th></th>
<th>Phase 1 VSPs (n=22)</th>
<th>Phase 1 survivors a (n=13)</th>
<th>Phase 2 survivors a (n=17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>42 (14.8)</td>
<td>33 (10.6)</td>
<td>28 (9.2)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>_ b</td>
<td>11 (85)</td>
<td>16 (94)</td>
</tr>
<tr>
<td>Men</td>
<td>— 2 (15)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Unknown</td>
<td>— 0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Race c, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>7 (32)</td>
<td>2 (15)</td>
<td>3 (18)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 (4)</td>
<td>4 (31)</td>
<td>2 (12)</td>
</tr>
<tr>
<td>White</td>
<td>14 (64)</td>
<td>11 (85)</td>
<td>13 (76)</td>
</tr>
<tr>
<td>Did not disclose</td>
<td>0</td>
<td>0</td>
<td>1 (6)</td>
</tr>
<tr>
<td>PCL-5 d, mean (SD)</td>
<td>—</td>
<td>47.01 (18.16)</td>
<td>58.12 (9.65)</td>
</tr>
<tr>
<td>Above cutoff, n (%)</td>
<td>—</td>
<td>10 (76.9)</td>
<td>17 (100)</td>
</tr>
<tr>
<td>AUDIT-C e, mean (SD)</td>
<td>—</td>
<td>6.08 (3.54)</td>
<td>6.88 (3.38) f</td>
</tr>
<tr>
<td>Above cutoff, n (%)</td>
<td>—</td>
<td>11 (84.6)</td>
<td>15 (88)</td>
</tr>
<tr>
<td>Drug use endorsed, n (%)</td>
<td>—</td>
<td>9 (69.2)</td>
<td>—</td>
</tr>
<tr>
<td>Marijuana use endorsed</td>
<td>—</td>
<td>6 (46.3)</td>
<td>—</td>
</tr>
<tr>
<td>Previous SUD g service use</td>
<td>—</td>
<td>4 (30.8)</td>
<td>—</td>
</tr>
<tr>
<td>Previous trauma-focused service use</td>
<td>—</td>
<td>8 (61.5)</td>
<td>—</td>
</tr>
<tr>
<td>Past-year SUD service use</td>
<td>—</td>
<td>—</td>
<td>6 (35.29)</td>
</tr>
<tr>
<td>Past-year mental health service use</td>
<td>—</td>
<td>—</td>
<td>14 (82.35)</td>
</tr>
</tbody>
</table>

aSurvivors: Survivors of intimate partner violence, sexual assault, or both.

bNot available.

cSome individuals identified as both Hispanic and White.

dPCL-5: Posttraumatic Stress Disorder Checklist for DSM-5.

eAUDIT-C: Alcohol Use Disorder Identification Test–Concise.

fAUDIT-C data was missing for 1 participant in phase 2.

gSUD: substance use disorder.

In the qualitative analysis, content from the focus groups and interviews was organized according to the evaluation categories for usability testing proposed by Baumel et al [48]. Results from survivors and providers are described within each theme (Table 2) and revised components of the intervention following these results are presented inTextbox 1.
Table 2. Themes and number of survivors and focus groups that discussed each theme (N=35).

<table>
<thead>
<tr>
<th>Theme</th>
<th>Survivors (n=13), n (%)</th>
<th>Focus groups (out of 3 groups; n=22), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td>11 (85)</td>
<td>2 (67)</td>
</tr>
<tr>
<td>Visual design</td>
<td>13 (100)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>User engagement</td>
<td>10 (77)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Content</td>
<td>12 (92)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Therapeutic persuasiveness</td>
<td>11 (85)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Therapeutic alliance</td>
<td>10 (77)</td>
<td>3 (100)</td>
</tr>
</tbody>
</table>

Textbox 1. Revised intervention components.

<table>
<thead>
<tr>
<th>Intervention component and contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Psychoeducation</td>
</tr>
<tr>
<td>• Rates of sexual assault (SA) and alcohol misuse</td>
</tr>
<tr>
<td>• Destigmatizing alcohol-involved SA</td>
</tr>
<tr>
<td>• Breathing exercise</td>
</tr>
<tr>
<td>• Assessment</td>
</tr>
<tr>
<td>• Values identification</td>
</tr>
<tr>
<td>• Alcohol use screener (Alcohol Use Disorder Identification Test–Concise)</td>
</tr>
<tr>
<td>• Past 3-month Drug Use Screener (National Institute of Drug Abuse Assist)</td>
</tr>
<tr>
<td>• Weekly money spent on alcohol and drugs</td>
</tr>
<tr>
<td>• Posttraumatic stress disorder primary care screener</td>
</tr>
<tr>
<td>• Identifying reasons for not drinking</td>
</tr>
<tr>
<td>• Personalized feedback</td>
</tr>
<tr>
<td>• Link between drinking to cope and alcohol-related problems</td>
</tr>
<tr>
<td>• Recommended drinking limits</td>
</tr>
<tr>
<td>• Comparison of money spent on alcohol and drugs with cost of common items</td>
</tr>
<tr>
<td>• Motivational interviewing</td>
</tr>
<tr>
<td>• Readiness rulers with value exercise</td>
</tr>
<tr>
<td>• Goal setting</td>
</tr>
<tr>
<td>• Coping skills identification</td>
</tr>
<tr>
<td>• Identifying how values relate to alcohol use</td>
</tr>
<tr>
<td>• Identify social support and coping strategies to use when having urges to drink</td>
</tr>
<tr>
<td>• Summary and plan</td>
</tr>
<tr>
<td>• Summary of feedback and recommended treatment referrals as indicated</td>
</tr>
</tbody>
</table>

Theme 1: Usability

Both survivors and providers rated the usability of CHAT very positively overall. VSPs and survivors reported that the intervention was easy to use and self-explanatory, stating it was “straight to the point,” “short, sweet, and user friendly,” and “clear and concise.” A survivor specifically mentioned the following:

“It’s easy to navigate, easy to understand, and it’s pretty much just a self-analysis. VSPs provided similar feedback about the usability of CHAT, with VSPs from 2 of the 3 focus groups stating positive comments. Specifically, VSPs made comments that CHAT was “very easy to understand.”
**Theme 2: Visual Design**

This theme describes the appearance of CHAT, including observations about the text, font size, pictures, colors, and look and feel of the activities within the application. Survivors discussed visual design slightly more than the providers, as visual design was the most discussed topic for survivors and the second most discussed topic (after content) for providers. Survivors made comments about visual design such as follows:

- *I like this. It kinda breaks down everything and gives you visual points to look at. Like you immediately go to what you are experiencing.*
- *I really like the diagram because it shows what you are doing...Drinking more than the single day limit.*

Another survivor described the following:

- *Picture is good too. Shows people drinking, gathering like it’s fun in a way, but reading those kind of wake up calls. Like it’s not really fun.*

In all VSP focus groups, the visual design of the intervention was discussed, with most comments about visual design being positive and only a few constructive recommendations about changes to visual design. Specifically, some comments that VSPs made on visual design included “I think the graphic is good,” “I like the colors,” and “I like this whole visual.” Recommendations about visual design included changing font sizes or adding pictures.

**Theme 3: User Engagement**

The theme, user engagement, describes how well the participant was able to engage with the application including how interactive and personalized the VSP and survivor participants felt that the intervention was. This included the users’ ability to engage in the activities embedded within the intervention and how often they felt that they would engage. More than half of the survivors commented on their potential engagement in the intervention, stating that they would be very willing to engage and that they would engage in the intervention often. Specifically, 1 survivor noted, “I would definitely use it. It’s easier.” Survivors also mentioned that they would be more likely to engage with the intervention than with people personally, as 1 survivor stated the following:

- *I feel like people would use it because it’d help. You may not even want to talk to your close people who are close to you. You may just want to be anonymous, and you probably would feel more comfortable with this.*

Another survivor said the following:

- *I guess it would be the ideal thing for somebody who didn’t want to come in and talk to anybody, and just wanna do it on...just looking for information on their own, looking for help on their own.*

In all the focus groups of VSPs, it was discussed that the survivors they worked with, as well as themselves, would engage with the intervention very frequently and that the intervention was easy to use. Specifically, 1 provider stated the following:

- *Wow! You can pull it up on any device. I would literally use this all the time [with different clients]*

Another provider stated the following:

- *It’s so great that the whole thing takes about 15 minutes. That’s nice.*

In describing ways that a patient might engage, a provider stated the following:

- *That’s really cool that you are able to print this off and put it on your fridge.*

**Theme 4: Content**

This theme described the quality and appropriateness of the content presented during the intervention. All survivors and VSPs in each focus group made positive statements about the intervention, stating that it was rooted in evidence, appropriate for themselves or their patients, and contained information that was often relevant to patients. Specifically, 1 survivor stated the following:

- *It think it really hits all the points. It even has God in here which is really something people don’t mention.*

Another survivor stated the following:

- *I actually wouldn’t add anything. I love this information.*

Throughout the interviews, survivors made comments about specific content that they found relevant or that they thought was important to include. In addition, VSPs discussed mostly positive comments about the content, with just a few suggestions or examples to add. For example, 1 VSP noted the strength in content by stating the following:

- *I love how there are so many short snippets. I thought those were all pretty powerful points.*

Another VSP discussed the comprehensive nature of the content by mentioning the following:

- *I like that it is fairly all-encompassing. I like that there are positive reasons and then neutral reasons.*

**Theme 5: Therapeutic Persuasiveness**

Therapeutic persuasiveness included comments regarding the suitability of the content provided by the intervention and whether the provider or survivor would see the content as useful in addressing substance use or trauma. All survivors described therapeutic persuasiveness and noted that the intervention would be useful in addressing these topics among survivors. Specifically, 1 survivor noted the following:

- *It makes you evaluate your life and see things that you maybe didn’t see while you were drinking or why you were drinking. It makes you really think about what you are doing.*

Another survivor described the following:

- *It helps you realize how much you really are drinking. I like it because it lets you know that [values] are important and can be accomplished by changing your substance use problem. This helps you put things into...*
Another VSP stated the following:

I really like the statistics. They kind of make you feel like you’re not the only one experiencing this. It’s okay that this is happening, it’s not something to be ashamed of, it’s something to deal with. It speaks to a lot of people.

**Theme 6: Therapeutic Alliance**

This theme describes whether users thought that the intervention provided support that could mirror, support, and take the place of human contact provided by an actual provider as well as whether they saw it as a resource to bolster ongoing therapy. Approximately half of the survivors discussed the therapeutic alliance gained by the intervention, with 1 survivor stating the following:

I love that the tool is helpful but also shows the places you can go to for actual, real help with your problems. Both are good.

VSPs in all 3 focus groups discussed this theme and described that the intervention would support the concepts taught in therapy and would serve as a useful addition to therapy. Specifically, 1 provider mentioned the following:

They can share the information without anybody knowing their situation. Because most times they really don’t want to share. So to be able to use this, versus verbalize it and have somebody judge them, that is much easier.

In addition, each VSP focus group discussed information on how to integrate the intervention into practice. For example, 1 VSP mentioned the following:

We get so many clients that have alcohol and drug histories. It’ll take several sessions for me to really get them to a point where they realize they are drinking/using drugs too much. But I feel like this [intervention] can get them to that point a lot quicker, where they’re more likely to at least think about [substance use] earlier on.

VSPs in each focus group also reported that this intervention would be helpful throughout treatment and potentially at different points for different clients. For example, 1 VSP described the following:

Honestly, it depends on what stage of change the client is in. Like, if they’re like “I’m not an addict or an alcoholic, I’m not going to listen to anything you say.”

Another VSP stated the following:

This [intervention] could be part of the assessment process. They come in for their mental health assessment, ask them to get there 20 minutes early to fill this out just to screen, then give them the tool to use if needed.

**Revised Intervention Content Aligned With Results**

Iterative changes to CHAT were made based on user experience during usability testing including fixing errors noted with logic branching, piping, spelling, grammar, increasing font size, and changing images based on participant suggestions. In addition, a page encouraging user to take 5 slow breaths was added based on VSPs’ suggestion that content may increase stress and adding breathing could help users engage with the intervention. Furthermore, options for gender-based content and language around recommended drinking limits and sex were adapted based on VSP recommendations. Gender-based content options were to view images of women, men, or nonbinary content. Drinking limits were explained based on biological sex because current recommended drinking limits are based on sex assigned at birth owing to physiological differences that change the way alcohol affects the body (eg, For people assigned female sex at birth, drinking more than 3 drinks on any day or 7 drinks per week is “at risk” or “heavy drinking.”).

**Phase 2: Demographics and Qualitative Results Regarding Acceptability**

**Overview**

Phase 2 survivors also reported high levels of alcohol use, traumatic stress symptoms, and previous substance use and mental health services use. In phase 2, 59% (10/17) of the survivors reported both types of exposure in their lifetime, and one quarter (4/17, 23%) of the survivors endorsed past-year SA and physical IPV. Phase 2 survivors endorsed high levels of alcohol use with an average AUDIT-C score of 6.88. More than one-third (6/17, 35%) of phase 2 survivors self-reported needing care for alcohol misuse, and one-quarter (4/17, 23%) self-reported needing care for drug use. A total of 17 survivors recruited through social media and the community viewed the web-based intervention and completed surveys. On a 5-point scale (1=completely disagree to 5=completely agree), survivor participants agreed that the intervention was, on average, acceptable (mean 4.1, SD 1.29).

**Qualitative Results**

Phase 2 participants’ (17/52, 32%) favorite aspects of the intervention were encapsulated by 4 themes: content and features (6/17, 35%) such as encouraging messages, focus on reduction rather than abstinence, and ability to download a personal plan; accessibility and ease of use (6/17, 35%); psychoeducation (6/17, 35%); and social connectedness (6/17, 35%) associated with forums. For example, 1 survivor wrote the following:

I liked that it made a personalized plan just for me whereas in a group setting it’s geared more towards everyone.

Approximately one-third (6/17, 35%) of the survivors denied disliking any aspect of the intervention. The aspects of CHAT that could be improved were encapsulated by 3 themes: content and features (6/17, 35%), including increased focus on broader mental health, describing prevalence of IPV and SA in a more sensitive manner, increased brevity, and increased...
personalization; connection to immediate services (4/17, 23%); and language access (1/17, 6%). Survivors suggested 5 dissemination methods: calls, emails, and texts (2/17, 12%), podcasts (1/17, 6%), posters and flyers (1/17, 6%), social media advertisements (10/17, 58%), websites (2/17, 12%), and referrals when seeking help for IPV, SA, or substance use (3/17, 18%). For example, 1 participant offered the following suggestion for dissemination:

If anyone entering a behavioral center, drug rehabilitation center, or any mental health office showing signs of sexual trauma or substance use, the tool could be suggested to them by a medical health provider and implemented through a computer at home or cell phone. If they don’t have a cellphone or computer at home, hopefully counties can provide the computer for them to complete the tool. Social media is always a good way to promote tools to help mental health, advertisements are how I found Better Help. I have found a lot of online counseling through Google as well.

Revised Intervention Content Based on Phase 2 Results
Iterative changes were made to the web-based intervention in response to participants’ suggestions (Figure 1). Images of the intervention were also updated, and images that were displayed were updated to match the preferred gender-based content selected by users at the beginning of the intervention. To improve brevity, we removed assessment of AUD symptoms and focused solely on alcohol misuse with the 3-item AUDIT-C [55]. In traditional SBIRT, the full AUDIT is used and people who report risky alcohol misuse receive brief intervention focused on reducing use and people who report more severe levels of alcohol misuse are referred to treatment. We decided to provide brief intervention and referral options based on alcohol misuse (rather than the full AUDIT) because the time after recent IPV and SA is a high-risk time for the escalation of substance use, and this minimizes the chance that we fail to provide treatment options to someone who could benefit from AUD treatment. In addition, to further personalize the intervention content to users’ drinking goals, additional content on harm reduction skills was added. Increased personalized feedback based on symptoms endorsed on the Primary Care PTSD Screen for DSM-5 was increased in the revised intervention to further interweave education about self-medication and links between alcohol misuse and traumatic stress. Finally, additional piping was used to incorporate more information about the users’ values as they relate to alcohol use. Words were also reduced for length and increased readability.

![MY PERSONAL PLAN](image)

Read the list of values and pick 3 that best represent the principles, standards, or qualities you consider MOST important in your life.

- Achievement: to have accomplishments I can be proud of
- Children: to take good care of my children
- Family: to have a happy, loving family
- Friendship: to have close, supportive friends
- I can share my goal with Jamie.

Figure 1. Final images of CHAT (Choices For Your Health After Trauma).
Discussion

Primary Findings

Survivors of IPV and SA report elevated rates of traumatic stress and alcohol misuse [4-6]. However, <35% recent survivors of IPV and SA seek services for mental health [29,30]. Higher levels of substance use are related to lower service use among survivors of recent trauma [7], which points to the importance of providing services for alcohol use recently after trauma exposure. This is especially important among recent survivors of IPV and SA given their uniquely high risk for alcohol misuse and drug use in this acute period [14,41]. We conducted an investigation of the usability and acceptability of a web-based SBIRT intervention for traumatic stress and alcohol misuse following an IPV and SA. Focus groups were conducted with IPV and SA service providers, and interviews were conducted with recent survivors of with the aim of gathering recommendations to improve the usability of the intervention. Following adaptation of the intervention based on this feedback, we examined the acceptability of the adapted SBIRT intervention in a second sample of recent survivors. Overall, our results suggested that service providers and recent survivors viewed the intervention as acceptable (as indicated by high ratings on the acceptability of intervention measure) and beneficial to integrate into services (as indicated by the qualitative feedback provided by VSPs and survivors). In this study, both providers and survivors commented that the intervention content was useful in addressing traumatic stress and alcohol misuse, including evaluating one’s alcohol use patterns in the context of trauma. It may be important to address the link between IPV and SA-related distress and alcohol use [16]. Although historically IPV and SA and alcohol use services have operated independently, our findings are in line with recommendations for integrated approaches that address trauma and alcohol misuse simultaneously [36,63,64].

Individuals with greater alcohol misuse may be less likely to seek care following IPV and SA [65], indicating the importance of addressing substance use–related barriers to service access. Some survivors in this study indicated that they would be more likely to engage with the intervention (rather than speak with another person) given the anonymity. This suggests that delivery of a self-directed, web-based SBIRT intervention may help circumvent common service barriers among survivors of IPV and SA related to shame, fear of the consequences of disclosure, and anticipatory stigma [10,66]. In addition, web-based delivery of SBIRT may help with barriers in clinical settings related to time constraints and staff availability [67] given the ease of access (ie, ability to use on any device), empowering content (eg, self-paced and power to decide what goals to select), and brevity (<20 minutes to complete) of the intervention. It will be important for future research to determine whether web-based SBIRT can help address some of these barriers to care if adopted into practice and to determine the most effective way of integrating this intervention into existing clinical practice.

It is imperative that the SBIRT intervention is delivered in a nonstigmatizing manner given the highly stigmatized nature of IPV, SA [68], and alcohol misuse [69]. Indeed, several survivors and providers in this study discussed the importance of ensuring that the content of the intervention is nonjudgmental and culturally sensitive. The high acceptability ratings of CHAT and previous SBIRT research [11] indicate that integrated interventions for trauma and substance misuse following recent IPV and SA can be delivered in a sensitive manner to address the important treatment needs of this population. It is crucial for future researchers developing interventions for survivors of IPV and SA to be mindful of using inclusive and nonjudgmental language. This may include explicit statements that the survivor is not to blame for the violence they have experienced, destigmatizing alcohol-involved SA by sharing statistics about the common nature of this type of assault, etc. Useful suggestions regarding design and content were made by the service providers and survivors that inform preferences for the web-based SBIRT intervention following IPV and SA focused on alcohol misuse. The intervention was modified in accordance with this feedback. The results underscored the significance of conducting usability testing with both providers and survivors. Neglecting to consider user perspectives and preferences is a key factor contributing to the low utilization of mobile mental health interventions [70,71]. Consistent with previous research indicating that privacy is an important concern for users of mobile health interventions [71], several providers and survivors in this study also discussed the need to ensure confidentiality. Therefore, interventions developed for this population may benefit from explicitly addressing issues of confidentiality. Finally, we examined the acceptability of the revised version of CHAT in a sample of recent survivors in phase 2 of this study. Research supports the effectiveness of early interventions in reducing trauma-related symptoms [72]; however, accessing health care in the weeks following assault is uncommon [73]. The current web-based SBIRT intervention was developed to address gaps in service provision after recent IPV and SA when risk for alcohol misuse and drug use is heightened owing to trauma-related distress. Survivors found the revised intervention to be acceptable and noted the personalization of the intervention as being important. Personalized care and delivery of individualized treatments is a priority for mental health care to improve the effectiveness of evidence-based interventions [74]. CHAT is an example of how a web-based SBIRT intervention can be personalized, and future research is needed to evaluate the effectiveness of this intervention in addressing needs related to substance use following IPV and SA. Taken together, our results support that, in general, recent survivors find it acceptable to be provided web-based SBIRT.

Strengths and Limitations

The strengths of this investigation include gathering feedback from both VSPs and survivors that informed revisions of the web-based SBIRT intervention. Furthermore, the intervention included multiple substances, and the motivational content of the SBIRT intervention was personalized to refer to the substances reported by survivors (ie, alcohol use, drug use, or substance use for use of both alcohol and drugs). The high acceptability of the web-based SBIRT intervention in this study encourages examination of its efficacy in future research trials. However, the sample sizes across the 2 study phases were small, and future research with larger sample sizes is needed to test...
the feasibility of the web-based intervention. Inclusion criteria for survivors in this study were based on reporting of alcohol misuse. Overall, 69% (9/13) of phase 1 survivors reported drug use, with the most common drug used being marijuana (6/13, 46%). Thus, it is possible that the findings would differ if survivors were recruited based on reporting drug use. Future research should compare the acceptability, feasibility, and efficacy of the interventions among survivors reporting various types of substance use. In addition, future research may test whether allowing people who use multiple substances to select a specific substance for the motivational interventions as in the Brief Spousal Assault Form for the Evaluation of Risk intervention in the study by Choo et al [11] would improve outcomes. It is also important to note that SBIRT is most effective for people who report alcohol misuse but do not yet meet the criteria for AUD. As many recent survivors of IPV and SA may already meet the AUD or PTSD criteria before the recent trauma, the SBIRT intervention may have limited utility as an early intervention for those individuals.

It should be noted that most survivors included in this study identified as female individuals, White, and not Hispanic, and sexual orientation information was not collected. Individuals with marginalized identities based on race, ethnicity, sexual orientation, gender, and immigration status experience additional barriers to accessing services following IPV and SA [10,33,73]. In addition, experiences of identity-based stigma and discrimination may contribute to substance use [75] and exacerbate trauma-related distress following interpersonal violence [76]. Thus, it is imperative for future research to examine the usability, acceptability, and efficacy of the SBIRT intervention among survivors with diverse intersecting identities and consider experiences of stigma and marginalization. Furthermore, given the gender differences in female and male exposure to IPV and SA as well as barriers to service engagement, this study should be replicated with greater male representation in the sample. Participation was also limited to English-speaking individuals as the intervention was only available in English, and future revisions of the intervention should address language as a barrier.

All service professionals included in this study provided services within IPV and SA advocacy centers, and therefore, future evaluation of the intervention should be conducted with providers in other types of settings (eg, primary care, emergency department, and law enforcement). Furthermore, although the intervention was not designed for a single outlet for dissemination, a strength of the intervention is its flexible nature and ability to be disseminated in a wide range of settings (ie, advocacy centers, web-based forums, and emergency departments). This may increase the reach of the intervention to survivors of IPV and SA in the community who do not seek formal services. The findings in phase 2 provided crucial data to inform better ways to reach survivors and disseminate the SBIRT intervention broadly on social media and should be tested for effectiveness in reaching survivors with lower levels of service use in future research. Further implementation research is needed to understand how to integrate CHAT into specific settings (eg, rape advocacy centers). Although SBIRT is intended to be an early intervention that addresses traumatic stress symptoms and alcohol misuse and prevents the development of AUD or PTSD (because the recent months following IPV and SA are a heightened period of risk for AUD and PTSD to develop), it is possible that CHAT could also benefit nonrecent survivors. Future research should expand to apply the SBIRT model to nonrecent survivors and examine whether a broader application is beneficial. In addition, the web-based intervention had limitations in terms of referral and treatment owing to its brevity. Moreover, it lacked a follow-up mechanism to assist individuals in overcoming barriers to accessing referrals. This is a weakness of the current SBIRT intervention. Future research should focus on increasing the robustness of the referral to treatment component. Adding a support person to follow-up with intervention users might be an important component for future development [77]. For example, 1 survivor participant recommended, “Definitely follow ups, like a call, would help.”

In conclusion, results from this study among VSPs and survivors support the usability and acceptability of a web-based SBIRT intervention designed for traumatic stress and alcohol misuse among recent survivors of IPV and SA. Future research should include samples with greater diversity and address the barriers related to English language proficiency. Overall, the findings encourage future examinations of the efficacy of web-based SBIRT for co-occurring traumatic stress and alcohol misuse among recent survivors.

Data Availability
Deidentified quantitative data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

References


57. IBM Corp. IBM SPSS statistics for windows, version 27.0. IBM Corp. 2020. URL: https://www.ibm.com/spss-statistics-27-no


**Abbreviations**

AUD: alcohol use disorder
AUDIT-C: Alcohol Use Disorder Identification Test–Concise
CHAT: Choices For Your Health After Trauma
COPE: Concurrent Treatment of PTSD and Substance Use Disorders Using Prolonged Exposure
HITS: Hurt, Insult, Threaten, Scream Measure
IPV: intimate partner violence
IRB: institutional review board
PCL-5: Posttraumatic Stress Disorder Checklist for DSM-5
PTSD: posttraumatic stress disorder
REDCap: Research Electronic Data Capture
SA: sexual assault
SBIRT: screening, brief intervention, and referral to treatment
VSP: victim service professional
Sodium Intake Estimation in Hospital Patients Using AI-Based Imaging: Prospective Pilot Study

Jiwon Ryu¹²*, MD, MSc; Sejoong Kim²³⁴*, MD, PhD; Yejee Lim¹², MD, PhD; Jung Hun Ohn¹², MD, PhD; Sun-wook Kim¹², MD, MSc; Jae Ho Cho¹², MD, MSc; Hee Sun Park¹², MD, MSc; Jongchan Lee¹², MD, MSc; Eun Sun Kim¹², MD, PhD; Nak-Hyun Kim¹², MD, PhD; Ji Eun Song⁵, MSN; Su Hwan Kim⁶, PhD; Eui-Chang Suh⁷, MD; Doniyorjon Mukhtorov⁸, MSc; Jung Hyun Park⁸, MSc; Sung Kweon Kim⁸, MSc; Hye Won Kim¹², MD, PhD

¹Hospital Medicine Center, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea
²Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea
³Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
⁴Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea
⁵Department of Nursing, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
⁶Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea
⁷Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
⁸LOAPI-Healthcare, AltheNutrigene, Seongnam-si, Republic of Korea

* these authors contributed equally

Corresponding Author:
Hye Won Kim, MD, PhD
Hospital Medicine Center
Seoul National University Bundang Hospital
Gumi-ro 173 Beon-gil 82, Bundang-gu
Seongnam-si, 13620
Republic of Korea
Phone: 82 7877638
Email: kimhwhw@gmail.com

Abstract

Background: Measurement of sodium intake in hospitalized patients is critical for their care. In this study, artificial intelligence (AI)–based imaging was performed to determine sodium intake in these patients.

Objective: The applicability of a diet management system was evaluated using AI-based imaging to assess the sodium content of diets prescribed for hospitalized patients.

Methods: Based on the information on the already investigated nutrients and quantity of food, consumed sodium was analyzed through photographs obtained before and after a meal. We used a hybrid model that first leveraged the capabilities of the You Only Look Once, version 4 (YOLOv4) architecture for the detection of food and dish areas in images. Following this initial detection, 2 distinct approaches were adopted for further classification: a custom ResNet-101 model and a hyperspectral imaging-based technique. These methodologies focused on accurate classification and estimation of the food quantity and sodium amount, respectively. The 24-hour urine sodium (UNa) value was measured as a reference for evaluating the sodium intake.

Results: Results were analyzed using complete data from 25 participants out of the total 54 enrolled individuals. The median sodium intake calculated by the AI algorithm (AI-Na) was determined to be 2022.7 mg per day/person (adjusted by administered fluids). A significant correlation was observed between AI-Na and 24-hour UNa, while there was a notable disparity between them. A regression analysis, considering patient characteristics (eg, gender, age, renal function, the use of diuretics, and administered fluids) yielded a formula accounting for the interaction between AI-Na and 24-hour UNa. Consequently, it was concluded that AI-Na holds clinical significance in estimating salt intake for hospitalized patients using images without the need for 24-hour UNa measurements. The degree of correlation between AI-Na and 24-hour UNa was found to vary depending on the use of diuretics.

Conclusions: This study highlights the potential of AI-based imaging for determining sodium intake in hospitalized patients.
artificial intelligence; AI; image-to-text; smart nutrition; eHealth; urine; validation; AI image; food AI; hospital; sodium intake; pilot study; imaging; diet; diet management; sex; age

Introduction

Overview

A low-salt diet is prescribed for patients with cardiovascular, kidney, or liver diseases. In these patients, sodium intake regulation is essential. High salt intake is a key modifiable risk factor for these diseases. Thus, monitoring the dietary salt intake of patients provides patients and clinicians with valuable information on dietary salt reduction advice [1].

For therapeutic purposes, the salt intake of hospitalized patients is assessed either indirectly by the prescription order of diet and food consumption questionnaires or directly from repeated 24-hour urinary sodium (UNa) excretion collections [2]. However, these methods have several limitations. Discrepancy with the prescribed diet order, the actual food intake of patients may vary because of poor compliance. Inaccurate answers provided during questionnaire surveys may result in biased results. Direct access to repeated 24-hour UNa collection, which is the standard method of salt intake marking, is an inconvenient and costly process because samples are to be sent to a laboratory, and flame photometry is typically required [3]. Furthermore, this method is inherently impractical because of the dependence on patient compliance and variability between collections.

For inpatients, a quick evaluation of salt intake would enable real-time advice and subsequent application to the next diet regimen. Therefore, a tool to objectively quantify a patient’s dietary intake and repetitively evaluate the sodium content of the diet of hospitalized patients is required.

Artificial intelligence (AI) technology has enabled image-based analyses of nutrition and ingredients. Picture-to-Amount, a deep learning architecture using a cross-modal image-to-text retrieval system, can predict the number of ingredients in a given food image [4]. Technology-assisted dietary assessment relies on AI to accurately group food pictures and measure food credit by assessing a cellphone food record [5]. AI models for dietary assessment have been verified for reproducibility and validity. A large-scale population survey verified the accuracy of one such AI model for nutrient analysis [6]. Carter et al [7] conducted a study to compare the nutritional intake recorded in the smartphone app called “My Meal Mate” and the nutritional intake using the 24-hour recall method. Ahmed et al [8] divided participants of their study into 2 groups, namely a meal diary group and a tablet application use, and compared their nutrient intakes [8]. A fully automatic monitoring system of nutrient intake by hospitalized patients was presented for medical use by processing Red Green Blue (RGB) depth image pairs before and after meal consumption using AI-based estimation [9]. However, most studies have focused on classifying food or identifying the content of protein, fats, and carbohydrates, whereas studies on specific nutrients such as salt are scarce. To further improve the assessment and monitoring of the salt content of inpatients’ diets, the validation of accurate estimation of the diet and consumption of salt by patients is still required.

Objectives

We evaluated whether sodium intake could be determined through AI methods using food photographs and known nutrition information in hospitalized patients. We also determined if it is significant to compare sodium intake with 24-hour UNa, which is the gold standard for measuring sodium intake.

Methods

Study Design and Population

This single-center prospective study was conducted from August to November 2021 and involved 54 hospitalized patients, recruited from the hospitalist-run acute care unit as well as nephrology and urology departments. The following criteria were used for inclusion: (1) adult patients aged 19 years or older, (2) patients who agreed to take photographs before and after meals, and (3) patients who were prescribed 10 g of salt (4 g of sodium) in their diet. Patients who were unable to eat because of conditions such as respiratory arrest requiring tracheal intubation, cardiac arrest, acute coronary syndrome or life-threatening arrhythmias, failure of more than 2 organs, recent trauma, or burns of the neck and face were excluded. People with alcohol addiction and pregnant patients were also excluded. Previous medical history, demographics, and laboratory data were retrieved from electronic medical records.

Nutrient Information and Acquisition of Food Images

A novel and dedicated image database was developed because models pretrained using the data of different regions and countries could not be used. Hence, curating a local food data set was necessary, particularly for measuring sodium intake in hospitalized patients. A 3-month diet was predetermined before the study was initiated. The Seoul National University Bundang Hospital Nutrition Department provided all the meals and nutritional information. To establish the new data set and salt estimation algorithm, photographs of cooked food, plates, and spoons from the selected menu were collected. The researchers took photographs of the patients’ meals before and after consumption for a day and prohibited the patients from eating snacks or any outside food, except for water (Figure 1A and 1B).
Food Image Analysis—Sodium Intake Measurement with AI (AI-Na)

We used the hybrid model for sodium intake measurement based on food image analysis. This model incorporates the You Only Look Once, version 4 (YOLOv4) model’s detection of the image areas and uses 2 classification approaches—Custom-101 and Hyperspectral imaging—to predict the food quantity estimation for sodium amount (Figure 3A and 3B). YOLOv4 has been used in several previous studies to detect and classify the food and dish areas in the images [10,11]. Multidish images were generated by cropped images using the boundary-box area to create a data set containing images with different sizes. Figure 3 shows data generated after converting single images into multidish images. A data set for a convolutional neural network with one kind of food was placed in dishes of different shapes. Then, all the food images were rotated to different angles, and the light and sharpness of these images were adjusted (Figure 3B). For the estimation of food quantity from the images, we predominantly used convolutional neural networks, with ResNet-101 being our primary model of choice. In our pursuit to select the optimal classification model, several alternative architectures were also evaluated. This included experimenting with models such as MobileNetV2, ResNet-18, Wide-ResNet-50, and InceptionV3. After rigorous testing and evaluation, ResNet-101 demonstrated superior performance, affirming its selection for our research objectives. The amount of food remaining was estimated, and the calorie and nutrition contents were estimated [12,13]. The food photographs were color-based images, and a hyperspectral image was used to clarify the differences between colors and improve image quality. Hyperspectral images have rich information and show superior performance when used for feature identification based on pixel intensity [14,15]. In our study, we used Custom-MST++ to facilitate the conversion of standard RGB images into hyperspectral images. Although a conventional RGB image is composed of 3 channels (ie, red, green, and blue), the reconstructed hyperspectral image boasts an enhanced structure, encompassing 31 distinct bands. This transition from a trichannel format to a multiband representation allows for a more nuanced and detailed analysis, critical for our research objectives. The following steps must be performed when using hyperspectral

Figure 1. Food photographs taken (A) before and (B) after meal consumption for sodium intake measurement. The red track indicates the recognition of the amount of food intake.

Figure 2. Main architecture of the hybrid model for food quantity estimation for and sodium amount estimation. RGB: Red Green Blue.
images for estimating food quantity: (1) reconstruction of the hyperspectral image from the RGB image, (2) preprocessing of the hyperspectral image thus obtained, (3) calculation of the pixel intensity per spectral band, and (4) classification with of the food quantity using random-forest regression of food quantity. Therefore, by using the aforementioned method and procedures, the types and quantities of food that the patients consumed were classified and estimated from the food images. In our research, our foremost goal was to devise an algorithm capable of estimating sodium intake solely from images. Initially, our challenge was to accurately discern the volume or weight of the food items depicted in these images. Leveraging cutting-edge AI methodologies, we crafted a model adept at both recognizing and quantifying various food items. As part of this implementation process, we used information regarding the selected menu’s nutrients, type of food, and amount of food based on the protocol of the hospital nutrition department. Once we established the food quantity, we cross-referenced it with the nutritional data pertinent to the identified items. By combining this food quantity data with the nutritional profiles, our algorithm adeptly calculated the sodium proportion in the given dishes, ultimately yielding the sodium intake value, which we have denoted as AI-Na (unadjusted). Subsequent adjustments were made to the algorithm by subtracting sodium content from administered fluids, resulting in a refined AI-Na (adjusted; Figure 2).

Figure 3. Sodium intake measurement using the artificial intelligence–based method. (A) Data generated after converting single images into multidish images; (B) Dataset for the convolutional neural network. Q‘ denotes quantity in the classification of food classes by portion size.

The 24-Hour UNa Collection as a Reference Value for Sodium Intake

The gold standard for estimating dietary sodium intake is the 24-hour UNa value [16]. Participants were asked to collect all urine during a 24-hour period, starting with the first urine sample on the morning of the day when they took the food photos and concluding with the second urine sample on the following morning. The urine aliquots were stored at −20 °C before transportation to the certified laboratory. An ion-selective electrode method (Modular DPE chemistry; Roche Diagnostics) was used to measure urinary sodium and potassium levels. The urinary creatinine (UCr) level was measured using the Jaffe reaction (kinetic colorimetric assay; Roche Diagnostics). The urine samples were excluded if any of the following were observed: (1) total volume of urine during the 24-hour period was <500 cc, (2) UCr was <0.6 g/day in men and <0.4 g/day in women, and (3) the self-reported spillage was more than 30 cc [17,18].

Statistical Analysis

The Mann-Whitney test and Brand–Altman method were performed to determine the extent to which the AI-Na values matched the 24-hour UNa as reference values. However, because diuretics affect the 24-hour UNa, we divided the participants into diuretic and nondiuretic groups for the analysis. As kidney function can also affect the 24-hour UNa, a regression equation using sodium intake, 24-hour UNa, and estimated glomerular filtration rate (eGFR) was developed, and an interaction term was used to evaluate the role of diuretics. The interaction term was used rather than the dose, depending on the use of diuretics, as the sample was too small to evaluate the relationship between diuretic dose and 24-hour UNa. In the baseline characteristics of the study population, continuous variables were expressed as median values, and categorical variables were described as frequencies and percentages. We considered 2-sided P values <.05 to be statistically significant. Statistical analyses were performed using the R 4.1.0 (R foundation for Statistical Computing).

Ethical Considerations

The study followed the general treatment policy for the underlying disease, and patients did not go beyond the scope of the standard treatment, except for capturing photographs of the prescribed meal. The study protocol complied with the Declaration of Helsinki and was approved by the Institutional Review Board of the Seoul National University Bundang Hospital (B2108-701-302). Written informed consent was obtained from all the patients.
Results

Estimating Sodium Intake From Food Images—Quantifying Results of Food Quantity Estimation

We collected 20,000 images for food quantity estimation and 1500 hospital images with sodium amount metadata. Table 1 and Table 2 show the results of food area detection and classification. In the context of our research, when we refer to YOLO versions, such as YOLOv3, YOLOv3-tiny, YOLOv3-tiny3l, YOLOv4, and YOLOv4-tiny, an important part of the highlight is that our choice of YOLOv4 signifies our preference for the best training and testing mean average precision among these versions. The decision was made after rigorous evaluation and comparative analysis, ensuring optimal performance for our specific use case. Moreover, similar to choosing the ResNet-101 architecture for its distinct advantages in certain applications, we used 5 alternative CNN models to get the best accuracy of classification. ResNet-101 got the highest F1-score and a lower train and validation loss in 50 epochs. Training and validation losses quantify the model’s prediction error. Lower values indicate higher accuracy and better generalization to new data. These losses are measured as dimensionless quantities derived from the chosen loss function, such as cross-entropy for classification tasks. Epoch count, such as the 50 used in our experiments, represents the total number of times the learning algorithm processes the entire data set, a critical factor in optimizing the model’s learning curve and preventing overfitting or underfitting.

Of the 54 participants enrolled in this study, 11 withdrew their consent because they could not wait for the photographs to be taken before the meal and because their general health deteriorated. Seven participants were excluded due to incomplete urine collection, and 11 were excluded due to inaccurate urine collection based on the 24-hour UCr value. Finally, the data of the 25 selected participants were analyzed (Figure 4).

The median age was 64 (IQR 53-74) years, and 68% (n=17) were men. The median values of serum creatinine and serum sodium were 1.0 (IQR 0.8-1.6) mg/dl and 138 (IQR 135.0-140.0) mEq/ml, respectively. The baseline characteristics of the study participants are presented in Table 3. Because the use of diuretics affects 24-hour UNa—the standard of sodium intake evaluation—the baseline characteristics were classified accordingly (Table 3). A total of 10 participants were treated with saline and parenteral nutrients. Because sodium in these fluids affects the 24-hour UNa, total intake was calculated by adding the amount of sodium administered (AI-Na [adjusted]).

The median sodium intake (AI-Na [unadjusted]) was estimated to be 1756.5 (IQR 1266.6-2273.2) mg when using the AI algorithm; further, the sodium in the fluids was included, resulting in a total sodium intake AI-Na [adjusted]) of 2022.7 (IQR 1396.2-2564.4) mg (Table 3). The 24-hour UNa was determined to be 2783.0 (IQR 1955.0-4922.0) mg. Depending on the effect of the diuretics, the value of 24-hour UNa varied and was 2599.0 (IQR 1771.0-4922.0) mg and 2921.0 (IQR 2231.0-4059.5) mg for the nondiuretic and diuretic groups, respectively (Table 4).

We compared the AI-estimated sodium intake values with those of 24-hour UNa and analyzed the degree of concordance and difference between the two.

Table 1. Results of food area detection.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Training mAP (%)</th>
<th>Testing mAP (%)</th>
<th>Weight size (MB)</th>
<th>Training time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>81.1</td>
<td>74</td>
<td>236</td>
<td>178</td>
</tr>
<tr>
<td>YOLOv3-tiny</td>
<td>90</td>
<td>85</td>
<td>33.7</td>
<td>65</td>
</tr>
<tr>
<td>YOLOv3-tiny_3l</td>
<td>87.6</td>
<td>77</td>
<td>33.7</td>
<td>65</td>
</tr>
<tr>
<td>YOLOv4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>98.9</td>
<td>95</td>
<td>245</td>
<td>207</td>
</tr>
<tr>
<td>YOLOv4-tiny</td>
<td>70</td>
<td>74</td>
<td>23</td>
<td>64.9</td>
</tr>
<tr>
<td>YOLOv4-tiny_3l</td>
<td>85.1</td>
<td>75</td>
<td>23</td>
<td>64.3</td>
</tr>
</tbody>
</table>

<sup>a</sup>mAP: mean average precision. Input size was 608x608, and iteration numbers were 50,000.

<sup>b</sup>YOLOv3: You Only Look Once, version 3.

<sup>c</sup>YOLOv4 was selected from multiple models listed in the Table.
Table 2. Classification of food quantity estimation.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Training $F_1$-score (%)</th>
<th>Training loss</th>
<th>Validation $F_1$-score (%)</th>
<th>Validation loss</th>
<th>Epoch</th>
<th>Time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MobileNetv2</td>
<td>95</td>
<td>0.15</td>
<td>93</td>
<td>0.7</td>
<td>50</td>
<td>0.8</td>
</tr>
<tr>
<td>ResNet-18</td>
<td>96</td>
<td>0.12</td>
<td>92</td>
<td>0.1</td>
<td>50</td>
<td>1.2</td>
</tr>
<tr>
<td>Wide-ResNet-50</td>
<td>96</td>
<td>0.2</td>
<td>94</td>
<td>0.1</td>
<td>50</td>
<td>1.3</td>
</tr>
<tr>
<td>ResNet-101$^a$</td>
<td>98</td>
<td>0.11</td>
<td>95.5</td>
<td>0.03</td>
<td>50</td>
<td>1.2</td>
</tr>
<tr>
<td>Inceptionv3</td>
<td>96</td>
<td>0.17</td>
<td>93</td>
<td>0.12</td>
<td>50</td>
<td>1.4</td>
</tr>
</tbody>
</table>

$^a$ResNet-101 was selected from multiple models listed in the Table.

Figure 4. Flow diagram of the study population.

Table 3. Comparison of the characteristics between the diuretic and nondiuretic groups.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=25)</th>
<th>Non-diuretics (n=14)</th>
<th>Diuretics (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR)</td>
<td>64 (53-74)</td>
<td>64.5 (43.0-72.0)</td>
<td>62.0 (56.5-79.5)</td>
</tr>
<tr>
<td>Sex (male), n (%)</td>
<td>17 (68)</td>
<td>12 (86)</td>
<td>5 (46)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$), median (IQR)</td>
<td>25.8 (23.7-28.8)</td>
<td>24.3 (23.7-26.3)</td>
<td>27.4 (26.1-29.1)</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>16 (64)</td>
<td>8 (57.1)</td>
<td>8 (73)</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td>12 (48)</td>
<td>6 (43)</td>
<td>6 (55)</td>
</tr>
<tr>
<td>Liver cirrhosis, n (%)</td>
<td>7 (28)</td>
<td>4 (29)</td>
<td>3 (27)</td>
</tr>
<tr>
<td>Congestive heart failure, n (%)</td>
<td>4 (16)</td>
<td>0 (0)</td>
<td>4 (36)</td>
</tr>
<tr>
<td>Coronary artery disease, n (%)</td>
<td>3 (12)</td>
<td>2 (14)</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Cerebrovascular disease, n (%)</td>
<td>3 (12)</td>
<td>1 (7)</td>
<td>2 (18)</td>
</tr>
<tr>
<td>Cancer, n (%)</td>
<td>5 (20)</td>
<td>4 (29)</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Chronic kidney disease, n (%)</td>
<td>6 (24)</td>
<td>4 (29)</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Hemoglobin (mg/dl), median (IQR)</td>
<td>11.1 (10.0-13.3)</td>
<td>11.6 (10.0-13.3)</td>
<td>10.5 (9.7-13.2)</td>
</tr>
<tr>
<td>Sodium (mEq/L), median (IQR)</td>
<td>138.0 (135.0-140.0)</td>
<td>138.0 (133.0-140.0)</td>
<td>138.0 (135.0-140.0)</td>
</tr>
<tr>
<td>Potassium (mEq/L), median (IQR)</td>
<td>4.0 (3.7-4.4)</td>
<td>4.0 (3.9-4.3)</td>
<td>4.0 (3.6-4.4)</td>
</tr>
<tr>
<td>Serum creatinine (mg/dl), median (IQR)</td>
<td>1.0 (0.8-1.6)</td>
<td>1.0 (0.7-1.4)</td>
<td>1.3 (0.9-1.8)</td>
</tr>
<tr>
<td>Albumin (mg/dl), median (IQR)</td>
<td>3.6 (3.3-4.1)</td>
<td>3.6 (3.3-4.0)</td>
<td>3.4 (3.1-4.2)</td>
</tr>
<tr>
<td>eGFR$^a$ (ml/1.73m$^2$), median (IQR)</td>
<td>70.41 (42.57-89.55)</td>
<td>83.82 (52.04-98.38)</td>
<td>54.06 (27.77-74.43)</td>
</tr>
</tbody>
</table>

$^a$eGFR: estimated glomerular filtration rate.
Table 4. Sodium input calculated with AI (AI-Na) and 24-hour urine sodium (UNa) excretion.

<table>
<thead>
<tr>
<th>Sodium and urine output metrics</th>
<th>Participants, median (IQR)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (N=25)</td>
<td></td>
</tr>
<tr>
<td>AI-Na (mg; unadjusted)</td>
<td>1756.5 (1266.6-2273.2)</td>
<td>.37</td>
</tr>
<tr>
<td>Sodium in the fluids (mg)</td>
<td>0 (0.0-177.1)</td>
<td>.98</td>
</tr>
<tr>
<td>AI-Na (mg; adjusted)</td>
<td>2030.0 (1396.2-2564.4)</td>
<td>.98</td>
</tr>
<tr>
<td>Total output (ml/day)</td>
<td>1170.0 (1120.0-2210.0)</td>
<td>.99</td>
</tr>
<tr>
<td>24-hour UNa (mg/day)</td>
<td>2783.0 (1955.0-4922.0)</td>
<td>.81</td>
</tr>
<tr>
<td>24-hour UCr (g/day)</td>
<td>0.9 (0.7-1.1)</td>
<td>.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sodium and urine output metrics</th>
<th>Nondiuretic (n=14)</th>
<th>Diuretic (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI-Na (mg; unadjusted)</td>
<td>1646.4 (1162.4-2244.9)</td>
<td>1756.5 (1620.9-2418.8)</td>
</tr>
<tr>
<td>Sodium in the fluids (mg)</td>
<td>177.1 (0.0-190.8)</td>
<td>0.0 (0.0-0.0)</td>
</tr>
<tr>
<td>AI-Na (mg; adjusted)</td>
<td>2034.2 (1266.6-2748.9)</td>
<td>2022.7 (1620.9-2418.8)</td>
</tr>
<tr>
<td>Total output (ml/day)</td>
<td>1750.0 (916.0-2468.0)</td>
<td>1650.0 (1210.0-1984.0)</td>
</tr>
<tr>
<td>24-hour UNa (mg/day)</td>
<td>2599.0 (1771.0-4922.0)</td>
<td>2921.0 (2231.0-4059.5)</td>
</tr>
<tr>
<td>24-hour UCr (g/day)</td>
<td>1 (0.9-1.2)</td>
<td>0.8 (0.6-0.9)</td>
</tr>
</tbody>
</table>

**Unadjusted with the amount of sodium in the administered fluids.**

**Adjusted with the amount of sodium in the administered fluids.**

**UCr: urine creatinine.**

The disparity between the 2 methods was not insignificant, and we assessed its impact on the concordance. Our analysis using the Bland-Altman method revealed a bias of $-1106.4$ mg, with a CI for the concordance limit ranging from $-5468.2$ mg to $3255.5$ mg. Given that the corresponding bias when converted to the amount of salt is approximately $2.76$ g and the concordance limit’s CI varies from $8.1$ g to $13.7$ g, it is challenging to draw conclusive inferences from these findings.

Table 5. Differences between the total sodium input values using an artificial intelligence (AI)–based method (adjusted by fluids) and 24-hour urine sodium (UNa) excretion.

<table>
<thead>
<tr>
<th>Sodium intake</th>
<th>Difference test, median (IQR)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total participants</td>
<td>Difference test, median (IQR)</td>
<td>.02</td>
</tr>
<tr>
<td>AI-Na (adjusted): mg</td>
<td>2022.7 (1369.2-2564.4)</td>
<td></td>
</tr>
<tr>
<td>24-hour UNa (mg)</td>
<td>2783.0 (1955.0-4922.0)</td>
<td></td>
</tr>
<tr>
<td>Participants with nondiuretics</td>
<td>Difference test, median (IQR)</td>
<td>.14</td>
</tr>
<tr>
<td>AI-Na (adjusted): mg</td>
<td>2034.2 (1282.1-2689.0)</td>
<td></td>
</tr>
<tr>
<td>24-hour UNa (mg)</td>
<td>2599.0 (1817.0-4566.0)</td>
<td></td>
</tr>
<tr>
<td>Participants with diuretics</td>
<td>Difference test, median (IQR)</td>
<td>.10</td>
</tr>
<tr>
<td>AI-Na (adjusted): mg</td>
<td>2023.0 (1621.0-2419.0)</td>
<td></td>
</tr>
<tr>
<td>24-hour UNa (mg)</td>
<td>2921.0 (2231.0-4060.0)</td>
<td></td>
</tr>
</tbody>
</table>

**Mann-Whiney test.**

**AI-Na: Al-estimated sodium intake.**

**Adjusted with the amount of sodium in the administered fluids.**

Interaction Model Using Regression Analysis

The difference between the 2 test values could be attributed to factors that affect 24-hour UNa excretion in a real-world setting, such as the use of diuretics as well as the patient’s gender, age, and renal function; this prompted us to derive a formula considering the aforementioned factors. The eGFR value using the Chronic Kidney Disease Epidemiology Collaboration equation was used as a variable, and the following regression equation was obtained using the interaction term for diuretics because gender, age, and renal function can all be calculated using eGFR (Table 6).

$$24h-UNa = 0.535 \times AI-Na \ [adjusted]-2292.009 \times I \ [diuretics] + 1.280 \times AI-Na \ [adjusted] \times I \ [diuretics] + 22.102 \times eGFR$$

$$24h-UNa = 2.355 \times AI-Na \ [adjusted] + 22.102 \times eGFR-2292.009 \ [diuretic group]$$

$$24h-UNa = 0.535 \times AI-Na \ [adjusted] + 22.102 \times eGFR \ [nondiuretic group]$$

In this equation, “I” is the interaction term.
In this regression equation, the AI-Na (adjusted) and 24-hour UNa exhibited a strong relationship. Additionally, eGFR was significantly related to 24-hour UNa. Because the impact of the eGFR value on the 24-hour UNa is the same regardless of the use of diuretics, which was negligible in this equation, it can be assumed that 2.355 times the total sodium input (AI-Na [adjusted]) corresponds to 24-hour UNa in the diuretic group, and 0.535 of the AI-Na corresponds to the measured 24-hour UNa in the non-diuretic group.

Table 6. Linear regression with the interaction term between sodium intake calculated by the artificial intelligence algorithm (AI-Na; adjusted) and 24-hour urine sodium (UNa) excretion (adjusted $R^2=0.739$; $F=18.7$; $p<0.001$).

<table>
<thead>
<tr>
<th>Linear regression with interaction term</th>
<th>Regression coefficients</th>
<th>SE</th>
<th>$P$ value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI-Na (mg; adjusted$^b$)</td>
<td>0.535</td>
<td>0.232</td>
<td>.03</td>
</tr>
<tr>
<td>Furosemide dose (mg)</td>
<td>-2292.009</td>
<td>2259.709</td>
<td>.32</td>
</tr>
<tr>
<td>eGFR$^c$ (ml/min/1.73m$^2$)</td>
<td>22.102</td>
<td>8.361</td>
<td>.02</td>
</tr>
<tr>
<td>AI-Na (furosemide dose; adjusted)</td>
<td>1.820</td>
<td>1.076</td>
<td>.11</td>
</tr>
</tbody>
</table>

$^aP$ value for the interaction term in the relationship.

$^b$Adjusted with the amount of sodium in the administered fluids.

$^c$eGFR; estimated glomerular filtration rate.

**Discussion**

**Principal Findings**

Our study may be the first study that compared an AI-based method with 24-hour UNa for measuring sodium intake in a real clinical field. Although the AI-Na and 24-hour UNa values were not the same, the various factors that affect the 24-hour UNa value, such as age, sex, renal function, diuretics, and even the underlying disease, cannot be ignored when using real-world data. Therefore, the 2 methods were worth evaluating using regression. The AI-Na values can be clinically considered as a significant indicator of sodium intake, although there were differences based on whether diuretics were used. Therefore, food images can be used to measure sodium intake to some extent, but this method is still inaccurate.

We calculated the sodium amounts in each test image as ground truth and used them for the AI sodium amount prediction model. Food amount served as one of the input values for Sodium amount prediction, as we used a multi-input fixture to predict sodium amount prediction and 24-hour UNa amount prediction using the collected data set from the hospital. As we collected metadata, it included food quantity, sodium amount, food classes, food intake, and patient information. The AI-based method predicted salt or sodium amount in food based on our multitask method.

Concerns regarding sodium intake have led to the development of several sodium measurement methods [16,19,20]. Numerous methods for measuring sodium intakes are available, ranging from 24-hour UNa, the most objective method, to single or concurrent dietary questionnaires. These methods describe the portion size or weight of the food consumed; therefore, these methods may not be accurate and may yield highly variable results because they rely solely on the information that the participants provide. Therefore, 24-hour UNa is generally used for the external validation of these methods. Although dietary records and 24-hour UNa are recognized as suitable methods for measuring sodium intake, challenges such as difficulty during measurement and variability in results persist. Therefore, there is a growing demand for a more accurate and convenient method to measure sodium intake.

The rising interest in health care has led to an increased demand for nutrient and diet management. With the widespread use of smartphones, apps, and advancements in AI, there have been efforts to integrate these technologies into a dietary management system. A study was conducted to evaluate the consistency of an app designed as a calorie measurement tool for weight loss [7]. The comparison focused on the results from inputting consumed food into the app with those from a phone recording. Another study compared the food recorded in the app with a written record to demonstrate the app’s validity [8]. These studies proposed a novel approach that deviates from traditional methods but still relies on patient self-recorded information. The need for objective methods of food and nutrient intake measurement has led to the emergence of AI-based studies that have classified and quantified food intake using food photos. AI-based techniques have advanced to a level at which they can be used to classify various objects or humans in photos, including the ability to identify the food on the plate and the different types of food in the image; they can also determine the food quantity [5,11,12,22]. Furthermore, the food’s ingredients can be distinguished, and even chemical and molecular information can be obtained using hyperspectral images [4,15]. These studies have focused on healthy populations, except for one that was conducted on hospitalized patients, similar to our study [9]. This study evaluated the food quantity, calories, carbohydrates, fat, and salt intake of hospitalized patients by analyzing the photos taken before and after meals. However, the reference value was calculated using the weight and nutrition information that the hospital provided, lacking the actual measurement of nutrient intake or clinical...
reference. Additionally, because the aforementioned study focused on the method of analyzing food photos, the experimental setting of this study did not reflect that of a real-world clinical one. Our study, on the other hand, considered several variables of an actual clinical environment and showed that salt intake measured by a photo-based AI algorithm had a significant relationship with the gold standard of sodium intake (ie, 24-hour UNa), thereby demonstrating that AI-Na value establishes the foundation for clinical use.

Accurately measuring food intake in hospitalized patients is crucial for determining their nutritional status, disease progression, or treatment option selection. However, traditional methods for evaluating food intake, such as visual estimation of the entire or 75% of the amount and patients’ subjective evaluations, often result in inaccuracies and an overestimation of up to 15% [23,24]. Therefore, a convenient and semi-automated digital method for food intake evaluation is required, and AI-based food photographs are expected to fulfill this need.

**Strengths and Limitations**

Our study’s strength lies in its evaluation of sodium intake in actual inpatients using cutting-edge technologies, food photos, and AI-based techniques as well as a comparison with established standards. Various variables of a real clinical setting were considered in this study, demonstrating the potential clinical usefulness of using AI in this domain. This study had some limitations. First, the sample size considered in this study was small. If more patients were enrolled, it would be possible to derive a more accurate formula using different variables. The second limitation was the inaccurate estimation of the food amount during food intake measurement. In addition to considering 24-hour UNa, it would have been useful if the food’s weight was considered both before and after consumption along with the corresponding nutrients. Third, the method proposed in this study failed to account for the potential loss of sodium during cooking and trimming processes, which may have further impacted the accuracy of the results.

Although we presented hyperspectral image reconstruction and employed a machine learning model for estimating food quantity, it is important to note that our approach did not directly detect sodium amounts from food images. Instead, it demonstrated a robust correlation between hyperspectral images and sodium amount estimation. This correlation provides a promising foundation for future research focused on refining methods for predicting sodium content.

In summary, our study aimed to incorporate AI into the clinical field; however, owing to limitations associated with AI, incomplete nutritional information, and the diversity of real-world treatments, comprehensive research planning is necessary for clinical use.

**Conclusions**

The method of measuring sodium intake using food photos was found to be inconsistent as compared with 24-hour UNa, which is widely used in clinical settings. However, the study results have clinical significance because variables of a real-world clinical setting, such as gender, age, diuretics, and fluid treatment, were considered. The findings also suggest that the formula derived in this study may not provide accurate estimates of the absolute sodium intake. However, if additional data are collected, it may be possible to develop a more useful formula in the future.

**Acknowledgments**

This study was supported by Seoul National Bundang University Hospital (grant 14-2021-0052).

**Data Availability**

The data set used in this research consists of food images and nutrition information. The data set is publicly available on AI Hub, a repository for AI-related data sets, maintained by the National AI Research and Development Agency of Korea. The data set can be accessed on AI Hub [25].

**Authors’ Contributions**

SK designed the study. JR and SK contributed to the conceptualization and methodology. YL, JHO, SK, JHC, HSP, JL, ESK, NHK, and JES contributed to data collection and validation. JR, HWK, SHK, ECS, JHP, SKK, and DM contributed to the formal statistical analysis. JR and HWK drafted and edited the manuscript. JHK, SKK, and DM contributed to the visualization. All authors read and agreed to the final version of the manuscript.

**Conflicts of Interest**

None declared.

**References**


25. AI Hub. URL: https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=74 [accessed 2024-02-02]

Abbreviations

- AI: artificial intelligence
- AI-Na: sodium intake calculated by the AI algorithm
- eGFR: estimated glomerular filtration rate
Human-Written vs AI-Generated Texts in Orthopedic Academic Literature: Comparative Qualitative Analysis

Hassan Tarek Hakam, BSc, MD; Robert Prill, PhD; Lisa Korte, PhD; Bruno Lovreković, MD; Marko Ostojić, MD; Nikolai Ramadanov, BSc, MD; Felix Muehlensiepen, PhD

1Center of Orthopaedics and Trauma Surgery, University Clinic of Brandenburg, Brandenburg Medical School, Brandenburg an der Havel, Germany
2Faculty of Health Sciences, University Clinic of Brandenburg, Brandenburg an der Havel, Germany
3Center of Evidence Based Practice in Brandenburg, a JBI Affiliated Group, Brandenburg an der Havel, Germany
4Center of Health Services Research, Faculty of Health Sciences, University Clinic of Brandenburg, Rüdersdorf bei Berlin, Germany
5Faculty of Orthopaedics, University Hospital Merkur, Zagreb, Croatia
6Department of Orthopaedics, University Hospital Mostar, Mostar, Bosnia and Herzegovina

Corresponding Author:
Hassan Tarek Hakam, BSc, MD
Center of Orthopaedics and Trauma Surgery
University Clinic of Brandenburg
Brandenburg Medical School
Hochstr 29
Brandenburg an der Havel, 14770
Germany
Phone: 49 03381 411940
Email: hassantarek.hakam@mhb-fontane.de

Abstract

Background: As large language models (LLMs) are becoming increasingly integrated into different aspects of health care, questions about the implications for medical academic literature have begun to emerge. Key aspects such as authenticity in academic writing are at stake with artificial intelligence (AI) generating highly linguistically accurate and grammatically sound texts.

Objective: The objective of this study is to compare human-written with AI-generated scientific literature in orthopedics and sports medicine.

Methods: Five original abstracts were selected from the PubMed database. These abstracts were subsequently rewritten with the assistance of 2 LLMs with different degrees of proficiency. Subsequently, researchers with varying degrees of expertise and with different areas of specialization were asked to rank the abstracts according to linguistic and methodological parameters. Finally, researchers had to classify the articles as AI generated or human written.

Results: Neither the researchers nor the AI-detection software could successfully identify the AI-generated texts. Furthermore, the criteria previously suggested in the literature did not correlate with whether the researchers deemed a text to be AI generated or whether they judged the article correctly based on these parameters.

Conclusions: The primary finding of this study was that researchers were unable to distinguish between LLM-generated and human-written texts. However, due to the small sample size, it is not possible to generalize the results of this study. As is the case with any tool used in academic research, the potential to cause harm can be mitigated by relying on the transparency and integrity of the researchers. With scientific integrity at stake, further research with a similar study design should be conducted to determine the magnitude of this issue.

(JMIR Form Res 2024;8:e52164) doi: 10.2196/52164

KEYWORDS
artificial intelligence; AI; large language model; LLM; research; orthopedic surgery; sports medicine; orthopedics; surgery; orthopedic; qualitative study; medical database; feedback; detection; tool; scientific integrity; study design
**Introduction**

Artificial intelligence (AI) is perhaps best defined as an algorithmic mechanism applied to machines, whereby solving challenges requires little to no human interaction [1]. Differentiating human-made and AI-generated work is becoming increasingly difficult with the rapid technological advancement of deep learning [2]. Deep learning is based on the replication of human thinking and the brain’s structure [3]. With the vast potential benefit that AI might bring to the table, extensive research has been conducted in the last decade with the purpose of finding potential solutions for health care–related problems [4]. The field of orthopedics, for example, might greatly benefit from AI image recognition capabilities to assist in the diagnosis of fractures or skin lesions. Other benefits can be drawn from AI’s capacity to analyze massive amounts of clinical information, which in turn presents benefits in clinical decision-making, risk assessment, and the generation of individualized care plans [5]. That is why an exponential increase in research on the topic of AI in the field of orthopedics has been noted, which has led to a subsequent increase in reviews trying to summarize the findings and give out recommendations [4].

Orthopedic sports medicine is the subspecialty of orthopedics that deals with pathologic conditions of the musculoskeletal system that arise from the practice of sports. This includes the prevention, diagnosis, and treatment of diseases. A particular challenge of sports medicine lies in the willingness of athletes to return to performance in a timely manner [4]. Through the use of deep neural networks, AI can assist specialists in various aspects of management. AI has shown to be especially advantageous for the diagnosis of fractures based on plain radiographs and computed tomography, with reviews reporting high accuracy, sensitivity, and specificity for the evaluation of plain radiographs [6] and computed tomography images [7]. With the evolution of convolutional neural networks and the increased capacity to integrate large amounts of written information, the patient’s medical records could serve as a basis for determining an individualized care plan as well as for making predictions for the best future course of treatment [8].

The influence of large language models (LLMs) on research in the field of orthopedics and sports medicine has not yet been well studied. AI is commonly used by researchers to help organize thought processes, obtain feedback, edit their work, and present their citations in the requested format. Consequently, AI has made academic work much more efficient [9]. However, considering that some of the most impactful journals allow the use of AI in composing or editing scientific texts, there are some ethical reservations regarding the authenticity and credibility of academic work [2]. Furthermore, some journals are actively involved in the development of tools to spot AI-generated texts [10]. In the light of this, the line where scientific research becomes fraudulent with regards to the use of AI must be determined. Different journals have adopted different guidelines for the use of AI.

The aim of this qualitative analysis is to determine the possibility that human researchers and AI-detection platforms can detect AI-generated texts. For this purpose, 4 researchers were recruited to participate in this study. As well as this, an AI-detection platform was used to assist in this endeavor.

**Methods**

This study adopted a similar method to previously conducted research on the matter [10].

**Recruitment**

For the purposes of the study, 4 participants were recruited. Two senior researchers in the fields of orthopedics and qualitative research, as well as 2 junior researchers in the same fields, expressed their interest in the subject at hand. All researchers were informed about the study’s objectives. The inclusion criteria for senior researchers were more than 10 years of research experience and having a doctoral degree in their field. Junior researchers were defined as students or physicians who had commenced their first project in the last 2 years.

**Ethical Considerations**

Due to the noninterventional nature of this study, as well as the anonymization of the included participants, local institutional and regulatory bodies did not require ethical approval. The methodology of the study and data collection were in line with the Geneva conventions. Informed consent was obtained from all participants involved in this study. The privacy and confidentiality of the involved participants has been protected by anonymizing their responses. No compensation was given to the participating individuals.

**Selection of Literature**

After searching PubMed for relevant material, 5 abstracts about meniscal injuries were selected for inclusion in the study [11-15]. The search strategy included the word “meniscus.” Subsequently, the first 5 articles published in reputable first quartile (Q1) or second quartile (Q2) journals were chosen to ensure the high quality of the articles. Abstracts that did not meet the criteria were excluded. This choice was made based on the fact that abstracts usually present a general overview of the topic at hand and communicate the main objectives of the paper. Although some treatment modalities are commonly applied to meniscal injuries, it is often impossible to completely restore the meniscal architecture, especially when the injury occurs in the middle, less vascularized portion [16]. Selecting meniscal injuries as a topic was, therefore, agreed upon by the research team as it is a common pathologic condition [17] and an area of extensive research [18].

**Involving AI**

Abstracts selected in the previous step were then rewritten by 2 AI platforms. One platform was the commonly used and extensively developed ChatGPT 3.4 (OpenAI) and the other was You.com. Using the instruction “rewrite the following in perfect academic English,” 5 new abstracts were generated by each AI. In the subsequent step, the command “write five abstracts on meniscal injuries” was used and 10 further abstracts were generated.
Randomization

The 25 resulting abstracts included the 5 original versions that were written by humans, the 5 rewritten versions that were generated by each AI, and the 5 newly generated versions that were composed by each AI platform. The abstracts were numbered from 1 to 25. These numbers were subsequently randomized using Microsoft Excel and the assigned abstracts were presented as a sheaf in the resulting order.

Evaluation

Evaluation of the abstracts was carried out using 2 methods. The first method of evaluation involved researchers with varying specialties and at different stages of their academic careers, while the second was based on the use of AI-detection software.

Participants were then asked to evaluate all the resulting abstracts using parameters that are commonly used for peer review. Suggested criteria that might aid in differentiating human-written from AI-generated literature included nuance, style, and originality [10]. Subtle phrasing and word choice might also be giveaways. A rating scale from 1 (very bad) to 5 (very good) was used for each parameter.

Participants were additionally asked whether they thought that the abstract was generated by a newer-generation AI, a more-developed AI, or a human. A short explanation was provided by each participant.

User Statistics

Descriptive statistics were used to investigate the correlation between the degree of academic experience and the number of correctly identified abstracts on one hand and between the previously mentioned parameters (eg, originality, grammatical soundness) and the correct identification of abstracts on the other. Furthermore, the correlation between the parameters and a researcher’s classification of an abstract was investigated. Interrater reliability was assessed by comparing the assessment of different articles by the same researcher, on the levels of both correct identification and assessed parameters. Intrarater reliability was assessed by comparing the assessments of different evaluators for both previously mentioned parameters.

The Mann-Whitney U test, the Wilcoxon W test, the Z test, and the asymptotic significance (2-tailed) P value were determined.

Results

The results of the analysis are presented in Tables 1-3. Further descriptive statistics are presented in Multimedia Appendix 1.

Table 1. The number of human-written and artificial intelligence (AI)–generated texts that were correctly or incorrectly identified by academics with different levels of academic expertise.

<table>
<thead>
<tr>
<th>Identified role of evaluator</th>
<th>Evaluations of original texts (n=5), n</th>
<th></th>
<th>Evaluations of texts rewritten by AI (n=20), n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human written (correctly identified)</td>
<td>AI–generated</td>
<td>Human written</td>
</tr>
<tr>
<td>Junior orthopedic surgeon</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Senior orthopedic surgeon</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Junior qualitative researcher</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Senior qualitative researcher</td>
<td>4</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2. This table details how authors judged manuscripts with artificial intelligence (AI)–generated abstracts with respect to whether an advanced large language model or a newer large language model was used.

<table>
<thead>
<tr>
<th>Identified role of evaluator</th>
<th>Evaluations of advanced AI-generated abstracts (n=10), n</th>
<th>Evaluations of newer unadvanced AI-generated abstracts (n=10), n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Advanced AI (correctly identified)</td>
<td>Newer unadvanced AI</td>
</tr>
<tr>
<td>Junior orthopedic surgeon</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Senior orthopedic surgeon</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Junior qualitative researcher</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Senior qualitative researcher</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3. This table represents how artificial intelligence (AI)–detector software judged the articles.

<table>
<thead>
<tr>
<th>Abstracts</th>
<th>Predicted to be human written, n</th>
<th>Predicted to be AI generated, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written by humans</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Rewritten by advanced AI</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Rewritten by newer unadvanced AI</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Completely generated by advanced AI</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Completely generated by newer unadvanced AI</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Discussion

Principal Results

The primary results of the study indicate that neither AI-detection software nor human critical appraisal can reliably distinguish AI-generated texts from human-written work. Regarding human detection of AI-generated texts, neither clinical experience nor area of expertise played a role in the evaluation of the presented material. The secondary results of the study indicate that criteria suggested by prior research, such as originality, style, and nuance, did not correlate with whether the researchers identified a text correctly or not. Furthermore, none of the criteria correlated with whether researchers judged a text as human written or AI generated. The qualitative analysis of the written answers did not provide any new insights on the subject in question. However, the junior orthopedic researcher was able to correctly identify texts according to the objectivity parameter. Whether this was due to correct interpretation or chance is unclear. Perhaps future studies with larger sample sizes can help in shedding light on this matter. Selecting the evaluators might have impacted the results of the study. Although the researchers were proficient published authors, English was not their primary language and this might have led to the inability to correctly identify the abstracts. However, the impact of this study is not reduced, as one might argue that scientific literature consumption is not restricted to researchers with English as their mother tongue. Furthermore, reading and publishing in English is becoming common practice, especially if research is considered to be relevant on the international level.

Comparison With Prior Work

Although AI is an evolutionary technology that presents an enormous potential for future research applications, the results of this study and previous studies with similar methodologies [10] are alarming. AI seems to have reached human-level writing skills, which in combination with its easy accessibility is able to threaten academic integrity. The findings of this analysis contradict previous claims for the ability to detect manuscripts generated by AI through model-agnostic and distribution-agnostic features [19]. Even though nonmalicious applications of AI, including grammatical corrections, reference style adjustment, and thought-process organization, represent plausible uses of AI models, potential fraudulent uses include the generation of complete texts from a simple command. Examples of malicious AI use might also include the rewriting of entire texts [20,21], as shown in this study. AI-generated texts can also be passed through AI-detection software by malicious users, who would then use the texts that passed the examination, making it even more difficult to subsequently detect fraudulent use.

Besides the ability to falsify results, AI presents researchers with the capacity to present false results in a plausible manner [22,23]. This also applies to inaccurate findings being reported confidently, which may be a misrepresentation that could lead to confusion, especially if the results are presented to unexperienced peers. Therefore, fact-checking the AI-generated statements and references will be essential when relying on such tools. AI also the capacity to generate images that can be used in the presentation of results [24]. In the area of orthopedic surgery, AI has already been proven to recognize patterns associated with multiple types of fractures [25]. Combined with its image-generation capacity, AI models will be able to create radiographic representations of fractures that are of no true scientific value but can be used to alter the results of a study.

Additionally, with the ever-increasing human inability to distinguish AI- and human-generated work, new rules must be written to ensure the scientific integrity of every published paper. Suggestions have included an increase in transparency in the design of AI models [26], as well as complete transparency in the use of AI by authors. This includes where and how LLMs were used in scientific projects [8,27]. Understanding the algorithms of these programs might aid in conceiving new and better programs to counteract fraud in its many forms. In an article in the journal Nature, the company Turnitin was reported to have incorporate AI-detection software [28]. Finally, and perhaps most importantly, the integrity of research is the most important aspect of the evolving discussion around the use of AI. Many previously conducted cross-examinations of academic publications revealed that research data obtained from prestigious academic institutions and published in equally prestigious academic journals were falsified. Whether these findings were intentionally corrupted or were errors of data collection is of little significance compared to the effects they might have on clinical and academic work. Thus, one can say that AI is just a tool, and its potential to cause good or harm is derived from individual motivations, experience level, and integrity [2]. Calls to completely ban AI from academic endeavors are, in the eyes of the authors, exaggerated, and future fraud can be minimized by optimizing self-regulatory mechanisms [29] and AI-detection models [30,31]. As well as this, the authors of this paper agree that detection of academic fraud is a responsibility of editors and journals, as a letter to Nature previously suggested [32]. However, the central role of researchers cannot be overemphasized.

Limitations

Limitations of this study include the inability to trace AI use in the original articles included in this study. However, we assumed that if AI were used, it would have been reported in the methodology or declarations sections. A second limitation of this study is that English is not the native language of the assessors. However, all the involved researchers have deep levels of proficiency, having published prior research in English. A third limitation is the small sample size of examined individuals and AI-recognition software, which does not allow us to draw definite conclusions on the matter at hand. However, as LLMs in the field of AI become more sophisticated, the recommendations that were made by previous authors and mentioned in this paper will still hold. The final limitation of this study is that a subset of articles dealing with meniscal injuries was chosen from the immense field of orthopedics. This is particularly important when considering the “hot topic” subset.
Conclusions
The statistical and qualitative analysis of the presented material showed that researchers were unable to differentiate human-written from AI-generated texts. Furthermore, the secondary finding of this study was that previously suggested criteria, such as originality and comprehension, did not aid in the differentiation of human-written and LLM-generated texts. Both findings show that humans and AI-detection software currently fail to properly identify the use of LLMs in the academic literature.

Furthermore, one can only speculate about the amount of undisclosed AI use in the academic literature. However, with the ever-increasing sophistication of LLMs, the integrity of future projects will be entirely dependent on scientists’ attitudes, as AI can serve as a facilitator and accelerator in publishing but can also be used with malicious intent. With regard to replicating this study, the authors strongly recommend that a larger sample size of articles with a larger number of researchers should be considered.

Data Availability
Data will be made available by the corresponding author upon request.

Authors’ Contributions
HTH was the main author of the manuscript and the principal investigator. RP, FM, and LK contributed to the design of the study. MO reviewed the scientific soundness of the included literature. BL and NR curated and analyzed the data.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental tables providing details of the statistical analysis.

References
9. Tools such as ChatGPT threaten transparent science; are we our ground rules for their use. Nature 2023 Jan;613(7945):612. [doi: 10.1038/s41586-023-00191-1] [Medline: 36694020]


Abbreviations
AI: artificial intelligence
LLM: large language model
Human-Written vs AI-Generated Texts in Orthopedic Academic Literature: Comparative Qualitative Analysis

Please cite as:
Hakam HT, Prill R, Korte L, Lovreković B, Ostojić M, Ramadanov N, Muehlensiepen F
Human-Written vs AI-Generated Texts in Orthopedic Academic Literature: Comparative Qualitative Analysis
JMIR Form Res 2024;8:e52164
URL: https://formative.jmir.org/2024/1/e52164
doi: 10.2196/52164
PMID: 38363631

© Hassan Tarek Hakam, Robert Prill, Lisa Korte, Bruno Lovreković, Marko Ostojić, Nikolai Ramadanov, Felix Muehlensiepen. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e52164

JMIR Form Res 2024 | vol. 8 | e52164 | p.2925
Investigating the Feasibility of Using a Wearable Device to Measure Physiologic Health Data in Emergency Nurses and Residents: Observational Cohort Study

Anish K Agarwal¹,²,³, MD, MPH, MS; Rachel Gonzales¹,²,³, MPH; Kevin Scott¹,², MD, MSEd; Raina Merchant¹,²,³, MD, MSHP

¹Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
²Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States
³Center for Health Care Transformation and Innovation, Penn Medicine, Philadelphia, PA, United States

Corresponding Author:
Anish K Agarwal, MD, MPH, MS
Perelman School of Medicine
University of Pennsylvania
423 Guardian Drive
410 Blockley Hall
Philadelphia, PA, 19104
United States
Phone: 1 2157465610
Email: anish.agarwal@pennmedicine.upenn.edu

Abstract

Background: Emergency departments play a pivotal role in the US health care system, with high use rates and inherent stress placed on patients, patient care, and clinicians. The impact of the emergency department environment on the health and well-being of emergency residents and nurses can be seen in worsening rates of burnout and cardiovascular health. Research on clinician health has historically been completed outside of clinical areas and not personalized to the individual. The expansion of digital technology, specifically wearable devices, may enhance the ability to understand how health care environments impact clinicians.

Objective: The primary objective of this pilot study was to assess the feasibility and acceptability of using wearable devices to measure and record physiologic data from emergency nurses and resident physicians. Understanding strategies that are accepted and used by clinicians is critical prior to launching larger investigations aimed at improving outcomes.

Methods: This was a longitudinal pilot study conducted at an academic, urban, level 1 trauma center. A total of 20 participants, including emergency medicine resident physicians and nurses, were equipped with a wearable device (WHOOP band) and access to a mobile health platform for 6 weeks. Baseline surveys assessed burnout, mental health, and expectations of the device and experience. Participants provided open-ended feedback on the device and platform, contributing to the assessment of acceptance, adoption, and use of the wearable device. Secondary measures explored early signs and variations in heart rate variability, sleep, recovery, burnout, and mental health assessments.

Results: Of the 20 participants, 10 consistently used the wearable device. Feedback highlighted varying experiences with the device, with a preference for more common wearables like the Apple Watch or Fitbit. Resident physicians demonstrated higher engagement with the device and platform as compared with nurses. Baseline mental health assessments indicated mild anxiety and depressive symptoms among participants. The Professional Fulfillment Index revealed low professional fulfillment, moderate workplace exhaustion, and interpersonal disengagement.

Conclusions: This pilot study underscores the potential of wearable devices in monitoring emergency clinicians’ physiologic data but reveals challenges related to device preferences and engagement. The key takeaway is the necessity to optimize device and platform design for clinician use. Larger, randomized trials are recommended to further explore and refine strategies for leveraging wearable technology to support the well-being of the emergency workforce.

(JMIR Form Res 2024;8:e51569) doi:10.2196/51569
KEYWORDS
digital health; emergency medicine training; wearable devices; burnout; mobile health; feasibility; wearable device; wearable; physiologic health data; nurse; resident; emergency department; acceptability; well-being

Introduction

Emergency departments (EDs) are a critical and frequently used resource within the US health care infrastructure (1 in 5 adults are treated in an ED annually) [1,2]. EDs are stressful environments [3-6]. Research has demonstrated the negative impact of the physical ED environment on patients [7-10], yet less is known about the effects on the health and well-being of the clinicians working within these EDs. Burnout and cardiovascular (CV) health remain threats to ED clinicians, their careers, and patient care [11-14]. Research on clinician health and burnout has historically been limited by retrospective studies which are often conducted outside of work. The rapid growth of digital technology, such as wearable devices and remote monitoring [15-17], offers new opportunities and challenges to investigate clinician health within health care environments and spaces.

Multiple occupational hazards have been attributed to the practice of emergency medicine (EM) [18]. EM clinicians have a 20% higher morbidity due to coronary artery disease, motor vehicle accidents, and impaired reproductive health [19-22]. Clinicians working night shifts, an essential practice in EM, have less restorative sleep, elevated blood pressure, and lower heart rate variability (HRV) [19,23-26]. Before the pandemic, the prevalence of stress, exhaustion, and burnout was alarmingly high in EM [27,28]. COVID-19 worsened these factors, resulting in workforce depletion, and making this an urgent and critical area of focus underscored by the Surgeon General and National Academy of Medicine [14,29].

A gap exists in understanding how clinicians identify and prioritize their health within the workplace. High-performance athletics provides a potential analogous framework whereby athletes track physiologic data (HRV, physical activity, and sleep) to guide their daily performance. There is an unrealized opportunity space for clinicians to understand and enhance care delivery and career longevity (reduce burnout and CV disease). Technological advancements provide a potentially unobtrusive and personalized method to collect individual data using wearable devices. Wearable device use has grown in popularity, with over 30% of Americans reporting they can obtain device ownership [30]. These wearable devices are typically wrist-worn and provide methods to measure health data such as step count, heart rate, and sleep [15]. What is less known, is if these devices and associated platforms are appealing to clinicians and can provide actionable insights to help inform strategies to support the workforce.

The objective of this study was to pilot test and evaluate the feasibility and acceptability of a wearable device and associated platform to measure and record emergency nurse and resident physician physiologic measures while they provide emergency care. This was a pilot study, investigating early barriers and facilitators to using these devices within health care settings for emergency nurses and residents.

Methods

Overview

Eligible EM resident physicians and emergency nurses included those providing 20 or more hours of patient care per week, having regular access to a smartphone, and providing consent to where a wrist-worn wearable device (WHOOP band [31]). Participants were recruited via email, completed informed consent, and were given a wrist-worn wearable device. Consenting participants completed a baseline survey assessing burnout and mental health (depression and anxiety), asked about their expectations of the study, and followed for 6 weeks. Validated instruments included the Patient Health Questionnaire (PHQ-8), General Anxiety Disorder (GAD-7), and the Professional Fulfillment Index (PFI) [32-36]. Over the 6 weeks, patients were also given access to a web-based platform that allowed participants to see their own physiologic data, access basic coaching videos, and connect to other users on the platform (Arena Strive [37]). At the completion of 6 weeks, participants were asked to complete a final survey exploring the feasibility and acceptability of the approach. Participants were also asked to provide free text commentary on the general approach and the specific device. The primary outcomes were use of the wearable device, acceptance of the device, and adoption of the device. Secondary measures included burnout, mental health symptoms, and physiologic measures recorded by the device including HRV and sleep.

Ethical Considerations

This was a longitudinal pilot feasibility study conducted at an urban, academic, level 1 trauma center in the northeastern United States. This study was reviewed and approved by the University of Pennsylvania Institutional Review Board (850371). All eligible participants completed informed consent forms and were informed that all data would be deidentified and aggregated for analysis. Participants received the wearable device at no cost and could keep the device following the completion of the study.

Data Analysis

Analysis was conducted using Stata SE (version 18; StataCorp). Descriptive statistics were used to summarize participant demographics and well-being survey results which included the PHQ-8, GAD-7, and the PFI. Single-sample 2-tailed t tests were used to investigate exploratory differences in physiologic measures.

Results

This was a pilot feasibility study and thus was not powered to detect individual health outcomes. A total of 20 participants were enrolled (13/20, 65% were female), 12 were EM resident physicians and 8 were emergency nurses. Of the 20 participants, 10 participants routinely wore the wearable device (6 resident physicians and 4 nurses).
Participants completing baseline mental health assessments reported mild anxiety as measured by the GAD-7 (mean score 5.07, SD 3.7), with 85% (n=17) reporting minimal or mild anxiety. Participants also reported mild depressive symptoms as measured by the PHQ-8 (mean 5.73, SD 2.9), with half reporting mild depressive symptoms. Participants completed the PFI to evaluate burnout and fulfillment. Individuals reported low professional fulfillment (mean 49.4, SD 16.9) moderate workplace exhaustion (mean 57.1, SD 24.4), and moderate interpersonal disengagement (mean 44.7, SD 20.1). Participants were asked via survey to comment on their early thoughts and goals with the pilot and the device. Notable themes emerged reflecting (1) technological features (eg, seeking a device with a watch face), (2) ways to integrate data into their personal lives and clinical roles, and (3) increasing self-awareness of the objective measures of stress related to clinical care.

Among participants who used the band consistently for 6 weeks, variation existed in their experience with the wearable and the data it generated. None of the participants were very likely to recommend the device to others. Two participants found the data interaction helpful and useful and overall, none commented on the platform being easy to use. When asked specifically, participants noted the band to be obtrusive given its lack of daily use features (eg, watch face and activity data) and odd charging mechanics. Several participants did comment positively that the data output and data generated was useful and empowering but needed to be collected using a more user-friendly design such as the more commonly used Apple Watch or FitBit. Participants sought the same application using these devices and expressed enthusiasm for those.

Of the 10 users who routinely used the device, we saw early variation in physiologic measures related to HRV, stress, and sleep (Table 1; Figures 1 and 2). While statistically significant differences are identified here, this remained a pilot study in feasibility, user input, and data collection methods. Early insights from these data suggest differences across roles between resident physicians and nurses, as well as across sex.

Table 1. Participant wearable device data.

<table>
<thead>
<tr>
<th></th>
<th>Overall (n=10); mean (SD)</th>
<th>Residents (n=6); mean (SD)</th>
<th>Nurses (n=4); mean (SD)</th>
<th>Women (n=4); mean (SD)</th>
<th>Men (n=6); mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days used</td>
<td>72.6 (49.7)</td>
<td>72.8 (46.9)</td>
<td>72.3 (61.1)</td>
<td>.99</td>
<td>50.5 (38.7)</td>
<td>87.3 (53.8)</td>
</tr>
<tr>
<td>Recovery score</td>
<td>59.5 (22.9)</td>
<td>60.7 (22.6)</td>
<td>57.7 (23.1)</td>
<td>.08</td>
<td>55.7 (25.7)</td>
<td>61.0 (21.5)</td>
</tr>
<tr>
<td>Resting heart rate</td>
<td>58.8 (7.4)</td>
<td>58.5 (5.6)</td>
<td>59.3 (9.4)</td>
<td>.16</td>
<td>63.3 (7.2)</td>
<td>57.1 (6.7)</td>
</tr>
<tr>
<td>Heart rate variability (ms)</td>
<td>50.4 (17.8)</td>
<td>47.3 (12.8)</td>
<td>55.1 (22.8)</td>
<td>&lt;.001</td>
<td>55.8 (29.1)</td>
<td>48.3 (10.0)</td>
</tr>
<tr>
<td>Sleep performance score</td>
<td>77.1 (19.1)</td>
<td>78.3 (19.4)</td>
<td>75.1 (18.6)</td>
<td>.03</td>
<td>78.2 (21.2)</td>
<td>76.6 (18.3)</td>
</tr>
<tr>
<td>Sleep disturbances</td>
<td>10.1 (5.8)</td>
<td>10.1 (5.5)</td>
<td>10.0 (6.3)</td>
<td>.82</td>
<td>13.1 (6.1)</td>
<td>8.9 (5.3)</td>
</tr>
</tbody>
</table>

Figure 1. An example, exploratory data synopsis at the participant level of heart rate variability (HRV) over the study period. The top pane depicts the HRV variation for an emergency nurse and the bottom for an emergency resident. This example snapshot can be viewed by participants on the mobile platform.
Discussion

Principal Findings

The physical and mental toll facing clinicians working in the ED continues to grow. Emergency nurses and resident physicians face a number of challenges impacting their health including, but not limited to, ED crowding, boarding, workplace violence, shifting schedules, and rising patient acuity. In the wake of the COVID-19 pandemic, an emphasis on supporting the workforce remains a priority. The evolving landscape of digital technology, including wearable devices, offers new opportunities for individuals to monitor their own health and potentially proactively identify physical or mental strain. This pilot study begins to examine if and how wearable devices can be used for emergency clinicians.

First, we found mixed enthusiasm for this approach given low interest in completing exit surveys and ongoing data interaction. It appears from this early feasibility study that resident physicians may be more engaged with this strategy and data collection method. Nurses in this pilot tended to be less engaged with the data tracking and follow-up mechanisms. Resident physicians were, in general, more enthusiastic before, during, and after the period ended. Though small in size, this pilot study does shed some initial interest from emergency clinicians and understanding key physiologic metrics such as sleep and physical activity. Key data insights remain physical activity as measured by step count, amount and quality of sleep, and HRV, which is an established physiologic measure relating to CV health [38]. The next steps to build on this pilot study included bringing it to a larger scale and designing it for nurse and physician preferences which we have learned to date.

The key finding of this pilot study is that the type of device and platform must be optimized for clinician use. In this study, we used a high-performance athletics device, which is designed primarily for physiologic measures [39]. This device does not have some traditional features that the average person may be accustomed to including a clock and the ability to send or receive messages. More traditional and popular bands such as the Apple Watch or Fitbit offer participants data tracking with features such as message-sending capability and a watch. These features are important when designing for scale, but were not previously known until we pilot-tested it, as other studies in more controlled environments have used the same device [39,40], underscoring the need to optimize design over technological capacity. The high-performance athletic band used in this pilot study offered a longer battery life and the ability to charge while wearing the device, which seemed less important to participants in this study. Future work needs to leverage the existing devices that clinicians wear in their everyday lives and incorporate those devices into this approach.

Finally, these devices and remote surveys highlight the persistent and real variation in professional disengagement, exhaustion, and burnout. In this small cohort, we do not identify significant amounts of anxiety or depression in screening assessments. We note some variation and physiologic measures across resin physicians and nurses as well as differences across individuals who identify as male vs female. These differences though statistically significant, represent only a small sample size, and follow-up studies need to be scaled at larger populations. Specific interests should investigate physiologic measures such as heart rate, HRV, and sleep.

Limitations

This study has several limitations. This was a single-center pilot study and had a small sample size, which was intentional by design. We see glimpses toward mechanisms to optimize digital technology and workforce sustainment. The signals identified
here represent only early pilot findings, and inherently there is also selection bias in individuals who opted to wear the band and complete surveys. Less is known about individuals who do not want to be part of this pilot, and future studies will need to be larger, randomized, and for a longer duration. Nonetheless, the study is among the first to begin to investigate the feasibility of using digital technology to support emergency physicians and nurses by helping them identify physiologic variations in their own health. The study represents the pilot beginnings to help identify proactive and much-needed new methods to mitigate strain that is related to physical and mental health.

Conclusion
This pilot study of emergency nurses and resident physicians investigating wearable devices to capture physiologic data from a cohort represents early signals toward feasible and acceptable programs. This pilot study identifies opportunities and interest in these mechanisms and a need to leverage more consumer-facing and potentially less sophisticated wearable devices for emergency clinicians. These methods can be further explored and larger, randomized trials can be conducted to investigate these strategies and how we support the workforce.

Acknowledgments
The authors would like to acknowledge the staff of the University of Pennsylvania Department of Emergency Medicine. This study was funded by the Emergency Medicine Foundation.

Data Availability
The deidentified data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


31. Your personal digital fitness and health coach. WHOOP. URL: https://www.whoop.com/ [accessed 2023-08-03]


37. Pioneering the field of high performance medicine. Arena Labs. URL: https://arenalabs.co/ [accessed 2023-08-03]


Abbreviations

CV: cardiovascular
ED: emergency department
EM: emergency medicine
GAD-7: General Anxiety Disorder
HRV: heart rate variability
PFI: Professional Fulfillment Index
PHQ-8: Patient Health Questionnaire

©Anish K Agarwal, Rachel Gonzales, Kevin Scott, Raina Merchant. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Feasibility of Guided Internet-Based Cognitive Behavioral Therapy for Panic Disorder and Social Anxiety Disorder in Japan: Pilot Single-Arm Trial

Seina Shinno1*, MMSc; Kazuki Matsumoto2,3*, PhD; Sayo Hamatani3,4, PhD; Yosuke Inaba5, PhD; Yoshihito Ozawa5, BSc; Yohei Kawasakii, PhD; Tomoki Ikai6, MD, PhD; Chihiro Sutoh3, MD, PhD; Hiroyuki Hayashi1,7, MD, PhD; Eiji Shimizu3,8, MD, PhD

1Department of Integrated Medical Sciences, Graduate School of Medicine, University of Fukui, Fukui, Japan
2Division of Clinical Psychology, Kagoshima University Medical and Dental Hospital, Kagoshima, Japan
3Research Center for Child Mental Development, Chiba University, Chiba, Japan
4Research Center for Child Mental Development, University of Fukui, Fukui, Japan
5Biostatistics Section, Clinical Research Center, Chiba University Hospital, Chiba, Japan
6Faculty of Medical Sciences, Division of Medicine Primary Health Care, University of Fukui, Fukui, Japan
7Department of Emergency and General Medicine, Fukui University Hospital, Fukui, Japan
8Department of Cognitive Behavioral Physiology, Graduate School of Medicine, Chiba University, Chiba, Japan
*these authors contributed equally

Corresponding Author:
Kazuki Matsumoto, PhD
Division of Clinical Psychology
Kagoshima University Medical and Dental Hospital
1-35-8 Sakuragaoka
Kagoshima, 8908520
Japan
Phone: 81 099 275 5707
Email: k2782199@kadai.jp

Abstract

Background: Cognitive behavioral therapy (CBT) is effective in treating anxiety disorders. Accessibility to CBT has been limited in Japan due to the shortage of therapists. While an open-source e-learning system can be used to create a simple internet-based cognitive behavioral therapy (ICBT) program, the safety and outpatient acceptance of this treatment approach have not been explored in Japan.

Objective: The aim of this study was to investigate whether outpatients with anxiety disorders could accept and successfully complete the ICBT program with guidance by CBT therapists when implementing therapeutic modules and CBT tasks. Due to being in the initial phase of a novel treatment in Japan, this study was intended for verification with a small sample size.

Methods: In total, 6 adults, including 4 male participants and 2 female participants, were enrolled in a single-arm trial. The intervention involved guided ICBT comprising 12 sessions, including CBT text, comprehension confirmation tests, and explanatory videos about cognitive behavioral models, accessible through a website. The therapist guided the participants in accessing the ICBT program and answering their questions using a chat tool. The primary outcome was anxiety severity assessed using the State-Trait Anxiety Inventory-Trait. Secondary outcomes included the Panic Disorder Severity Scale, Liebowitz Social Anxiety Scale (LSAS), Beck Anxiety Inventory (BAI), Patient Health Questionnaire–9, Generalized Anxiety Disorder–7, and Working Alliance Inventory–Short Form (WAI-SF). Statistical analyses were performed using paired 2-tailed t tests to assess the changes in clinical symptoms. The total WAI-SF score at the final session was used to evaluate the therapeutic alliance. For statistical analyses, mean changes for total State-Trait Anxiety Inventory-Trait, BAI, Panic Disorder Severity Scale, LSAS, Patient Health Questionnaire–9, and Generalized Anxiety Disorder–7 scores were analyzed using the paired 2-tailed t test. The 2-sided significance level for hypothesis testing was set at 5%, and 2-sided 95% CIs were calculated.

Results: Most participants diligently engaged with the ICBT program. No adverse events were reported. The mean total scores for the primary outcome decreased by 11.0 (SD 9.6) points (95% CI –22.2 to 0.20; Hedges g=0.95), but it was not statistically
significant. The mean total scores for the secondary outcomes that assess clinical symptoms decreased, with a significant reduction observed in the BAI of 15.7 (SD 12.1) points (95% CI –28.4 to –3.0; \( P=0.03; \) Hedges \( g=1.24 \)). The mean total scores for PDSS and LSAS decreased significantly, by 12.0 (SD 4.24) points (95% CI –50.1 to 26.1; \( P=0.16; \) Hedges \( g=1.79 \)) and 32.4 (SD 11.1) points (95% CI –59.7 to –4.3; \( P=0.04; \) Hedges \( g=1.38 \)), respectively. Of the participants, 67% (n=4) showed treatment response, and 50% (n=3) achieved remission after the intervention. The therapeutic alliance, measured using the WAI-SF, was moderate.

**Conclusions:** Guided ICBT may be feasible for the treatment of outpatients with panic disorder and social anxiety disorder in Japan.

**Trial Registration:** University Hospital Medical Information Network Clinical Trials Registry UMIN0000038118; https://center6.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000043439

(JMIR Form Res 2024;8:e53659) doi:10.2196/53659

**KEYWORDS**
cognitive behavioral therapy; internet intervention; panic disorder; social anxiety; feasibility trial; adult; adults; anxiety disorder; internet-based; e-learning; Japan; statistical analyses; therapist; therapists; intervention; severity; symptoms; therapeutic alliance; mobile phone

**Introduction**

Anxiety disorders such as social anxiety disorder (SAD) and panic disorder (PD) involve prolonged and significant anxiety, leading to substantial impairment in daily functioning. In Japan, the 12-month prevalence of anxiety disorders is 5.3%, making them the most commonly diagnosed mental disorders [1]. Cognitive behavioral therapy (CBT) has been proven to be effective in treating anxiety disorders, particularly SAD and PD [2,3]. As per the clinical guidelines [4,5], CBT is the primary treatment option for patients with SAD or PD. However, CBT has rarely been integrated into the Japanese psychiatric clinical practice. The limited availability of CBT therapists and high implementation costs have been major obstacles significantly impacting the accessibility of CBT [6,7]. According to a 2018 report, the implementation rate of CBT in Japanese psychiatric clinics was 6.2% [7]. When CBT is applied, it is rarely based on an evidence-based cognitive behavioral model. In most CBT practices, therapists independently select and combine CBT techniques [7].

Since the latter half of the 1990s, self-help CBT programs have been provided in rural areas via the internet. CBT delivered through the internet is known as internet-based cognitive behavioral therapy (ICBT) [8]. In the ICBT approach, therapists guide patients via phone or email to enhance their treatment experience. Previous studies of guided ICBT have demonstrated significant therapeutic effects on anxiety disorders. According to 2 systematic reviews with meta-analyses, guided ICBT is likely as effective as face-to-face CBT [9,10]. Most studies of guided ICBT in these reviews were conducted in Europe, particularly, Northern Europe [10-13]. Due to the varying internet infrastructure and ICT literacy across different countries, evidence from clinical trials in Eastern countries is crucial.

Recently, several studies on ICBT with treatment programs adapted to specific cultures have been conducted in China, Korea, and Pakistan, yielding results supporting the effectiveness of ICBT [14-16]. Although some culturally adapted ICBT programs for mental disorders have begun to emerge in Japan [17-20], there is a lack of evidence of clinical trials that specifically target anxiety disorders [21]. A previous single-arm study in Japan demonstrated the feasibility of CBT via videoconference as another remote format for 20 adults with SAD or PD [22]. In this study, remote therapists provided individual CBT in real time. The results indicated that Japanese individuals with anxiety disorders find remote interventions preferable and readily acceptable. However, in the guided ICBT format, the absence of face-to-face interaction with the therapist is fundamental, and it is anticipated that the treatment experience may qualitatively differ, potentially leading to less favorable patient acceptance and treatment responses.

To carefully assess the feasibility of guided ICBT, a novel treatment in Japan, we conducted a rigorously controlled clinical trial with a small sample size, monitoring for the risk of deterioration and unexpected adverse events. The participants’ guided ICBT experiences were also summarized in a brief case series included as Multimedia Appendix 1.

**Methods**

**Study Design**

This prospective single-arm open trial was conducted between September 2019 and March 2020 at 2 university hospitals and a psychiatric clinic in Japan. Patients with PD or SAD were recruited between September and November 2019 through posters and leaflets placed at medical institutions in the 2 prefectures and on the university’s home page. The participants continued to receive treatment from general medical practitioners during the study period, and their authorization was obtained before enrollment. Personal IDs and passwords to the treatment program and chat tools were provided to participants.

**Ethical Considerations**

This study project and the present clinical trial protocol adhered to the Ethical Guidelines for Medical and Biological Research Involving Human Subjects in Japan [23]. This study was approved by the institutional review board of Chiba University Hospital (G2019004). Written informed consent was obtained from all participants after they were thoroughly informed of the study protocol. No compensation was provided for participating in the research, although the guided ICBT was provided for...
This pilot single-arm trial was registered at the University Hospital Medical Information Network Clinical Trials Registry (UMIN0000038118).

**Recruitment**

Figure 1 illustrates the flow of participants throughout the study. Initially, 8 patients were recruited in the study. The screening through psychological assessments was conducted by 3 clinical psychologists (SS, KM, and SH), and it was ultimately confirmed by 2 physicians (HH and ES) whether the participants met the diagnostic criteria of the *Diagnostic and Statistical Manual of Mental Disorder, Fifth Edition* [24]. After the screening, 2 patients were excluded either because they did not meet the inclusion criteria or because of pending cancellations as part of the participant management process. In total, 6 patients underwent a baseline assessment and were subsequently enrolled in the study. Following enrollment, 1 patient disagreed with the exposure in session 6 during the intervention. Notably, there were no dropouts in this trial, and no adverse events were reported during treatment. Of the 6 patients, 5 (83%) successfully completed the treatment modules and participated in the telephonic posttreatment assessment conducted by an independent assessor. One patient with PD who could not be assessed over the phone received all symptom rating scales via mail except for the Panic Disorder Severity Scale (PDSS).

**Figure 1.** Participant flow.

**Inclusion Criteria**

Inclusion criteria were (1) age between 18 and 65 years, (2) a primary diagnosis of PD or SAD according to the Mini-International Neuropsychiatric Interview [25], (3) moderate symptoms of PD or SAD: PDSS total score >9 or Liebowitz Social Anxiety Scale (LSAS) total score >70 [26-29], and (4) the ability to send emails and access the e-learning system.

**Exclusion Criteria**

Exclusion criteria were (1) psychosis and bipolar disorder, (2) current high risk of suicide, (3) substance abuse or dependence diagnosed within the past 12 months (antisocial personality disorder), and (4) IQ<80 measured by the Japanese Adult Rating Test [30].

**Primary Outcome Measure**

The primary outcome was a change in anxiety symptoms assessed using the Japanese version of the State-Trait Anxiety Inventory-Trait (STAI-T) [31]. The STAI-T is a self-reported scale comprising 40 items that quantitatively measures anxiety and can be evaluated using 2 scales [32]. The STAI-T comprises 20 items rated on a 4-point scale, and the total score ranges from 20 to 80, divided into 5 levels (I=very low, II=low, III=normal, IV=high, and V=very high). The validity and reliability of the Japanese version of the STAI-T have been demonstrated [33].

**Secondary Outcome Measures**

In addition to the primary outcome, several secondary outcomes with established reliability and validity were selected based on the research objectives. The Beck Anxiety Inventory (BAI) was...
used to quantify comprehensive anxiety symptoms [34]. The BAI is a 21-question multiple-choice self-report inventory used to assess anxiety severity [34,35]. Other secondary outcome measures included the PDSS for participants with a primary PD diagnosis [26,27] and the LSAS for those with a primary SAD diagnosis [28,29].

The PDSS is a 7-item clinical interview rating scale that evaluates the core PD characteristics and demonstrates good psychometric properties [36]. As the PDSS is meant to be administered by professionals knowledgeable of the clinical manifestations of PD, an independent assessor (SH) was used to conduct and interpret this scale. The treatment response and remission rates were calculated using the LSAS and PDSS. For SAD, treatment response was defined as a 31% or greater reduction in the total LSAS score, and remission was defined as a final LSAS score of \( \leq 36 \) [29]. For PD, treatment response was defined as a 40% or greater reduction in the total PDSS score, and remission was defined as a final PDSS score of \( \leq 5 \) [37].

The trial also assessed the psychological bond between therapists and participants using the Working Alliance Inventory-Short Form (WAI-SF) [38], depressive symptoms using the Patient Health Questionnaire–9 (PHQ-9) [39,40], and generalized anxiety symptoms using the Generalized Anxiety Disorder–7 (GAD-7) [39,41]. The WAI-SF assesses the strength of the therapeutic alliance between the therapist and patient through 12 items rated on a scale of 1 (never) to 7 (always). The total score ranges from 12 to 84, reflecting the overall strength of the therapeutic bond between therapist and patient.

Both the PHQ-9 and GAD-7 were scored on a 4-point scale (0=none, 1=a few days, 2=more than half, and 3=almost daily). The PHQ-9 scores range from 0 to 27, with the cutoff value for clinically significant depressive symptoms set at 10. Symptomatology was categorized as follows: 0-4=none, 5-9=mild, 10-14=moderate, 15-19=moderate to severe, and 20-27=severe depressive state. The Japanese version of the PHQ-9 has demonstrated adequate validity [40]. GAD-7 scores range from 0 to 21, with a cutoff value for clinically significant generalized anxiety set at 10. Symptomatology was defined as follows: 0-4=none, 5-9=mild, 10-14=moderate, and 15-21=severe general anxiety. The GAD-7 has demonstrated good reliability and validity [42].

**Intervention**

We developed ICBT programs based on the Clark and Wells [43] model for SAD and Seki and Shimizu's model for PD [44]. Each ICBT program consisted of 12 modules. Therapeutic guidance was provided by a clinical psychologist (KM) with extensive experience in face-to-face CBT for patients with PD and SAD. The therapist, a clinical psychologist with a PhD, had completed the CBT training course [45]. After each session, the therapist conducted peer supervision with the third author. The intervention involved guided ICBT comprising 12 sessions, including CBT text, comprehension confirmation tests, and explanatory videos about cognitive behavioral models, accessible through a website. The therapist guided the participants in accessing the ICBT program and answering their questions using a chat tool.

The ICBT program for PD included the following modules: guidance for ICBT program (session 0), psychological education and case conceptualization (week 1), review of safe behaviors (week 2), modification of self-image (week 3), attention shift training (week 4), behavioral experiments with internal sensory exposure (week 5), staged exposure (week 6 and 7), intervention on memory of the first panic attack (week 8), examination of others’ interpretations of panic attacks (week 9), stop dwelling on the panic attacks (week 10), reconstruction of remaining beliefs (schema work; week 11), and relapse prevention (week 12).

The ICBT program for SAD included the following modules: guidance for ICBT program (session 0), psychological education and case conceptualization (week 1), examination of safety behavior (week 2), video feedback (week 3), attention shift training (week 4), behavioral experiment 1 (week 5), behavioral experiment 2 (week 6), creation of anxiety hierarchy form and graded exposure (week 7), validating negatively rated expectations (week 8), stop dwelling on things (week 9), rewriting the meaning of memories linked to self-image (week 10), reconstruction of remaining beliefs (schema work; week 11), and relapse prevention (week 12).

The participants in this study were generally instructed not to initiate, discontinue, or modify their pharmacotherapy after the intervention. The participants were required to inform the research team if there were any changes to their medication regimen.

**Hardware and Software**

The participants used personal PCs, tablet PCs, or smartphones. The software used included LearningBox (Tatsuno Information System Co) and MediLine (ShareMedical Co), a medical chat service for delivering ICBT. LearningBox is an e-learning system enabling administrators to create and manage educational materials, handle member groups, and record and assess grades. Although e-learning systems store and manage user results, they do not include personal or sensitive information.

MediLine is a medical chat service (medical social networking service) designed to replace email and phone calls. It incorporates robust encryption to prevent military-level information leaks and operates in a double-encrypted state, following the Japanese government guidelines. End-to-end encryption is performed in real time on the server during communication, and temporary memory is used in the terminal. The lectures for each CBT session were recorded as video footage, uploaded to YouTube [46,47], and shared with the participants.

**Data Setting**

SH was responsible for collecting outcomes at baseline (week 1) and postintervention (week 12). The collected data were registered on a server and managed by the Data Management Office of Chiba University Hospital.

**Statistical Analysis**

Statistical analyses were conducted according to the intention-to-treat principle by a team of statistical analysts (YI, YO, and YK). For screening assessments at baseline, summary...
statistics were generated, including proportions for categorical data and means and SDs for continuous variables. All outcomes that could be expressed as continuous variables were analyzed using paired 2-tailed \( t \) tests before and after the intervention. Specifically, total STAI-T, BAI, PDSS, LSAS, PHQ-9, and GAD-7 scores were analyzed. The 2-sided significance level for hypothesis testing was set at 5\%, and 2-sided 95\% CIs were calculated. The total WAI-SF score, which measures the strength of the therapeutic alliance, was submitted as a raw score after the completion of ICBT. One male participant with PD did not submit PDSS data after the intervention. Imputation of missing values was not performed.

### Results

#### Demographic Data and Clinical Characteristics

In total, 6 participants (4 male participants and 2 female participants) with a mean age of 41 (SD 8.2; range 26-51) years were enrolled in the clinical trial. **Table 1** shows the demographic data and the participants’ clinical characteristics. All participants continued their pharmacotherapy during the trial with specific medications such as paroxetine hydrochloride hydrate (n=3), venlafaxine hydrochloride (n=1), sertraline hydrochloride (n=1), and a combination of sertraline hydrochloride and ezilam (n=1). The participants’ estimated IQs tended to be higher than the mean (110, SD 4.2).

**Table 1.** Demographic characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Participant 1</th>
<th>Participant 2</th>
<th>Participant 3</th>
<th>Participant 4</th>
<th>Participant 5</th>
<th>Participant 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td>Male</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>37</td>
<td>44</td>
<td>26</td>
<td>51</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td><strong>Education (years)</strong></td>
<td>16</td>
<td>12</td>
<td>9</td>
<td>16</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td>Full time</td>
<td>Full time</td>
<td>Unemployed</td>
<td>Full time</td>
<td>Unemployed</td>
<td>Unemployed</td>
</tr>
<tr>
<td><strong>Age at onset (years)</strong></td>
<td>16</td>
<td>20</td>
<td>12</td>
<td>51</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td><strong>Primary diagnosis</strong></td>
<td>SAD(^a)</td>
<td>SAD</td>
<td>SAD</td>
<td>PD(^b)</td>
<td>PD</td>
<td>PD</td>
</tr>
<tr>
<td><strong>Comorbidity</strong></td>
<td>None</td>
<td>None</td>
<td>MDD(^c)</td>
<td>AP(^d)</td>
<td>AP</td>
<td>AP</td>
</tr>
<tr>
<td><strong>Estimated IQ</strong></td>
<td>110</td>
<td>104</td>
<td>108</td>
<td>110</td>
<td>112</td>
<td>116</td>
</tr>
</tbody>
</table>

\(^a\)SAD: social anxiety disorder.  
\(^b\)PD: panic disorder.  
\(^c\)MDD: major depressive disorder.  
\(^d\)AP: agoraphobia.

#### Adverse Events

No mental or physical adverse events were reported after the intervention.

#### Evaluation Outcomes

**Table 2** provides a detailed overview of the outcomes. In the primary outcome, the mean total score on the STAI-T decreased by 11.0 (SD 9.6) points, although this did not reach statistical significance (95\% CI –22.2 to 0.20; \( P = .05 \); Hedges \( g = 0.95 \)). However, several secondary outcomes demonstrated significant improvements. The mean total BAI score significantly decreased by 15.7 (SD 12.1) points (95\% CI –28.4 to –3.0; \( P = .03 \); Hedges \( g = –1.24 \)). The mean total PDSS score also decreased significantly, by 12.0 (SD 4.2) points (95\% CI –50.1 to 26.1; \( P = .16 \); Hedges \( g = 1.79 \)). Furthermore, the mean total LSAS score significantly decreased by 32.0 (SD 11.1) points (95\% CI –59.7 to –4.3; \( P = .04 \); Hedges \( g = 1.38 \)).
Table 2. Outcomes from pre- to postintervention (N=6).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Preintervention, mean (SD)</th>
<th>Postintervention, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAI-T&lt;sup&gt;a&lt;/sup&gt;</td>
<td>56.8 (12.5)</td>
<td>45.8 (10.5)</td>
<td>.05</td>
</tr>
<tr>
<td>BAI&lt;sup&gt;b&lt;/sup&gt;</td>
<td>34.2 (13.9)</td>
<td>18.5 (11.1)</td>
<td>.03</td>
</tr>
<tr>
<td>PDSS&lt;sup&gt;c&lt;/sup&gt; (n=3)</td>
<td>15.3 (6.03)</td>
<td>3.0 (4.24)</td>
<td>.16</td>
</tr>
<tr>
<td>LSAS&lt;sup&gt;d&lt;/sup&gt; (n=3)</td>
<td>88.3 (17.0)</td>
<td>56.2 (28.2)</td>
<td>.40</td>
</tr>
<tr>
<td>PHQ-9&lt;sup&gt;e&lt;/sup&gt;</td>
<td>12.5 (5.61)</td>
<td>8.7 (5.61)</td>
<td>.16</td>
</tr>
<tr>
<td>GAD-7&lt;sup&gt;f&lt;/sup&gt;</td>
<td>9.17 (6.85)</td>
<td>6.0 (4.5)</td>
<td>.27</td>
</tr>
</tbody>
</table>

<sup>a</sup>STAI-T: State-Trait Anxiety Inventory-Trait.
<sup>b</sup>BAI: Beck Anxiety Inventory.
<sup>c</sup>PDSS: Panic Disorder Severity Scale.
<sup>d</sup>LSAS: Liebowitz Social Anxiety Scale.
<sup>e</sup>PHQ-9: Patient Health Questionnaire–9.
<sup>f</sup>GAD-7: Generalized Anxiety Disorder–7.

Regarding depressive symptoms, the mean total PHQ-9 score decreased from 12.5 (SD 5.61) to 8.7 (SD 5.61) points, but the reduction was not statistically significant (95% CI –9.3 to 1.6; \( P=0.13 \); Hedges \( g=0.68 \)). Similarly, for generalized anxiety symptoms, the mean total GAD-7 score decreased from 9.17 (SD 6.85) to 6.0 (SD 4.47) points, but the reduction was not statistically significant (95% CI –9.8 to 3.5; \( P=.27 \); Hedges \( g=0.55 \)).

At the postintervention assessment (week 12), the treatment response rates were 67% (n=2) for PD and 67% (n=2) for SAD. The remission rates were 67% (n=2) for PD and 33% (n=1) for SAD.

**Acceptance of Guided ICBT**

Of the 4 male participants, 3 completed all treatment modules. These 3 participants completed the treatment without asking questions to the therapist and were able to understand the content and execute therapeutic tasks independently. A female participant with SAD avoided even minimal communication with therapists through chat tools due to severe social anxiety symptoms. While she accepted the self-help format of CBT, the chat with the therapist itself became burdensome rather than beneficial. A female participant with PD and fear of panic attack reported being encouraged by the therapist’s empathy and encouragement. A male participant with PD self-discontinued the treatment module and subsequently lost contact with the therapist. He frequently struggled to control his emotions, making it challenging to address his concerns effectively. For a brief case series of guided ICBT, please refer to Multimedia Appendix 1.

**Discussion**

**Principal Results**

This study explored the feasibility of guided ICBT in outpatients with PD and SAD in Japan. In total, 5 of the 6 participants successfully completed all modules, and no adverse events were reported throughout the study period. While anxiety symptoms decreased in most patients after the intervention, with a substantial effect size, the primary outcome (STAI-T) did not show a significant difference. However, significant improvements were observed in the secondary outcome, as measured by the BAI. Our results suggest that guided ICBT has a positive impact on anxiety symptoms, as evidenced by the notable changes in BAI scores. Most participants diligently engaged with the ICBT programs. Meanwhile, patients who considered themselves to have enough understood the materials understood the materials hardly contacted the therapists.

**Limitations**

When interpreting our study, it is important to consider the following limitations. First, the small sample size prevented us from determining whether the findings were due to a type II (\( \beta \)) error. Second, this was a single-arm trial without a control group. More robust and statistically rigorous results could be obtained through a clinical randomized controlled trial in which the sample size is determined based on the effect sizes observed in our study. A randomized controlled trial in Japan would allow for better investigation of the effectiveness of ICBT in Japanese patients with anxiety disorders.

Additionally, we did not assess the extent to which the participants implemented ICBT program strategies in their daily lives. While the therapist encourages participants to practice what they learned through chat tools, phone calls, and in-person interactions, patients might face challenges in implementing these changes or coping with symptoms [48]. In this study, patients with PD had difficulty applying learned techniques in daily life. Future studies should evaluate the impact of practice on symptom improvement to gain a more comprehensive understanding of the therapeutic processes.

**Comparison With Prior Work**

Our results showed that participants engaged in video lectures explaining the treatment modules, potentially enhancing their understanding and motivation for treatment. Despite the small sample size, guided ICBT with these lecture videos showed significant effectiveness as reflected in a treatment response rate of 67% (n=4) and a remission rate of 50% (n=3). These
results are consistent with a representative meta-analysis examining treatment response predictors in 1162 patients undergoing ICBT [9]. The reduction in the total score on the primary symptom rating scales before and after the trial, along with the effect size, underscored the efficacy of guided ICBT that incorporated lecture videos. Notably, the within-group effect sizes for the primary symptoms in this study (Hedges g=1.38 for LSAS and 1.79 for PDSS) were as substantial as those observed in previous representative studies [49-51].

In this study, the dropout rate was 17% (1 of 6 participants), which is comparable to face-to-face CBT (26%) [52]. In contrast, another study of guided ICBT reported an average dropout rate of 32% for depression [53]. Notably, one dropout from our trial was a male patient with a significantly lower therapeutic alliance score (12 points on the WAI-SF total score). Consistent with previous studies on dropout predictors, patients with low therapeutic alliance scores tended to drop out [54]. When the therapeutic alliance is weak, therapists and patients should engage in discussions regarding treatment goals and tasks. However, rebuilding therapeutic alliances in guided ICBT contexts may be more challenging than in face-to-face contexts.

Participants with anxiety disorders might willingly participate in CBT programs but may not always practice cognitive behavioral skills [8]. For instance, one participant with PD (participant 5) experienced severe panic symptoms related to employment status. Conversely, some patients demonstrate independent therapeutic progress without extensive therapist interactions. Overall, our findings align with the existing research, indicating that severe anxiety, coexisting depression, and low socioeconomic status are associated with poor outcomes [55]. Please refer to Multimedia Appendix 1 for details regarding each participant’s treatment course.

In the case of SAD, all 3 participants successfully applied the acquired skills to their daily lives. However, none of the participants with SAD used a communication tool with their therapist. This suggests that using a chat tool may be burdensome for individuals with social anxiety, potentially reducing the accessibility of ICBT, as these patients often dislike high-intensity communication. However, the therapists’ guidance may have encouraged them to engage in self-help activities [56]. For patients with social anxiety who find it challenging to ask questions or express their thoughts, therapists could consider expressing their ability to provide answers as needed instead of engaging in routine email interactions.

It is worth noting that even patients with social anxiety who avoid chatting may still be willing to receive treatment, as indicated by a guided ICBT trial conducted in China, in which patients with SAD completed more modules of the treatment program than the nonclinical group [15]. However, if the symptoms are severe, patients may struggle to practice cognitive behavioral skills independently in their daily lives. Participants who showed symptom improvement had relatively high social functioning, continued employment, and direct community involvement (participants 1, 2, 4, and 6).

Conversely, patients with severe disorders (participant 3) and those resistant to exposure-mediated interventions (participant 5) may require a significant amount of therapeutic attention. For instance, individual face-to-face CBT might be a more suitable option for someone like participant 3, who has SAD. Similarly, for participant 5, a personalized face-to-face CBT approach in which the therapist helps analyze the gains and losses of exposure and safety behaviors and demonstrates internal sensory exposure techniques could be beneficial. Clinicians and therapists who are going to conduct CBT should consider disclosing treatment modules to patients and assess the level of support required before providing ICBT.

**Conclusions**

Guided ICBT may be a feasibility treatment approach for patients with PD and SAD in Japan. If the patients have high motivation and the ability to understand and practice treatment modules, supports might not be strictly necessary during the treatment from the CBT therapist. The empathetic words from the therapist and encouragement for program implementation may assist some patients in overcoming fear and engaging in the assigned tasks.

**Acknowledgments**

This study was funded by the Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Research (grant 18K03130).

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

**Authors’ Contributions**

SS played a significant role in study design, recruitment, and paper writing. KM contributed to all aspects of the research process. KM, SH, and ES were involved in the development and implementation of the cognitive behavioral therapy course. TI, HH, and ES contributed to the study design, participant recruitment, and the overall administration of the clinical trial. YI, YO, and YK were responsible for conducting the statistical analyses. All authors thoroughly reviewed the paper before submission and granted their approval for publication.
Conflicts of Interest

None declared.

Multimedia Appendix 1
Case series of the 6 participants.

References


**Abbreviations**

- **BAI**: Beck Anxiety Inventory
- **CBT**: cognitive behavioral therapy
- **GAD-7**: Generalized Anxiety Disorder–7
- **ICBT**: internet-based cognitive behavioral therapy
- **LSAS**: Liebowitz Social Anxiety Scale
- **PD**: panic disorder
- **PDSS**: Panic Disorder Severity Scale
- **PHQ-9**: Patient Health Questionnaire–9
- **SAD**: social anxiety disorder
- **STAI-T**: State-Trait Anxiety Inventory-Trail
- **WAI-SF**: Working Alliance Inventory-Short Form
©Seina Shinno, Kazuki Matsumoto, Sayo Hamatani, Yosuke Inaba, Yoshihito Ozawa, Yohei Kawasaki, Tomoki Ikai, Chihiro Sutoh, Hiroyuki Hayashi, Eiji Shimizu. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Six-Month Pilot Testing of a Digital Health Tool to Support Effective Self-Care in People With Heart Failure: Mixed Methods Study

Alison Keogh1,2, BSc, MSc, PhD; Carol Brennan3, BSc, MSc; William Johnston1, BSc, PhD; Jane Dickson4,5, BSc; Stephen J Leslie6, BSc, PhD; David Burke5,7, BSc, MSc; Peter Megyesi1,3, BSc; Brian Caulfield1,3, BSc, MSc, PhD

1Insight Centre Data Analytics, University College Dublin, Dublin, Ireland
2School of Medicine, Trinity College Dublin, Dublin, Ireland
3School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
4Physiotherapy Department, Beacon Hospital, Dublin, Ireland
5Cardiology, Beacon Hospital, Dublin, Ireland
6Cardiac Unit, Raigmore, Inverness, United Kingdom
7School of Medicine, University College Dublin, Dublin, Ireland

Corresponding Author:
Alison Keogh, BSc, MSc, PhD
Insight Centre Data Analytics
University College Dublin
Belfield
Dublin, D04V1W8
Ireland
Phone: 353 17167777
Email: Alison.keogh@ucd.ie

Abstract

Background: Digital tools may support people to self-manage their heart failure (HF). Having previously outlined the human-centered design development of a digital tool to support self-care of HF, the next step was to pilot the tool over a period of time to establish people’s acceptance of it in practice.

Objective: This study aims to conduct an observational pilot study to examine the usability, adherence, and feasibility of a digital health tool for HF within the Irish health care system.

Methods: A total of 19 participants with HF were provided with a digital tool comprising a mobile app and the Fitbit Charge 4 and Aria Air smart scales for a period of 6 months. Changes to their self-care were assessed before and after the study with the 9-item European HF Self-care Behavior Scale (EHFScBS) and the Minnesota Living with HF Questionnaire (MLwHFQ) using a Wilcoxon signed rank test. After the study, 3 usability questionnaires were implemented and descriptively analyzed: the System Usability Scale (SUS), Wearable Technology Motivation Scale (WTMS), and Comfort Rating Scale (CRS). Participants also undertook a semistructured interview regarding their experiences with the digital tool. Interviews were analyzed deductively using the Theoretical Domains Framework.

Results: Participants wore their devices for an average of 86.2% of the days in the 6-month testing period ranging from 40.6% to 98%. Although improvements in the EHFScBS and MLwHFQ were seen, these changes were not significant ($P=0.10$ and $P=0.70$, respectively, where $P>.03$, after a Bonferroni correction). SUS results suggest that the usability of this system was not acceptable with a median score of 58.8 (IQR 55.0-60.0; range 45.0-67.5). Participants demonstrated a strong motivation to use the system according to the WTMS (median 6.0, IQR 5.0-7.0; range 1.0-7.0), whereas the Fitbit was considered very comfortable as demonstrated by the low CRS results (median 0.0, IQR 0.0-0.0; range 0.0-2.0). According to participant interviews, the digital tool supported self-management through increased knowledge, improved awareness, decision-making, and confidence in their own data, and improving their social support through a feeling of comfort in being watched.

Conclusions: The digital health tool demonstrated high levels of adherence and acceptance among participants. Although the SUS results suggest low usability, this may be explained by participants uncertainty that they were using it fully, rather than it being unusable, especially given the experiences documented in their interviews. The digital tool targeted key self-management behaviors and feelings of social support. However, a number of changes to the tool, and the health service, are required before it
can be implemented at scale. A full-scale feasibility trial conducted at a wider level is required to fully determine its potential effectiveness and wider implementation needs.

(JMIR Form Res 2024;8:e52442) doi: 10.2196/52442

KEYWORDS
digital health; heart failure; cardiology; self-care; behavior change; eHealth; mHealth; mobile health; mobile app; mobile phone; elderly; self-care; self-management; digital tools; digital tool; human-centered design; app; apps; applications; wearables; wearable; Fitbit; usability; adherence; feasibility; congestive heart failure; cardiac failure; myocardial failure; heart decompensation

Introduction

Heart failure (HF) is a major global cause of disability associated with high morbidity and mortality, frequent hospitalization, high health care costs, impaired functional status, and poor quality of life [1-3]. Defined as “a clinical syndrome with symptoms and/or signs caused by a structural and/or functional cardiac abnormality and corroborated by elevated natriuretic peptide levels and/or objective evidence of pulmonary or systemic congestion” [4], symptoms include shortness of breath (dyspnea), fatigue, pulmonary edema, and a reduced ability to complete activities of daily living [2]. Self-care behaviors are critical components of the long-term management of HF [3,5-8].

Self-care is an overarching concept formed of the key concepts of (1) self-care maintenance (eg, taking or adjusting medication as prescribed, engaging in physical activity, and adhering to a healthy diet), (2) self-care monitoring (eg, regular weighing), and (3) self-care management (eg, changing diuretic dose in response to symptoms) [5]. Adequate self-care requires patients to understand what they need to do, have the skills to implement advice given to them, and adjust their behaviors according to how they feel and their symptoms. This is a complex, multicomponent set of behaviors that is constantly evolving, which can result in patients finding it difficult to successfully undertake [2,3,9,10].

Consequently, recent years have seen growing interest in the development of new and novel ways to support patients in their self-care behaviors. In particular, digital health options have begun to be widely used as the ubiquitous use of smartphone apps, and wearable devices appear to be adopted to support health care [3,11]. Used independently, or in combination, such digital health tools (DHTs) may support patients to monitor key behaviors, support improved autonomy in their own care, and provide pathways for patients and health care professionals (HCPs) to communicate [2,9]. However, despite the promise of DHTs in the management of HF, results regarding its potential effectiveness are mixed [3,12,13]. The pace of development of digital tools has resulted in them only being tested in small numbers or over a short duration, and critically, these tools are rarely developed with clear clinical or patient perspectives embedded within them [3]. Recently, the use of human-centered design approaches has led to the successful development of digital health technologies designed to support chronic disease management [14]. As such, implementing a human-centered design approach to the development of DHTs designed to promote self-care behaviors in patients with HF may positively influence the impact their condition has on their quality of life.

Therefore, we used a human-centered design process, as described in the International Organization for Standardization 9241-210:2019 regulations [15], to design and develop a DHT to support effective self-care behaviors in people with HF over a medium-term duration of 6 months to test for the potential of participant fatigue with the DHT (ie, reduced usage over time, poor acceptability, etc). The full process and approach taken were previously outlined in detail elsewhere and pilot-tested over a 2 week period [16]. Specifically, a consumer grade device was used to understand whether such ubiquitous tools can be used to empower people with HF to self-monitor their condition, bridging the gap between them and their clinicians. However, this tool required a more robust assessment of the longitudinal impact on self-care behaviors. Additionally, we wished to get an initial indication of its potential use from the perspective of the HCPs. Therefore, the aim of this study was to conduct a 6-month observational pilot test of this DHT in practice to examine its usability and the participant’s adherence to the system.

Methods

Recruitment

Participants were recruited from a private hospital in Dublin, Ireland, between July and October 2021 and had previously been diagnosed with HF. Purposive sampling was used using the patient lists of Beacon Hospital Cardiology to facilitate the assessment of the acceptability and usability of the DHT. Participants were deemed eligible if they could provide written informed consent; were previously diagnosed with HF; were under the care of Beacon Hospital Cardiology (aged ≥18 years); were under New York Heart Association classification 1-3; were open to the use of technology in the promotion of HF self-care; had access to an internet connection or mobile data; and were intellectually, visually, and auditorily capable of communicating with the investigator and understanding and complying with the requirements of the study. Participants were deemed ineligible if they were medically unstable or undergoing medical treatment judged not to be medically compatible by the investigator (eg, undergoing treatment for cancer), or if they had any skin condition that may affect the integrity of their skin when wearing the activity tracker. Participants (n=43) were approached directly by the members of the cardiology team to determine their interest and eligibility in the study.

Ethical Considerations

The study received ethical approval from the Beacon Hospital Research Ethics Committee (BEAO114 and BEAO151), and written informed consent was obtained from all participants.
before commencing the study. No financial compensation was provided to participants taking part.

**Digital Health Tool**

The DHT used in this study was designed using human-centered design steps and included behavior change techniques, as previously outlined in detail by Johnston et al [16]. Briefly, the system was designed to comprise a cross-platform (iOS or Android) mobile app capable of linking to a consumer activity tracker and smart scales, specifically, the Fitbit Charge 4 and Aria Air smart scales (information also available in [17]). The mobile app was broadly divided into five sections: (1) advice, (2) symptom reporting, (3) activity tracker and scale data (exercise, weight, heart rate, and sleep), (4) medication reminders, and (5) other vital sign tracking—all targeted through the inclusion of specific behavior change techniques (Figure 1) [16,17]. During the design of this DHT, an initial prototype was trialed with participants with HF for a period of 2 weeks, where positive feedback and adherence were seen [16]. The system was considered easy to use, positively affected their motivation to engage in key self-care behaviors, provided them with skills and perceived knowledge that made them more aware of the importance of self-care behaviors, positively influenced their confidence, and facilitated help seeking. After this, aspects of the system that needed to be improved were identified. These changes were implemented before deployment in this longer study. Specifically, the scaling in the app was adjusted to support larger fonts, the ability to input decimal points for vital signs was inputted, screens were not allowed to take a time-out while videos were playing, the information button was made more visible, technical issues surrounding daily data were addressed, and the ability to visualize within day heart rate data was implemented [16].

**Study Methods**

The recruited participants were invited to an initial setup session at the hospital (Figure 2). Demographic data such as age, sex, and the highest level of education were collected at the beginning of the session. The participants then completed the 9-item European HF Self-care Behavior Scale (EHFScBS) [18,19] and the Minnesota Living with HF Questionnaire (MLwHFQ) [20,21] to evaluate self-care behaviors in patients with HF and the effect of HF treatments on the quality of life. Following a setup and familiarization session with WJ (approximately 40 minutes), participants were asked to use the system as part of their usual daily routine for the following 6 months. Depending on their recruitment date, participants were using the DHT for a 6-month period between July 2021 and April 2022. During this period, patients were asked to wear the Fitbit Charge 4 activity tracker on their wrist, take their weight every morning using the Fitbit Aria Air scales, and interact with the developed mobile app. A check-in symptom questionnaire was completed once a month [16]. This same questionnaire was also triggered to be sent to participants once any of their monitored components (ie, heart rate, sleep, weight, or physical activity) changed by 2 SDs from their baseline level in the previous 7 days. In the event that a trigger occurred, the

Figure 1. Screenshots from the mobile app detailing the main menu, advice section, symptom report, and screens [16].
questionnaire was sent to participants as an alert in the app, informing them of a change and asking them to complete the questionnaire. The questionnaire was then sent, along with a trigger, to the cardiology team of the Beacon Hospital who would telephone the participant to determine whether any further medical action or intervention was required.

At the end of the 6-month period, individual semistructured interviews were completed over the phone and recorded with each participant. Open-ended questions were used to explore their perceptions of the acceptability, usability, and practicality of the DHT; understand their experiences pertaining to the impact of the DHT on their self-care behaviors; identify usability and user experience issues; and identify aspects that could improve the DHT (Multimedia Appendix 1). Participants also completed 3 usability questionnaires: System Usability Scale (SUS), a questionnaire designed to measure system usability [22]; Wearable Technology Motivation Scale (WTMS), a questionnaire based on the intrinsic needs listed within self-determination theory [23]; and the Comfort Rating Scale (CRS), a questionnaire designed to assess the comfort of wearable devices across the dimensions of emotion, attachment, harm, perceived change, movement, and anxiety [24]. They also repeated the EHFScBS and MLwHFQ to indicate whether a change in their behaviors or quality of life occurred.

After the completion of the study, semistructured interviews were also conducted with 2 members of the clinical team in Beacon Hospital Cardiology to explore their perceptions of the DHT (Multimedia Appendix 2).

Data Analysis
The recorded interviews were transcribed verbatim and anonymized. The coding of participants’ experiences with the DHT over the 6-month period was done using the Theoretical Domains Framework (TDF). The TDF was used within the development study to map the target behaviors of the tool; thus, it was used as a lens through which to view the usability, adherence, and potential behavior changes noted during this 6-month study. A deductive content analysis was undertaken whereby transcripts were coded according to the components of TDF [25,26] using a critical realist approach. This approach posits that a reality exists independent of our construction of it, while maintaining that our knowledge of it is interpretive, partial, and fallible [27]. In taking this approach, we recognize that our own experiences influence our insights but that we maintain our objective to view the situation as it occurs. To elaborate on our experiences, AK is a research physiotherapist with over 5 years of experience in digital health research and a PhD in behavior change. CB is undertaking a PhD in behavior change, with expertise in the use of the TDF and publications using the same. Together, these experiences influence and enrich their interpretation of the data.

AK coded all transcripts. First, they familiarized themselves with the data by reading and rereading all transcripts and generating initial notes on the data. Meaningful phrases were highlighted and assigned codes according to the domains of the TDF [28,29]. CB then acted as a critical friend to the coding, reviewing 20% of transcripts and providing critical feedback to improve the interpretation of findings and discussion of codes. How these barriers and facilitators align to each of the key skills of self-management was considered to identify areas for further development. Finally, given that only 2 HCPs were interviewed regarding their experiences, information from their transcripts was summarized narratively.

Questionnaire data were analyzed using SPSS Statistics for Mac (version 27; IBM Corp). The questionnaire data were scored using the appropriate standardized procedure for each
questionnaire. The MLwHFQ is scored by summing each of the components, resulting in a score ranging from 0 to 105 (whereby higher scores indicate higher impairment) [20]. The EHFScBS 9-item is scored by reversing the responses to the questionnaire and standardizing them [30]. This results in a score ranging from 0 to 100 (where a higher score indicates good self-care, and <30 is deemed as inadequate) [18]. Changes in results before and after the study were measured using a Wilcoxon signed rank test. The P value was set at .05; however, this was adjusted with a Bonferroni adjustment whereby P<.03 was significant. The SUS is scored out of 40 but converted to a 0-100 scale as per the standard procedure, with >68 deemed acceptable and >80 considered excellent [31]. Each item of the CRS is scored from 0 to 20 (whereby higher scores equate to worse comfort) [24]. The median of the 6-item questionnaire was calculated. Finally, the WTMS is scored by calculating the average score across the different components for each participant, resulting in a score ranging from 0 to 7 (whereby 7 indicates higher intrinsic motivation) [23]. In addition, adherence was determined by identifying the number of days a user wore the Fitbit device throughout the day and recorded their weight.

### Results

#### Participants

A total of 43 people were contacted by the clinicians in the Beacon Hospital to invite them to participate. Of these, 10 (23%) were not contactable, 1 (2%) did not satisfy the inclusion or exclusion criteria, 6 (14%) declined, and 7 (16%) did not use a smartphone. A total of 19 (44%) participants were recruited, of whom 17 completed the poststudy follow-up session (Table 1). One participant withdrew in the middle of the study because of increased health concerns, and 1 participant was unable to be contacted at the end of the study. Results regarding participants acceptability and changes to their self-care routines are based on the 17 who completed the entire study. Participants wore their devices for an average of 86.2% of the days in the testing period ranging from 40.6% to 98% (average of 157 days out of a possible 184 days). Furthermore, participants weighed themselves for an average of 73.7% of the potential testing days, ranging from 4.9% to 100% (average of 134 days out of a possible 184 days).

### Table 1. Participant demographics (N=17).

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>11 (65)</td>
</tr>
<tr>
<td>Female</td>
<td>6 (35)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD; minimum-maximum)</strong></td>
<td>72 (9.9; 54-81)</td>
</tr>
<tr>
<td><strong>BMI, mean (SD; minimum-maximum)</strong></td>
<td>28.3 (6.1; 19.3-40.4)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Did not complete second level</td>
<td>2 (12)</td>
</tr>
<tr>
<td>Completed second level</td>
<td>4 (24)</td>
</tr>
<tr>
<td>Third level education (any)</td>
<td>7 (41)</td>
</tr>
<tr>
<td>Not reported</td>
<td>4 (24)</td>
</tr>
</tbody>
</table>

#### Patient-Reported Outcomes

Although improvements in the EHFScBS were seen between baseline and the completion of the study, these changes were not significant (P>.03; Table 2, Figures 3 and 4). No participants were deemed to have inadequate self-care according to this scale. Similarly, although improvements were seen in the MLwHFQ results, these were not significant. Nonetheless, the median score suggests moderate quality of life, with 7 participants demonstrating good quality scores (41%) and 5 listing poor scores (29%) [32].

### Table 2. Changes in participants’ reported self-care of their heart failure (HF).

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Median prestudy results (IQR; minimum-maximum)</th>
<th>Median poststudy results (IQR; minimum-maximum)</th>
<th>Z score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-item European HF Self-care Behavior Scale (0-100)</td>
<td>52.8 (38.0-47.0; 22-89)</td>
<td>72.2 (41.0-52.0; 42-94)</td>
<td>−1.61</td>
<td>.11</td>
</tr>
<tr>
<td>Minnesota Living with HF Questionnaire (0-100)</td>
<td>31.0 (10.0-40.0; 0-77)</td>
<td>26.0 (15.0-53.0; 2-97)</td>
<td>−0.39</td>
<td>.70</td>
</tr>
</tbody>
</table>

With regard to the acceptability of the DHT, Table 3 lists the results from the SUS, WTMS, and CRS. The SUS score considered the Fitbit and mobile app as a whole system. Results suggest that the usability of this system was not acceptable to participants as the median score was 58.8 (55.0-60.0; 45.0-67.5). Indeed, no participant scored the system above 68, which is considered to be the threshold of acceptability. In contrast, participants demonstrated a strong motivation to use the system according to the WTMS (median 6.0, IQR 5.0-7.0; range 1.0-7.0), whereas the Fitbit was considered very comfortable as demonstrated by the low CRS results (median 0.0, IQR 0.0-0.0; range 0.0-2.0).
Figure 3. Pre- and post-results per participant for the European Heart Failure Self-care Behavior Scale.

Figure 4. Pre- and post-results per participant for the Minnesota Living with Heart Failure Questionnaire.
Table 3. Participant reported acceptability of the digital health tool.

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Median results (IQR; minimum-maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Usability Questionnaire (0-100)</td>
<td>58.8 (55-60; 45.0-67.5)</td>
</tr>
<tr>
<td>Wearable Technology Motivation Scale (0-7)</td>
<td>6.0 (5-7; 1.0-7.0)</td>
</tr>
<tr>
<td>Comfort Rating Scale (0-20)</td>
<td>0.0 (0.0-0.0; 0.0-2.0)</td>
</tr>
</tbody>
</table>

Participant Interviews

A total of 13 of the TDF domains were viewed as barriers or facilitators to the self-management of HF, whereas just 1 domain of the TDF did not match data (“optimism”). However, alongside “goals” this domain was not originally mapped to the key behaviors of self-management during the development of the mobile app [16]. A limited number of codes were noted for “goals” and “reinforcement,” suggesting that neither domain plays an important role in the facilitation of self-management in HF; therefore, they were excluded from the analysis. Of the remaining domains, 5 were generally viewed as facilitators of self-management (knowledge; social role and identity; memory, attention, and decision processes; social influences; and behavioral regulation). One (environmental context) was a barrier to self-management, whereas 5 were neither (skills, beliefs about capabilities, beliefs about consequences, intentions, and emotions; Figure 5).

Figure 5. Demonstrated impact of each Theoretical Domains Framework component on self-management of heart failure (green = facilitator, orange = unclear, red= barrier).

<table>
<thead>
<tr>
<th>Determinants of self-management behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDF domain</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
</tr>
<tr>
<td>Skills</td>
</tr>
<tr>
<td>Memory, attention and decision processes</td>
</tr>
<tr>
<td>Environmental context and resources</td>
</tr>
<tr>
<td>Beliefs about consequences</td>
</tr>
<tr>
<td>Beliefs about capabilities</td>
</tr>
<tr>
<td>Reinforcement</td>
</tr>
<tr>
<td>Social role &amp; identity</td>
</tr>
<tr>
<td>Emotion</td>
</tr>
<tr>
<td>Behavioral regulation</td>
</tr>
<tr>
<td>Social influences</td>
</tr>
</tbody>
</table>

Knowledge

Participants noted that the information provided in the mobile app was simple to understand and provided them with a greater general awareness of important elements to them and the management of their heart condition, specifically heart rate and weight. Indeed, people noted that, if anything, the information was too simple and they would have liked more personalized information, information about dietary requirements and blood pressure. Thus, overall, the knowledge provided by the app was considered useful in the management and awareness of their HF.

Well it was very simple to work. It kept me up to date. I kept an eye on my weight every morning which I usedn’t do. And then I would check the heart rate as well and of course what I found very helpful was the questionnaire that you sent on every month. [P10, male, 80]

From my own personal point of view, blood pressure, since I had open heart surgery, my blood pressure has been very low, so I suffer on a daily basis with light-headedness and all of that. So, I suppose I would have found it useful to know blood pressure readings. [P17, female, 67]

Well, like as I said, it wouldn’t give me what the problem was with me, like the blood flow through the heart, it wasn’t doing that. It’s just monitoring the heart. [P2, male, 74]

Memory, Attention, and Decision Processes

This domain is a complex element that included participants’ memory of completing self-management behaviors and their awareness and understanding of the data provided to them for
some, they were unclear as to what the green area within the graphs of the mobile app represented; thus, they paid little attention to it. Furthermore, participants remarked that they sometimes forgot to look at the app and thus were passively monitoring themselves. However, participants were positive about the fact that the app provided them with a way to become more aware of their daily patterns. This, they believed, supported their self-management as it helped them to know they were remaining consistent. Specifically, they were comfortable if they went outside of their normal range for 1 or 2 days but knew to contact someone if that became persistent. This suggests that people did not explicitly change their behavior, but simply being aware that everything was “normal” was sufficient for them to feel confident in themselves.

I knew that I would be contacted if there were discrepancies or going outside the zone, so I knew that, but I mean if for example my weight had started to increase, I would have been very aware of checking up on fluid retention and all that sort of thing, yeah. It was good to be able to see that, I wouldn’t, honestly, have been aware, unless my ankles, I’ve never had particular puffiness or anything like that or similar, so, but, I think there was a slight bit of that whenever I was in hospital in October, although it wasn’t obvious, it wasn’t visible but obviously it would have been obvious from a daily weight reading. [P17, female, 67]

**Behavioral Regulation**

The use of the digital tool facilitated participants to weigh themselves (daily, weekly, or biweekly). This was the most active behavior change that appeared to occur as a result of the system. Some reported being motivated to go for a short walk if they felt that they had not reached enough steps according to what their DHT was; however, overall, people reported not changing anything substantially. As previously mentioned, an awareness of their normal appeared to be the greatest benefit to them. This may be because other features such as monitoring their sleep and heart rate were mostly passive in nature. The components that required the most active engagements (ie, medication adherence and other methods of monitoring) were not used by any participant as they felt that they already had a method to track their medication that worked for them. Essentially, the digital tool appeared to integrate easily into their existing methods of regulating their HF, supporting their ability to monitor their key metrics without requiring too much additional effort on their part.

I’ll certainly keep the Fitbit up…wearing the Fitbit, I will keep that going definitely, so that I can keep an eye on things. I won’t say you can become obsessed with it but it gives you a good handle, I wouldn’t be checking my heart rate every couple of hours or anything like that but every couple of days I would do or if I did something strenuous I would just have a look and see how am I and what I am doing or anything else like that. [P3, male, 68]

**Social Influences**

Participants continue to rely on HCPs to lead the management of their condition and initiate topics of discussion. Furthermore, they were motivated to join the study as a result of their doctor asking if they were interested in it. Their perception was that if the doctor felt that it might be useful or interesting, then they were happy to participate. Despite this deference to HCPs, some felt that the information from the DHTs would empower them to talk to their doctors about their progress. However, the strongest element of support that they received from the tool was the comfort they gained in knowing they were “being watched.” Participants appreciated the calls that they received from HCPs if the system was triggered. Far from feeling intruded on, they instead felt supported and “minded” from afar, and had no privacy concerns regarding these triggers.

I have to say, I really appreciated it very, very much. I can’t really think of a particular negative. It really helped me and really reassured me because I had two episodes last year where I went into atrial fibrillation and you know I think it was good to know that, you know, if I did go off the baseline and someone contacted me, that they would then offer that I could speak to someone on the cardiac team, because sometimes it’s very difficult to access even your cardiologist, you know? Even getting past the secretary can be very difficult, so that was, yeah, I really liked that. I just felt that, you know, if something did go drastically wrong that someone was there picking up on it. [P17, female, 67]

I think it’s nice the idea that if something goes badly wrong that somebody rings up like a few times my heartbeat has changed and I’ve got phone calls to ask am I ok and trying to figure out why. [P4, male, 64]

**Social Role and Identity**

Participants perception of themselves and their own identity appeared to facilitate their self-management behaviors. This was closely linked to knowledge and memory, attention, and decision processes. Specifically, being aware of their data and their patterns of behavior sparked participants to reflect on “how lazy I am one day and how much I am doing the next day type of thing.” This suggests that they saw it as important to understand what they can do to help themselves, which was facilitated by the DHT. Furthermore, participants described being diligent about going to the doctor regularly and “doing everything they were saying I should do,” in order to help themselves manage their condition. Thus, they recognized the role that they had to play in their own condition, even if that is being led by the cardiologist.

**Skills**

Although skills were related to participants’ ability to self-manage, in this study, they were also required to interact with the digital tool in order to do so. Indeed, the biggest barrier to them being able to self-manage with technology was the technology itself. Specifically, for those less familiar with technology, it took them a while to settle into its use. Some
faced issues with syncing and connectivity, for which they required support from family members or members of the study team to overcome. However, as they got used to it, and as problems were solved, they then found that the DHT was an enabler to their self-management as it became one place for them to see all of their information.

Technology, you know it was always a bit beyond me a little bit. So, I thought it worked out fine … I became a little more comfortable with it. There were a few times alright….and again with the weighing scales, there was a problem and I made contact with one of your colleagues. And simple things like for example on the tablet, I tend to just … just press the on/off button until the screen goes out you see. [P16, male, 81]

Beliefs About Consequences

The key for this was whether participants believed that there was a link between their self-management behaviors and their HF. For example, 1 participant did not see the relevance of weighing themselves as they could not see how it impacted their HF. In contrast, others were using the app to help support their perception for how they were feeling. In general, people were able to look at their data and explain any discrepancies as a result of their recent behavior; thus, the system appeared to support building this belief. However, it was not always clear how this linked explicitly to their heart condition, as opposed to their overall health.

Well I’ll tell you once or twice it went down [their weight] and I said ‘Oh my God am I not well’ and then the next day is would come back up. Now I am only talking about a pound here or a pound there but if it did go up oh I would, I’d certainly have to watch, cut out maybe eating a bar of chocolate which I eat every so often or a sweet or whatever. I would be aware of it. [P8, female, 81]

Beliefs About Capabilities

Some participants felt they did not possess the skills to get the most out of the DHT. However, their uncertainty was also related to their ability to self-manage themselves in general. One person was wary about traveling in case something went wrong; another simply listed being unmotivated to self-manage. In contrast, others felt that the DHT improved their capabilities by providing them with the information to make informed choices and to speak to their doctor if needed.

I probably am not tech-savvy enough to have gotten everything I needed to get from both of those or everything they could give me, and I’m also going to, in the early part of the interview, put my hands up and say that I didn’t probably put enough effort into that. [P18, male, 68]

I feel good at the minute, I feel ok, and everything’s going well, and I’m going swimming, I’m going walking, I don’t feel any different up in the heart…..I wouldn’t have known all this stuff if I hadn’t of had the monitor on me and the whole lot, so I think with all this it’s good. [P14, male, 64]

Emotion

The biggest change in emotions reported by participants was the confidence they had in managing themselves as a result of the DHT, and the reassurance they received knowing that they were being monitored. A feeling of safety was reported as a result of this. Despite this, some participants reported moments of anxiety or alarm if the DHT sent them a trigger unexpectedly, or if their data were outside of their normal. For one person, it simply took them some time to realize that a certain amount of change is considered normal; thus, it ultimately led to them feeling reassured once they learnt this. Others though would feel guilty if their activity levels were low, or if they were not losing weight.

Well, I am not as fretful now as it was at the beginning and I do think that wearing the Fitbit has been part of that – knowing that there is something there and also the fact that I was contacted on several occasions that my baselines had changed in a few things. [P7, male 54]

It made me anxious because I looked back on a night like that and say well, I can’t see anything that that’s gone terribly astray here. I would like to have known what triggered it. [P9, female, 71]

Environmental Context and Resources

The focus of this domain was on factors within the environment that either supported or hindered participants’ self-management. In relation to the DHT, having to charge the Fitbit was a downside, but not a burden. Issues connecting the scales were a greater burden as it may result in incorrect readings and were difficult for participants to fix. Other elements that acted as barriers to self-management were typically related to life events, for example, other illness, bereavement, the weather, relaxing their diet while on holidays, and the area where they live. In addition, participants’ medication was noted as negatively impacting their sleep or weight. It was unclear whether this was medication specific to their HF or whether it was related to other comorbidities, but regardless, it influenced elements of their self-management behaviors that they were tracking.

I suppose it was a bad time of the year in the sense, by the time I’d get home from work in the evening, it was pitch dark. So, I wasn’t getting out for a walk and that. We only get kind of a half an hour lunch break. So, you didn’t even have time to exercise. [P15, female, 55]

Well, weather wise and then, you know, we had an awful lot to do after the funeral, you know… things had to be sorted, well they are not even sorted yet but anyway, you know, things had to be done. [P1, female, 81]

HCP Experiences

The 2 HCPs (1 consultant cardiologist and 1 physiotherapist) were broadly positive around the potential for this DHT to support HF management; however, they felt that this would require a number of changes both to the system itself and to how it is implemented before its utility is realized.
Both HCPs were surprised at the level of “hands-on” work required to manage the DHT and the triggers sent through it. Essentially, they expected that the trigger system would reduce the work of clinicians, whereas, in reality, it simply transferred some of the work to behind the scenes data management, or even increased their work. Specifically, the system of triggers required them to spend a number of hours each week calling participants and following up on those who did not answer. However, as the purpose of this study was to explore the participant perspective, the HCP did not monitor the number of triggers received or the number that resulted in a hospital visit. Nonetheless, it was felt that future iterations of this tool need to set different criteria for when a participant was called or not.

They felt that it was only feasible to recruit participants with greater health and digital literacy, and those who were based locally to support them in the initial setup of the DHT. However, neither of these elements were explicitly listed as inclusion or exclusion criteria within this study. Recruitment was undertaken using a purposive, pragmatic approach whereby participants were recruited from the existing patient list of cardiologists within the Beacon Hospital. The HCPs acknowledged that their perceptions of a person’s awareness of their condition and their ability to use technology were born out of their interactions with these patients and not from objectively measured assessments. In general, they did not approach patients who they considered to be more frail and older, who appeared to have greater anxieties around technology, and who were less aware of their condition.

Ultimately, it was felt that the DHT has great potential, but that it would require significant changes within the health care service if it were to be implemented into practice. Specifically, adjusting the trigger thresholds, altering the protocol for calling people, and ensuring that the HCPs have dedicated time in their schedules to manage these digital data are critical for its progress.

Discussion

Principal Findings

This 6-month observational study assessed the feasibility and utility of our human-centered designed DHT to evaluate user engagement and observationally evaluate whether it may affect end points such as quality of life and self-care behaviors. Similar to the results of our development paper [16], participants demonstrated high adherence to the DHT and reported increased awareness of their behaviors and increased confidence in themselves. Questionnaires demonstrated that it was a comfortable device and that their motivation to wear the Fitbit was high, although the usability of the DHT was considered poor. However, interviews suggest that this result may be linked to their low confidence regarding the use of technology. Participants were most impressed at the trigger function of the DHT, which provided them with a great sense of safety and comfort.

Comparison With Previous Research

Self-care refers to “performing the daily activities that serve to maintain or restore health and well-being, prevent illness, and manage chronic illness” [33] and includes the knowledge, skills, self-efficacy, and attitudes required to effectively manage signs and symptoms as they arise [10]. Similar to previous research, participants in this study reported that they take their medication as prescribed, have lower levels of physical activity than desired, and monitor their weight regularly but not daily [6]. Notably, their awareness of their sleep, weight, activity, and heart rate increased; however, they felt that they did not change anything substantially as the result of this information, while their EHFScBS and MLwHrQ results remain unchanged. This may be because the included participants were not considered adequate in their self-management at the start of the study; thus, it should be questioned whether they needed to change their behavior unless they began feeling unwell. Furthermore, although changes were not significant, this may be the result of the small sample size, and nonetheless, positive changes in both outcome measures were noted. However, it should also be considered whether no changes occurred because participants were not clear as to why they were assessing certain elements. For example, many seemed to feel that monitoring weight was necessary because they needed to lose it, not because they were monitoring water retention, a finding similar to previous research [34]. Furthermore, skill building requires more than information alone and instead should focus on deficits and managing unique situations as they arise [34]. It should be considered whether, by relying on a trigger system, we are removing this skill rather than building it. However, given the complexity of such decisions [34,35], and the relief provided to participants as a result of the triggers, further thought should be given as to how skills can be developed alongside the support provided by triggers.

When we mapped the participant experiences to the elements of the TDF, the domains of knowledge, social influences, and social identity appear to be particularly strong in facilitating self-management. Although the study was meant to improve participants’ own abilities and motivation to integrate behaviors into daily living, according to the results of “social influences,” it is possible that they took part in the study and remained so adherent because of their desire to follow doctors’ wishes. This brings up a persistent issue with self-management and the motivation to complete it. Theories such as self-determination theory posit that people will initiate and maintain a behavior if it supports intrinsic motivation [36]. Participants began to note that they were able to keep an eye on themselves and that it was up to them to manage what they could control, suggesting that there is certain amount of intrinsic motivation generated as a result of being able to monitor themselves, and this is further supported by the results of the WTMS. However, relying on doctors’ advice to participate and adhere to an intervention is externally motivated behavior that risks poor engagement over time [36]. There is a fine balance between following advice in a partnership and following instructions [37,38]. Ultimately, the most effective and valued element of the intervention was the link that participants had with their HCP. Future studies should consider how to train participants to lead conversations.
in this area with their clinicians, so that they are more empowered to conduct effective self-care, rather than simply rely on triggers being sent to them. Furthermore, the emotive elements of self-management appear to have played a strong role in participants’ experience of the study. Specifically, the benefits were confidence and reassurance, whereas guilt and anxiety also prevailed. Previous research has suggested that emotional reactions such as fear or anxiety, which tend to be viewed as maladaptive coping strategies, may also have a positive influence on self-care [39]. Many patients describe action-based strategies such as learning how to “pace” their activities or “listen to their bodies” to help optimize their ability to maintain physical activity [39]. Seeking guidance and advice from trusted sources is an integral component of self-care, and knowing when to ask for help is considered to be a tactical skill [34]. Thus, there is a need to also emphasize other domains such as skills and emotion further so that the identified barriers are targeted for reduction in a future clinical trial.

Important aspects regarding the feasibility of this system were shown in this study. A promising result was the acceptability of the system to participants. Specifically, they noted no issue with comfort, a high motivation to wear it. However, they also noted poor usability according to the SUS. This is likely to be the result of their concern regarding the system and their ability to use it. Nonetheless, they demonstrated high levels of adherence and engagement with the tool compared with other research in HF or the population in general where abandonment is high [40-44], and importantly, they felt great comfort and reassurance that someone was looking out for them. Additionally, almost 50% of those contacted to take part in the study agreed to participate in it, suggesting that there was interest and acceptability in the idea. However, because of the purposive sampling method, it is unclear whether people with lower levels of literacy or higher levels of disease severity will demonstrate the same levels of acceptability. Among a sample of patients with HF, 96% owned a mobile phone and 32% relied on the mobile phone for internet access, searched health information, and reported moderate self-confidence in using mobile apps [3]; thus, the DHT should be tested in a wider population. However, in the future, when expanding this tool for use within multiple settings, the impact on HCPs needs to be considered. Specifically, the requirement for HCPs to monitor the triggers themselves, and follow up with patients, was perceived as an increased burden. In the future, HCPs whose role is to manage patient-driven data insights will be needed, but this will require additional funding and training to be implemented successfully. The volume of data required to train such models effectively was a surprise to the HCPs who had expected the system to be “smarter” than it was, demonstrating that there is a need to set expectations with HCPs in any future feasibility studies. This study was not powered or intended to assess the predictive capacity of the trigger system. Indeed, if comparing it with the Medical Research Council guidelines, this was the test phase required before evaluation [35]. However for future studies, models need to be refined until the thresholds per person are identified. Adjusting the threshold to become more manageable to HCPs should be considered, but within the context that participants valued feeling safe.

Limitations
A key limitation of this study was that no objective assessment of participants’ behavior change (or lack thereof) was assessed. As a pilot study, the objective of this study was not to evaluate participants’ objective behavior change, and indeed, the study was not powered for that. Nonetheless, it limits our assessment of the potential for this DHT to patient perceptions only. Another limiting factor of this work is the recruitment methods employed, which are likely to have limited the generalizability of the findings with regard to representativeness of the sample. Specifically, purposive sampling was used whereby the HCPs contacted those on their list who they believed would be interested and capable of participating. Consequently, without objectively testing for health and digital literacy, the result of this is likely to be a bias toward people who were most likely to be able to use it. For DHTs to be evaluated fully, it must be offered to all, irrespective of their perceived suitability or not. Thus, future iterations of this tool should be trialed in as wide a range of participants as possible to test its ability to respond to patient needs and abilities. Finally, future studies need to assess feasibility more from the perspective of the health care service. Outcomes such as the number of trigger calls sent to the hospital, the time spent managing the data, the time spent by clinicians setting up the DHT with participants, and the number of visits to the clinic compared with the calls logged should be evaluated to better understand the impact of the DHT on the service.

Conclusions
The DHT demonstrated high levels of adherence and acceptance among participants. The DHT targeted key self-management behaviors and feelings of social support. However, a number of changes to the DHT, and the health service, are required before it can be implemented at scale. A full-scale feasibility trial conducted at a wider level is required to fully determine its potential effectiveness and wider implementation needs.

Acknowledgments
The authors would like to thank the participants of this study for their time and commitment to this body of work. This paper was produced as part of IVA5034 ECME research project and supported by the European Union’s INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB).

Data Availability
The data sets generated or analyzed during this study are not publicly available as we do not have ethical approval to share them. The anonymized questionnaires are available from the corresponding author on reasonable request.
Conflicts of Interest

None declared.

References


Abbreviations
- CRS: Comfort Rating Scale
- DHT: digital health tool
- EHFScBS: European Heart Failure Self-care Behavior Scale
- HCP: health care professional
- HF: heart failure
- MLwHFQ: Minnesota Living with Heart Failure Questionnaire
- SUS: System Usability Scale
- TDF: Theoretical Domains Framework
- WTMS: Wearable Technology Motivation Scale

© Alison Keogh, Carol Brennan, William Johnston, Jane Dickson, Stephen J Leslie, David Burke, Peter Megyesi, Brian Caulfield. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Pharmacy Students’ Attitudes Toward Distance Learning After the COVID-19 Pandemic: Cross-Sectional Study From Saudi Arabia

Saud Alsahali1*, PhD; Salman Almutairi1*, PharmD; Salem Almutairi1*, PharmD; Saleh Almofadhi1*, PharmD; Mohammed Anaami1*, PhD; Mohammed Alshammari1*, PhD; Suhaj Abdulsalim1*, PhD; Yasser Almogbel1*, PhD

Department of Pharmacy Practice, College of Pharmacy, Qassim University, Buraydah, Saudi Arabia

*all authors contributed equally

Corresponding Author:
Saud Alsahali, PhD
Department of Pharmacy Practice
College of Pharmacy, Qassim University
Madinah Road
Unaizah, Qassim
Buraydah, 56219
Saudi Arabia
Phone: 966 163011503
Email: s.alsahali@qu.edu.sa

Abstract

Background: Electronic learning refers to the use of assistive tools in offline and distance learning environments. It allows students to access learning tools and materials anytime and anywhere. However, distance learning courses depend on several factors that affect the quality of learning, which consequently affect students’ preferences in the settings and tools used to deliver educational materials.

Objective: This study aimed to evaluate students’ preferences for continuing distance learning after the pandemic and to assess the distance educational environment after the pandemic. It also aimed to identify the factors affecting distance learning and evaluate students’ preferences regarding modes of communication with instructors.

Methods: A web-based survey was used to conduct this cross-sectional study. The target participants of this study were students in the doctor of pharmacy program at Unaizah College of Pharmacy, Qassim, Saudi Arabia. All students enrolled from December 2022 to January 2023 received an invitation with a link to the web-based survey.

Results: The survey was completed by 141 students (58 female students and 83 male students). The research results showed that most students (102/141, 72.3%) did not wish to continue distance education for laboratory courses, and 60.3% (85/141) did not wish to continue taking distance team-based learning after the pandemic. Additionally, 83.7% (118/141) of the students indicated that distance courses were simple. More than half of the participants (79/141, 56%) stated that having a camera on during class negatively impacted their learning, and only 29.1% (41/141) of the students stated that nonvisual communication with their fellow students impacted their learning. A large proportion of students (83/141, 58.9%) reported impairment of social engagement on campus, 44% (62/141) in-person interactions during classes, and 73.7% (104/141) were relieved that their classes were not disrupted.

Conclusions: Similar to all types of education, distance learning is characterized by advantages and disadvantages, as reported by students. Students felt that the course material was intelligible, and the distance course was uncomplicated. Moreover, they expressed relief that their studies were not disrupted. However, they also reported the loss of face-to-face contact during courses as the most significant drawback of distance learning versus face-to-face learning, followed by a lack of social connection on campus.

(JMIR Form Res 2024;8:e54500) doi:10.2196/54500

KEYWORDS

distance learning; e-learning; pharmacy education; team based learning; educational experience.
Introduction

In December 2019, unexpected cases of severe respiratory illness were found in Wuhan, the capital city of China’s Hubei province [1]. The World Health Organization recognized a new coronavirus, termed SARS-CoV-2, which caused the outbreak of COVID-19 [2]. The initial COVID-19 instance in Saudi Arabia was detected on March 3, 2020. Consequently, Saudi Arabia placed most of its public and private facilities under lockdown and instituted statewide limits on population movement. Due to the widespread mitigation efforts implemented by several nations, the COVID-19 pandemic disrupted the daily lives of millions of people [3]. On March 8, 2020, the Saudi Arabian government announced a conversion from regular classroom education to distance education. In the following days, all Saudi Arabian colleges and institutions implemented this approach. Saudi Arabia was among the first nations in the Middle East to enforce full quarantine and the urgent switch to distance learning [4].

The use of assistive tools in both offline and distance learning environments is referred to as e-learning [5] and includes synchronous and asynchronous modes. Synchronous distance learning permits live interactions between educators and students through chats and video conferences, while asynchronous distance learning is based on emails and recorded videos [6]. Distance learning can be used in place of, or in addition to, traditional education [5]. Before the pandemic, some colleges had used distance education using learning management systems such as Blackboard (Anthology Inc) and Zoom (Zoom Video Communications, Qumu Corporation). Distance education delivery abruptly and entirely replaced traditional classroom education, resulting in problems such as incomplete examinations and difficulties with courses requiring hands-on training or laboratory activities [4]. After the COVID-19 outbreak forced the closure of universities in Saudi Arabia, the Saudi Ministry of Education created an alternative policy to support distance education and provided public universities with resources to aid distance learning and virtual classrooms. This was done to prevent interruptions in the educational process. Blackboard is a web-based distance learning tool used by universities and other higher education institutions that enables students to continue their studies [7]. It is one of the most widely used web-based platforms and is currently used by all Saudi Arabian universities [8].

Most telecommunications firms in Saudi Arabia (eg, Saudi Telecom Company, Mobily, and Zain Saudi Arabia) aided in efficiently delivering distance learning and health care during the pandemic. These companies offered free data services to the most widely used telehealth and health applications, as well as educational platforms [3]. In addition, the Saudi Ministry of Education introduced distance learning programs to safely and effectively continue the educational process. Within days, all universities, including medical schools, switched to distance mode of education delivery. Medical schools altered their teaching approach as a result of this tremendous, unforeseen shift from traditional learning to an exclusively distance learning environment [9].

The components of distance learning delivery include ease of information access, material distribution, updating, and standardization. Distance learning allows for rapid and easy content revision to achieve learning objectives [10]. Furthermore, it provides several opportunities for students to improve their autonomous learning and enables students to access learning through distance learning tools and materials anytime and anywhere. Moreover, distance education provides excellent opportunities for students to enhance their connection with instructors [11]. Nevertheless, the appropriate conditions are necessary for the creation of a productive learning atmosphere. In addition, the development of technical infrastructure, the purchase and maintenance of tools and equipment, the instruction of human resources, and the production of distance content are costly and time-consuming processes. The institution, administrators, teachers, and students require information technology resources and skills. Poor time management and a lack of self-control have been reported as the 2 major factors affecting students’ perception of distance learning [12].

A lack of in-person, face-to-face interactions in the classroom or workplace characterizes distance learning classes. Interpersonal communication is vital for numerous students and organizations. For instance, in-person interaction with lecturers and group discussions are crucial motivational activities and learning techniques [6]. Team-based learning (TBL) is a small-group, active learning method that incorporates rapid feedback, teamwork, and solo work to help students apply conceptual knowledge [13]. These activities significantly affect students’ class behavior and comprehension of lectures. However, the effectiveness of distance learning courses depends on individual factors, such as the student’s home setting, socioeconomic factors, and the educational level of their parents [6].

Most universities effectively transitioned their teaching operations from traditional to distance learning in response to the pandemic. However, several problems emerged, including poor teaching quality, work overload, lack of accessibility, student nervousness and dissatisfaction, fundamental questions concerning the use of distance education in the future, the requirement for suitable educational staff support and education to handle the inescapable adjustments required for implementing contemporary educational methods in a distance learning setting, and appropriate teacher preparation. Students expressed concern about the ability of distance education to replace human interaction in terms of emotion [14].

COVID-19 affected the reception and delivery of education at all levels [15], with educational authorities in most countries implementing distance learning strategies [11]. However, the sudden conversion to distance learning affected the performance of students and teachers in all learning environments, potentially influencing the efficacy of distance learning. Many academic staff members had no previous experience teaching using distance learning. Students may have been more anxious than usual due to worries regarding infection, timely graduation, finances, and employment, which could have had a detrimental effect on their academic achievement [16]. Synchronized technology enables teachers and pupils to interact “live.”
Particularly under uncertain global conditions, such as pandemics, distance learning can result in more efficient and straightforward access to knowledge [9]. Distance education can also serve as a tool to mitigate social isolation and assist in outbreak control. It also provides a flexibility that allows for learning regardless of location or time, which is essential for education during periods of crisis. In addition, the success of the technical infrastructure at schools and colleges depends on adequate maintenance [17]. This study aims to assess pharmacy students’ perceptions of distance learning and to evaluate the postpandemic distance learning environment. We also try to identify the factors affecting distance learning and to assess the emotions experienced by students during their period of distance education.

Methods

Study Design and Setting

A web-based survey was used to conduct this cross-sectional study. This study’s intended participants were students enrolled in the doctor of pharmacy (PharmD) program at Unaizah College of Pharmacy, Qassim University, Saudi Arabia, during the 2022-2023 academic year. PharmD students and interns who completed 1 year of distance learning during the COVID-19 pandemic using distance learning platforms provided by Qassim University were eligible for this study. The research did not include students who were attending other colleges. In addition, we excluded students who had not completed 1 year of distance learning at the university, similar to first-year students in the PharmD program, since they were not enrolled in the university during the COVID-19 pandemic.

Sample Size

The target population in this study comprised male and female third-year to sixth-year (internship year) PharmD students from Unaizah College of Pharmacy, Qassim University. According to the college’s students’ academic affairs office, the number of enrolled students was 327. Raosoft (Sample Size Calculator; Raosoft, Inc) was used to compute the sample size. With a 50% response rate, a 5% margin of error, and a CI level of 95%, the sample size should not be fewer than 141 students [18].

Study Tool

The questionnaire used in this study was created based on a prior study by Kedraka et al [14], and the authors obtained permission to use the questionnaire. A survey composed of multiple-choice questions was created using a web-based survey creator. That survey consisted of 34 questions assessing attitudes toward distance learning [14]. We slightly modified some questions to comply with the teaching methods used at Unaizah College of Pharmacy. Subject experts validated this modified questionnaire for clarity, understandability, and applicability. The final survey had 6 sections. The first section included 5 questions to assess the necessary abilities of students in meeting learning objectives and measuring progress. Student preferences also play a crucial role in the learning process. The second section comprised 9 questions that could serve as recommendations for developing a distance learning environment and measuring student performance in distance classrooms. Performance is greatly influenced by technical knowledge, comfort level, and attitudes toward using the internet for learning. Using a 5-point scale (1: not at all to 5: very much), we evaluated the opinions of the participants regarding the effectiveness of remote learning in terms of meeting the needs of the course, satisfaction with communication with the instructor, level of interest in the new method of teaching, level of course participation, interactions between students and teachers, and interactions between students. These factors may provide insight into the perceptions of students regarding the distance learning environment.

The third section, which included 4 questions, aimed to assess remote learning quality by identifying influential factors and motivating elements. The fourth section included 4 questions used to assess successful communication between students and teachers in distance courses, recommend steps to enhance interaction, and lessen the isolation associated with distance courses. The fifth section consisted of 6 questions intended to identify elements lost due to the shift from face-to-face interaction to distance learning, improve course delivery, and understand the factors that constitute a successful distance learning experience. Distance learning requires a critical understanding of the effects of interactions on student engagement. The sixth section consisted of 6 questions focusing on the motives and sentiments of the students toward their participation in the educational program. All eligible students were asked to participate in this study, and the questionnaire was disseminated web-based using WhatsApp (WhatsApp LLC).

Ethical Considerations

The ethics permission for this study was granted by Qassim University’s Health Research Ethics Committee in Saudi Arabia (22-17-11). The participation in this study was voluntary and informed consent to take part in this study was obtained at the beginning of the survey and the statements on the anonymity of responses or confidentiality of data were maintained throughout this study. No compensation or reward was offered to this study’s participants.

Statistical Analysis

The data were analyzed and student replies were compiled using SPSS (version 20.0; IBM Corp). The survey results were compiled using descriptive statistics comprising frequencies and percentages. Further reported were the mean scores for male and female students.

Results

Demographic Data

The survey was completed by 141 students (female: n=58, 41%; male: n=83, 58.9%) from the Unaizah College of Pharmacy. The greatest number of participants were recruited from the fifth year (51/141, 36.2%), followed by the third year (36/141, 25.5%), fourth year (30/141, 21.3 %), and sixth year (24/141, 17%).
Preference for Continuing Distance Education After the Pandemic

As shown in Table 1, most of the students (102/141, 72.3%) were unwilling to continue distance education (male: 50/83, 60.2%; female: 52/58, 89.6%; not at all or a little) after the pandemic for courses that included laboratory activities. In contrast, the majority of the students, 68% (96/141), had a favorable view regarding the continuation of distance education for elective courses (male: 53/83, 63.8%; female: 43/58, 74.1%; much or very much). Regarding TBL examinations, the majority of students, 60.3% (85/141), did not wish to continue taking distance examinations after the pandemic (male: 38/83, 45.8%; female: 47/58, 81%).

Regarding lectures, less than half of the students (63/141, 44.6%) did not wish to continue taking the lectures via distance platforms (male: 36/83, 43.4%; female: 27/58, 46.5%). Approximately 46.8% (66/141) of students wished to continue receiving their education distance for seminars (male: 41/83, 49.4%; female: 25/58, 43.1%).

Table 1. Pharmacy students’ preference for continuing distance education after the pandemic.

<table>
<thead>
<tr>
<th>Type of course and sex</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Quite, n (%)</th>
<th>Much, n (%)</th>
<th>Very much, n (%)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>20 (24)</td>
<td>16 (19.3)</td>
<td>11 (13.3)</td>
<td>16 (19.3)</td>
<td>20 (24.1)</td>
<td>3.0 (1.5)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>10 (17.2)</td>
<td>17 (29.3)</td>
<td>8 (13.9)</td>
<td>14 (24.1)</td>
<td>9 (15.5)</td>
<td>2.9 (1.4)</td>
</tr>
<tr>
<td>Elective courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>11 (13.3)</td>
<td>7 (8.4)</td>
<td>12 (14.4)</td>
<td>13 (15.7)</td>
<td>40 (48.2)</td>
<td>3.8 (1.5)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>4 (6.9)</td>
<td>3 (5.2)</td>
<td>8 (13.8)</td>
<td>14 (24.1)</td>
<td>29 (50)</td>
<td>4.1 (1.2)</td>
</tr>
<tr>
<td>Laboratory courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>41 (49.4)</td>
<td>9 (10.8)</td>
<td>8 (9.7)</td>
<td>6 (7.2)</td>
<td>19 (22.9)</td>
<td>2.4 (1.7)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>40 (69)</td>
<td>12 (20.7)</td>
<td>2 (3.4)</td>
<td>2 (3.4)</td>
<td>2 (3.4)</td>
<td>1.5 (1.0)</td>
</tr>
<tr>
<td>Team-based learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>28 (33.7)</td>
<td>10 (12)</td>
<td>13 (15.7)</td>
<td>14 (16.9)</td>
<td>18 (21.7)</td>
<td>2.8 (1.6)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>33 (56.9)</td>
<td>14 (24.1)</td>
<td>4 (7)</td>
<td>2 (3.4)</td>
<td>5 (8.6)</td>
<td>1.8 (1.2)</td>
</tr>
<tr>
<td>Seminars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>19 (22.9)</td>
<td>6 (7.2)</td>
<td>17 (20.5)</td>
<td>6 (7.2)</td>
<td>35 (42.2)</td>
<td>3 (1.6)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>20 (34.5)</td>
<td>8 (13.8)</td>
<td>5 (8.6)</td>
<td>12 (20.7)</td>
<td>13 (22.4)</td>
<td>2.8 (1.6)</td>
</tr>
</tbody>
</table>

Assessment of the Distance Educational Environment

As shown in Table 2, most participants had positive opinions regarding the former’s educational potential of the distance versus the conventional learning environment. Male students exhibited a more favorable attitude (higher average) compared with female students in all statements that assessed the distance learning environment compared to conventional learning on campus. Using a 5-point scale (1: not at all to 5: very much), we evaluated the opinions of the participants regarding the effectiveness of remote learning in terms of meeting the needs of the course (male: 3.5; female: 2.9), satisfaction with communication with the instructor (male: 3.6; female: 3.1), level of interest in the new method of teaching (male: 3.5; female: 2.6), level of course participation (male: 3; female: 2.9), interactions between students and teachers (male: 3.2; female: 2.7), and interactions between students (male: 3; female: 2.8).

Less than half of the students (61/141, 43.3%; male: 39/83, 47%; female: 22/58, 37.9%) believed that the course content was understandable. Additionally, 83.7% (118/141) of the students indicated that taking the course online was simple (male: 70/83, 84.3%; female: 48/58, 82.7%; much or very much). As for developing new skills, 38.3% (54/141) of the students thought that they developed new skills related to distance education (male: 40/83, 48.2%; female: 14/58, 24%).
Table 2. Pharmacy students’ attitudes about the remote learning environment.

<table>
<thead>
<tr>
<th>Assessment and sex</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Quite, n (%)</th>
<th>Much, n (%)</th>
<th>Very much, n (%)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The content of the course is comprehensible</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>5 (6)</td>
<td>7 (8.4)</td>
<td>32 (38.6)</td>
<td>20 (24.1)</td>
<td>19 (22.9)</td>
<td>3.5 (1.1)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>6 (10.3)</td>
<td>11 (19)</td>
<td>19 (32.8)</td>
<td>14 (24.1)</td>
<td>8 (13.8)</td>
<td>3.1 (1.2)</td>
</tr>
<tr>
<td><strong>New skills related to distance education are being developed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>6 (7.2)</td>
<td>12 (14.5)</td>
<td>25 (30.1)</td>
<td>23 (27.7)</td>
<td>17 (20.5)</td>
<td>3.4 (1.2)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>13 (22.4)</td>
<td>11 (19)</td>
<td>20 (34.5)</td>
<td>12 (20.7)</td>
<td>2 (3.4)</td>
<td>2.6 (1.2)</td>
</tr>
<tr>
<td><strong>The teaching method of distance learning covers the prerequisites of the course</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>10 (12)</td>
<td>7 (8.4)</td>
<td>19 (23)</td>
<td>26 (31.3)</td>
<td>21 (25.3)</td>
<td>3.5 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>13 (22.4)</td>
<td>11 (19)</td>
<td>9 (15.5)</td>
<td>18 (31)</td>
<td>7 (12.1)</td>
<td>2.9 (1.4)</td>
</tr>
<tr>
<td><strong>The new mode of teaching is interesting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>8 (9.6)</td>
<td>9 (10.8)</td>
<td>20 (24.2)</td>
<td>24 (28.9)</td>
<td>22 (26.5)</td>
<td>3.5 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>16 (27.6)</td>
<td>10 (17.2)</td>
<td>15 (25.9)</td>
<td>14 (24.1)</td>
<td>3 (5.2)</td>
<td>2.6 (1.3)</td>
</tr>
<tr>
<td><strong>Communication with the teacher is satisfactory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>6 (7.2)</td>
<td>11 (13.3)</td>
<td>18 (21.7)</td>
<td>22 (26.5)</td>
<td>26 (31.3)</td>
<td>3.6 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>9 (15.5)</td>
<td>13 (22.4)</td>
<td>8 (13.8)</td>
<td>19 (32.8)</td>
<td>9 (15.5)</td>
<td>3.1 (1.3)</td>
</tr>
<tr>
<td><strong>Participation in class is great</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>8 (9.6)</td>
<td>16 (19.3)</td>
<td>18 (21.7)</td>
<td>18 (21.7)</td>
<td>23 (27.7)</td>
<td>3.4 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>13 (22.4)</td>
<td>13 (22.4)</td>
<td>10 (17.3)</td>
<td>13 (22.4)</td>
<td>9 (15.5)</td>
<td>2.9 (1.4)</td>
</tr>
<tr>
<td><strong>Attendance is easy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>5 (6)</td>
<td>3 (3.6)</td>
<td>5 (6.1)</td>
<td>19 (22.9)</td>
<td>51 (61.4)</td>
<td>4.3 (1.1)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>1 (1.7)</td>
<td>3 (5.2)</td>
<td>6 (10.4)</td>
<td>18 (31)</td>
<td>30 (51.7)</td>
<td>4.3 (1.0)</td>
</tr>
<tr>
<td><strong>Interaction between teacher and students is great</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>13 (15.7)</td>
<td>9 (10.8)</td>
<td>29 (34.9)</td>
<td>13 (15.7)</td>
<td>19 (22.9)</td>
<td>3.2 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>16 (27.6)</td>
<td>10 (17.2)</td>
<td>16 (27.6)</td>
<td>9 (15.5)</td>
<td>7 (12.1)</td>
<td>2.7 (1.4)</td>
</tr>
<tr>
<td><strong>Interaction among students is great</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>17 (20.5)</td>
<td>9 (10.8)</td>
<td>10 (12.1)</td>
<td>19 (22.9)</td>
<td>28 (33.7)</td>
<td>3.4 (1.5)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>13 (22.4)</td>
<td>13 (22.4)</td>
<td>12 (20.7)</td>
<td>13 (22.4)</td>
<td>7 (12.1)</td>
<td>2.8 (1.3)</td>
</tr>
</tbody>
</table>

Factors Affecting e-Learning

As shown in Table 3, more than half of the participants (79/141, 56%) reported that having the camera on during class impacted their learning (male: 49/83, 59%; female: 30/58, 51.7%). Further, only 29.1% (41/141) of the students indicated that nonvisual communication with their fellow students would impact their learning (male: 22/83, 26.5%; female: 19/58, 32.7%). About 29.7% (42/141) of the students reported that the teacher’s insufficient knowledge concerning the handling of the platform impacted distance learning (male: 30/83, 36.1%; female: 12/58, 21%), while 44.6% (63/141) indicated that the inability to cooperate with their fellow students would affect distance learning (male: 44/83, 53%; female: 19/58, 32.7%).
Table 3. Factors affecting e-learning.

<table>
<thead>
<tr>
<th>Factors and sex</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Quite, n (%)</th>
<th>Much, n (%)</th>
<th>Very much, n (%)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The teacher’s camera should be turned on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>12 (14.5)</td>
<td>9 (10.8)</td>
<td>13 (15.7)</td>
<td>18 (21.7)</td>
<td>31 (37.3)</td>
<td>3.56 (1.4)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>1 (1.7)</td>
<td>12 (20.7)</td>
<td>15 (25.9)</td>
<td>18 (31)</td>
<td>12 (20.7)</td>
<td>3.48 (1.1)</td>
</tr>
<tr>
<td>Teacher’s deficiencies in knowledge of handling the platform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>15 (18.1)</td>
<td>11 (13.3)</td>
<td>27 (32.4)</td>
<td>14 (16.9)</td>
<td>16 (19.3)</td>
<td>3.06 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>8 (13.8)</td>
<td>20 (34.5)</td>
<td>18 (31)</td>
<td>7 (12.1)</td>
<td>5 (8.6)</td>
<td>2.67 (1.1)</td>
</tr>
<tr>
<td>Nonvisual communication with my fellow students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>24 (28.9)</td>
<td>12 (14.5)</td>
<td>25 (30.1)</td>
<td>9 (10.8)</td>
<td>13 (15.7)</td>
<td>2.69 (1.4)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>10 (17.2)</td>
<td>18 (31)</td>
<td>11 (19.1)</td>
<td>9 (15.5)</td>
<td>10 (17.2)</td>
<td>2.84 (1.4)</td>
</tr>
<tr>
<td>Noncooperation with my fellow students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>9 (10.8)</td>
<td>9 (10.8)</td>
<td>21 (25.4)</td>
<td>22 (26.5)</td>
<td>22 (26.5)</td>
<td>3.46 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>17 (29.3)</td>
<td>9 (15.5)</td>
<td>13 (22.4)</td>
<td>12 (20.7)</td>
<td>7 (12.1)</td>
<td>2.70 (1.4)</td>
</tr>
</tbody>
</table>

Preferred Modes of Communication With Instructors

As shown in Table 4, the participants’ preferred method of communication was the course chat (96/141, 68%; male: 59/83, 71.1%; female: 37/58, 63.8%). The second preferred method of communication for male students was a microphone (47/83, 56.6%), followed by face-to-face communication in the classroom (47/83, 56.6%) and groups on Zoom (37/83, 44.6%) throughout the course. The second preferred method of communication for female students was face-to-face communication in the classroom (36/58, 62%), followed by groups on Blackboard or Zoom (26/58, 44.8%). The least preferred option was a microphone (25/58, 43.1%).

Table 4. Pharmacy students’ preferred mode of communication with the instructor.

<table>
<thead>
<tr>
<th>Ways of communication and sex</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Quite, n (%)</th>
<th>Much, n (%)</th>
<th>Very much, n (%)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>On chat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>5 (6)</td>
<td>5 (6)</td>
<td>14 (16.9)</td>
<td>26 (31.3)</td>
<td>33 (39.8)</td>
<td>3.9 (1.2)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>9 (15.5)</td>
<td>8 (13.8)</td>
<td>4 (6.9)</td>
<td>22 (37.9)</td>
<td>15 (25.9)</td>
<td>3.4 (1.4)</td>
</tr>
<tr>
<td>Speaking on the internet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>6 (7.2)</td>
<td>9 (10.8)</td>
<td>21 (25.4)</td>
<td>22 (26.5)</td>
<td>25 (30.1)</td>
<td>3.6 (1.2)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>8 (13.8)</td>
<td>9 (15.5)</td>
<td>16 (27.6)</td>
<td>17 (29.3)</td>
<td>8 (13.8)</td>
<td>3.1 (1.2)</td>
</tr>
<tr>
<td>Internet-based room</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>8 (9.6)</td>
<td>12 (14.5)</td>
<td>26 (31.3)</td>
<td>14 (16.9)</td>
<td>23 (27.7)</td>
<td>3.4 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>8 (13.8)</td>
<td>8 (13.8)</td>
<td>16 (27.6)</td>
<td>17 (29.3)</td>
<td>9 (15.5)</td>
<td>3.2 (1.3)</td>
</tr>
<tr>
<td>In the classroom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>9 (10.8)</td>
<td>9 (10.8)</td>
<td>18 (21.8)</td>
<td>22 (26.5)</td>
<td>25 (30.1)</td>
<td>3.5 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>4 (6.9)</td>
<td>9 (15.5)</td>
<td>9 (15.6)</td>
<td>22 (37.9)</td>
<td>14 (24.1)</td>
<td>3.6 (1.2)</td>
</tr>
</tbody>
</table>

Traditional Education Components not Present in Distance Learning

As shown in Table 5, a large proportion of students, 58.9% (83/141), reported impaired social engagement around the campus (male: 45/83, 54.2%; female: 38/58, 65.5%), while 45.4% (64/141) of the students reported that they missed their classmates (male: 36/83, 43%; female: 28/58, 48.3%), in-person interactions during classes (62/141, 44%; male: 33/83, 39.8%; female: 29/58, 50%), and interactions in the classroom (60/141, 42.5%; male: 31/83, 37.3%; female: 29/58, 50%). Only 31.2% (44/141) of students stated that they missed their professors (male: 27/83, 32.5%; female: 17/58, 29.3%), while 37.5% (53/141) of the students stated that they missed being able to visit the library (male: 21/83, 25.3%; female: 32/58, 55.2%).
Table 5. Components of traditional learning that are currently missing from remote education.

<table>
<thead>
<tr>
<th>Element and sex</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Quite, n (%)</th>
<th>Much, n (%)</th>
<th>Very much, n (%)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>14 (16.9)</td>
<td>11 (13.3)</td>
<td>31 (37.3)</td>
<td>17 (20.5)</td>
<td>10 (12)</td>
<td>3.0 (1.2)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>9 (15.5)</td>
<td>19 (32.8)</td>
<td>13 (22.4)</td>
<td>15 (25.9)</td>
<td>2 (3.4)</td>
<td>2.7 (1.1)</td>
</tr>
<tr>
<td>Fellow students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>10 (12)</td>
<td>7 (8.4)</td>
<td>30 (36.2)</td>
<td>16 (19.3)</td>
<td>20 (24.1)</td>
<td>3.3 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>6 (10.3)</td>
<td>16 (27.6)</td>
<td>8 (13.8)</td>
<td>15 (25.9)</td>
<td>13 (22.4)</td>
<td>3.2 (1.4)</td>
</tr>
<tr>
<td>Library</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>30 (36.1)</td>
<td>10 (12)</td>
<td>22 (26.6)</td>
<td>11 (13.3)</td>
<td>10 (12)</td>
<td>2.5 (1.4)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>7 (12.1)</td>
<td>10 (17.2)</td>
<td>9 (15.5)</td>
<td>15 (25.9)</td>
<td>17 (29.3)</td>
<td>3.4 (1.4)</td>
</tr>
<tr>
<td>Classrooms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>18 (21.7)</td>
<td>12 (14.5)</td>
<td>22 (26.4)</td>
<td>19 (22.9)</td>
<td>12 (14.5)</td>
<td>2.9 (1.4)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>9 (15.5)</td>
<td>11 (19)</td>
<td>9 (15.5)</td>
<td>10 (17.2)</td>
<td>19 (32.8)</td>
<td>3.3 (1.5)</td>
</tr>
<tr>
<td>In-person communication during lessons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>10 (12)</td>
<td>12 (14.5)</td>
<td>28 (33.7)</td>
<td>17 (20.5)</td>
<td>16 (19.3)</td>
<td>3.2 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>10 (17.2)</td>
<td>5 (8.6)</td>
<td>14 (24.2)</td>
<td>14 (24.1)</td>
<td>15 (25.9)</td>
<td>3.3 (1.4)</td>
</tr>
<tr>
<td>Social interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>9 (10.8)</td>
<td>4 (4.8)</td>
<td>25 (30.2)</td>
<td>22 (26.5)</td>
<td>23 (27.7)</td>
<td>3.6 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>4 (6.9)</td>
<td>7 (12.1)</td>
<td>9 (15.5)</td>
<td>17 (29.3)</td>
<td>21 (36.2)</td>
<td>3.8 (1.3)</td>
</tr>
</tbody>
</table>

Emotions Experienced During Distance Education

As shown in Table 6, a significant majority of the students, 73.7% (104/141), expressed relief that their classes had not been disrupted (male: 59/83, 71.1%; female: 45/58, 77.6%). However, there was still much interest (74/141, 52.5%) in how the studies would go (male: 50/83, 60.2%; female: 24/58, 41.3%), and much enthusiasm among students for their first distance learning experiences (82/141, 58.1%; male: 50/83, 60.2%; female: 32/58, 55.2%). Additionally, the majority of students, 80.1% (113/141; male: 67/83, 80.7%; female: 46/58, 79.3%), liked not having to go to their lessons, whereas only 40.4% (57/141) of students were satisfied that the distance courses would continue (male: 34/83, 40.9%; female: 23/58, 39.6%).
Table 6. Emotions experienced by students during distance education.

<table>
<thead>
<tr>
<th>Emotion and sex</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Quite, n (%)</th>
<th>Much, n (%)</th>
<th>Very much, n (%)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy at classes not being held</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>14.0 (16.9)</td>
<td>9.0 (10.8)</td>
<td>26.0 (31.3)</td>
<td>19.0 (22.9)</td>
<td>15.0 (18.1)</td>
<td>3.1 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>9.0 (15.5)</td>
<td>9.0 (15.5)</td>
<td>17.0 (29.3)</td>
<td>12.0 (20.7)</td>
<td>11.0 (19)</td>
<td>3.1 (1.3)</td>
</tr>
<tr>
<td>Pleasure at not having to commute to attend classes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>2.0 (2.4)</td>
<td>3.0 (3.6)</td>
<td>11.0 (13.3)</td>
<td>17.0 (20.5)</td>
<td>50.0 (60.2)</td>
<td>4.3 (1.0)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>2.0 (3.4)</td>
<td>5.0 (8.6)</td>
<td>5.0 (8.7)</td>
<td>12.0 (20.7)</td>
<td>34.0 (58.6)</td>
<td>4.2 (1.1)</td>
</tr>
<tr>
<td>Relief at not losing the semester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>1.0 (1.2)</td>
<td>5.0 (6)</td>
<td>18.0 (21.7)</td>
<td>21.0 (25.3)</td>
<td>38.0 (45.8)</td>
<td>4.1 (1.0)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>2.0 (3.4)</td>
<td>4.0 (6.9)</td>
<td>7.0 (12.2)</td>
<td>14.0 (24.1)</td>
<td>31.0 (53.4)</td>
<td>4.2 (1.1)</td>
</tr>
<tr>
<td>Enthusiasm for the new experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>7.0 (8.4)</td>
<td>6.0 (7.2)</td>
<td>20.0 (24.2)</td>
<td>19.0 (22.9)</td>
<td>31.0 (37.3)</td>
<td>3.7 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>8.0 (13.8)</td>
<td>11.0 (19)</td>
<td>7.0 (12)</td>
<td>12.0 (20.7)</td>
<td>20.0 (34.5)</td>
<td>3.4 (1.5)</td>
</tr>
<tr>
<td>Disappointment because the new educational environment does not work for me</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>15.0 (18.1)</td>
<td>6.0 (7.2)</td>
<td>30.0 (36.1)</td>
<td>15.0 (18.1)</td>
<td>17.0 (20.5)</td>
<td>3.2 (1.3)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>9.0 (15.5)</td>
<td>9.0 (15.5)</td>
<td>16.0 (27.6)</td>
<td>15.0 (25.9)</td>
<td>9.0 (15.5)</td>
<td>3.1 (1.3)</td>
</tr>
<tr>
<td>Curiosity regarding the mood of studies in the future</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=83)</td>
<td>1.0 (1.2)</td>
<td>5.0 (6)</td>
<td>27.0 (32.5)</td>
<td>32.0 (38.6)</td>
<td>18.0 (21.7)</td>
<td>3.7 (0.9)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>5.0 (8.6)</td>
<td>7.0 (12.1)</td>
<td>22.0 (37.9)</td>
<td>13.0 (22.4)</td>
<td>11.0 (19)</td>
<td>3.3 (1.2)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

In this study, we attempted to comprehend student opinions, attitudes, and worries regarding distance learning and the barriers and difficulties they faced. Data were gathered through a web-based survey questionnaire between December 2022 and January 2023. The results of our study demonstrate that the participants showed good attitudes to the continued delivery of lectures and elective courses via distance learning. These findings support previous research showing that 63% of students preferred distance learning for the lectures compared to 29% for offline lectures [19]. In contrast, most had negative attitudes toward distance learning for laboratory courses and TBL examinations. These findings support previous research showing that students had negative perceptions of distance learning in many ways of teaching. According to Kedraka et al [14], most of the students at 2 participating universities (University of Patras: 84.1%; Democritus University of Thrace: 98%) did not wish to continue their education for laboratory courses distance beyond the epidemic. A study conducted by Stoian et al [20] reported that 38.8% preferred using face-to-face education, which seemed the most beneficial for their professional development, followed by a distance education rate of 34.3%, then both (face-to-face and distance education) at 27%.

Students’ perceptions of the distance learning environment were positive, with 83.7% (118/141) of the students stating that taking a course through distance learning was easy and 43.3% (61/141) finding the course content understandable. Our findings are in agreement with Bani Hani et al [10], who revealed that more than half of the students participating in the study (615/999) stated they gained the same or even better knowledge than what they gained before the pandemic and around half (458/999) of all the students recognize the university’s e-learning website as available for easy access. A Saudi study [21] showed that 49.2% of students exhibited positive attitudes toward the provided distance learning, and 34% of students identified some barriers to the provision of distance learning.

The majority of participants, 56% (79/141), felt that having a camera on during class had a detrimental influence on their learning. In a survey conducted by Gherheş et al [22], more than half of students claimed that they were unwilling to have their cameras on during distance sessions. Furthermore, 76.5% of participants stated that they did not use webcams when communicating via social media [23]. In contrast, in a study performed by Pullan et al [24], 51% of participants reported that having a camera turned on did not change their ability to engage, while 38% thought that having a camera on improved their engagement.

More than half of the students, 58.9% (83/141), claimed to have impaired social contacts on and around campus, and 44% (62/141) reported having impaired in-person interactions with teachers during class, with 42.5% (60/141) reporting poor interactions in the classroom. These findings support previous research. Muthuprasad et al [25] reported that 60% of respondents agreed with the assertion that contact with the instructor is less effective in distance classrooms as compared to face-to-face classes. Another study [26] revealed that a lack of live communication reduces the effectiveness of distance learning (52.1%). Furthermore, a study by Dodd et al [27]
revealed that 84.6% of the students found it difficult to interact with other students, teachers 74.6%, and 74.7% reported more difficulty in distance learning than face-to-face.

The preferred method of communication (96/141, 68%) for the students in our study was the course chat, a result that supports earlier studies. Coman et al [28] found that 52.4% of students preferred to communicate with teachers through writing on chat. A study [29] revealed that most students (87.4%) preferred synchronized learning sessions for group discussions. In addition, 61.7% of students disagreed that distance learning provides similar learning satisfaction to classroom learning.

As for feelings, 73.7% (104/141) of students were relieved that their courses had not been interrupted and 58.1% (82/141) had enthusiasm for the new experience. However, there was still much interest, 52.5% (74/141), in how the studies would go in the future. Murphy et al [30] reported that their participants expressed uncertainty (59.5%), anxiety (50.7%), and nervousness (41.2%) regarding the switch to distance courses. In addition, Guse et al [31] revealed in a survey conducted in May 2020 on students that 45.5% of dental students noticed a decline in their excitement to study compared to 30.7% of medical students. Worsening mental health was reported by 36.4% (16/44) and 29.5% (26/88) of dental and medical students, respectively.

As this study was conducted in a single institution, its conclusions may not apply to other Saudi Arabian or global academic institutions. The findings of this study may be used to design a more effective and user-friendly distance course.

As the study design is cross-sectional, the results only reflect the situation of distance learning during the data collection.

Conclusions
Similar to all types of education, distance learning is, as reported by the students, characterized by advantages and disadvantages, an understanding of which will aid educational institutions in developing plans for the more effective distribution of instructional materials to students. According to the results, the students felt that the course material was intelligible and that the distance courses were uncomplicated. Moreover, they expressed relief that their studies were not disrupted. These benefits facilitate the design of courses that meet the requirements of students and may lead to advancements in distance learning techniques.

Students in this study identified loss of face-to-face contact during their courses as the greatest drawback of remote education, followed by a lack of social connection on campus. This lack of peer interaction is particularly notable, for it is crucial to understand the importance of peer connection for students. Therefore, these variables should be considered for the improvement of e-learning. In summary, face-to-face communication is an indispensable part of the educational process.

Acknowledgments
The authors gratefully acknowledge the authors of Katerina Kedraka et al [14] for permission to reuse their survey in this study.

Data Availability
The data sets used for this study are available upon reasonable request.

Conflicts of Interest
None declared.

References
4. Alsahali et alJMIR FORMATIVE RESEARCH
7. Prasetyo YT, Tumanan SAR, Yarte LAF, Ogoy MCC, Ong AKS. Blackboard E-learning system acceptance and satisfaction among Filipino high school students: an extended Technology Acceptance Model (TAM) approach. : IEEE; 2020 Presented
at: 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); December 14-17, 2020; Singapore p. 1271-1275. [doi: 10.1109/ieem45057.2020.9309876]


Abbreviations

TBL: team-based learning
Nurse-Led Brief Intervention for Enhancing Safe Sex Practice Among Emerging Adults in Hong Kong Using Instant Messaging: Feasibility Study

Sharon Hoi Lam Pak1, MPhil; Man Ping Wang1, PhD; Anne M Teitelman2,3, PhD; Janet Yuen Ha Wong4, PhD; Daniel Yee Tak Fong1, PhD; Edmond Pui Hang Choi1, PhD

1School of Nursing, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
2School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
3College of Nursing, Thomas Jefferson University, Philadelphia, PA, United States
4School of Nursing and Health Studies, Hong Kong Metropolitan University, Hong Kong, China (Hong Kong)

Corresponding Author:
Janet Yuen Ha Wong, PhD
School of Nursing and Health Studies
Hong Kong Metropolitan University
11/F, 1 Sheung Shing Street, Homantin, Kowloon
Hong Kong
China (Hong Kong)
Phone: 852 39702988
Fax: 852 24062375
Email: jyhwong@hkmu.edu.hk

Abstract

Background: The incidence of sexually transmitted infections has been increasing throughout the world. Additionally, substantial changes in emerging adults’ attitudes toward sex and the popularization of premarital sex could further affect the diagnosis and treatment of sexually transmitted infections. With the high acceptability and effectiveness of instant messaging (IM) interventions for health promotion, there is potential for such interventions to improve condom use knowledge and promote safer sex practice.

Objective: The study evaluates the feasibility of a nurse-led IM intervention to promote safer sex practices in emerging adults.

Methods: A 30-minute adaptive IM intervention and a 5-day booster dose of daily messages after 2 weeks through WhatsApp (Meta Platforms, Inc) were conducted with emerging adults in local universities in Hong Kong aged between 18 and 29 years with previous sexual experience. A questionnaire was distributed 1 week after the intervention that measured the consistency in condom use, the change in condom use knowledge and attitudes, and the acceptability of the intervention. The feasibility of the intervention was assessed by Bowen’s feasibility framework.

Results: A total of 20 participants completed the intervention and questionnaire. Results showed (1) high satisfaction level (mean satisfaction score: 9.10/10), (2) high demand of the intervention (retention rate: 95%), (3) smooth implementation of the intervention, (4) high practicality (13/20, 65% of the participants viewed IM to be an effective means of intervention), (5) potential integration of the intervention, and (6) significant mean increase in condom use knowledge and attitudes (mean increase 9.05; \( t_{19}=3.727; \) 95% CI 3.97-14.13; \( P=.001 \)).

Conclusions: The IM intervention was feasible, acceptable, and had potential impacts on improving safer sex practices. These findings will support the future development of IM interventions in the arena of sexual health promotion.

(JMIR Form Res 2024;8:e52695) doi:10.2196/52695

KEYWORDS
condom use; emerging adults; HIV prevention; IM intervention; mHealth; nurse-led intervention; safer sex practice; sexual health; sexually transmitted infections; text-messaging
Introduction

Reducing sexually transmitted infections (STIs) has been identified as a global health sector strategy by the World Health Organization (WHO) [1]. The WHO reported that there are more than 1 million STIs acquired every day and that estimated that there are around 374 million new cases each year [2]. Globally, almost 57 million disability-adjusted life years are lost to STIs [2]. Among all STIs, human papillomavirus (HPV) is the most common infection. In severe cases, this will further lead to invasive cervical cancer [3].

In Hong Kong, there was nearly a 3-fold increase in the incidence of diagnosed HIV from 250 cases in 2011 to 725 cases in 2015 [4]. After reaching its peak in 2015, the number of cases gradually decreased to 505 in 2020. A sharp decline in STI cases was noted from 2020 to the third quarter of 2021. However, the decrease could be due to reduced screening or testing during the COVID-19 pandemic. Although no specific data were found in the context of Hong Kong, studies in the United States and Europe showed a relationship between COVID-19 cases and STI-reported cases. Results reflected that the reported cases were much lower than the expected cases in both countries [5,6]. Hence, the situation is still alarming, as the pandemic could have concealed the real size of the population with STI infection.

In addition, the asymptomatic nature of STIs and a substantial change in Hong Kong people’s attitudes toward sex may further delay the diagnosis and treatment of STIs [7]. For instance, premarital sex and high-risk sexual behaviors are becoming more common. According to the Youth Sexuality Study by the Family Planning Association of Hong Kong, the proportion of youth practicing safe sex (use of condoms during sexual intercourse) decreased from 1996 to 2011 [8]. Among male respondents, the number decreased significantly, from 66.4% in 1996 to 21.5% in 2011, while among female respondents, the number decreased gradually from 55.2% in 1996 to 45.4% in 2011 [8]. The AIDS Concern conducted a survey from 2016 to 2017 and also reported that only 16%, 22%, and 29% of young women who had sex with their boyfriends, regular sex partners, and irregular sex partners, respectively, used condoms every time during sex in the past 6 months [9]. The WHO reported a positive relationship between unsafe sex practices and the incidence of STIs [10]. Hence, the burden caused by STIs is avoidable by minimizing unsafe sex, such as having consistent condom use.

A few systematic reviews have demonstrated that mobile health (mHealth) interventions through interactive instant messaging (IM) are effective for improving health behaviors, for example, smoking cessation [11], medication adherence [12], weight management [13], and blood pressure control [14,15]. In the sexual health arena, to the best of our knowledge, no study has yet evaluated the effectiveness of a personalized and brief intervention delivered through IM for enhancing condom use knowledge and attitude, leading to the primary prevention of STIs. In addition, from our experience in evaluating an interactive computer-based intervention in female university students by a multisite randomized controlled trial, the study implementation led us to understand that a specific, personalized, and tailored intervention to cater to sexual activity recommendations and relationship problems is needed [16]. In the Chinese cultural context of pervasive sexual conservatism, it is unclear whether an IM-delivered brief intervention to promote condom use would be acceptable and welcomed by university students. Therefore, the research question of this study is whether an IM-delivered brief intervention by nurses is feasible to educate about safer sex practice among sexually active university students in Hong Kong.

The study aims to investigate the feasibility of an IM-delivered brief intervention to promote safer sex practices among sexually active university students in Hong Kong. The objectives of the study are to design a nurse-led brief intervention based on Information, Motivation, and Behavioral skills (IMB) model of Fisher and Fisher [17] and to investigate the feasibility of an IM intervention for improving safer sex practices using the feasibility framework reported by Bowen et al [18].

Methods

Study Design

The CONSORT (Consolidated Standards of Reporting Trials) eHealth checklist was followed as a reporting guideline to report and appraise the IM-based intervention [19]. Implementation and results of the web-based questionnaire were reported according to the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist [20]. The feasibility study was conducted from August to October 2021.

Recruitment

Subjects were recruited through snowball sampling at local universities between September and November 2021. An electronic flyer was sent through IM (ie, WhatsApp; Meta Platforms, Inc) to inform potential participants about the purpose of the study, participation eligibility, time commitment, and incentives for study participation. The inclusion criteria were being an emerging Chinese adult aged between 18 and 29 years, being able to read Chinese, having a smartphone with WhatsApp installed, and having had sexual activities in the past 12 months. The exclusion criteria were being pregnant, having a psychiatric illness, and having received information related to contraceptives and STIs from universities, hospitals, clinics, or nongovernmental organizations in the past 12 months. The sample selection process is shown in Figure 1. Eligible participants were asked to sign a web-based consent form for study enrollment and were invited to provide 3 available time slots for receiving the 30-minute nurse-led brief intervention.
Nurse-Led IM Intervention

Intervention Development
The development of the intervention was done by our team, who are experts in the field of sexual health and the development of IM-based interventions. Our nurse-led brief intervention was designed based on the Fisher and Fisher [17] IMB model. A systematic review found that brief interventions designed according to the IMB model were successful in increasing condom use [21]. The model consists of 3 constructs that influence behavioral adherence: that are information, motivation, and behavioral skills [17]. A total of 3 goals were developed accordingly: (1) provide information about STIs and condom use; (2) motivate participants to use condoms with motivational interviewing techniques; and (3) facilitate behavioral skill acquisition to increase condom use, lubricant application, and their purchase.

Regarding the intervention content, we addressed favorable and unfavorable views, including condom use reliability and effectiveness; the pleasure or excitement of condom use; the stigma of condom use; the embarrassment of condom negotiation; condom purchasing; and sexual coercion prevention related to condom use.

Implementation Process
A registered nurse who is an expert in sexual health promotion oversaw the intervention delivery throughout the study. The IM conversations were first initiated by providing personalized feedback on the participant’s consistency in condom use, as reported on the baseline survey. Second, further IM engagement explored their reasons for condom use and emphasized their decision to protect themselves by using a condom for the prevention of STIs and pregnancy. Third, the nurse identified the underlying reasons for their consistent and inconsistent condom use and then provided personalized advice, myth clarification, and further motivated participants by emphasizing the importance of condom use at every sex, even with a stable partner. Fourth, continued IM exchanges provided a menu of strategies and conversation techniques to use in condom use negotiations or in verbalizing delaying sex whenever unwanted. Finally, further IM prompts provided support on final tips about safe sex and enhancing self-efficacy for condom use, lubricant use, and its purchase. Overall, the IM messages were phrased in a way to communicate a sense of optimism to use condoms for safety and comfort in the dating relationship. It was conducted in an empathetic, respectful, and nonjudgmental manner. The whole real-time conversation lasted for around 30 minutes through WhatsApp. They were reminded to text the nurse again for any inquiries. A standardized guideline was developed as a reference for the nurse to derive the intervention under the IMB model (Multimedia Appendix 1).

IM was considered for intervention delivery because of its real-time, private, and persuasive nature [22]. IM allows real-time interactions, the use of stickers and emojis, and the delivery of URL links as resources. Also, the personalization of IM conversations is suitable to the needs of those participants, who are passive and sensitive when it comes to sex education. In addition, we provided flexibility for participants to receive the intervention at the most suitable moment; that the intervention time was scheduled after the completion of the baseline questionnaire.

Apart from the 30-minute brief intervention conversations, a booster dose of 5 daily messages was also provided 2 weeks after the brief intervention delivery to further reinforce safer sex knowledge and positive sexual health attitudes. An example of an interaction between the nurse and a participant on WhatsApp is provided in Multimedia Appendix 2.
In order to protect the privacy and confidentiality of the participants during the IM intervention, the messages should not reveal personal identification or other sensitive information, such as HIV status. Only the principal investigator and personnel for data analysis were permitted to access the raw data and study content. All the downloaded messages were stored with passcode protection and will be destroyed after a set period.

**Data Collection**
A baseline questionnaire containing the number of sexual intercourses, the number of protected sexual intercourses in the past 12 months, and the Multidimensional Condom Attitudes Scale (MCAS) was given on the entry of the study (T0). Another questionnaire with the same set of measurements at baseline and questions evaluating acceptability, demand, implementation, practicality, integration, and limited efficacy was distributed 1 week after the completion of the IM intervention (T1). The first author downloaded all the WhatsApp chat data for feasibility evaluation. Incentives of US $18 cash were given for compensating participants for their time and efforts in the study evaluation.

**Demographic Data**
Individual characteristics such as age, sexual orientation, relationship status, and gender of the sex partner were obtained at baseline. A history of childhood sexual abuse experience and sexual health were also asked at baseline to ensure personalized conversation and advice could be made during the intervention. The telephone numbers of the participants were collected to facilitate the delivery of the IM intervention.

**Feasibility Evaluation Criteria**
The feasibility of the brief intervention was investigated according to the recommendations of Bowen et al’s [18] feasibility framework, which includes the focal areas of (1) acceptability, (2) demand, (3) implementation, (4) practicality, (5) integration, and (6) limited efficacy.

**Acceptability**
We examined participants’ perceived appropriateness and satisfaction level with the intervention. Questions included “What is your satisfaction with the intervention on a scale of 1-10?” “Do you think the intervention is appropriate for young adults aged between 18 and 29 years?” “Do you find it embarrassing when discussing sexual health issues during the intervention?” “Do you have any disagreement with the contents?” and “Whether you would recommend the intervention to friends?” Participants could also voice out recommendations for improvement.

**Demand**
We examined the actual use of the intervention, which was reflected in the number of subjects who consented to join the study and its retention rate. All the text messages downloaded were also reviewed for content classification. Participants’ views on their interest in and intention to use were also considered. Questions included “Have you ever asked any questions via the IM app? If yes, is the information useful for you?” and “Will you use the intervention again in the future?”

**Implementation**
We investigated the extent of the intervention that can be successfully delivered to the participants. The number of messages delivered and received was measured. We would be interested to know if the information is appropriate to be understood. Questions such as “Do you have more understanding of condom use?” and “Have you tried to communicate with a sexual partner about condom use?” were asked. In addition, participants’ thoughts on using the intervention through an IM app were also asked to obtain qualitative information. For example, “What do you remember the most?”

**Practicality**
We examined the extent to which the intervention could be carried out by using existing resources. The total number of hours for the whole intervention, including the booster dose, was calculated to estimate the resources needed for implementation. In addition, participants were asked to choose the most effective means of promoting sexual health information to emerging adults.

**Integration**
We investigated the extent to which participants were using the intervention and whether the intervention could be integrated into the current system. The perceived fit with infrastructure, perceived sustainability, and costs to the organization and policy bodies were considered. Questions included “Will you use the intervention again if you come up with any questions about sexual health?” and “What do you think about using IM intervention for sexual health promotion?”

**Limited Efficacy**
We examined whether the intervention could result in a potentially promising outcome and its intended effects. The outcome evaluation was conducted by assessing (1) consistency in condom use and (2) condom use knowledge and attitudes. We compared both outcomes by measuring them before and after the intervention.

Consistency in condom use was measured by the percentage of unprotected sexual intercourse with all partners in the past month. Condom use knowledge and attitudes were measured by the 25-item MCAS. A previous study has shown acceptable validity and reliability in the Chinese population [23]. The Cronbach α of the study was >.7 [23]. It contained 5 domains, including the reliability and effectiveness of condoms, the sexual pleasure associated with condom use, the stigma associated with people proposing or using condoms, embarrassment about negotiating and using condoms, and embarrassment about purchasing condoms [24]. The items were answered on a 7-point Likert-scale, with a total score ranging from 7 to 175. The higher the score, the more positive were knowledge, attitude, norms, and self-efficacy related to condom use.

**Data Analysis**
Descriptive statistics were used to determine the participants’ characteristics. Due to the small sample size of the study, a quantile-quantile plot was used to assess the normality of the distribution. A 2-tailed paired sample t test was done to compare pre- and postintervention performances for outcome measures.
with a normal distribution, while the Wilcoxon signed rank test was used for outcome measures without a normal distribution. The content of all the downloaded messages was screened by the first author, and important data were noted. Another member of the team counterchecked all content extracted from WhatsApp to ensure treatment fidelity [25]. Statistical analysis was conducted using SPSS Statistics (IBM Corp).

**Ethical Considerations**

The study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW-20-299). Web-based informed consent forms were sent and completed by eligible participants. All messages collected should not reveal personal identification or other sensitive information to protect the privacy and confidentiality of the participants during the IM intervention.

**Results**

**Demographics**

Among the 21 participants who completed the baseline questionnaire, 20 (retention rate=95%) of them completed the T0 and T1 questionnaires. The age of the participants ranged from 18 to 25 years, with a mean age of 22 years. All the participants were university students. In terms of sexual orientation, 16/21 (76%) were heterosexual, and 5/21 (23%) were either bisexual or homosexual. The participants’ characteristics can be seen in Table 1.
Table 1. Characteristics of the participants (n=21).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>22 (2.02)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Female</td>
<td>17 (81)</td>
</tr>
<tr>
<td>Sexual orientation, n (%)</td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>16 (76)</td>
</tr>
<tr>
<td>Bisexual or homosexual</td>
<td>5 (24)</td>
</tr>
<tr>
<td>Relationships, n (%)</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>3 (14)</td>
</tr>
<tr>
<td>In a relationship</td>
<td>13 (62)</td>
</tr>
<tr>
<td>Cohabitation</td>
<td>5 (24)</td>
</tr>
<tr>
<td>Sex of sexual partner, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16 (76)</td>
</tr>
<tr>
<td>Female</td>
<td>5 (24)</td>
</tr>
<tr>
<td>Sexual health history, mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Number of sex partners in the past 3 months</td>
<td>1.2 (0.95)</td>
</tr>
<tr>
<td>Number of times sex happened in the past 3 months</td>
<td>9.9 (8.16)</td>
</tr>
<tr>
<td>Number of condoms used in the past 3 months</td>
<td>6.45 (7.07)</td>
</tr>
<tr>
<td>History of childhood sexual coercion, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1 (5)</td>
</tr>
<tr>
<td>No</td>
<td>20 (95)</td>
</tr>
<tr>
<td>History of pregnancy, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No</td>
<td>21 (100)</td>
</tr>
<tr>
<td>Diagnosed with STIs(^a), n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No</td>
<td>21 (100)</td>
</tr>
<tr>
<td>Ever had an STI test or pap smear test, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1 (5)</td>
</tr>
<tr>
<td>No</td>
<td>20 (9)</td>
</tr>
</tbody>
</table>

\(^a\)STI: sexually transmitted infection.

Feasibility Evaluation Criteria

Acceptability

Among the 20 participants, the level of satisfaction was high (mean 9.10, SD 1.25). All of them considered that the intervention was appropriate for emerging adults aged between 18 and 29 years. A total of 19/20 (95%) participants did not feel embarrassed when discussing sexual health issues during the intervention. Participants claimed that through the use of emoji and stickers in WhatsApp, they were able to express their emotions and opinions dynamically, which made them feel more comfortable and casual during the chatting session. No participants disagreed with the content in the chat sessions, and all participants would recommend the intervention to friends. Some of the participants also provided practical recommendations. First, participants suggested the provision of more information on the myths about condom and contraceptive use, condom use decision-making and communication with partners, and the availability of condoms in Hong Kong. In fact, during the booster dose, we received messages to ask further questions related to condom use, STIs, and HPV vaccination, which could be included in the content of future interventions. Second, it was suggested to have a mutual sharing personal sexual health information between participants and the nurse. For instance, sharing of personal experiences and problems encountered in dating relationships or during sexual intercourses. The nurse can further ask about...
individual sexual activities, preferences, and habits. These could make the participants more relatable to the nurse during the intervention and enhance personalized information.

**Demand**

During the recruitment, 39 potential participants were approached, and 21 consented after screening for eligibility and willingness to join the study, with a response rate of 54%. Around 20 of them were retained in the intervention and completed the posttest questionnaire (ie, retention rate of 95%), with only 1 participant lost to follow-up. Among all downloaded messages, the chat session content covered condom use, STIs, HPV vaccination, relationship issues, same-sex activities, refusal of sexual activities, and women’s health checkups. All of them found the message content useful and would use it in the future.

**Implementation**

The chat sessions sent 75 instant messages to each participant on average and received 41 messages from each participant on average with content on lubricant use, contraceptives, STIs, HPV vaccination, and refusal of sexual activities. All participants had a greater understanding of condom use, as reflected in the significant improvement in the MCAS scores. Additionally, 19/20 (95%) tried to communicate with their partners about the use of condoms after the intervention. For qualitative information, we asked participants about the most memorable content. Participants indicated that they particularly remembered the messages about condom use (8/20, 40%) and STIs (7/20, 35%).

**Practicality**

In this study, we had 1 nurse manage all chat sessions, including the 30-minute brief intervention, and 5 daily messages as a booster dose 2 weeks after the brief intervention. The total number of hours spent was around 10 hours. A participant reflected that the response speed during the chat session varied, which might be due to the fact that a nurse had to handle 2 participants at the same time. Nevertheless, 13/20 (65%) of the participants suggested that smartphone apps are an effective way to promote sexual health information, while all of them indicated that mass media (ie, Instagram and Facebook) could be an effective way for promotion.

**Integration**

All participants indicated that they would use the IM intervention again if they came up with any sexual health questions in the future. Participants also claimed that mobile phones are a convenient and comfortable way to receive information about sexual health issues.

**Limited Efficacy**

Due to the small sample size of the study, a quantile-quantile plot was used to assess whether the data set followed a normal distribution (Multimedia Appendix 3). The primary outcome of MCAS showed a normal distribution, and the results of the pre- and posttest are shown in Table 2. The results of the paired-samples t test showed a significant increase in the total mean score of condoms use knowledge and attitudes measured by MCAS from 124 (SD 10.32) to 133 (SD 11.90; \( t_{19}=3.727; P=.001 \)). The Cohen effect size was large (\( d=.81 \)). The mean increase in the MCAS score was 9.05 (95% CI 3.97-14.13). A total of 3 MCAS subscales showed significant results. There was a mean increase in reliability and effectiveness of condoms (\( P=.003 \)), sexual pleasure associated with condom use (\( P<.001 \)), and embarrassment about purchasing condoms (\( P=.03 \)). On the contrary, stigma associated with people proposing or using condoms and embarrassment about negotiating and using condom were not significant.

Table 2. The change on condom use knowledge and attitudes by Multidimensional Condom Attitudes Scale (MCAS).

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Pretest, mean (SD)</th>
<th>Posttest, mean (SD)</th>
<th>t test (df)</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condom use knowledge and attitudes (total MCAS score)</td>
<td>124.05 (2.31)</td>
<td>133.10 (2.66)</td>
<td>3.73 (19)</td>
<td>3.97 to 14.13</td>
<td>.001</td>
</tr>
<tr>
<td>Reliability and effectiveness of condoms</td>
<td>26.85 (1.01)</td>
<td>29.55 (1.01)</td>
<td>3.38 (19)</td>
<td>1.03 to 4.37</td>
<td>.003</td>
</tr>
<tr>
<td>Sexual pleasure associated with condom use</td>
<td>17.80 (0.76)</td>
<td>21.05 (0.76)</td>
<td>4.56 (19)</td>
<td>1.76 to 4.74</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stigma associated with people proposing or using condoms</td>
<td>31.30 (0.83)</td>
<td>32.05 (0.83)</td>
<td>0.66 (19)</td>
<td>−1.62 to 3.12</td>
<td>.52</td>
</tr>
<tr>
<td>Embarrassment about negotiating and using condoms</td>
<td>29.10 (1.06)</td>
<td>29.05 (1.06)</td>
<td>−0.06 (19)</td>
<td>−1.81 to 1.71</td>
<td>.95</td>
</tr>
<tr>
<td>Embarrassment about purchasing condoms</td>
<td>19.00 (1.80)</td>
<td>21.40 (1.80)</td>
<td>2.42 (19)</td>
<td>0.32 to 4.48</td>
<td>.03</td>
</tr>
</tbody>
</table>

On the other hand, condom use consistency did not show a normal distribution in the normality testing. Therefore, nonparametric testing could be a better representation of the results. The Wilcoxon signed rank test was used to compare the condom use consistency before and after the IM intervention. The increase in the mean score of the condom use consistency was not significant (\( Z=-1.41; P=.16 \)), indicating that the intervention was not effective in improving condom use consistency in the study.

**Discussion**

**Principal Findings**

This study demonstrated that a nurse-led IM-delivered brief intervention is feasible, content-relevant, and acceptable to participants.
sexually active Chinese emerging adults for promoting sexual health. In addition, this study demonstrated preliminary efficacy in promoting greater condom knowledge and more favorable attitudes about condoms. These findings are encouraging since, according to the IMB model, improvements in knowledge and attitude are important precursors to behavior changes [17].

This study is the first to use smartphone IM apps for delivering sex education in Hong Kong. The rapid rise in smartphone use provides an opportunity to use such mHealth technology for important health promotion interventions. The IM-delivered intervention has been reported to be highly acceptable in young people aged between 16 and 29 years in Australia, as it is more personal and informal [26]. It can also engage a larger number of individuals personally for a low cost [26]. Another qualitative study, targeting young people aged between 16 and 24 years in the United Kingdom, found consistent results that the delivery of knowledge through IM is friendly and convenient [27]. The findings are in line with our data that the nurse-delivered text messages were able to engage this particular age group regarding sexual health topics in a comfortable and private manner.

Although there are findings that suggest potential benefits of such interventions, many mHealth interventions were found to have low certainty in the evidence [28]. Also, studies evaluating IM-delivered intervention for improving contraceptive-related outcomes have had mixed results when implemented for primary prevention [29]. A total of 2 studies provided considerable effects on sexual health outcomes, such as the percentage of protected sex acts and use of effective contraception, after the delivery of the IM-delivered intervention [30,31]. Another 2 studies showed significant improvements in both reproductive health knowledge and the odds of using condoms in young adults after receiving the IM-delivered intervention [32,33]. Differences in the results could be due to the design and duration of the intervention. In terms of the intervention design, a 2-way interactive component can be more effective in promoting sexual health outcomes [33]. In terms of the intervention duration, it can range from weeks to months [27]. A study suggested that the participants identified redundancies when receiving messages over a long intervention period (ie, 2 years) [30].

Around 20% of the participants in another study in Australia found the 12-weekly messages annoying [26]. Hence, future research on evaluating the IM-delivered interventions in terms of appropriateness of design, optimal intervention length, message delivery frequency, and effectiveness in promoting sexual health knowledge and safe sex practice is necessary.

Our results were analyzed using Bowen et al’s [18] systematic feasibility framework. Condom use knowledge and attitudes were significantly increased. In addition, the results regarding the acceptability of the intervention were perceived to be appropriate according to participants’ needs. Participants were satisfied with the intervention and content. We achieved a retention rate of 95.2%. Most participants showed interest in the intervention, with an average satisfactory score of 9. Also, all of them would recommend the intervention to others. Furthermore, the intervention did not create embarrassment in the discussion due to its private nature. With reflection on the implementation process and results of this feasibility study, we recognized that (1) the IM-delivered intervention was feasible and created no embarrassment during discussions, (2) discussing sexual health issues directly helped to clarify myths, (3) screening for condom use consistency and sexual orientation helped tailor conversations, (4) condom use inconsistency was primarily due to the displeasure of using a condom and alternative contraceptives being used for pregnancy prevention, which also suggested a lack of awareness of STI prevention, and (5) most conversations could use such gender-neutral terms as “they or them” and “partner,” except for discussions of condom or protective barrier practical use tips to fit both heterosexual and lesbian, gay, bisexual, transgender, queer, plus (LGBTQ+) relationships, and (6) using WhatsApp stickers in a conversation helped to create a comfortable atmosphere and establish rapport to facilitate further discussions. This study highlighted the usefulness of using Bowen et al’s [18] framework for systematically evaluating the feasibility of an innovation.

The booster dose stimulated participants’ interest in asking further questions about sexual health-related issues. The booster can not only provide new information to participants but also deliver similar intervention message content to refresh participants’ memories. This helps maintain behavioral changes over time [34]. Previous studies containing booster interventions indicated that participants were able to revisit the content and reinforce their knowledge in these sessions [21,35]. In our booster session, diverse topics such as sexual needs, frequency of sexual activities, delay of sexual intercourse, and STIs were covered. Some participants also raised personal and private questions related to STIs during the booster session. This may indicate that participants needed time to digest the information received initially, which fostered the trust and courage needed to further ask personal questions. Therefore, booster intervention is essential for some participants to assimilate information and share their information at their own pace [27].

**Implications**

Sex education is essential as a primary prevention strategy for emerging adults due to the underestimation of the risk of unprotected sex [9]. A survey from AIDS Concern found that nearly 40% of surveyed young women underestimated their risk of STIs, while 80% of them were assessed as the high-risk population [9]. Although Hong Kong has a relatively low level of youth sexual activity, recent studies have brought up concerns about the inadequacies of sex education in Hong Kong [36]. Conventional sex education programs or school-based sex education are some common means of promoting condom use in Hong Kong, however, they are still not mandatory. In addition, the teaching hours are largely insufficient. Half of the schools in Hong Kong offered 5 hours or less of sex education in a school year [37]. The conservative culture on sexual health issues also shaped Hong Kong people’s views and attitudes [36]. Although Hong Kong is an internationalized city that accepts cultural diversity and open-mindedness, the general public is still reserved and implicit when it comes to sex education [36]. This study supports the idea that an IM-delivered intervention is potentially effective in filling the gap due to its personalized and proactive delivery mode.
Further efforts should be made to scale up the implementation of the intervention in the community. In total, 1494 messages were sent and received between the nurse and 20 participants during the 30-minute brief intervention and a 5-day booster dose of daily messages 2 weeks later. Also, a participant reflected that the response speed during the chat session varied. The reduced instantaneity may affect the intervention effect. Compared with face-to-face talk and counseling, this intervention was delivered with mHealth technology through free smartphone apps, which were more cost-effective in terms of delivery. We did not conduct a cost-effectiveness analysis in this feasibility study. Previous studies have been done to evaluate the cost-effectiveness of IM interventions targeting smoking cessation [38] and weight management [39]. Smartphone-based interventions using SMS text messaging or IM have demonstrated good cost-effectiveness due to their low cost and scalability. However, a cost-effectiveness assessment on this specific topic has not been done. The main concern of a nurse-led IM intervention could be the requirement of more manpower if it is conducted on a large scale. Having said that, we would suggest developing an intervention manual and training a group of sexual health counselors for implementation efficiency at a large scale. In addition, with the advancement of technology, the training manual can be used for the further development of chatbots, which can assist in intervention delivery automatically. A systematic review reported a significant effect of a chatbot intervention in reducing depression [40]. While a randomized controlled trial also found chatbot intervention to be effective in promoting fertility awareness and preconception health [41]. An intervention protocol with standard answers should be well-designed to maintain a high standard of IM chat session content.

Strengths and Limitations
The strength of the study was in providing a realistic assumption on the feasibility of an IM-delivered brief intervention on promoting safer sex practices in the community. Our evaluation was conducted systematically based on Bowen et al’s [18] framework. The intervention was able to create a friendly, comfortable, nonjudgemental atmosphere that allowed participants to think and ask any questions about sexual health. The participants were also able to gain knowledge and more favorable attitudes about condom use.

The limitation of this study was the small sample size, which might affect the generalizability of the results. In addition, a comparison group was not established, so caution is needed when interpreting its findings. However, it is appropriate to have a small sample size to determine whether the intervention is feasible before carrying out a large-scale study [18]. Additionally, it could be cost-ineffective to increase the sample size without conducting a feasibility study. Another limitation was the short follow-up period. It is very likely to take longer than 2 weeks postintervention to see changes in behavior [21]. Future research with a longer time frame would be able to more fully assess the outcome of this intervention.

Conclusion
An IM-delivered brief intervention was feasible and acceptable. The intervention was also demonstrated to be practical and effective in enhancing participants’ general knowledge and attitudes toward condom use. The intervention was able to minimize embarrassment when discussing sensitive topics. Although intervention implementation can be time-consuming and labor-intensive, the study findings informed IM-based sexual health intervention development or modifications to enhance sexual health promotion among sexually active university students. More research is needed to evaluate the effectiveness and cost-effectiveness of such an intervention at a larger scale with trained counselors and evaluate its long-term effects by using a rigorous design, such as a randomized controlled trial.

Acknowledgments
The authors thank the study participants for their participation. This study is not supported by grants or funding.

Authors’ Contributions
SHLP, AMT, and JYHW conceived the study. JYHW delivered the intervention. SHLP and JYHW analyzed the data. JYHW, AMT, MPW, DYT, and EPHC provided guidance throughout the study. SHLP wrote the first draft of the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
A guideline for the nurse to derive the intervention under the Information, Motivation, and Behavioral skills model during the intervention.

[DOCX File, 18 KB - formative_v8i1e52695_app1.docx ]

Multimedia Appendix 2
Examples of the interaction between participants and the nurse during the brief intervention on WhatsApp. Multimedia content (blue, right) and answers from participants (orange, left) are shown.
Multimedia Appendix 3
QQ plots for Multidimensional Condom Attitudes Scale total score and condom use consistency.

References
15. Tam HL, Wong EML, Cheung K, Chung SF. Effectiveness of text messaging interventions on blood pressure control among patients with hypertension: systematic review of randomized controlled trials. JMIR Mhealth Uhealth 2021;9(9):e24527 [FREE Full text] [doi: 10.2196/24527] [Medline: 34550078]


Abbreviations

**CHERRIES:** Checklist for Reporting Results of Internet E-Surveys

**CONSORT:** Consolidated Standards of Reporting Trials

**HPV:** human papillomavirus

**IM:** instant messaging

**IMB:** Information, Motivation, and Behavioral skills

**LGBTQ+:** lesbian, gay, bisexual, transgender, queer, plus

**MCAS:** Multidimensional Condom Attitudes Scale

**mHealth:** mobile health

**STI:** sexually transmitted infection

**WHO:** World Health Organization

©Sharon Hoi Lam Pak, Man Ping Wang, Anne M Teitelman, Janet Yuen Ha Wong, Daniel Yee Tak Fong, Edmond Pui Hang Choi. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Efficient Machine Reading Comprehension for Health Care Applications: Algorithm Development and Validation of a Context Extraction Approach

Duy-Anh Nguyen1; Minyi Li2, PhD; Gavin Lambert3,4, Prof Dr; Ryszard Kowalczyk5,6, Prof Dr; Rachael McDonald3, Prof Dr; Quoc Bao Vo1, PhD

1School of Software and Electrical Engineering, Swinburne University of Technology, Hawthorn, Australia
2School of Computing Technologies, RMIT, Melbourne, Australia
3Iverson Health Innovation Research Institutes and School of Health Sciences, Swinburne University of Technology, Hawthorn, Australia
4Baker Heart & Diabetes Institute, Melbourne, Australia
5STEM, University of South Australia, Adelaide, Australia
6Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Corresponding Author:
Duy-Anh Nguyen
School of Software and Electrical Engineering
Swinburne University of Technology
John St
Hawthorn, 3122
Australia
Phone: 61 392148444
Email: anhngd93@gmail.com

Abstract

Background: Extractive methods for machine reading comprehension (MRC) tasks have achieved comparable or better accuracy than human performance on benchmark data sets. However, such models are not as successful when adapted to complex domains such as health care. One of the main reasons is that the context that the MRC model needs to process when operating in a complex domain can be much larger compared with an average open-domain context. This causes the MRC model to make less accurate and slower predictions. A potential solution to this problem is to reduce the input context of the MRC model by extracting only the necessary parts from the original context.

Objective: This study aims to develop a method for extracting useful contexts from long articles as an additional component to the question answering task, enabling the MRC model to work more efficiently and accurately.

Methods: Existing approaches to context extraction in MRC are based on sentence selection strategies, in which the models are trained to find the sentences containing the answer. We found that using only the sentences containing the answer was insufficient for the MRC model to predict correctly. We conducted a series of empirical studies and observed a strong relationship between the usefulness of the context and the confidence score output of the MRC model. Our investigation showed that a precise input context can boost the prediction correctness of the MRC and greatly reduce inference time. We proposed a method to estimate the utility of each sentence in a context in answering the question and then extract a new, shorter context according to these estimations. We generated a data set to train 2 models for estimating sentence utility, based on which we selected more precise contexts that improved the MRC model’s performance.

Results: We demonstrated our approach on the Question Answering Data Set for COVID-19 and Biomedical Semantic Indexing and Question Answering data sets and showed that the approach benefits the downstream MRC model. First, the method substantially reduced the inference time of the entire question answering system by 6 to 7 times. Second, our approach helped the MRC model predict the answer more correctly compared with using the original context ($F_1$-score increased from 0.724 to 0.744 for the Question Answering Data Set for COVID-19 and from 0.651 to 0.704 for the Biomedical Semantic Indexing and Question Answering). We also found a potential problem where extractive transformer MRC models predict poorly despite being given a more precise context in some cases.
Conclusions: The proposed context extraction method allows the MRC model to achieve improved prediction correctness and a significantly reduced MRC inference time. This approach works technically with any MRC model and has potential in tasks involving processing long texts.

(JMIR Form Res 2024;8:e52482) doi: 10.2196/52482

KEYWORDS

question answering; machine reading comprehension; context extraction; covid19; health care

Introduction

Health professionals and the general public have a high demand for information and knowledge in the health care domain. However, traditional information retrieval (IR) systems such as PubMed do not provide precise responses to queries because they only return a list of relevant abstracts or full-text publications, which the user must read and interpret themselves. Thus, question answering (QA) systems that provide direct answers are preferred to enable the best use of evidence in clinical care [1]. Machine reading comprehension (MRC) is the task of predicting an answer to a question based on an input context. A large body of MRC research is based on extractive methods, where the answer is a text span from the input context that best answers the question [2]. Although extractive answers are confined to the input context, they are grounded and sensible. Current approaches to MRC tasks that achieve state-of-the-art performance on open-domain data sets such as the Stanford Question Answering Data Set (SQUAD) [2] have been shown to not work well when the input is long [3]. When the context is very long, it is difficult for the MRC models to compute attention scores for the context; as a result, answer prediction is poor. In biomedical QA applications, the context of a question is often embedded within a scientific article, which may comprise hundreds of sentences, such as in the Question Answering Data Set for COVID-19 (COVID-QA) [4] and the Biomedical Semantic Indexing and Question Answering (BioASQ) data sets [5], where the context is an entire scientific article. The length of the contexts in the COVID-QA data set is much greater than that of the contexts in the SQUAD data set (6119 vs 153 tokens or words), and the answers in COVID-QA are also longer than those in the SQUAD data set (14 vs 3 tokens). Consequently, existing MRC models typically predict less accurately and more slowly in the biomedical domain.

The work by Min et al [6] proposed a sentence selection–based method for extracting minimal context from documents in QA. Their strategy was to find the exact sentence that contained the answer. Experiments on the TriviaQA [7], NewsQA [8], and SQUAD [2] data sets showed improvement in both the inference speed and answer $F_1$-score. However, one could argue that this strategy might not be suitable for complex domains such as health care, as the information provided in the ground truth answer sentence might not be sufficient to make accurate predictions.

Other studies have proposed techniques aimed at reducing the length of input context by identifying only the relevant sentences necessary for answering the question within the document. Context extraction has been used in QA and other natural language processing tasks such as machine translation and text summarization [6,9-11]. Instead of having the model process a long document, which can be inefficient in some domains, the context extraction task aims to focus on relevant parts of the document and to confine the main model to working solely on those parts. By shortening the input, the inference time and task accuracy can be improved.

Yang et al [12] proposed combining an IR model (Anserini) with a Bidirectional Encoder Representations from Transformers (BERT) [13] MRC model, where the IR model selected the paragraphs most similar to the question, and the selected texts were passed to the BERT model for more accurate answer predictions. Although the method was effective on the SQUAD data set, the simple IR strategy for context extraction based on textual similarity is less likely to work well on more complex domains where there is little overlap between the question and the context.

A previous study [14] proposed a sentence classification approach to predict which sentences from the document (context) constituted the best answer. The sentence selection approach was also used in the study by Min et al [6] to shorten the context before passing it as an input to the MRC model. The approaches in the studies by Kang et al [9] and Wang and Jin [10] used reinforcement learning models to learn to select a context from a long document for MRC and machine translation tasks, respectively. The approaches in the studies by Min et al [6], Wang and Jin [10], and Wang and Jin [15] proposed several techniques for selecting a short context from a long document before passing it to an MRC model. These approaches relied on a sentence selection mechanism to construct a new context. The common rationale behind the sentence selector was based on the semantic similarity between the sentence and the ground truth answer. Thus, some studies [6,9,10] have trained a model that predicted the existence of the ground truth answer in a sentence. In complex domains, information from multiple sentences may be needed to answer a question, although most of the important sentences have little semantic similarity with the answer. Similarly, although the sentence classification method proposed in the study by Min et al [6] works well for less complicated domains where one-sentence contexts are sufficient to answer questions, the method may struggle to find contexts consisting of multiple sentences in more complex domains because it does not consider the information expressed in multiple sentences. The reinforcement learning–based approach in the studies by Kang et al [9] and Wang and Jin [10] takes into account the currently selected sentences when considering another sentence; however, the models were trained to recognize the existence of the ground truth answer in the selected sentences.

https://formative.jmir.org/2024/1/e52482
In this study, we introduce a novel approach to context extraction based on sentence selection. We estimate the usefulness of each sentence within its surrounding context for answer prediction. We showed that our approach can select the correct context while keeping it much shorter than the original article.

Methods

Designing the Baselines

Intuitively, a shorter input context reduces the MRC inference time, but does it improve the prediction as well, assuming that the shortened context contains the relevant information and answer? We used the fine-tuned Robustly Optimized BERT Pretraining Approach (RoBERTa) model in the study by Möller et al [4] and the COVID-QA data set to test this hypothesis and develop the baselines for our proposed approach.

MRC task prediction correctness was measured using the $F_1$-score, which is the harmonic mean of precision and recall of the prediction. Precision is obtained by dividing the number of correctly predicted tokens by the number of predicted tokens. Recall is equal to the number of correct tokens divided by the number of ground truth tokens. Another metric that may be used to evaluate an MRC task is exact match (EM), which is the percentage of the test cases that are exactly the same as the ground truth.

We selected three types of input context as baseline approaches, representing different quality levels of the context:

1. Original article: Worst-case scenario in which the MRC model takes the longest time to predict.
2. Paragraph retriever: We used the paragraph retriever to select the top-k paragraphs (k=6) and concatenate them to obtain the selected context. This strategy can be considered as a simple context extraction method. In most cases, the paragraph retriever can extract the relevant information to the question while keeping the extraction considerably shorter than the original article; details of the retrieval model are provided in the Cosine Similarity for Paragraph Retrieval and Sentence Utility section.
3. Target paragraph: We extracted the paragraph containing the ground truth answer sentence and used it as the input context for the MRC model. Intuitively, the target paragraph contains the relevant information as well as the answer to the question, so it can be considered the near-perfect solution. We assumed that the target paragraph is close to the best context selected by humans.

The results in Table 1 confirm the hypothesis that a more precise input context leads to better MRC performance. As the input length is reduced, inference times (s) of the paragraph retriever baseline and the target paragraph baseline are significantly lower than those of the original article baseline. In contrast, while keeping the relevant information, the more precise input context makes it easier for the MRC model to predict the answer, as can be seen from the target paragraph baseline $F_1$-scores and EM scores. The paragraph retriever baseline did not show any clear improvement in terms of the prediction correctness.

Table 1. The $F_1$-scores and inference time of the different context extraction baselines on the Question Answering Data Set for COVID-19.

<table>
<thead>
<tr>
<th></th>
<th>COVID-QA</th>
<th>Exact match score</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original article</td>
<td>0.724</td>
<td>0.462</td>
<td>54.6</td>
</tr>
<tr>
<td>Paragraph retriever</td>
<td>0.724</td>
<td>0.47</td>
<td>8.8</td>
</tr>
<tr>
<td>Target paragraph</td>
<td>0.757</td>
<td>0.494</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Estimating Sentence Utility Based on Answer Correctness and Confidence Score

A previous study by Min et al [6] proposed to predict whether each sentence contains the answer and extracted the highly probable ones as the input context. However, we show that this logic is not suitable for more complex domains such as health care. The example in Table 2 was taken from the COVID-QA data set and tested using the fine-tuned RoBERTa MRC model. Different input contexts were fed to the RoBERTa model, and the output prediction and confidence scores were observed. In the first 3 cases, although the contexts were all relevant to the question, the MRC model could not make a prediction because insufficient information was given ($F_1$-score=0), especially in the third case where the input context contained the ground truth answer. Only in the later cases, where the input contains both the answer sentence and the sentences with relevant information, the MRC model could make correct predictions (high $F_1$-score of 0.92). In the last case, although the MRC model made a highly accurate prediction, the input context clearly contained excessive redundant information. We present a method to assess the usefulness of a sentence within a given context for answering a question. The method aims to assign high utility to sentences that are either relevant to the question or contain the answer, while assigning low utility to irrelevant sentences.
In Table 2, we show the correlation between the prediction correctness, the MRC model’s confidence score, and the input context. Intuitively, if adding a sentence to the input context leads to an improved prediction, that sentence is useful, and its usefulness is more or less proportional to the increase in prediction correctness. In Table 2, when adding the sentence in case 3 to the sentences in cases 1 and 2 as the input context, the MRC model made a big improvement in the $F_1$-score (cases 5-7). This indicates that the sentence in case 3 was important for answering the question (because it contained the ground

<table>
<thead>
<tr>
<th>Number</th>
<th>Input context</th>
<th>Predicted answer</th>
<th>$F_1$-score</th>
<th>Confidence score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recently, LAB(^a) presenting influenza virus antigens have been studied [3,18,19].</td>
<td>No answer</td>
<td>0</td>
<td>2.12</td>
</tr>
<tr>
<td>2</td>
<td>For mucosal immunization, LAB is a more attractive delivery system than other live vaccine vectors, such as Shigella, Salmonella, and Listeria [20,21].</td>
<td>No answer</td>
<td>0</td>
<td>0.84</td>
</tr>
<tr>
<td>3</td>
<td>It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity [18,22,23].</td>
<td>No answer</td>
<td>0</td>
<td>2.75</td>
</tr>
<tr>
<td>4</td>
<td>Recently, LAB presenting influenza virus antigens have been studied [3,18,19]. For mucosal immunization, LAB is a more attractive delivery system than other live vaccine vectors, such as Shigella, Salmonella, and Listeria [20,21].</td>
<td>For mucosal immunization</td>
<td>0.14</td>
<td>3.70</td>
</tr>
<tr>
<td>5</td>
<td>For mucosal immunization, LAB is a more attractive delivery system than other live vaccine vectors, such as Shigella, Salmonella, and Listeria [20,21]. It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity [18,22,23].</td>
<td>It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity</td>
<td>0.92</td>
<td>5.53</td>
</tr>
<tr>
<td>6</td>
<td>Recently, LAB presenting influenza virus antigens have been studied [3,18,19]. It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity [18,22,23].</td>
<td>It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity</td>
<td>0.92</td>
<td>3.48</td>
</tr>
<tr>
<td>7</td>
<td>Recently, LAB presenting influenza virus antigens have been studied [3,18,19]. For mucosal immunization, LAB is a more attractive delivery system than other live vaccine vectors, such as Shigella, Salmonella, and Listeria [20,21]. It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity [18,22,23].</td>
<td>It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity</td>
<td>0.92</td>
<td>11.18</td>
</tr>
<tr>
<td>8</td>
<td>Recently, LAB presenting influenza virus antigens have been studied [3,18,19]. For mucosal immunization, LAB is a more attractive delivery system than other live vaccine vectors, such as Shigella, Salmonella, and Listeria [20,21]. It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity [18,22,23].</td>
<td>It is considered safe and exhibits an adjuvant-like effect on mucosal and systemic immunity</td>
<td>0.92</td>
<td>11.81</td>
</tr>
</tbody>
</table>

\(^a\)LAB: lactic acid bacteria.  
\(^b\)pgsA: phosphatidylglycerol phosphate synthase.  
\(^c\)sM2: Hepatic Fibrosis Susceptibility Due To Schistosoma Mansoni Infection.  
\(^d\)CTA1: cholera toxin A1.
truth answer). In contrast, this also shows that the sentence in case 3 alone was insufficient, and some utility existed in the sentences in cases 1 and 2.

The confidence score is output by the MRC model alongside the predicted answer, which represents how strongly the model believes that the prediction is the correct answer. Intuitively, the confidence score also reflects the quality of the context; it increased more when important sentences were added to the input context and did not increase or increased slightly when redundant sentences were added (Table 2 shows the confidence score changing with different input contexts in a similar way with prediction correctness).

On the basis of these observations, we realize a strong relationship between the usefulness of a sentence and the increase in answer correctness and MRC model confidence. Thus, we propose 2 methods to calculate the utility value, the usefulness of a sentence within a context, as follows:

\[ u_1 = F_1(q, D) - F_1(q, D \setminus c) \] (1)
\[ u_2 = \text{Conf}(q, D) - \text{Conf}(q, D \setminus c) \] (2)

where \( q \) is the question, \( c \) is the sentence from which the utility is being calculated, \( D \) is the context containing \( c \) (usually the article or document), \( \text{Conf} \) is the confidence score output by the MRC model, and \( F_1 \) is the \( F_1 \)-score of the predicted answer for a question-context pair.

Calculating \( u_1 \) requires a known ground truth answer, and calculating \( u_2 \) requires running the MRC model multiple times. Therefore, to obtain utility scores at inference time, we train a model for approximating the utility scores. From the COVID-QA data set, we randomly selected 1519 questions to generate training data for the proposed utility model, and the remaining 500 questions were used as the test set. From the training set, 500 questions were randomly selected as the validation set, which was used for the parameter optimization. Approximately 120,000 triplets of question-sentence-context were sampled, and the fine-tuned RoBERTa model [4] was used to calculate 2 types of utility values according to equations 1 and 2, which are the training signals of the model. The contexts were selected according to the following strategy to simulate different levels of quality:

- The context contained sufficient information to answer the question, and the target paragraph was chosen for this scenario.
- The context contained relevant but insufficient information to answer the question. To simulate this scenario, the context was composed of 1 to 5 sentences around the ground truth sentence, except itself.
- The context contained somewhat relevant information but was not needed to answer the question. We selected the paragraph adjacent to the target paragraph as the context for this case.
- The context was completely irrelevant to the question, which was simulated by choosing the paragraph furthest from the target paragraph or from another article as the input context.

We selected the sentences such that the utility values covered a wide range, based on the intuition that the closer the sentence is to the ground truth sentence, the higher its utility.

The utility model architecture is shown in Figure 1, and 2 versions of the utility model were trained: \( F_1 \) based and confidence based.

**Figure 1.** Proposed utility model structure.

### Cosine Similarity for Paragraph Retrieval and Sentence Utility

The articles in the COVID-QA data set are structured into paragraphs. The information necessary for answer prediction is often contained within 1 paragraph. On the basis of this observation, we trained a simple retrieval model based on BERT [16] and cosine similarity for selecting paragraphs relevant to the question. The retrieval model is a biencoder (Figure 2) that estimates the cosine similarity between the encodings of the question and a paragraph. On the basis of the similarity scores, paragraphs were ranked, and the top-k paragraphs were selected. The cosine similarity score can also be calculated for a pair of question-sentence, which also represents the relevance of a sentence. As a result, we used the cosine similarity score as the third type of sentence utility \( u_3 \).

https://formative.jmir.org/2024/1/e52482

---

Nguyen et al

XSL-FO RenderX
Figure 2. Cosine similarity model structure. BERT: Bidirectional Encoder Representations from Transformers.

**Context Extraction Strategy**

The paragraph retriever can produce a similarity score between the question and 1 sentence from the article; thus, it can be used as the third type of utility value \( u_3 \). Then, we combined the F1-based utility values, confidence-based utility values, and cosine similarity scores in an ensemble manner:

\[
    u = (\omega_1 u_1 + \omega_2 u_2 + \omega_3 u_3) / (\omega_1 + \omega_2 + \omega_3) \quad (3)
\]

where \( u_1, u_2, \) and \( u_3 \) are utility scores produced by the F1-based utility model, confidence-based utility model, and retrieval model, respectively; \( \omega_1, \omega_2, \) and \( \omega_3 \) are the weights of the utility models, respectively.

Textbox 1 shows the algorithm for context extraction using sentence utility. The rationale was to find potential locations in the article that might contain the necessary information for answer prediction. Once trained, the utility model can estimate the usefulness of sentences to new questions and contexts. Naturally, the ground truth sentence often has the highest utility value in the article; therefore, at test time, the utility model is used to estimate the utility of the sentences to locate potential useful context positions. After estimating the utility value of each sentence in an article, one should focus on the peaks as they are likely close to the ground truth sentence. As not all peaks are useful (e.g., the peaks marked in circles in Figure 3 are likely irrelevant to answering the question), only the highest peaks are of interest. Thus, we set parameter \( h \) as the threshold percentage of the selected peaks compared with the tallest one. After the peaks were identified, we selected the neighboring \( w \) sentences of each peak (\( w \) sentences before and \( w \) sentences after the peak sentence).

**Textbox 1.** Context extraction algorithm.

- Input: question \( q \) and article \( D \)
- Output: context \( c \)
- Parameters:
  - \( h \in [0,1] \): threshold percentage of peaks’ values compared to the tallest
  - \( w \in N^* \): number of sentences to be selected from the peak
  - \( k \in N^* \): number of retrieved paragraphs
  - \( \omega_1, \omega_2, \omega_3 \in (0,1) \): utility model weights

1. Use paragraph retriever to select top-\( k \) paragraphs in \( D \).
2. Estimate the utility and cosine similarity scores for each sentence in the retrieved paragraphs using the F1-based and confidence-based utility models, and the paragraph retriever, calculate the ensemble utility score using equation 3.
3. Select all peak sentences, whose values are \( \geq h \max(scores) \).
4. Select surrounding sentences of the peaks which are within a distance of \( w \) sentences.
5. Context \( c \) is all the selected sentences.
Parameter Optimization

There are 6 parameters in our proposed context extraction algorithm: $k$, $w$, $h$, $\omega_1$, $\omega_2$, and $\omega_3$. We applied the Bayesian optimization method to the validation data set to determine the best combination of parameters. As it takes considerable time to obtain answer $F_1$-score for the validating data set, it is impractical to use answer $F_1$-score as an optimization objective. Instead, we constructed a loss function based on 2 subobjectives:

1. Ground truth sentence selection accuracy ($obj_1$): The percentage of samples in which the selected context contained the ground truth sentence. As our algorithm is based on identifying sentences with peak utility scores, we considered finding the ground truth sentences as the most important factor in our algorithm.

   - Context $F_1$-score ($obj_2$): In addition to locating ground truth sentences accurately, the algorithm needs to limit the length of the selected context to reduce inference time and potentially increase answer prediction correctness. For this, we chose to use the target paragraph (the paragraph containing the ground truth sentence) as the second subobjective to optimize the algorithm. Intuitively, the target paragraph is a near-ideal scenario, as shown in the next section, allowing the MRC model to achieve higher prediction correctness and fast inference time. We compared the selected context with the target paragraph and calculated its context $F_1$-score as follows:

   \[
   \text{context precision} = \frac{\# \text{ correct words}}{\# \text{ words in selected context}} \quad (4)
   \]

   \[
   \text{context recall} = \frac{\# \text{ correct words}}{\# \text{ words in target paragraph}} \quad (5)
   \]

   \[
   \text{context } F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)
   \]

   Thus, we constructed the loss function for the optimization method as follows:

   \[
   L = - (\alpha \times \text{obj}_1) + (1 - \alpha) \times \text{obj}_2 \quad (7)
   \]

   where $\alpha$ is a hyperparameter signifying the weight of the first objective.

   With each $\alpha$ value in range $[0.05, 0.1, ..., 0.95]$, a Bayesian optimization process [17] was applied to obtain a set of optimized parameters. Finally, the lowest objective value $L$ was achieved with $\alpha=.95$, and we chose the corresponding parameters for our context extraction algorithm.

Ethical Considerations

This study did not involve any data nor methods that require ethical considerations.

Results

Data

The COVID-QA data set [4] contains 2019 question-answer pairs, 1500 (74.29%) of which were used to generate training data for the proposed utility model, and the remaining (n=519, 25.71%) were used for evaluation. We also evaluated our approach on the BioASQ data set [5]. We extracted 342 factoid-type questions from the BioASQ 7b-10b data sets to evaluate our proposed context extraction method. The MRC model used for this data set was the RoBERTa model fine-tuned on BioASQ data [18]. Table 3 presents the results.
Table 3. The $F_1$-score and inference time of the different context extraction settings; the Robustly Optimized BERT Pretraining Approach model was fine-tuned separately on the Question Answering Data Set for COVID-19 (COVID-QA) and Biomedical Semantic Indexing and Question Answering (BioASQ) data set.

<table>
<thead>
<tr>
<th></th>
<th>COVID-QA</th>
<th>BioASQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_1$-score</td>
<td>Exact match score</td>
</tr>
<tr>
<td>Original article</td>
<td>0.724</td>
<td>0.462</td>
</tr>
<tr>
<td>Paragraph retriever</td>
<td>0.724</td>
<td>0.47</td>
</tr>
<tr>
<td>Target paragraph</td>
<td>0.757</td>
<td>0.494</td>
</tr>
<tr>
<td>Our approach</td>
<td>0.744</td>
<td>0.498</td>
</tr>
</tbody>
</table>

Experiments on the COVID-QA data set were performed on a GTX1050 graphics processing unit (GPU), and experiments on the BioASQ data set were performed on an A100 GPU.

**Paragraph Retrieval Accuracy**

The performance of the paragraph retriever was measured by top-$k$ accuracy, which is the percentage of test cases where the paragraph containing the ground truth answer was among the $k$ highest-scoring paragraphs.

Paragraph retrieval accuracy of the test set with $k = 1, 2, ..., 10$ are 0.904, 0.95, 0.972, 0.978, 0.986, 0.992, 0.992, 0.994, 0.994, respectively.

**MRC Performance With Context Extraction**

The context extraction model’s performance was measured on the test set using different metrics: answer $F_1$-score, EM, and inference time. According to our experiments, the inference times of the paragraph retrieval model and the utility model were 0.02 seconds and 0.3 seconds, respectively.

We used the parameters selected from the optimization method described in the Parameter Optimization subsection of the Methods section for our context extraction algorithm, which was responsible for selecting a new shorter context from the original article and then passing it to the RoBERTa model [4].

The results in Table 3 show the performance of the fine-tuned RoBERTa model using our proposed context extraction method and other baselines. When the original article was the input context, the MRC model performed the worst, with both the lowest $F_1$-scores and EM score on both data sets, and it also had the longest inference time. The target paragraph setting had the best $F_1$-score and EM score on both the COVID-QA and BioASQ data sets as well as the fastest inference time on the COVID-QA data set. On the BioASQ data set, however, the exact position of the answer was not provided; therefore, we aggregated all paragraphs containing the answer as the target paragraph. As a result, the target paragraphs in the BioASQ data sets were longer than those in the paragraph retriever and context extraction methods, resulting in a longer inference time. Our proposed context extraction method for the COVID-QA data set achieved overall performance second only to the target paragraph setting. For the BioASQ data set, our method also achieved the second-best results in $F_1$-scores and EM scores and had a slightly longer inference time than the paragraph retriever setting.

In several COVID-QA test cases, our context extraction method allowed the MRC model to predict significantly more accurately than the other baselines (Table 4). In the first example, the paragraph retriever and target paragraph methods were able to predict part of the answer, whereas the MRC model only made a somewhat relevant prediction with the original article setting. In contrast, the context extraction method allowed the MRC model to predict almost the ground truth answer (missing only the last sentence). In the second and third examples, none of the 3 baseline methods were able to predict the answer at all. However, the context extraction method enabled the MRC model to achieve high accuracy in prediction.

As mentioned in the third paragraph of the MRC Performance With Context Extraction subsection of the Results section, the BioASQ data set did not provide the exact position of the answer, and the answers were generally short (1-3 tokens) and appeared in many positions in the article. As a result, our token-matching method for identifying answer sentences not only returned answer sentences but also nonanswer sentences. Therefore, the algorithm parameters were not as optimal as the COVID-QA case, and the context extraction approach $F_1$-score was slightly better than that of the paragraph retriever method.

In terms of the EM score, our approach showed a more significant improvement compared with the original article and paragraph retriever methods. Table 5 shows 2 examples in which our method outperformed the 3 baselines by a large margin. In the first example, only the paragraph retriever method made a relatively correct prediction, whereas the other 2 baselines’ predictions were completely wrong. In the second example, all 3 baselines predicted incorrectly. Our method enabled the MRC model to predict correctly in both cases.

We further investigated the test cases where applying context extraction led to a significant change in answer prediction correctness (the difference in $F_1$-scores was >0.5) and then counted the number of cases where context extraction led to worse or better predictions compared with the original article setting. We found that among the cases in which context extraction led to a lower $F_1$-score, there were many cases in which the ground truth sentence as well as its surrounding sentences were selected, which means that the worse prediction was not caused by the context extraction algorithm (Table 6). Upon investigating these cases, we found two main reasons for the drop in prediction correctness:

1. Annotated ground truths are imperfect: As $F_1$-score measures the overlap between the ground truth and
prediction, when the prediction is too long or too short, it receives a low score, although it may express the same thing as the ground truth (examples 1 and 2 in Table 7).

2. Faulty behavior from the MRC model: In some cases, the MRC model was unable to predict an answer when inputting only the context (or paragraph) containing the ground truth sentence, whereas it was able to predict accurately when inputting the entire article (examples 3 and 4 in Table 7).

The second reason was interesting: reducing the context (retaining the ground truth sentence and its adjacent ones) caused the MRC model to predict significantly worse or to be unable to predict at all. We replaced the RoBERTa MRC model with the BERT for Biomedical Text Mining model [19] (BERT model fine-tuned on biomedical data) and performed the same experiments with the context extraction algorithm. Tables 8 and 9 show that the trends are consistent with those of the BERT for Biomedical Text Mining MRC model, thereby supporting our observation that the MRC model worked poorly in short and precise contexts in some cases.

Table 4. Examples where the proposed context extraction method significantly outperformed other methods in the Question Answering Data Set for COVID-19. The predicted answer by the machine reading comprehension model changes with different input contexts; the $F_1$-score is calculated based on the predicted answer and the ground truth answer.

<table>
<thead>
<tr>
<th>Original article</th>
<th>Paragraph retriever</th>
<th>Target paragraph</th>
<th>Our approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question: What is the filamentous phage variom made of?</td>
<td>Prediction: particle is enclosed by a rod-like protein capsid. $\sim$1000 nm long and 5 nm wide, made up almost entirely of overlapping pVIII monomers, each of which lies $\sim$27 angstroms from its nearest neighbor and exposes 2 amine groups as well as at least 3 carboxyl groups</td>
<td>$F_1$-score: 0.286</td>
<td>$F_1$-score is calculated based on the predicted answer and the ground truth answer.</td>
</tr>
<tr>
<td>Ground truth: made up of $\sim$2500-4000 overlapping copies of the 50-residue major coat protein, pVIII, arranged in a shingle-type lattice. Each monomer has an array of chemically addressable groups available for bioorthogonal conjugation, including 2 primary amine groups (shown in red), 3 carboxyl groups (show in blue), and 2 hydroxyl groups (show in green). The 12 N-terminal residues generally exposed to the immune system for antibody binding</td>
<td>$F_1$-score: 0.0</td>
<td>$F_1$-score: 0.868</td>
<td></td>
</tr>
<tr>
<td>Question: Which 4 studies were included?</td>
<td>Prediction: investigation of antivirals, interferon atomization, darunavir and cobicistat, abidol, and remdesivir use for patients with 2019-nCoV</td>
<td>$F_1$-score: 0.0</td>
<td>$F_1$-score: 0.875</td>
</tr>
<tr>
<td>Ground truth: phase I clinical trials on SARS or MERS$^a$ vaccines</td>
<td>Prediction: no answer</td>
<td>$F_1$-score: 0.0</td>
<td></td>
</tr>
<tr>
<td>Question: What are examples of viral vectors for delivering vaccines?</td>
<td>Prediction: anthrax, hepatitis B, HIV-1, influenza, measles, SARS, malaria, and tuberculosis M. Saxena</td>
<td>$F_1$-score: 0.077</td>
<td></td>
</tr>
<tr>
<td>Ground truth: recombinant vaccines are based on both DNA viruses (such as fowlpox virus–based vaccines that target avian influenza virus and fowlpox virus, or vaccinia virus–based vectors against the rabies virus in wildlife) and RNA viruses (such as Newcastle disease virus–based vaccines to be used in poultry or YFV$^b$–based vaccines to be used in horses against the West Nile virus)</td>
<td>Prediction: Salmonella and adenovirus</td>
<td>$F_1$-score: 0.032</td>
<td></td>
</tr>
</tbody>
</table>

$^a$MERS: Middle East respiratory syndrome.

$^b$YFV: yellow fever virus.
Table 5. Examples where the proposed context extraction method significantly outperforms other methods in the Biomedical Semantic Indexing and Question Answering data set. The predicted answer by the machine reading comprehension model changes with different input contexts; \( F_1 \)-score is calculated based on the predicted answer and the ground truth answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Original article</th>
<th>Paragraph retriever</th>
<th>Target paragraph</th>
<th>Our approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question: Which ultraconserved element is associated with embryonic stem cells’ self-renewal?</td>
<td>Prediction: IncrNA(^b) ( F_1 )-score: 0.0</td>
<td>Prediction: T-UCstem1 KD(^c) ( F_1 )-score: 0.667</td>
<td>Prediction: miR-9(^d) ( F_1 )-score: 0.0</td>
<td>Prediction: T-UCstem1 ( F_1 )-score: 1.0</td>
</tr>
<tr>
<td>Ground truth: T-UCstem1(^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question: Which protein phosphatase has been found to interact with the heat shock protein, HSP20(^e)?</td>
<td>Prediction: PKA(^f) ( F_1 )-score: 0.0</td>
<td>Prediction: PKA ( F_1 )-score: 0.0</td>
<td>Prediction: PKA ( F_1 )-score: 0.0</td>
<td>Prediction: PP1(^g) ( F_1 )-score: 1.0</td>
</tr>
<tr>
<td>Ground truth: “protein phosphatase 1” and “PP1”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)T-UCstem1: transcribed ultraconserved stem1.
\(^b\)IncrNA: long non-coding RNA.
\(^c\)T-UCstem1 KD: T-UCstem1 knockdown.
\(^d\)miR-9: microRNA-9.
\(^e\)HSP20: heat shock protein 20.
\(^f\)PKA: protein kinase a.
\(^g\)PP1: protein phosphatase 1.

Table 6. Number of cases of significant change in predictions after context extraction in the testing data of the Question Answering Data Set for COVID-19, compared with original article baseline.

<table>
<thead>
<tr>
<th></th>
<th>Better (number of cases)</th>
<th>Worse (number of cases)</th>
<th>Worse because of poorly selected context (number of cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paragraph retriever</td>
<td>25</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>Target paragraph</td>
<td>37</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Our approach</td>
<td>33</td>
<td>21</td>
<td>4</td>
</tr>
</tbody>
</table>

Nguyen et al

https://formative.jmir.org/2024/1/e52482
Table 7. Examples of bad predictions caused by other reasons. Each row shows the model’s predictions with original article baseline and proposed context extraction method as well as the difference of input context lengths between the 2 approaches.

<table>
<thead>
<tr>
<th>Number</th>
<th>Example</th>
<th>Model prediction</th>
<th>Extracted context to article length ratio (in sentences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>• Question: Where was HTNV(^a) isolated from?</td>
<td>• Ground truth: from the striped field mouse <em>Apodemus agrarius</em></td>
<td>1:19</td>
</tr>
<tr>
<td></td>
<td>• Baseline prediction: striped field mouse <em>A. agrarius</em></td>
<td>• Context extraction prediction: striped field mouse <em>A. agrarius</em>, detected in part by the binding of antibodies from patient serum samples to the lung tissues of healthy, wild-caught field mice</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Extracted context to article length ratio (in sentences): 1:19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>• Question: What is the ultimate destination for N, for its assembly into viral particles?</td>
<td>• Ground truth: the Golgi</td>
<td>1:35</td>
</tr>
<tr>
<td></td>
<td>• Baseline prediction: the Golgi</td>
<td>• Context extraction prediction: the Golgi, and it traffics there via the endoplasmic reticulum-Golgi intermediate complex, also known as vesicular-tubular cluster</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Extracted context to article length ratio (in sentences): 1:35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>• Question: What method is useful in administering small molecules for systemic delivery to the body?</td>
<td>• Ground truth: intranasal</td>
<td>1:45</td>
</tr>
<tr>
<td></td>
<td>• Baseline prediction: intranasal</td>
<td>• Context extraction prediction: no answer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Extracted context: intranasal entry has long been used to administer small molecules, such as proteins, for systemic delivery. Because the nasal mucosa is highly vascularized, delivery of a thin epithelium of medication across the surface area can result in rapid absorption of the medication into the blood. Therefore, siRNAs administered intranasally might be deposited in the nose, and some of them may be unable to reach the lower respiratory tract. In fact, it has been reported that intranasal application of unformulated siRNAs(^b) resulted in lower delivery efficiency and homogeneous pulmonary distribution than that achieved with intratracheal application [31]. The intranasal method is suitable for some lung diseases, such as upper respiratory infection by RSV(^c), and it also has potential for systemic delivery rather than pulmonary delivery of siRNAs. Therefore, it is important to consider the route of administration in animal studies when assessing the delivery and therapeutic efficacy of a formulation for pulmonary delivery. Careful choice of efficient delivery in response to the condition of lung diseases is necessary.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Extracted context to article length ratio (in sentences): 1:45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>• Question: What past research has been done on severe, single-wave pandemics?</td>
<td>• Ground truth: after a new influenza virus (H7N9) was identified in China in 2013, a series of modeling articles described the effect of, and level of preparedness for, a severe, single-wave pandemic in the United States.</td>
<td>1:40</td>
</tr>
<tr>
<td></td>
<td>• Baseline prediction: a series of modeling articles described the effect of, and level of preparedness for, a severe, single-wave pandemic in the United States.</td>
<td>• Context extraction prediction: no answer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Selected context: Is the world ready for a respiratory virus with high transmissibility and severity? After a new influenza virus (H7N9) was identified in China in 2013, a series of modeling articles described the effect of, and level of preparedness for, a severe, single-wave pandemic in the United States. 7 In scenarios that used clinical attack rates (the proportion of individuals who become ill with or die from a disease in a population initially uninfected) of 20% to 30% (for comparison the clinical attack rate was 20% in the first year of the 2009 H1N1 pandemic), depending on severity there would be an estimated 669,000 to 4.3 million hospitalizations and an estimated 54,000 to 538,000 deaths without any interventions in the United States. The models suggested that without a vaccine, school closures would be unlikely to affect the pandemic; an estimated 35,000 to 60,000 ventilators would be needed, up to an estimated 7.3 billion surgical masks or respirators would be required; and perhaps most important, if vaccine development did not start before the virus was introduced, it was unlikely that a significant number of hospitalizations and deaths could be averted owing to the time it takes to develop, test, manufacture, and distribute a vaccine.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Extracted context to article length ratio (in sentences): 1:40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)HTNV: Hantaan orthohantavirus.  
\(^b\)siRNA: small interfering RNA.  
\(^c\)RSV: Respiratory Syncytial Virus.

Table 8. Bidirectional Encoder Representations from Transformers for Biomedical Text Mining model performance with different baselines and proposed context extraction method for data of the Question Answering Data Set for COVID-19.

<table>
<thead>
<tr>
<th></th>
<th>(F_1)-score</th>
<th>Exact match score</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original article</td>
<td>0.54</td>
<td>0.312</td>
<td>161.6</td>
</tr>
<tr>
<td>Paragraph retriever</td>
<td>0.591</td>
<td>0.344</td>
<td>29.3</td>
</tr>
<tr>
<td>Target paragraph</td>
<td>0.644</td>
<td>0.368</td>
<td>7.5</td>
</tr>
<tr>
<td>Our approach</td>
<td>0.597</td>
<td>0.346</td>
<td>18.7</td>
</tr>
</tbody>
</table>
Table 9. Number of cases of significant change in predictions after context extraction with Bidirectional Encoder Representations from Transformers for Biomedical Text Mining in the testing data of the Question Answering Data Set for COVID-19, compared with original article baseline.

<table>
<thead>
<tr>
<th></th>
<th>Better (number of cases)</th>
<th>Worse (number of cases)</th>
<th>Worse because of poorly selected context (number of cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paragraph retriever</td>
<td>43</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Target paragraph</td>
<td>62</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Our approach</td>
<td>45</td>
<td>16</td>
<td>2</td>
</tr>
</tbody>
</table>

In summary, our proposed context extraction method helped the MRC model to predict several times faster on both the COVID-QA and BioASQ data sets. Although the $F_1$-score did not improve significantly, we observed that our context extraction method extracted good-quality context in most cases.

Discussion

Principal Findings

In the result analysis, we demonstrated that our proposed context extraction method was able to extract useful context from the original lengthy article, thus slightly improving the prediction correctness of the MRC model and greatly reducing inference time. Upon further investigation, we found 2 elements that may cause the MRC model to predict poorly although the input context was well selected. Although the first reason—imperfect annotations—is hard to avoid, the second reason, which is the RoBERTa model’s strange behavior, remains uncertain.

In example 3 in Table 7, we discovered that the MRC model relied on certain phrases to find the answer. In particular, if the extra sentence “Intranasal delivery is another common method of pulmonary drug application in animal studies.” was included in the extracted context, the MRC model was able to predict “Intranasal.” We believe that the MRC model needed the phrase “Intranasal delivery” from the extra sentence to make a prediction, which was unnecessary to the human response. Similarly, in example 4, a sentence outside the extracted context was needed for the MRC model to make the prediction. In particular, when prepending one of these sentences from the original article to the extracted context, the MRC model was able to make a highly accurate prediction: “With the emergence of MERS-CoV in the Middle East, a preparedness plan was developed that included a surveillance plan, laboratory testing, and contact tracing guidance.” and “Despite the high case-fatality rate (an important measure of severity), MERS cases can be asymptomatic and mild (25% in one outbreak).” However, none of those sentences contained useful information for answering the question. We found that both the extra sentences were not relevant to the question.

In summary, this discovery presented an interesting behavior of the extractive MRC models, particularly the RoBERTa model. In some cases, the model was unable to predict the answer, although the selected context was correct; however, it predicted the answer accurately with the original lengthy article. The behavior was counterintuitive; it appeared that in these cases, the MRC model used relevant but unimportant information in the article to predict a correct answer incidentally. As a result, when the context extraction algorithm excluded such information, the MRC model was unable to find the answer. This shows that there is some difference between machine and human comprehension. It would be worthwhile to investigate this matter with other MRC models that are not a variant of BERT, such as generative MRC models. We demonstrated that our proposed context extraction method consistently extracted high-quality contexts in most cases. However, in some instances, the resulting predicted answers were poor owing to other factors.

Conclusions

In this study, we propose a novel method for context extraction tasks in MRC in a complex domain, where multiple sentences from a long context provide the information required to answer the question accurately. We demonstrated that our proposed context extraction method works on the COVID-QA and BioASQ data sets, showing that it assists the MRC model in achieving improved prediction correctness and, more importantly, reducing the total inference time. We also observed an intriguing behavior of the MRC model: in some cases, the model performs better when presented with longer input contexts.

Data Availability

The data sets generated and analyzed during this study are available in the GitHub repository [20] and the main page of the BioASQ challenge [21].

Conflicts of Interest

None declared.

References


Abbreviations

- **BERT**: Bidirectional Encoder Representations from Transformers
- **BioASQ**: Biomedical Semantic Indexing and Question Answering
- **COVID-QA**: Question Answering Data Set for COVID-19
- **EM**: exact match
- **IR**: information retrieval
- **MRC**: machine reading comprehension
- **QA**: question answering
- **RoBERTa**: Robustly Optimized BERT Pretraining Approach
- **SQUAD**: Stanford Question Answering Data Set
Performance of ChatGPT on the India Undergraduate Community Medicine Examination: Cross-Sectional Study

Aravind P Gandhi1*, MD; Felista Karen Joesph2*, MBBS; Vineeth Rajagopal3, MD; P Aparnavi4, MD; Sushma Katkuri5, MD; Sonal Dayama5, MD; Prakasini Satapathy6,7, PhD; Mahalaqua Nazli Khatib8, MD; Shilpa Gaidhane9, MD; Quazi Syed Zahiruddin10, MD; Ashish Behera11, MD

1Department of Community Medicine, All India Institute of Medical Sciences, Nagpur, Maharashtra, India
2Melmaruvathur Adhiparasakthi Institute of Medical Sciences and Research, Melmaruvathur, India
3Department of Community Medicine and School of Public Health, Postgraduate Institute of Medical Education and Research, Chandigarh, India
4Department of Community Medicine, KMCH Institute of Health Sciences and Research, Coimbatore, India
5Department of Community Medicine, ESIC Medical College & Hospital, Sanathnagar, Hyderabad, India
6Center for Global Health Research, Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India
7Medical Laboratories Techniques Department, AL-Mustaqbal University, Hillah, Babil, Iraq
8Division of Evidence Synthesis, Global Consortium of Public Health and Research, Datta Meghe Institute of Higher Education, Wardha, India
9Centre for One Health Education, Research & Development, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education, Wardha, India
10Global Health Academy Division of Evidence Synthesis, School of Epidemiology and Public Health and Research, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education and Research, Wardha, India
11Department of Internal Medicine, Postgraduate Institute of Medical Education and Research, Chandigarh, India

*these authors contributed equally

Corresponding Author:
Aravind P Gandhi, MD
Department of Community Medicine
All India Institute of Medical Sciences
Room 420 Department of Community Medicine
Plot 2, Sector 20, MIHAN
Nagpur, Maharashtra, 441108
India
Phone: 91 9585395395
Email: aravindsocialdoc@gmail.com

Abstract

Background: Medical students may increasingly use large language models (LLMs) in their learning. ChatGPT is an LLM at the forefront of this new development in medical education with the capacity to respond to multidisciplinary questions.

Objective: The aim of this study was to evaluate the ability of ChatGPT 3.5 to complete the Indian undergraduate medical examination in the subject of community medicine. We further compared ChatGPT scores with the scores obtained by the students.

Methods: The study was conducted at a publicly funded medical college in Hyderabad, India. The study was based on the internal assessment examination conducted in January 2023 for students in the Bachelor of Medicine and Bachelor of Surgery Final Year–Part I program; the examination of focus included 40 questions (divided between two papers) from the community medicine subject syllabus. Each paper had three sections with different weightage of marks for each section: section one had two long essay–type questions worth 15 marks each, section two had 8 short essay–type questions worth 5 marks each, and section three had 10 short-answer questions worth 3 marks each. The same questions were administered as prompts to ChatGPT 3.5 and the responses were recorded. Apart from scoring ChatGPT responses, two independent evaluators explored the responses to each question to further analyze their quality with regard to three subdomains: relevancy, coherence, and completeness. Each question was scored in these subdomains on a Likert scale of 1-5. The average of the two evaluators was taken as the subdomain score of the question. The proportion of questions with a score 50% of the maximum score (5) in each subdomain was calculated.
Results: ChatGPT 3.5 scored 72.3% on paper 1 and 61% on paper 2. The mean score of the 94 students was 43% on paper 1 and 45% on paper 2. The responses of ChatGPT 3.5 were also rated to be satisfactorily relevant, coherent, and complete for most of the questions (>80%).

Conclusions: ChatGPT 3.5 appears to have substantial and sufficient knowledge to understand and answer the Indian medical undergraduate examination in the subject of community medicine. ChatGPT may be introduced to students to enable the self-directed learning of community medicine in pilot mode. However, faculty oversight will be required as ChatGPT is still in the initial stages of development, and thus its potential and reliability of medical content from the Indian context need to be further explored comprehensively.

(JMIR Form Res 2024;8:e49964) doi:10.2196/49964

KEYWORDS
artificial intelligence; ChatGPT; community medicine; India; large language model; medical education; digitalization

Introduction

Artificial intelligence (AI), generally defined as the simulation of human intelligence by a machine, was developed in the late 20th century. ChatGPT, which debuted a preview version in November 2020, is a new entrant in the AI field with enhanced, user-friendly, and near human–like attributes [1,2]. ChatGPT belongs to the category of large language models (LLMs) as a deep-learning AI trained by imbibing large volumes of texts to produce human-like outcomes. ChatGPT is the successive version of GPT-3 and draws 96 million monthly visitors [3], including 13 million unique visitors daily during January 2023 [4], with more than 100 million users [3] registered within the short span of its launch. More recently, the search giant Google released their LLM-based chatbot “Gemini” (formerly known as “Bard”) to the public [5].

The challenges of medical education [6] and the perceived ineffectiveness of traditional teaching methods such as lectures [7] are obstacles to the effective learning of medical students for which technology poses as a solution [6]. Community medicine—a vital subject taught over the 3 out of 4 years of medical school, with an additional 3 months of internship training-deals with public health topics in the undergraduate medical curriculum in India [8]. The community medicine curriculum plays a crucial role in training medical students to understand the community’s public health needs and develop the necessary skills to promote health and prevent disease [9].

AI-assisted learning and teaching practices are already widely used [10]. LLMs have also been integrated into allopathic and alternate systems of medical education through various aspects such as by solving multiple-choice questions with reasoning, answering queries, providing interactive practice cases, and facilitating differential diagnosis [11-13]. The potential scope for LLMs in medical education also includes curriculum development, personalized study plans, and program evaluation and monitoring [14]. ChatGPT’s inherent property of the “transformer model” has been reported to assist in writing review articles on health [15]. The interactive digital interface and prompt-based responses of GPT models with access to a multidomain database [16] offer feasibility for students to use these tools during the “interested learner” stage of self-directed learning (SDL) [15], which is an essential principle of adult learning.

ChatGPT has been shown to complete exams from varied domains such as business administration [17], the medical licensing exam for medical graduates [18], and the parasitology exam at the undergraduate medical level [19], with mixed results. ChatGPT’s ability to answer and explain the questions administered to medical students in their evaluations indicate its potential role as a learning assistant to students. In due course, students may use multidisciplinary LLMs such as ChatGPT in their learning. This is particularly relevant for a subject such as community medicine, which requires real-world application of knowledge beyond medicine to multidisciplinary determinants of health. However, when such tools are used, it is essential that the responses given are relevant and accurate. Thus, assessing the LLM’s capacity to accurately answer and explain the community medicine examination questions and concepts administered to students is imperative. Furthermore, such evaluations will provide a better understanding of how language models can be used in medical education and the challenges that must be overcome to fully realize their potential. In this background, the primary aim of this study was to validate the ability of ChatGPT to answer the undergraduate medical community medicine examination. We also compared the scores obtained on the examination by ChatGPT with those obtained by the students.

Methods

Design, Setting, and Population

This was a retrospective study based on secondary data conducted in February 2023 at a publicly funded medical college in Hyderabad, India. There are approximately 450 medical undergraduate students currently being trained at the institute. We enrolled all 94 students in the Bachelor of Medicine and Bachelor of Surgery (MBBS) Final Year–Part I curriculum. The Final Year–Part I program comprises three subjects: otorhinolaryngology, ophthalmology, and community medicine. Students are taught and trained in practical and theoretical aspects in these subjects for a period of 12 months during the course of their MBBS studies.

Study Tool

Two question papers from the syllabus of community medicine for the MBBS requirements were set with a total of 100 marks (India uses a “mark” scoring system to rate exam answers, with marks having a similar meaning as points) distributed among
40 questions in two papers with 20 questions each. There are three sections in both papers, with different weightage of marks for each section: section 1 had 2 long essay–type questions worth 15 marks each, section 2 had 8 short essay–type questions worth 5 marks each, and section 3 had 10 short-answer questions worth 3 marks each.

Evaluation of ChatGPT Responses
The same questions from the internal assessment exam for the students conducted by the Department of Community Medicine, ESIC Medical College and Hospital, Hyderabad in January 2023 were input as prompts to ChatGPT by one author (APG) on January 25, 2023. The author has a registered open-access account with OpenAI, which has released ChatGPT.

The answers given by ChatGPT 3.5 to each question were recorded and evaluated by three evaluators who had previously been involved in assessing the answer sheets of the students. The internal examination paper of the students is assessed by faculty at both the Assistant Professor and Professor levels; hence, we included two evaluators of different cadres to eliminate potential bias between a senior (SK) and junior (SD) faculty member in their evaluations owing to differences in experience. Despite this difference in experience, both evaluators have the same educational background (MD in community medicine). Adhering to the double-evaluation guidelines of the university to which the institute is affiliated, if there was a difference in marks of more than 33% awarded to a question by the two evaluators, a third evaluator (APG; assistant professor, MD in community medicine) was called in and their evaluation mark was considered to be the final score for that question.

Qualitative Evaluation of ChatGPT Responses
Apart from scoring for the content, the evaluators assessed the responses to each question under three subdomains to further analyze the quality of ChatGPT responses: relevancy (“Is the answer relevant to the prompt?”), coherence (“Is the description in the answer internally coherent?”), and completeness (“Does the answer sufficiently address all parts of the prompt?”). Each question was scored in these subdomains on a Likert scale of 1-5. The average of the two evaluators was taken as the subdomain score of the question.

Comparison Between ChatGPT and Student Scores
ChatGPT scores were compared with the average score of the Final Year–Part I students obtained during the internal examination. To calculate the average score of the students, their deidentified internal examination scores were obtained from the existing departmental registers. A score of ≥50% is considered the minimum passing percentage for the community medicine subject of the MBBS Final Year–Part I requirements.

Data Analysis
Data analysis was performed in Microsoft Excel. The scores obtained by ChatGPT are expressed as percentages. The overall mean score of the MBBS students was calculated by adding the individual scores of all students and dividing by the number of students who had attended the internal assessment examination. Feedback from the evaluators (SK and SD) was obtained to explore their perception toward the ChatGPT responses.

Ethical Considerations
As this study was based on the use of secondary data without identifiers, on application (reference 799/U/IEC/ESICMC/F522/03/2023) to the Institutional Ethics Committee of ESIC Medical College and Hospital, Hyderabad, the study was determined to be exempted from ethical review.

Results
Scores of Students and ChatGPT Responses
The student responses were scored separately and then the prompts to ChatGPT were delivered. The average score of the students was 43% (43/100) on paper 1 and 45% (45/100) on paper 2 (Multimedia Appendix 1). A sample response given by ChatGPT is shown in Figure 1. Complete answers given by ChatGPT, along with the prompts given by the investigators, for all questions are provided in Multimedia Appendices 2 and 3.
Figure 1. Sample ChatGPT response to a community medicine examination question.

The mean score of the responses for the internal assessment community medicine exam questions by ChatGPT was 133 (66.7%); ChatGPT scored 72% (72/100) on paper 1 and 61% (61/100) on paper 2.

**Qualitative Evaluation of ChatGPT Responses**

In terms of the relevancy of the responses given by ChatGPT, 35 out of 40 questions (88%) were rated to have more than 50% of the maximum score (more than 3 out of 5), while 36 of the 40 questions (90%) had a coherence score greater than 50% of the maximum score. Regarding completeness, ChatGPT’s responses to 32 of the 40 questions (80%) were rated to have more than 50% of the maximum score (Multimedia Appendix 1).

Qualitative feedback was obtained from the evaluators regarding the responses to the questions given by ChatGPT. Medical and health technical terminologies were used relatively less frequently by ChatGPT in comparison with the student responses. Figurative representations (images/diagram) were lacking in the free version of ChatGPT 3.5. This is an inherent limitation of the tool since ChatGPT cannot generate images. ChatGPT did not split the long paragraphs into subheadings, which was implemented in some of the student responses as a strategy to ease understanding. Since the subject was primarily focused on the Indian context, the evaluators also opined that ChatGPT did not cover this context adequately in the responses. The more Western-oriented responses might be because ChatGPT is mainly trained based on English content available on the internet. Supporting this hypothesis, OpenAI has also noted that the role of ChatGPT in a classroom setting is currently biased toward Western settings, while non-Western perspectives are inequitably incorporated within the model [20]. The scope for expansion of the perspectives beyond borders is possible in the future.

The responses to the epidemiological questions need to be elaborately discussed with more clarity using examples and illustrations. Regarding long essay–type descriptive answers, ChatGPT did not go beyond the length limit of 1000-1200 words, even after the prompt was fine-tuned to specify to answer the questions in at least 3000 words. ChatGPT responded negatively when asked whether it has any word limits for a single response. However, the issue of a smaller number of words also persisted in subsequent attempts.

**Discussion**

**Principal Findings**

In this study, ChatGPT scored 72.3% on paper 1 and 61% on paper 2, while the mean scores of the students were 43% and 45%, respectively. The responses of ChatGPT were also rated to be satisfactorily relevant (88%), coherent (90%), and complete (80%) for most of the questions. Therefore, ChatGPT passed the Indian undergraduate-level community medicine exam with an overall score of 66.7%, which is above the minimum pass criterion for students under the university norms (50%), and hence can be considered satisfactory.

Smarter and faster AI tools have been massively introduced in academia and health care during the 21st century. Almost all disciplines worldwide harness AI’s rapidly emerging principles and applications. Medical education can also leverage the potential of AI in providing engaged and self-directed learning.
for students [16,21]. This efficacy might be due to the extensive training database of ChatGPT, which comprised 300 billion words spanning multiple disciplines and the “transformer model.” Earlier learning models were restricted to a specific task (ie, narrow AI). After the “transformer model” was introduced in 2017, a more comprehensive, cross-learning tool using a transfer-learning strategy emerged. Transfer learning enables the tool/software to apply the learnings from one task to execute another [22].

Previous studies have shown mixed results regarding ChatGPT’s ability to answer questions for various subjects. ChatGPT cleared the Master of Business Administration (MBA) examination with a B to B– grade at the University of Pennsylvania, administered under research mode by the business school professor [17]. The multiple-choice question–based test in family medicine in Belgium was cleared by ChatGPT, with better performance found on the negatively worded questions compared to that of the students [23]. Other studies have demonstrated the at or near passing ability of ChatGPT in cracking the 3-step United States Medical Licensing Exam (USMLE) [18,24]. Although ChatGPT’s ability to answer parasitology questions was not comparable to that of the medical students of Korea, the responses provided by ChatGPT had an acceptable explanation for the questions [19]. In contrast, in this study, we found that ChatGPT surpassed the mean score of the students. This might be due to the varied methodology applied compared to that adopted in previous studies. The Korean study was specific to the field of parasitology, and the medical students were administered the exam immediately after completing the module. In contrast, in this study, students were answering questions related to the entire syllabus of community medicine that they had been learning for 3 years. In the current setting, the internal examinations only serve as trials for the university examinations and do not determine the student’s qualification per se. Hence, the motivation factors also vary. Thus, the quantum of the syllabus, timing of the examination, and motivations might have contributed to the students’ relatively low scores in the current setting. The median time for the students to answer each paper (20 questions) was 3 hours, whereas ChatGPT provided faster responses (5-10 minutes for each paper).

GPT-4, the latest version of ChatGPT, has scored more than 80% in all 3 steps of the USMLE [25] and demonstrated an accuracy rate of 76.4% in understanding and answering the surgery board exam questions from Korea [26]. The Korea study also observed a significant difference in the accuracy between the original version (3.5) and the advanced version (4.0) of ChatGPT, indicating constant enhancement of the tool [26]. A study from Australia reported that despite the ability of GPT-4, the latest version of ChatGPT, to provide answers that were on par with those of the first-year plastic surgery service residents, it could not match the performance of residents with advanced years of training [30]. The responses of ChatGPT in this study were considered to be satisfactorily relevant, coherent, and complete for most of the questions (>80%). Previous studies have reported relatively more relevant, accurate, and congruent answers being provided by ChatGPT than the earlier AI systems [24,28]. Given the better performance of ChatGPT on the student exams, aspersions on the existing evaluation techniques of medical students have been cast [31]. The need to reevaluate the existing assessment tools with more critical thinking–based training and testing, along with less memory-based evaluation, has also been proposed [31]. ChatGPT-assisted framing of questions and clinical scenarios may also be undertaken for student assessments, thus aiding faculty [21]. The phenomenon of hallucination, wherein ChatGPT provided a wrong answer but with a confident explanation, has also been reported in the literature [23].

**Strengths and Limitations**

This study represents an early attempt to systematically evaluate ChatGPT’s capacity to answer examination questions on an undergraduate subject covering public health concepts, involving two independent subject experts for assessment. We further evaluated the responses regarding relevancy, coherence, and completeness. However, this study was not without limitations. To maintain uniformity and avoid subjectivity, we adopted a similar prompt system for the students and ChatGPT while administering the paper. However, ChatGPT requires more specific and detailed prompts than the students since students are focused on a specific domain, whereas ChatGPT is a multidisciplinary tool. The ability of ChatGPT in terms of word limits is also ambiguous. This might have restricted ChatGPT from exceeding certain word limits, affecting its ability to provide an adequately descriptive response. At the time of the study, information in the ChatGPT database had been updated up to 2021; hence, the tool could not answer or understand the scope of the latest advances in public health. In addition, it was not feasible to blind the evaluators to ensure they remained unbiased to the responses of students and AI. This bias could be in either direction (favorable or against ChatGPT), according to the subjective opinion of the evaluators. We tried to adjust for this potential bias by having two independent evaluators. Moreover, the comparison was based on a single internal assessment conducted among students from a single institute, which limits the generalizability of the findings. Ethical issues surrounding the use of these LLM tools in medical and health care workers’ education, especially in terms of students using the tool to generate assignments, also remain to be considered while undertaking related research and implementation [32,33].
Conclusion
In conclusion, ChatGPT effectively answered questions related to the Indian undergraduate-level community medicine curriculum. ChatGPT demonstrated adequate comprehension, relevancy, and correctness in answering the community medicine–related questions, indicating its potential as an assistive tool in medical education in India. However, the need for continuous improvement in accuracy and the challenges associated with a contextual understanding of ChatGPT must also be considered. Additionally, efforts should be made to address the potential biases and limitations of language models, ensuring their ethical and responsible use in educational settings.

ChatGPT may be introduced to students to enable the SDL of community medicine in a pilot mode under faculty oversight, as its development remains in the initial stages, while the potential and reliability of medical content from the Indian context need to be more comprehensively explored. Integrating ChatGPT as a complementary tool in the learning process could contribute to a more novel, interactive, and engaging educational experience for medical students, ultimately enhancing their understanding and application of community medicine principles. Operational studies exploring the feasibility, experience, and effectiveness of LLMs such as ChatGPT in the actual learning of Indian medical students should be undertaken in the future.

Acknowledgments
The authors acknowledge the role of ChatGPT (January 9, 2023 version) in providing information about the artificial intelligence tool and its workings. We would also like to acknowledge the Global Center for Evidence Synthesis for their support with the research and dissemination.

All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided APF support for the publication of this article.

Data Availability
All data generated or analyzed during this study are included in this published article and supplementary files.

Conflicts of Interest
None declared.

Multimedia Appendix 1
[XLSX File (Microsoft Excel File), 21 KB - formative_v8i1e49964_app1.xlsx ]

Multimedia Appendix 2
ChatGPT answers for the questions in paper 1.
[PDF File (Adobe PDF File), 179 KB - formative_v8i1e49964_app2.pdf ]

Multimedia Appendix 3
ChatGPT answers for the questions in paper 2.
[PDF File (Adobe PDF File), 207 KB - formative_v8i1e49964_app3.pdf ]

References


20. Educator considerations for ChatGPT. OpenAI. URL: https://platform.openai.com/docs/chatgpt-education [accessed 2023-08-29]


Abbreviations
- **AI**: artificial intelligence
- **LLM**: large language model
- **MBA**: Master of Business Administration
- **MBBS**: Bachelor of Medicine and Bachelor of Surgery
- **SDL**: self-directed learning
- **USMLE**: United States Medical Licensing Exam

©Aravind P Gandhi, Felista Karen Joesph, Vineeth Rajagopal, P Aparnavi, Sushma Katkuri, Sonal Dayama, Prakasini Satapathy, Mahalaqua Nazli Khatib, Shilpa Gaidhane, Quazi Syed Zahiruddin, Ashish Behera. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e49964

JMIR Form Res 2024 | vol. 8 | e49964 | p.3002

Van Lewis King Jr¹, MD; Gregg Siegel², MSc; Henry Richard Priesmeyer³, PhD; Leslie H Siegel², MFA; Jennifer S Potter¹, PhD

¹Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center San Antonio, San Antonio, TX, United States
²Biomedical Development Corporation, San Antonio, TX, United States
³Department of Management and Marketing, St. Mary's University, San Antonio, TX, United States

Abstract

Background: Self-management of opioid use disorder (OUD) is an important component of treatment. Many patients receiving opioid agonist treatment in methadone maintenance treatment settings benefit from counseling treatments to help them improve their recovery skills but have insufficient access to these treatments between clinic appointments. In addition, many addiction medicine clinicians treating patients with OUD in a general medical clinic setting do not have consistent access to counseling referrals for their patients. This can lead to decreases in both treatment retention and overall progress in the patient’s recovery from substance misuse. Digital apps may help to bridge this gap by coaching, supporting, and reinforcing behavioral change that is initiated and directed by their psychosocial and medical providers.

Objective: This study aimed to conduct an acceptability, usability, and utility pilot study of the KIOS app to address these clinical needs.

Methods: We developed a unique, patient-centered computational software system (KIOS; Biomedical Development Corporation) to assist in managing OUD in an outpatient, methadone maintenance clinic setting. KIOS tracks interacting self-reported symptoms (craving, depressed mood, anxiety, irritability, pain, agitation or restlessness, difficulty sleeping, absenteeism, difficulty with usual activities, and conflicts with others) to determine changes in both the trajectory and severity of symptom patterns over time. KIOS then applies a proprietary algorithm to assess the individual’s patterns of symptom interaction in accordance with models previously established by OUD experts. After this analysis, KIOS provides specific behavioral advice addressing the individual’s changing trajectory of symptoms to help the person self-manage their symptoms. The KIOS software also provides analytics on the self-reported data that can be used by patients, clinicians, and researchers to track outcomes.

Results: In a 4-week acceptability, usability (mean System Usability Scale-Modified score 89.5, SD 9.2, maximum of 10.0), and utility (mean KIOS utility questionnaire score 6.32, SD 0.25, maximum of 7.0) pilot study of 15 methadone-maintained participants with OUD, user experience, usability, and software-generated advice received high and positive assessment scores. The KIOS clinical variables closely correlated with craving self-report measures. Therefore, managing these variables with advice generated by the KIOS software could have an impact on craving and ultimately substance use.

Conclusions: KIOS tracks key clinical variables and generates advice specifically relevant to the patient’s current and changing clinical state. Patients in this pilot study assigned high positive values to the KIOS user experience, ease of use, and the appropriateness, relevance, and usefulness of the specific behavioral guidance they received to match their evolving experiences. KIOS may therefore be useful to augment in-person treatment of opioid agonist patients and help fill treatment gaps that currently
exist in the continuum of care. A National Institute on Drug Abuse–funded randomized controlled trial of KIOS to augment in-person treatment of patients with OUD is currently being conducted.

*(JMIR Form Res 2024;8:e48068) doi:10.2196/48068*

**KEYWORDS**

opioid use disorder; digital health; behavioral medicine; KIOS; mHealth; substance use disorder; substance use treatment; self-management; opioid misuse; substance use; social support; KIOS app; KIOS application; software; patient-centered; opioid

**Introduction**

The United States is experiencing an opioid epidemic [1]. In 2021, 106,699 people died from overdoses involving opioids, including both prescription and illicit opioids, and the number of deaths continues to increase [1]. Severe opioid use disorder (OUD) is a lifelong and costly illness affecting millions of people worldwide. The socioeconomic impact of OUD affects every aspect of a person’s life; repercussions from long-term opioid use often include legal problems, damage to personal relationships, and significant morbidity and mortality [2].

One of the most effective treatments for severe OUD typically involves opioid agonist treatment [3,4]. While the benefits of medication treatment often occur rapidly, positive behavioral changes for those with more severe OUD can take months and years to develop. Patients receive professional treatment in the clinic, but the struggle with addiction occurs largely in the setting of their everyday lives. OUD treatment often involves periods of exacerbation and remission, and the vulnerability to recurrent drug use remains a life-long risk for many patients. Treatment adherence is often intermittent, and the potential for a recurrence of substance use disorder symptoms carries with it the chronic risk of overdose, trauma, suicide, and infectious diseases [2,5].

Self-management of OUD is a critical component of recovery. Behavioral and supportive therapy used in conjunction with medication is often helpful in changing deeply embedded behaviors that can lead to recurrent drug use [5,6]. However, patients often face a shortage of mental health services, which limits access to care, education, and counseling and impedes their ability to develop effective patient self-management behaviors [7]. Ongoing coaching and feedback are important for recovery for many patients but difficult and expensive to implement, and generally inaccessible during the time of greatest need.

Digital therapeutic software apps are intended to help address this challenge with a variety of approaches being used, and many patients are open to these treatment enhancements [8]. Recognizing the potential of this research and the benefits of evaluating multiple treatment approaches, the National Institutes of Health has increasingly supported digital health behavior research during the past 15-20 years [9,10]. The National Institute on Drug Abuse Clinical Trials Network has conducted several digital health studies examining digital therapeutics for treatment and other technologies for screening, assessment, and a range of other uses [11].

There are growing numbers of increasingly sophisticated apps to address various substance use problems [12-16]. These apps use diverse approaches [17,18], and there is evidence of benefits and cost-effectiveness associated with their use such as improved retention and reduced drug use [18-20].

This study examines the usability, acceptability, and utility of KIOS, a patient-centered computational software system [21]. The KIOS app uses a mobile technology platform to provide daily symptom monitoring and on-demand, individualized feedback. Feedback is based on changes in key treatment variables between successive self-reports to teach and reinforce healthy practices, foster self-management, and promote adherence to treatment plans for patients enrolled in methadone maintenance treatment for OUD. Since engagement with the digital device is key to determining possible efficacy, this study was conducted to make a preliminary evaluation of the KIOS software and client interface in preparation for a randomized controlled trial in a treatment clinic setting.

**Methods**

**Overview**

This study used a mixed model design to demonstrate technical feasibility; patient engagement; and the acceptability, usability, and utility of KIOS with individuals receiving methadone maintenance for OUD.

**Ethical Considerations**

The focus group protocol was approved by BioMed IRB (approval number 18-1-100-OUD). Participants provided informed written consent and were paid for their participation. In addition, the KIOS use study protocol was approved by BioMed IRB and participants provided signed informed consent and were enrolled in a 4-week, single-group, pre-post evaluation of the KIOS app. Participants were paid US $75 after completing orientation and training. They were paid an additional US $75 for participating in a debriefing or focus group. All patient responses were stored in a confidential fashion in a password-protected computer database and only available to approved study personnel. All data were deidentified prior to analysis.

**Developing the User Interface**

Contextual design [22], an industry-standard, user-centered methodology, was used to design the user experience. Individuals receiving methadone maintenance were recruited for focus groups at a private treatment clinic (CleanSlate Addiction Treatment Center, n=5) and the local public treatment clinic (The Center for Healthcare Services, n=9). Participants provided informed written consent for their participation. Feedback from participants was solicited about the perceived
usefulness of the tool, functionality, benefits, and incentives for using the software.

Based on the focus group feedback, a functional user interface (UI) was created. The web-based KIOS UI includes assessment screens for each of the clinical variables (craving, depressed mood, anxiety, irritability, pain, agitation or restlessness, difficulty sleeping, absenteeism, difficulty with usual activities, and conflicts with others; each rated on a 1-7 scale from none to severe). The determination of the clinical variables is described in the Results section. Inside KIOS, the patient rates the severity of these 10 symptoms and logs self-reported opioid use, alcohol use, and recreational drug use in an important behavior checklist. Participants also reported regular exercise, avoidance of triggers or high-risk situations, and sleep habits (not reported here). The KIOS app analyzes the changes in symptoms during the time between patient self-reports and then delivers advice specific to the patient’s evolving symptoms. On the advice screen of the app, KIOS contextually suggests behaviors that are associated with improving self-report variables, including reaching out to support people or treatment providers, if indicated. In these circumstances, the app encourages the integrated use of comprehensive treatment resources (eg, for increasing depression, patients are directed to consult a medical provider). The results of both the symptom changes and the reporting on important behaviors could also be reviewed by the patient’s individual counselor. This offers the counselor access to real-time data over time on both symptoms and behaviors. It offers the ability to reinforce behavioral treatment planning and maximize the limited counseling time available at each appointment. KIOS also contains a home and menu screen, 6 graphs for tracking symptoms over time, a primary advice screen, a second supplemental advice screen if patients want additional advice, a calendar-based journal, and help screens for technical support, feedback, and emergency contacts for help or crisis situations.

KIOS Use Study Description

The primary outcome was user satisfaction as measured by acceptability, usability, and utility. Secondary outcomes included self-reporting of drug and alcohol use. The study was conducted exclusively on a web-based interface and via telephone. Participants were recruited from Community Medical Services, a private opioid treatment organization. Staff at Community Medical Services provided potential participants with study information and individuals contacted study staff if they wanted to participate.

Inclusion criteria were (1) interest in study participation; (2) 18 years or older; (3) receiving methadone maintenance for OUD for ≥4 weeks; (4) ability to access KIOS via computer, smartphone, or tablet; and (5) no unmanaged major psychiatric illness or suicidality.

Participants attended a web-based KIOS training session. They were asked to complete KIOS assessments on a web-based interface at least 3 times per week but no more than once daily. KIOS could still be accessed by participants as often as desired to review advice and graphs or access the journal. All assessments were logged on the KIOS server to tabulate the frequency of use.

After the trial, participants completed an 18-question KIOS survey. Participants also completed the System Usability Scale (SUS)-Modified, a single-factor 10-item self-report scale that was used to evaluate participants’ subjective experience using KIOS [23,24].

User data were logged on the KIOS server for the 10 variables from each assessment. At the end of the study, a final web-based debriefing focus group was conducted to gather user feedback not otherwise collected during the study. Participants described their experience in the study and provided their impressions, suggestions, and critiques of KIOS in an open dialogue.

Analysis

Means, medians, and SDs were calculated for usage, SUS, and patient questionnaire data. KIOS variable scores were averaged over each week for each participant to account for differing numbers of assessments. Calculations were similarly made for an overall KIOS index score (sum of all variables except craving) and for emotional (depressed mood, anxiety, and irritability), physical (pain, agitation or restlessness, and difficulty sleeping), and social (absenteeism, difficulty with usual activities, and conflicts with others) subscales. The craving variable was kept separate due to a consistent body of literature linking it to substance use outcomes [25-27]. The changes in mean values for each week were calculated relative to baseline measures, and 2-tailed Student t tests were conducted for significance between baseline and the weekly means. A Pearson correlation was performed by comparing the KIOS index score to the craving scores averaged over the study period.

Results

Identifying Key Treatment Variables

Managing recovery from OUD is a complex process characterized by constantly changing intra- and interpersonal circumstances. It was essential to identify a limited number of relevant clinical variables that would both effectively reflect an individual’s subjective behavioral state as well as provide sufficient information to permit the assignment of appropriate behavioral advice. Deidentified data from large-scale randomized OUD clinical trials (the POATS trial: n=653 and the X:BOT trial: n=570) [28,29] were analyzed to provide relevant clinical variables. The KIOS software requires clinical variables that change frequently to operate optimally. Therefore, data obtained from instruments in these studies were analyzed to determine how frequently variables changed from one assessment to the next.

An expert panel (Table 1) selected a subset of 32 variables that were most relevant to OUD recovery. The members of the panel were selected for their expertise in behavioral interventions in OUD and to impart a multicenter perspective to the study. To achieve panel consensus, a modified Delphi process was used [30]. A web-based survey tool and web-based meetings were used to expedite the process. The Delphi process was iterated until consensus was achieved. The decision rule for the inclusion of variables into the system called for the variable to be selected by a majority of the panel members.
Table 1. Expert panel.

<table>
<thead>
<tr>
<th>Panel member</th>
<th>Emphasis</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jennifer S Potter, PhD, MPH</td>
<td>SUD/OUD/pain/CTB</td>
<td>Professor of Psychiatry and Behavioral Sciences, University of Texas Health Science Center, San Antonio</td>
</tr>
<tr>
<td>Van L. King, MD</td>
<td>SUD/OUD treatment</td>
<td>Professor of Psychiatry and Behavioral Sciences, University of Texas Health Science Center, San Antonio</td>
</tr>
<tr>
<td>Roger Weiss, MD</td>
<td>SUD clinical guidelines</td>
<td>Chief, Division of Alcohol and Drug Abuse; Director, Alcohol and Drug Abuse Clinical Research Program, Professor of Psychiatry, Harvard Medical School</td>
</tr>
<tr>
<td>Nathanial Katz, MD, MS</td>
<td>Pain/OUD</td>
<td>CEO, Analgesic Solutions, Natick, MA</td>
</tr>
<tr>
<td>Mary Velasquez, PhD</td>
<td>SUD/motivational interviewing</td>
<td>Professor and Director of the Health Behavior Research and Training Institute at The University of Texas at Austin Steve Hicks School of Social Work</td>
</tr>
<tr>
<td>Erin Finley, PhD, MPH</td>
<td>Evidence-based practices/OUD</td>
<td>Assistant Professor, University of Texas Health Science Center, San Antonio</td>
</tr>
<tr>
<td>Kathleen Carroll, PhD</td>
<td>SUD/web-based CBT</td>
<td>Albert E Kent Professor of Psychiatry; Director of Psychosocial Research, Division of Addictions, Yale</td>
</tr>
<tr>
<td>Sharon Walsh, PhD</td>
<td>SUD/OUD pharmacotherapies</td>
<td>Professor of Behavioral Science, Psychiatry; Director of the Center on Drug and Alcohol Research, University of Kentucky</td>
</tr>
</tbody>
</table>

aSUD: substance use disorder.
bOUD: opioid use disorder.
cCBT: cognitive behavioral therapy.

Independently, each panel member selected variables with the criteria that a change in the variable is important in determining the patient’s current clinical state. If consensus was not achieved, the panel was queried until the decision rule was met. The variables proposed by the expert panel were then evaluated in a correlation matrix to eliminate redundant variables. Ten individual variables (craving, depressed mood, anxiety, irritability, pain, absenteeism, agitation or restlessness, difficulty sleeping, difficulty with usual activities, and conflicts with others) were selected to capture a person’s clinical state most comprehensively while minimizing the number of questions required of the patient at any assessment.

Next, the panelists proposed combinations of 2 or more variables that provided the most valuable information for describing the current state of the patient and that would be particularly sensitive to identifying transitional worsening or improvement in OUD. This grouping of variables created the underlying structure that mapped the patient trajectory and enabled the development of interventions. Reports were generated by the KIOS software describing all the possible changes for the selected set of interacting variables. The panel supplemented these descriptions with interpretations and recommended interventions.

Appropriate behavioral interventions incorporating cognitive behavioral therapy (CBT) and mindfulness, using the Behavior Change Technique Taxonomy [31] as a reference, were applied to each patient state. These change techniques have been shown effective in multiple studies [32] (Table 2). All recommendations were phrased in patient-friendly language and subjected to a rubric specifying reading level and text length, and that the advice is consistent with a CBT approach. In a previous study, the research team successfully followed a similar approach to chart trajectories and develop self-management interventions for bipolar disorder [21].
### Table 2. Sample intervention strategy for 4-dimensional variable grouping.

<table>
<thead>
<tr>
<th>KIOS logic pathway</th>
<th>Patient state</th>
<th>Behavioral intervention taxonomy</th>
<th>Advice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Anxiety: unchanged</td>
<td>• 2.2 Feedback on behavior</td>
<td>“You’ve reported more difficulty sleeping and a rise in depression. Since these could be related, your first task is to work on sleep hygiene. Consider what you can adjust in your sleep environment to help you rest better at night. Can you reduce noise or light? Start dimming lights earlier in the evening? Do you need earplugs, or do you need to adjust room temperature? Brainstorm what might make you comfortable and restful and try changing up a few things in your sleep environment to help.”</td>
</tr>
<tr>
<td></td>
<td>• Depression: increased</td>
<td>• 2.3 Self-monitoring of behavior</td>
<td>“Be sure you’re also following your medication guidelines. If things continue on a downward track, talk to your care provider or someone in your support system. It’s important to pay attention to depressive thoughts and to reach out for help if they don’t resolve.”</td>
</tr>
<tr>
<td></td>
<td>• Agitation: decreased</td>
<td>• 3.1 Social support</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Difficulty sleeping: increased</td>
<td>• 11.1 Pharmacological support</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 12.1 Restructuring the physical environment</td>
<td></td>
</tr>
</tbody>
</table>

### Participants

In total, 19 individuals signed informed consent. Three did not complete orientation. One stopped participating after 11 days and was considered lost to follow-up. The median time in methadone treatment at the beginning of the study was 9 months: 4 patients had been in treatment for ≤6 months, 6 patients had been in treatment for 6-12 months, and 5 had been in treatment for >12 months. Participants (N=19) were 47% (n=9) White (non-Hispanic), 21% (n=4) White (Hispanic), 11% (n=2) more than 1 race, 11% (n=2) American Indian or Alaska Native, 5% (n=1) Black or African American, and 5% (n=1) Asian. The mean age of all participants was 34.4 (SD 7.3) years, and 63% (n=12) were women.

### KIOS Use and Acceptability

In total, 15 of 16 participants who logged in to KIOS at least once finished the trial. Those 15 participants completed 191 total assessments, averaging 12.7 (SD 3.8) per user (median 13, IQR 11-14). During the poststudy debriefings, several participants requested continued access to KIOS after the trial ended without further compensation; 6 participants used KIOS poststudy and 1 participant used the app for over 10 months.

The final debriefing focus group generated many comments that described the individualized experience of using the app. All participants who attended the group stated it was helpful and worth using. For example, focus group attendees noted that KIOS gave timely and responsive advice, helped with adherence to treatment goals, and promoted reflection and motivation. Some commented on how natural the advice seemed, almost like a counselor was responding in real time.

Some participants mentioned features that could improve KIOS such as a customized reminder system to specify the time of day to use the app, features to connect directly with counseling staff or peer support, gamified content (such as daily challenges, more content that included pictures), positive reinforcement (such as digital trophies, badges, or monetary rewards), and expanded and more specific description of exercise or meditation practices beyond those currently included in the app.

### KIOS Usability

The mean SUS-Modified score was 89.5 (SD 9.2; median 92.5, IQR 85-95). Higher scores indicate greater usability. The psychometric properties of the SUS have been validated and replicated, and the score for KIOS falls slightly below the top score, “best imaginable=90.9” and above “excellent=85.5” on an adjective rating of the SUS [23,24] (Table 3).
Table 3. System Usability Scale-Modified scores (1=strongly disagree to 5=strongly agree).

<table>
<thead>
<tr>
<th>Raw score, mean (SD)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.40 (0.74)</td>
<td>I think I would like to use this system frequently</td>
</tr>
<tr>
<td>1.40 (0.83)</td>
<td>I found the system unnecessarily complex</td>
</tr>
<tr>
<td>4.93 (0.26)</td>
<td>I thought the system was easy to use</td>
</tr>
<tr>
<td>1.13 (0.52)</td>
<td>I thought I would need the support of a technical person to use this system</td>
</tr>
<tr>
<td>4.40 (0.83)</td>
<td>I found the various functions in this system were well integrated</td>
</tr>
<tr>
<td>1.20 (0.56)</td>
<td>I thought there was too much inconsistency in this system</td>
</tr>
<tr>
<td>4.87 (0.35)</td>
<td>I would imagine that most people would learn to use this system very quickly</td>
</tr>
<tr>
<td>2.40 (1.35)</td>
<td>I found the system very cumbersome to use</td>
</tr>
<tr>
<td>4.80 (0.41)</td>
<td>I felt very confident using the system</td>
</tr>
<tr>
<td>1.47 (0.74)</td>
<td>I needed to learn a lot of things before I could get going with this system</td>
</tr>
</tbody>
</table>

**KIOS Utility**

The KIOS questionnaire was developed by the research team to evaluate the utility of the KIOS app. The mean response to all KIOS questionnaire data was 6.32 out of a maximum of 7.0 (SD 0.25), indicating an overall strong agreement with most statements. The highest rated statement was “I see benefits using KIOS,” which averaged 6.73 out of 7 (SD 0.46); each respondent rated this item as either 6 or 7. The lowest rated item was “KIOS helped identify some personal triggers,” which was still generally well agreed upon and rated 5.73 (SD 1.62; Table 4).

Table 4. KIOS satisfaction survey (1=disagree to 7=agree).

<table>
<thead>
<tr>
<th>Value, mean (SD)</th>
<th>KIOS satisfaction survey question</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.67 (0.62)</td>
<td>I was able to read and understand the advice</td>
</tr>
<tr>
<td>6.20 (1.08)</td>
<td>The advice felt personal and fitting to how I was feeling</td>
</tr>
<tr>
<td>6.33 (0.98)</td>
<td>I found the advice helpful</td>
</tr>
<tr>
<td>6.33 (0.98)</td>
<td>I learned something new from the advice</td>
</tr>
<tr>
<td>6.67 (0.62)</td>
<td>I liked the tone and style of the written advice</td>
</tr>
<tr>
<td>6.53 (0.83)</td>
<td>The length of the advice was appropriate</td>
</tr>
<tr>
<td>6.33 (0.98)</td>
<td>Overall, I found the graphs useful</td>
</tr>
<tr>
<td>6.47 (1.13)</td>
<td>Overall, the graphs were easy to understand</td>
</tr>
<tr>
<td>6.13 (1.06)</td>
<td>Overall, the graphs were visually appealing</td>
</tr>
<tr>
<td>6.60 (0.63)</td>
<td>The important behaviors checklist is helpful</td>
</tr>
<tr>
<td>6.73 (0.46)</td>
<td>I see benefits from using KIOS</td>
</tr>
<tr>
<td>6.13 (0.74)</td>
<td>Using KIOS improved my self-management skills</td>
</tr>
<tr>
<td>6.20 (0.77)</td>
<td>KIOS helped me to better understand my recovery process</td>
</tr>
<tr>
<td>6.47 (0.64)</td>
<td>Using KIOS helped me to become more aware of subtle shifts in my physical, mental, and social health</td>
</tr>
<tr>
<td>6.47 (0.83)</td>
<td>Using KIOS helped me to become more aware of practices that can restore or maintain balance in my physical, mental, and social health</td>
</tr>
<tr>
<td>5.73 (1.62)</td>
<td>Using KIOS helped me to identify some personal triggers</td>
</tr>
<tr>
<td>5.93 (1.10)</td>
<td>Using KIOS increased my overall satisfaction with my health status</td>
</tr>
<tr>
<td>6.13 (0.92)</td>
<td>Using KIOS adds value beyond the treatment that I typically receive</td>
</tr>
<tr>
<td>6.34 (0.25)</td>
<td>All questions</td>
</tr>
</tbody>
</table>

**KIOS Assessment Data**

The Healthy Behavior Checklist allows patients to check yes or no for engaging in the behavior at each assessment. The behaviors include opioid use, recreational drug use, alcohol use, exercising, good sleep habits, and avoiding triggers and high-risk situations. Only 1 participant self-reported 1 occurrence of unprescribed opioid use. This occurred at the baseline assessment, and no other illicit opioid use was reported by any participant during the study. Four participants reported 18
occurrences of alcohol use and 3 others reported 24 instances of recreational drug use (nonspecified); no participant reported both alcohol and recreational drug use.

**Figures 1 and 2** demonstrate how KIOS data can be used to track patient reporting over time. In **Figure 1**, the KIOS score and craving score were highly correlated, indicating that the symptom-reporting scales appear to be closely related to self-reported craving scores over time (Pearson $r=0.550$; $P<.001$). Two KIOS subscales were significantly lower at the end of the trial. Emotional (week 4; $P=.01$) and social (week 4; $P=.02$) subscales were significantly lower compared to the baseline (**Figure 2**).

**Figure 1.** KIOS score and craving score. Craving scores were significantly lower at all time points after baseline (week 4; $P=.013$). The KIOS score, an index comprised of the sum of all assessment variables except for craving, was also significantly lower (week 4; $P=.006$).
Discussion

Principal Findings

The KIOS app is an adjunctive tool to OUD treatment that gives patients 24/7 access to behavioral intervention strategies and supportive and reinforcing behavioral guidance that potentially could augment and improve response to routine clinic-based counseling interventions. In this study, participants rated the primary outcomes of acceptability, usability, and utility very highly. The scores on the SUS-modified were in the excellent range. The mean response to all KIOS questionnaire data was 6.32 out of a maximum of 7.0 (SD 0.25), indicating an overall strong agreement. Participants used KIOS on average about 3 times weekly, which is a very good indication of sufficient use of the app to generate relevant advice.

KIOS Acceptability

Participants on average used KIOS 3 times per week, which is the minimum requested in the study. This allowed sufficient time for relevant behavioral change between self-reports to generate pertinent recovery advice to the user. However, this may indicate a need to make the KIOS app more engaging, since participants could use the app daily if desired. Focusing the content on participants earlier in OUD treatment who may want more recovery advice or improving the experience by augmenting personalization could make the app more engaging [33,34]. Revisions to the app will include more explicit recommendations to continue in OUD treatment and more varied suggestions about mindfulness-oriented and positive self-care activities. Patients commented favorably on the individualized and specific nature of the advice and how this motivated and kept them on track in recovery. Some even commented that it seemed a counselor was responding to their self-report assessments in real time. This sense of copresence, of someone concurrently participating in the intervention with a psychological connection to the app user, was associated with improved satisfaction and efficacy in studies of other software apps [35,36].

KIOS Usability

The SUS-Modified is a well-validated instrument [23,24]. KIOS was rated as very usable for these patients who were recruited from a community-based methadone maintenance treatment program. One item stood out as less positive: how cumbersome the app was to use. Since ease of use is clearly important, revisions to the app will involve attention to this aspect of use (eg, ease of moving between the home screen and different app functions). The revised KIOS app will be used on a smartphone platform to improve this aspect of usability.

KIOS Utility

Utility was measured by a questionnaire specifically developed to evaluate the KIOS app. Participants strongly agreed on most questions related to usefulness, applicability, and relevance to recovery. Items that were less strongly endorsed were related to identifying personal triggers and improved health status. Since about one-third of the participants had been in treatment over 12 months, it is possible that identifying personal triggers...
was less of a treatment concern, and therefore, the mean score for this item was somewhat lower. In addition, since the study was only 4 weeks in duration, there was not much time to notice changes in general health status. Self-report of drug and alcohol use was modest in this sample, where only a minority of the participants was in treatment for less than 6 months. Much of the content of the app is aimed at changes in symptoms related to substance use problems; yet, participants who had minimal drug use problems and few mental health concerns rated the experience as helpful and relevant to their recovery. The randomized controlled study of KIOS will have a duration of 6 months and only focus on patients who are early in the treatment process. This will result in greater changes in drug use behavior and a longer period of time to capture potential changes in personal triggers and general health status. Revisions to the app will include more focus on positive behavioral changes to improve patient self-management and social support for recovery.

KIOS Assessment Self-Reports

The scores derived from the KIOS self-report generated potentially useful data for both the KIOS user and the treatment team. As expected, the KIOS score and craving score were highly correlated. The correlation between these 2 measures suggests that the assessment variables, which collectively form the KIOS score capture relevant emotional, social, and physical phenomena that correspond with self-reported craving. Figures 1 and 2 demonstrate how KIOS data can usefully track patient reporting over time. These results could also be reviewed by the patient’s counselor to reinforce behavioral treatment planning and maximize the limited counseling time available at each appointment to assist the patient with this aspect of their care. The KIOS journal feature is available to help the patient record and organize pertinent thoughts and reflections, and the behavioral checklist helps to organize and track drug use and positive behaviors associated with improved recovery outcomes.

KIOS advice is designed to track patient self-report items related to OUD recovery over time. The KIOS app modifies the advice given to the user based on the person’s changes in self-report at each assessment. Advice is then directed specifically to the symptoms of most concern to the user. KIOS can draw the patient’s attention to triggers and potential problem behaviors with helpful and supportive advice in real time to reinforce recovery activities and goals at times of vulnerability to drug use. It may advise reaching out to support people or treatment providers depending on the type of a variable or increasing intensity of symptom reporting (eg, for increasing depression and insomnia consulting a medical provider), further encouraging integrated use of comprehensive treatment resources. It gives specific, evidence-informed suggestions for various symptoms including sleep hygiene, stress management, avoiding triggers, and pain reduction behaviors.

Expanded access to digital behavioral interventions has the potential to bridge a significant treatment gap due to the lack of counseling resources in many OUD treatment settings. However, there are very few digital behavioral apps that specifically address OUD [13,17,37,38]. Due to the nearly uniformly serious nature of this disorder, it is unlikely that a digital app by itself would be potent enough to help a person manage this disorder without professional treatment participation. The available apps focus primarily on connecting users to treatment services and peer support, making available CBT educational modules, using self-report check-ins, and using contingency management interventions. Their use has demonstrated significant improvements in some treatment outcomes [39]. For example, studies of reSET-O showed efficacy regarding opioid use or abstinence and in reinforcing and increasing treatment activity frequency in controlled trials [40]. The web-based CBT4CBT app improved retention and reduced drug use when combined with office-based treatment as usual [41]. It is one of the few digital interventions that has shown improved efficacy compared to in-person CBT in randomized controlled trials and demonstrates the potential power of these types of treatment augmentations [41,42]. The results of a randomized controlled trial of the A-CHESS digital app for use as an adjunct treatment in OUD did not demonstrate significant benefits compared to the control condition [16,43]. However, some results indicated that for specific subpopulations there could be benefits.

Other apps use a therapeutic relational approach using a conversational agent, such as those using a chatbot platform [12,44]. Some of these interventions have demonstrated efficacy [45]. They are currently used to help patients with a variety of mental health issues, although few have an evidence base [13,45-47]. The use of the W-SUDs modification of the Woebot app [12,13] has been associated with self-reported reductions in drug use and urges to use drugs and improvement in anxiety in users with substance use concerns [13]. The Woebot gives the impression of a therapeutic coach in the conversation generated by the chatbot. There is evidence from prior studies that some users are more likely to disclose information [48] to nonhuman apps, and that a strong therapeutic alliance can form [49-51]. Although not measured in this study, some participants commented on how KIOS responses seemed like real-time exchanges with a therapist. Future studies of KIOS should examine this aspect of the UI.

The study limitations include the short study duration in this trial to primarily test the acceptability, utility, and usability of the software. There was no control group, and it was not compared to another treatment condition or substance use treatment app to determine the specific benefits of KIOS compared to other apps or interventions for persons with substance use problems. The small sample size limits generalizability in several ways: to other opioid agonist treatment clinics or office-based OUD treatment, other patient samples that may be less interested in using digital apps, and in the ability to detect statistical differences between study variables. The patient focus group comments reflected the views of persons who were interested in using the app and potentially predisposed to having a positive user experience. KIOS relies on patient self-reporting to generate advice, so if a patient did not report accurate information, then they would not get accurate advice. This could be a problem if the patient’s counselor had access to the patient’s data. For example, if the patient was not ready to reveal their substance use, they might not give an
accurate self-report. This should also be examined in future studies of the app.

Conclusions
The KIOS app offers a different approach to the digital app for augmenting outpatient treatment for OUD. Potentially, the advice generated from the app may be more specific to the user because self-reported symptom clusters are used to identify appropriate behavioral interventions selected from evidence-based literature. It also has the potential advantage of using the therapeutic relational coaching approach that many users find more engaging.

Current plans for the development of KIOS include adding varied advice features, incorporating and reinforcing more positive recovery behaviors (eg, positive social support), and pursuing Food and Drug Administration approval as a prescription medical device. Although KIOS was studied in methadone maintenance clinics, it may be a very helpful tool in office-based buprenorphine addiction medicine treatment settings that often also have limited counseling availability. A National Institute on Drug Abuse–funded randomized controlled trial of KIOS is currently being conducted to evaluate its efficacy.

Acknowledgments
The authors thank the patients and staff of CleanSlate, The Center for Health Care Services, and Community Medical Services in San Antonio, TX, and Cedar Park, TX, for their assistance and participation in the study and to the consultants on the panel who assisted in developing the KIOS software app. The authors gratefully acknowledge Charles L Bowen, MD, the principal investigator on the grant that supported this work. Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under award number R43MH119672 (principal investigator: Charles Bowden). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to the proprietary nature of the device. Reasonable requests from the KIOS inventor, HRP, should be addressed to him specifically.

Conflicts of Interest
GS and LHS work for or own stocks or stock options of the company that developed or owned the app, product evaluated, and discussed in the paper. HRP, GS, and LHS were involved in the development of the innovation, software, and app they have been evaluating. VLK and JSP have no financial conflicts of interest.

References


**Abbreviations**

CBT: cognitive behavioral therapy
Behavioral Activation Mobile App to Motivate Smokers to Quit: Feasibility and Pilot Randomized Controlled Trial

Belinda Borrelli, BA, MA, PhD; Y Kiera Bartlett, BS, MSc, PhD; Daniel Fulford, BA, PhD; Greg Frasco, BS; Christopher J Armitage, BA, PhD; Alison Wearden, BSc, MSc, PhD

1Center for Behavioral Science Research, Henry M. Goldman School of Dental Medicine, Boston University, Boston, MA, United States
2Manchester Centre for Health Psychology, School of Health Sciences, University of Manchester, Manchester, United Kingdom
3Department of Psychological & Brain Sciences, Boston University, Boston, MA, United States
4Sargent College of Health & Rehabilitation Sciences, Boston University, Boston, MA, United States
5Rafik B. Hariri Institute for Computing and Computational Science & Engineering, Boston University, Boston, MA, United States
6Greater Manchester Patient Safety Research Collaboration, National Institute for Health and Care Research, Manchester, United Kingdom

Corresponding Author:
Belinda Borrelli, MA, PhD
Center for Behavioral Science Research
Henry M. Goldman School of Dental Medicine
Boston University
Floor 3
560 Harrison Ave
Boston, MA, 02118
United States
Phone: 1 617 358 3358
Fax: 1 617 638 6381
Email: belindab@bu.edu

Abstract

Background: Behavioral activation (BA) is an evidence-based treatment for depression that fosters engagement in values-based activities to increase access to positive reinforcement. Depressed mood has been shown to hinder smoking cessation.

Objective: This study determined the feasibility and preliminary efficacy of a mobile app to motivate smokers to quit by using BA and integrating motivational messages to quit smoking.

Methods: Adult smokers (N=56; mean age 34.5, SD 9.52 years) who were not ready to quit smoking within 30 days were recruited from advertisements and randomized to either 8 weeks of the BA app (set 2 values-based activities per week+motivational messages+feedback on changes in smoking, mood, and values-based activities) or the control group (no app; received resources for quitting smoking). All participants completed the baseline and end-of-treatment web-based questionnaires. Controls also completed weekly web-based assessments, and BA app participants completed assessments through the app.

Results: There were no dropouts and only 2 participants in each condition did not complete the end-of-treatment questionnaire. The results demonstrated that it is feasible to recruit smokers who are unmotivated to quit into a smoking cessation induction trial: 86% (57/66) of eligible participants were randomized (BA app: n=27; control: n=29). Participants reported high levels of satisfaction: 80% (20/25) of participants said they would recommend the BA app, there were moderate-to-high scores on the Mobile App Rating Scale, and 88% (22/25) of participants rated the app 3 stars or higher (out of 5). There were high levels of BA app engagement: 96% (26/27) of participants planned activities, and 67% (18/27) of participants planned 7 or more activities. High engagement was found even among those who were at the highest risk for continued smoking (low motivation to quit, low confidence to quit, and high negative affect). The results provided support for the hypothesized relationships between BA constructs: greater pleasant activity completion was associated with greater positive affect (b=0.37, SE 0.21; 95% CI –0.05 to 0.79; P=.08), and greater positive affect tended to predict fewer cigarettes smoked the next day (b=-0.19, SE 0.10; 95% CI –0.39 to 0.01; P=.06). Additionally, a greater number of activities planned was associated with lower negative affect (b=-0.26, SE 0.15; 95% CI –0.55 to 0.04; P=.09). Overall, 16% (4/25) of BA app participants set a quit date versus 4% (1/27) among controls, and there were promising (but not significant) trends for motivation and confidence to quit.
Conclusions: The findings suggest that a mobile app intervention can be made appealing to smokers who are unmotivated to quit by focusing on aspects most important to them, such as mood management. This theory-based intervention has shown some initial support for the underlying theoretical constructs, and further efficacy testing is warranted in a fully powered trial.

(KEYWORDS smoking cessation; mobile app; motivation; depressed mood; depression; behavioral activation; negative affect; positive affect; quit smoking; health behavior change

Introduction

Currently available evidenced-based treatments are specifically designed for smokers who are ready to quit within 30 days, which constitute only 12% of smokers [1]. Evidence-based strategies are needed to reach the remaining population of unmotivated smokers and accelerate their motivation to quit before smoking-related illness occurs (or worsens).

Because unmotivated smokers are not likely to seek treatment, the majority of research has focused on integrating smoking cessation into infrastructures that are frequented by smokers, such as health care or worksites [2-6]. Mobile apps are another channel to reach smokers who are unmotivated to quit. However, while there are studies examining the effectiveness of mobile apps for smoking cessation [7], to our knowledge, there are no studies that focus on using mobile apps for motivating smokers who are not ready to quit. This is likely because it is difficult to encourage unmotivated smokers to use a smoking cessation app because the content is not personally relevant to them. Given prior research showing that the use of mobile apps increases with greater personal relevancy and perceived utility [8], it is important to find out what is important to unmotivated smokers, provide information on that topic using a mobile app, and then gradually weave in a smoking cessation intervention to increase intervention engagement and motivate quitting. This is consistent with the “foot-in-the-door” technique from social psychology [9]. For example, in a 1 smoking cessation induction trial, smokers were offered asthma education on how to prevent asthma attacks in their children with asthma who were recently admitted to urgent care for an asthma exacerbation. The intervention largely focused on what they cared about most (preventing future asthma attacks in their children), but smoking cessation and how to quit were gradually woven into the intervention over time [10-13].

We previously conducted a series of focus groups with smokers who were unmotivated to quit and found that the most important thing on their minds was managing mood and stress, and that they would be therefore interested in a mobile app that could help them with mood management [14]. This is consistent with results from our large prior quantitative study with smokers unmotivated to quit, which showed a high prevalence of depression and depressed mood [15]. Therefore, focusing on things that are most important to unmotivated smokers (such as stress and mood) may increase app engagement and prime them to consider quitting smoking. To date, smoking cessation interventions using smartphone apps alone have not been successful [7]. One reason for this could be the lack of focus on content that would most interest unmotivated smokers.

Therefore, we developed a mobile app based on behavioral activation (BA) [16-18], hypothesizing that increases in positive affect and decreases in negative affect could prime unmotivated smokers to consider quitting. BA is an evidenced-based treatment for depression and depressed mood that has been successfully used in both psychiatric and medical populations [19-22] and has been shown to be superior to medication [19]. BA has also been shown to be effective for depression when delivered via a web-based platform [23]. BA purports that depression results from chronically low levels of, or loss of, positive environmental reinforcement. The lack of both pleasant events and positive reinforcement is associated with depression and negative avoidant coping behaviors (eg, smoking) [17]. BA improves mood through strategies to increase activation (eg, setting goals to engage in activities that are consistent with personal values and goals), which in turn increases access to natural sources of positive reinforcement (eg, other people, hobbies, and activities) [24].

In a series of prior studies, we iteratively and collaboratively developed a BA-based mobile app with smokers who were unmotivated to quit, through focus groups, a “think aloud study,” and a 1-month single-arm usability trial [14]. This current study aimed to conduct a pilot randomized trial of the BA app with smokers who were unmotivated to quit in order to assess (1) the feasibility (eg, enrollment) of the BA app, (2) satisfaction and engagement with the BA app, (3) if unmotivated smokers complete BA app intervention activities despite risk factors for continued smoking (low motivation to quit, low confidence to quit and high negative affect), and (4) preliminary effectiveness of the BA app on smoking behavior and on motivation and confidence to quit smoking. We hypothesized that (1) smokers who are unmotivated to quit will engage with the BA app, (2) engagement with the BA app will be associated with lower negative affect and greater positive affect, and (3) those who are randomized to the BA app will show positive changes in smoking (eg, set quit date and reduce the number of cigarettes) and in smoking-related variables (greater motivation and confidence to quit).
smoking to be in the study and that the study focused on improving the well-being of smokers. They were also told that the study would provide them with tips and resources to quit smoking in case they became motivated to quit smoking. Eligible participants who completed a web-based consent form were asked to complete a web-based baseline questionnaire. The inclusion criteria were (1) aged 18 years or older, (2) current smoker (>100 in lifetime and ≥3 cigarettes per day), (3) not planning to quit smoking within 30 days, and (4) owned and regularly used an Android or iPhone and had previously downloaded and used an app. Exclusion criteria were (1) inability to read, speak, or understand English to a sufficient level to understand the information sheet and app; (2) self-report that they do not wish to stop smoking at any point in the future; and (3) current involvement in any other research about smoking.

**Ethical Considerations**

The study was approved by the University of Manchester Research Ethics Committee 5 (reference 2017-0128-2605). Participants were sent both the participant information sheet and the consent statements to consider during the recruitment process (Multimedia Appendix 1). All participants then completed a web-based consent form prior to completing the baseline questionnaire. Data were collected and stored in accordance with the Data Protection Act and the university-approved Data Management Plan. All data are anonymous, identified only by a participant number. Participants received a £20 (US $25.30) love2shop voucher for completing the follow-up questionnaire (valid at a range of high street stores and retailers; Multimedia Appendix 2).

**Procedure**

**Design**

Participants were randomized to either the BA app or to the control condition using a parallel 1:1 allocation ratio. The random sequence of the BA app or control (represented by a 1 or a 2) was generated by GraphPad [25] in blocks of 10 participants. As participants consented, they were allocated to the next available number in the sequence. Study staff could not be masked to treatment conditions because some participants needed assistance with downloading the app. These staff were not involved in collecting the outcome data. Participants were allocated to either use the BA app for 8 weeks or to receive an information sheet with resources related to quitting smoking (control). Both groups completed weekly assessments and received the same information and resources about quitting (either through the app for the BA app group or as an email attachment for the control group). Participant flow is outlined in Figure 1.
BA App Intervention

Programming for the mobile app was conducted by the Software and Innovation Lab at Boston University. If needed, research assistants guided participants with app downloads from the Google Play Store or Apple Store (via phone consultation). Once downloaded, participants were led through the app’s tutorial on its features and on the intervention to ensure that staff were not involved in the intervention. The BA app helped participants identify, schedule, and complete values-based activities. Specifically, participants identified their values (eg, doing things with others, volunteer work, hobbies, being spiritual, and community activism) through the app, which allowed the app to suggest activities consistent with these values (algorithms available upon request). The app then encouraged participants to set at least 1 activity goal in each of 2 categories once per week: “pleasurable activities” and “challenging and meaningful activities.” Pleasurable activities were defined as those that provide a short, pleasant break from the pace of life, such as reading, taking baths, or watching movies [26]. Challenging activities were defined as those that absorb the mind, require concentration, and promote a sense of accomplishment (eg, engaging in a hobby or skill) and a feeling that “time flies” when engaged with that activity [26].

Meaningful activities were defined as those in which people feel like they are making a difference or activities that involve spiritual endeavors (eg, volunteer work and attending religious services) [26]. Each week, participants scheduled their chosen activities using the calendar feature in the app. The app also allowed participants to set “custom” activities (outside of the ones suggested by the app) and subgoals if the activity was too large of a goal for the week (eg, “learn a new language”). Participants received app notification “reminders” on the days they had scheduled the activities and were also prompted to record whether or not they completed the activity on the scheduled day. Participants received “trophies” for progress on activities, which were displayed on a rewards screen on which the number of cigarettes they smoked that week and mood level was also displayed.

The identification, scheduling, and completion of values-based activities were used as a “foot-in-the-door” approach to provide motivational messages about smoking, previously developed from focus groups with smokers who were unmotivated to quit and designed to address barriers to quitting and myths about smoking cessation medications [14]. The messages were delivered through “app notifications,” and the frequency of delivery was calibrated to participants’ level of motivation to
quit, which was assessed weekly through the app. If participants were planning on quitting within 30 days, they were sent notifications to review the “Resources to Quit” sections within the app which listed evidenced-based smoking cessation apps, quitlines, and local resources for quitting. Participants who were motivated to quit within 30 days were additionally sent smoking cessation messages from a different message bank within the app that focused on helping them plan a quit day and strategies for quitting.

Control
Control participants received an information sheet with resources related to quitting smoking, for example, links to smoking cessation apps and information about stop-smoking medications. This was the same information that was available to the BA app group, on the resources page with the app.

Measures
Overview
All participants were asked to complete a web-based baseline and end-of-treatment questionnaires for which they received one £20 (US $25.30) gift card. Both groups completed brief weekly surveys: control participants were emailed a link to a web-based survey and BA app participants answered the daily and weekly questions through the app.

Participant Characteristics
Gender, age, level of education, ethnicity, household income, and employment status were assessed. Nicotine dependence was assessed with a single item from the Fagerström Test for Nicotine Dependence [27], which is highly correlated with the full scale [28].

Satisfaction With the BA App
The Mobile App Rating Scale (MARS), which has shown excellent internal consistency and reliability [29], was used to assess app quality, perceived impact of the app, and app satisfaction. The MARS app quality scale consisted of 4 subscales (engagement, functionality, aesthetics, and information), each rated on a 1-5 scale (higher scores=increasing satisfaction). The MARS Perceived Impact subscale was measured on a 1-5 scale (1=strongly disagree to 5=strongly agree) and was adapted to the content of the BA app (eg, the app increased my knowledge of quitting smoking and the app increased my motivation to quit smoking). Participants were asked to give their “star rating” as a measure of their satisfaction on a 1-5 scale (1=one of the worst apps I have used through 5=one of the best apps I have used), as well as report if they would recommend the app to people who might benefit from it (1=not at all to 5=definitely).

Engagement With the App
Participants set weekly values-based activity goals through the app (“planned activities”) and reported on whether or not those activities were completed, also through the app on a weekly basis. The type of activity was also assessed (pleasant or challenging and meaningful) by the app. The app prompted participants to enter the above data and sent several notification reminders to do so.

Predictors of Engagement
Depressed mood was measured with the 10-item Center for Epidemiologic Studies Depression Scale (CES-D-10) [30,31]. Scores of 10 or higher indicate the presence of significant depressive symptoms [32]. Mood was also measured with the Positive and Negative Affect Schedule [33], which has 20 items that describe feelings and emotions (10 positive and 10 negative), each measured on a 5-point scale (not at all to extremely). The above variables were given at baseline, weekly, and at the end of treatment for both groups. In the BA app, the following variables were assessed on a daily basis: the number of cigarettes per day, motivation to quit smoking (1-10), confidence to quit smoking (1-10), and level of positive mood (happy, excited; 1-10 scale; higher=more positive) and negative mood (eg, sad, bored, or angry; 1-10 scale; higher=more negative).

Smoking-Related Variables
The following variables were assessed at baseline, weekly, and at the end of treatment: the number of cigarettes smoked per day, 7-day point prevalence (have you smoked any cigarettes at all in the past 7 days, even a puff?), motivation to quit (1-10 scale; 1= not at all motivated to quit, through 10=very much want to quit), and confidence to quit (1-10 scale; 1 is “not at all” confident to quit through 10 “very confident to quit”).

Data Analysis
Descriptive analyses were conducted, and distributions of variables were inspected for violations of normality. Group differences were evaluated using chi-square and 2-tailed t tests (in all instances), for categorical and continuous variables, respectively. For the multilevel models, lead variables were created for each relevant outcome (eg, number of cigarettes) to assess prospective associations at the day and week level. Pleasant, meaningful and challenging, and the combined sum activities (ie, summed pleasant, meaningful, and challenging activities) were aggregated at the day and week level.

Due to the small sample of this pilot study, our primary analyses focused on describing trends (eg, examination of the direction of coefficients) rather than focusing on statistical associations. Analyses were conducted in R (version 4.2.3; R Foundation for Statistical Computing) using the stats package for 2-tailed t tests (in all instances), ANOVAs, and linear regressions and the lme4 package [34] for multilevel model analyses. Group differences between the BA app and controls on baseline variables were examined with independent 2-tailed t tests (in all instances). Then, descriptive statistics were examined for the engagement and satisfaction of those in the BA app condition. While an intention-to-treat approach was not used given the preliminary nature of the study, differences in baseline variables of interest between participants who provided at least 33% complete weekly data for between-group analyses and 33% complete daily data for within-group analyses were examined.

Analyses first investigated satisfaction with the BA app through descriptive statistics and then investigated engagement with the BA app through both descriptive statistics and through prospective linear regressions focused on whether or not those
at greatest risk for continued smoking at baseline (ie, negative affect, low confidence, and low motivation) would engage with the intervention (activity completion). Second, multilevel modeling, which accounts for the clustering of repeated observations, was used to examine the within-person effect or the degree to which the outcome of interest was predicted by a given person’s deviation from their typical level. Among those within the BA app group, concurrent (same day or week) and prospective (next day or week) within-person analyses of affect, motivation, and confidence to quit; activity planning and completion; and smoking were examined. Lastly, changes in key outcome variables (eg, cigarettes smoked, motivation, confidence, and affect) from pre- to postintervention were examined by group using general linear models.

Results

Preliminary Analyses

As shown in Figure 1, smokers who were not motivated to quit were willing to enroll in our trial: of those who were eligible, 86% (57/66) of participants consented and were randomized to either the BA app (n=27) or to the control group (n=29). Only 2 people in each condition did not complete the end-of-treatment questionnaire (Figure 1). Participant characteristics are displayed in Table 1. Participants smoked an average of 13.4 cigarettes per day, and 28 (50%) out of 56 participants reported smoking their first cigarette within 30 minutes of waking. Participants had low levels of both motivation to quit (mean 4.3, SD 2.2) and confidence to quit (mean 3.8, SD 2.4), as well as high levels of depressed mood (CES-D-10: mean 9.0, SD 5.4). There were no significant differences between groups on any baseline variable (all \( P > .05 \)), and 7 (12%) out of 56 participants had \( > 33\% \) missing weekly data. There were no significant differences in demographics or smoking behavior between those with and without missing data. There were no reported side effects, technical problems, or privacy breaches during the course of the study.

Table 1. Participant demographics, smoking behavior, and mood.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control (n=29)</th>
<th>BA(^a) app (n=27)</th>
<th>Overall (N=56)</th>
<th>Tests of difference, (N=56)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t test (df=54)</td>
<td>Chi-square (df=1)</td>
<td>( r ) test (df=54)</td>
<td>( P ) value</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>34.24 (8.99)</td>
<td>34.85 (10.23)</td>
<td>34.54 (9.52)</td>
<td>N/A(^b) 0.24 0.02</td>
</tr>
<tr>
<td>Female, n (%</td>
<td>21 (72)</td>
<td>20 (74)</td>
<td>41 (73)</td>
<td>N/A</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (7)</td>
<td>3 (11)</td>
<td>5 (9)</td>
<td>0.01 N/A</td>
</tr>
<tr>
<td>Asian</td>
<td>2 (7)</td>
<td>1 (4)</td>
<td>3 (5)</td>
<td>0.01 N/A</td>
</tr>
<tr>
<td>Black</td>
<td>24 (83)</td>
<td>20 (74)</td>
<td>44 (79)</td>
<td>0.63 N/A</td>
</tr>
<tr>
<td>White</td>
<td>1 (3)</td>
<td>3 (11)</td>
<td>4 (7)</td>
<td>0.35 N/A</td>
</tr>
<tr>
<td>Other</td>
<td>11 (38)</td>
<td>11 (41)</td>
<td>22 (39)</td>
<td>0.05 N/A</td>
</tr>
<tr>
<td>Cigarettes smoked per day, mean (SD)</td>
<td>12.14 (6.93)</td>
<td>14.74 (10.39)</td>
<td>13.39 (8.79)</td>
<td>N/A 1.11</td>
</tr>
<tr>
<td>Confidence to quit, mean (SD)</td>
<td>3.93 (2.53)</td>
<td>3.78 (2.38)</td>
<td>3.86 (2.44)</td>
<td>N/A 0.23</td>
</tr>
<tr>
<td>Motivation to quit, mean (SD)</td>
<td>4.62 (2.29)</td>
<td>4.21 (2.17)</td>
<td>4.32 (2.23)</td>
<td>N/A 1.04</td>
</tr>
<tr>
<td>First cigarette within 30 minutes of waking, n (%)</td>
<td>15 (52)</td>
<td>13 (48)</td>
<td>28 (50)</td>
<td>0.72 N/A</td>
</tr>
<tr>
<td>CES-D(^c), mean (SD)</td>
<td>8.55 (4.72)</td>
<td>9.48 (6.14)</td>
<td>9 (5.42)</td>
<td>N/A 0.64</td>
</tr>
<tr>
<td>CES-D ( \geq 10 ), n (%)</td>
<td>9 (31)</td>
<td>12 (44)</td>
<td>21 (38)</td>
<td>1.07 N/A</td>
</tr>
<tr>
<td>PANAS(^d) positive mood, mean (SD)</td>
<td>31.47 (8)</td>
<td>31.18 (7.21)</td>
<td>31.33 (7.54)</td>
<td>N/A 0.14</td>
</tr>
<tr>
<td>PANAS negative mood, mean (SD)</td>
<td>20.06 (7.02)</td>
<td>20.98 (8.68)</td>
<td>20.20 (7.80)</td>
<td>N/A 0.44</td>
</tr>
</tbody>
</table>

\(^{a}\)BA: behavioral activation.

\(^{b}\)N/A: not applicable.

\(^{c}\)CES-D: Centers for Epidemiological Studies, Depression Scale.

\(^{d}\)PANAS: Positive and Negative Affect Schedule.
Mobile App Satisfaction and Engagement

**Satisfaction**

Satisfaction scores on the MARS subscales were within the “acceptable to very good” range (Table 2). Participants indicated moderate to strong agreement regarding the perceived impact of the app on awareness, knowledge, positive attitude toward quitting, intention to quit, further help seeking, and willingness to reduce their smoking (Table 2). The majority of the participants (21/25, 80%) said they would recommend the app to several people, many people, or “everyone” (Table 2). Three participants rated the app as 2 stars or less, while 10 participants gave it a 3-star rating, 11 gave it a 4-star rating, and 1 gave it a 5-star rating.

<table>
<thead>
<tr>
<th>Table 2. Satisfaction with the BA app.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td><strong>App quality subscale, mean (SD)</strong></td>
</tr>
<tr>
<td>Engagement</td>
</tr>
<tr>
<td>Functionality</td>
</tr>
<tr>
<td>Aesthetics</td>
</tr>
<tr>
<td>Information</td>
</tr>
<tr>
<td>Overall</td>
</tr>
<tr>
<td><strong>Perceived impact of the app, mean (SD)</strong></td>
</tr>
<tr>
<td>The app increased my awareness of the importance of setting values-based activities to help me think about quitting smoking</td>
</tr>
<tr>
<td>The app increased my knowledge of quitting smoking</td>
</tr>
<tr>
<td>The app has made my attitude toward quitting smoking more positive</td>
</tr>
<tr>
<td>The app has increased my motivation to quit</td>
</tr>
<tr>
<td>The app would encourage me to seek further help to quit smoking (if I decided to quit)</td>
</tr>
<tr>
<td>Use of this app will decrease my smoking</td>
</tr>
<tr>
<td><strong>Recommend to people, n (%)</strong></td>
</tr>
<tr>
<td>Not at all</td>
</tr>
<tr>
<td>Very few people</td>
</tr>
<tr>
<td>Several people</td>
</tr>
<tr>
<td>Many people</td>
</tr>
<tr>
<td>Everyone</td>
</tr>
<tr>
<td><strong>Overall star rating, n (%)</strong></td>
</tr>
<tr>
<td>1 Star</td>
</tr>
<tr>
<td>2 Stars</td>
</tr>
<tr>
<td>3 Stars</td>
</tr>
<tr>
<td>4 Stars</td>
</tr>
<tr>
<td>5 Stars</td>
</tr>
</tbody>
</table>

*aBA: behavioral activation.*

**Engagement With the App and Predictors of Engagement**

Participants demonstrated high levels of engagement with the app, despite being not motivated to quit smoking. All participants in the BA app group planned at least 1 activity. A total of 253 activities were planned (132 pleasant and 121 meaningful and challenging); 26 (96%) out of 27 participants planned at least 1 activity after the first week and 19 (70%) out of 27 participants planned 7 or more activities over the course of the 8-week intervention. A total of 138 activities were completed (68 pleasant and 70 meaningful and challenging). Four participants did not complete any activity (either pleasant or challenging and meaningful).

Next, we explored whether or not those at risk for continued smoking (lower motivation and confidence to quit; high negative affect) would actually engage with the intervention (eg, complete values-based activities; Table 3). Lower motivation to quit at baseline was prospectively associated with more total activities completed during the intervention ($b=-0.71$, SE 0.26; 95% CI –1.23 to –0.20; $P=0.01$), more pleasant activities completed during the intervention ($b=-0.30$, SE 0.14; 95% CI –0.57 to –0.02; $P=0.04$), and more meaningful and challenging activities.
completed during the intervention \((b=-0.42, \text{ SE } 0.14; 95\% \text{ CI } -0.70 \text{ to } -0.14; P=.01)\). Lower confidence to quit at baseline was prospectively associated with a greater number of total activities completed during the intervention \((b=-0.53, \text{ SE } 0.25; 95\% \text{ CI } -1.03 \text{ to } -0.04; P=.03)\) and more pleasant activities completed during the intervention \((b=-0.33, \text{ SE } 0.12; 95\% \text{ CI } -0.57 \text{ to } -0.08; P=.01)\), but not more meaningful and challenging activities completed. Lower negative affect at baseline was prospectively associated with a greater number of total activities completed during the intervention \((b=-1.47, \text{ SE } 0.69; 95\% \text{ CI } -2.82 \text{ to } -0.12; P=.03)\), more meaningful and challenging activities completed during the intervention \((b=-0.86, \text{ SE } 0.38; 95\% \text{ CI } -1.60 \text{ to } -0.12; P=.02)\), and a trend for more pleasant activities completed during the intervention \((b=-0.61, \text{ SE } 0.36; 95\% \text{ CI } -1.36 \text{ to } 0.11; P=.10; \text{ not statistically significant})\).

### Table 3. Prediction of intervention engagement from baseline variables \((n=25)\).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Total number of activities completed</th>
<th>Number of pleasant activities completed</th>
<th>Meaningful and challenging activities completed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(b)</td>
<td>(\text{SE})</td>
<td>95% CI</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------</td>
<td>--------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Total number of activities completed</td>
<td>Baseline positive affect</td>
<td>0.02</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>Baseline negative affect</td>
<td>(-1.47)</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Baseline motivation</td>
<td>(-0.71)</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Baseline confidence</td>
<td>(-0.53)</td>
<td>0.25</td>
</tr>
</tbody>
</table>

### Relationships Between Activities and Predictors (BA App Group Only)

Participants who completed a greater number of pleasant activities tended to report greater positive affect on the same day \((b=0.37, \text{ SE } 0.21; 95\% \text{ CI } -0.05 \text{ to } 0.79; P=.08)\). Participants who reported higher levels of positive affect on 1 day tended to smoke fewer cigarettes the next day \((b=-0.19, \text{ SE } 0.10; 95\% \text{ CI } -0.39 \text{ to } 0.01; P=.06)\). On days in which individuals experienced greater negative affect (ie, worsened negative affect), they smoked more cigarettes the next day \((b=0.16, \text{ SE } 0.08; 95\% \text{ CI } 0.01-0.32; P=.04)\). A greater number of activities planned in a given week, compared to an individual’s typical level, tended to predict lower negative affect (ie, more favorable negative affect) the same week \((b=-0.26, \text{ SE } 0.15; 95\% \text{ CI } -0.55 \text{ to } 0.04; P=.09)\), although this was not significant. The total number of activities and activity types did not significantly predict the number of cigarettes smoked the next day.

### Smoking-Related Variables and Group Differences

Table 4 displays changes in smoking-related variables from baseline to follow-up for both groups. There was a trend for fewer cigarettes smoked per day across the intervention period \((b=-0.19, \text{ SE } 0.11; P=.07; \text{ not statistically significant})\) for both groups. There was no group main effect \((b=1.67, \text{ SE } 1.25; P=.19)\) or difference in slope between groups \((b=0.02, \text{ SE } 0.17; P=.91)\). Across both groups, motivation to quit increased significantly over time \((b=0.19, \text{ SE } 0.06; P=.01)\). There was also a main effect of group \((b=1.61, \text{ SE } 0.42; P=.01)\), indicating that participants in the BA app group had higher levels of motivation to quit across the intervention period. The slope for group was not significant, however, suggesting no group differences in the rate of increase in motivation over time \((b=-0.05, \text{ SE } 0.09; P=.63)\). While there was a trend for higher confidence to quit across the intervention period in the BA app group \((b=0.90, \text{ SE } 0.51; P=.08; \text{ not statistically significant})\), confidence did not change significantly during the intervention period between groups \((b=0.07, \text{ SE } 0.05; P=.19)\).

At the end of the intervention, 4 of 25 (16%) participants in the app group and 1 participant of 27 (3.7%) in the control group set a quit date \(\chi^2=2.25; P=.13\). Two (25%) of 8 participants in the app group and 1 (8%) of 12 participants in the control group reported that they planned to quit in the next 30 days \(\chi^2=1.05; P=.31)\).
that there is sufficient evidence to warrant further testing in a larger trial.

The novel aspect of the app is that it takes a “foot-in-the-door” approach by focusing on aspects that we found matter most to unmotivated smokers (stress and mood management) with the idea that addressing these risk factors for continued smoking will increase smokers’ receptivity to pushed content on smoking cessation. While it may seem counterintuitive to enroll unmotivated smokers into a smoking cessation app, we were able to demonstrate that our approach is feasible, given our high rate of enrollment and no dropout. One other study has also demonstrated high enrollment of unmotivated smokers by focusing on aspects that matter most to them, and then weaving in smoking cessation messages [12]; however, there was no digital component in this study. Additionally, our sample was comprised of smokers with high levels of depressed mood (21/56, 38% scored above the cutoff for depression on the CES-D) and low motivation and confidence to quit, indicating that we were able to attract smokers who are at risk for continued smoking.

Our results showed moderate to high levels of satisfaction and high levels of engagement with the intervention (planning and completing activities). High levels of engagement with the app also mean that these smokers, who were unmotivated to quit, received our intermittent, pushed messages about smoking cessation; messages that they would not otherwise receive. The messages focused on key barriers to quitting and were based on state-of-the-art motivational approaches [12,36] as well as our previous mixed methods research with the target population [14]. We believe that this approach was successful because there were no study dropouts.

Although counterintuitive, those with greater risk factors for continued smoking (low motivation to quit, low confidence to quit, and high negative affect) completed a greater number of activities over the course of the study (ie, engaged with the hypothesized active ingredients of the intervention). These findings have several implications. First, just because...
smokers are unmotivated to quit, or have high risk factors for continued smoking, does not mean that they are not willing to engage with an intervention. This finding lends support to continue with future research in developing digital behavior change solutions for those who are unmotivated to change. Second, these associations were prospective and provided support for the theoretical model underpinning the intervention. Finally, this finding dovetails with a previous study with smokers, where smokers who were unmotivated to quit were more likely to quit with an intensive motivational interviewing intervention than a less intensive one [13].

Consistent with BA theory, we hypothesized that completing a greater number of activities would lead to higher positive affect and lower negative affect and that positive changes in mood would be associated with less smoking. We found support for some but not all of these associations. We found that pleasant activities were associated with greater positive affect and that positive affect predicted smoking fewer cigarettes the next day. Although significance was borderline due to the small sample, the fact that these trends were based on a theoretical model specified a priori give support for future study of these relationships in a larger sample. Planning more activities was also associated with lower levels of negative affect (although only a trend), but lower levels of negative affect were significantly and prospectively associated with smoking fewer cigarettes. This set of findings indicates support for key relationships with the BA model and therefore warrants further investigation.

As with any pilot feasibility study, there are limitations in what can be concluded from the study. For example, the effectiveness of the app on smoking behavior cannot be determined due to lack of power, although there were promising trends in the hypothesized direction regarding intentions to quit. While we can conclude that the results are promising and warrant a larger trial, we cannot conclude that the cross-sectional and prospective relationships between hypothesized variables will be demonstrated in a larger trial. In general, our effects were small in size and should be interpreted with caution. The trial was also conducted in the United Kingdom with a relatively racially and ethnically homogeneous population, so it is unclear if smokers who are unmotivated to quit in other countries and smokers of other races and ethnicities would find the app equally satisfactory or if they would engage with the app to the same extent. Additionally, participants and study staff could not be masked to treatment conditions, given the study design and digital platform. Finally, although satisfaction scores were in the “moderate to high” range, future studies should do qualitative work postintervention to learn about how to increase satisfaction with the app.

These limitations should be viewed in the context of innovation. Innovation regarding intervention approaches for unmotivated smokers has stalled. The vast majority of digital interventions include content and features that are tailored for those who are motivated to quit. On a broader scale, digital interventions in general have not been sufficiently leveraged to target those who are not motivated to change. This study presents progress in this direction and could potentially serve as a paradigm for others to blend social psychology principles (eg, foot-in-the-door approach) with evidenced-based treatments (BA) to motivate change across different areas of health using digital platforms.

Acknowledgments
This research was funded by Cancer Research UK and they did not have a role in the analysis or writing of the paper. All authors contributed significantly to the manuscript and have been personally and actively involved in substantive work leading to the report. All authors have seen and approved the final version of the manuscript and have agreed to the submission.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
DF is a subject matter expert for Click Therapeutics and Boehringer Ingelheim.

Multimedia Appendix 1
Study protocols.
[DOCX File, 292 KB - formative_v8i1e54912_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 2999 KB - formative_v8i1e54912_app2.pdf]

References


13. Borrelli B, Endrighi R, Hammond SK, Dunsgier S. Smokers who are unmotivated to quit and have a child with asthma are more likely to quit with intensive motivational interviewing and repeated biomarker feedback. J Consult Clin Psychol 2017;85(11):1019-1028 [FREE Full text] [doi: 10.1037/ccp0000238] [Medline: 29083219]


25. Randomly assign subjects to treatment groups. GraphPad. URL: https://www.graphpad.com/quickcalcs/randomize1/ [accessed 2024-02-27]


Abbreviations

BA: behavioral activation
CES-D-10: 10-item Center for Epidemiologic Studies Depression Scale
MARS: Mobile App Rating Scale
Evaluation of a Pilot Program to Prevent the Misuse of Prescribed Opioids Among Health Care Workers: Repeated Measures Survey Study

Stephen Hebard\textsuperscript{1*}, PhD; GracieLee Weaver\textsuperscript{2*}, PhD; William B Hansen\textsuperscript{1*}, PhD; Scarlett Ruppert\textsuperscript{2*}, MPH

\textsuperscript{1}Prevention Strategies, Greensboro, NC, United States
\textsuperscript{2}Department of Public Health Education, University of North Carolina Greensboro, Greensboro, NC, United States

\*all authors contributed equally

Corresponding Author:
GracieLee Weaver, PhD
Department of Public Health Education
University of North Carolina Greensboro
PO BOX 26170
Greensboro, NC, 27402-6170
United States
Phone: 1 336 334 5000
Email: gmweaver@uncg.edu

Abstract

Background: Overprescription of opioids has led to increased misuse of opioids, resulting in higher rates of overdose. The workplace can play a vital role in an individual’s intentions to misuse prescription opioids with injured workers being prescribed opioids, at a rate 3 times the national average. For example, health care workers are at risk for injuries, opioid dispensing, and diversion. Intervening within a context that may contribute to risks for opioid misuse while targeting individual psychosocial factors may be a useful complement to interventions at policy and prescribing levels.

Objective: This pilot study assessed the effects of a mobile-friendly opioid misuse intervention prototype tailored for health care workers using the preparation phase of a multiphase optimization strategy design.

Methods: A total of 33 health care practitioners participated in the pilot intervention, which included 10 brief web-based lessons aimed at impacting psychosocial measures that underlie opioid misuse. The lesson topics included: addiction beliefs, addiction control, Centers for Disease Control and Prevention guidelines and recommendations, beliefs about patient-provider relationships and communication, control in communicating with providers, beliefs about self-monitoring pain and side effects, control in self-monitoring pain and side effects, diversion and disposal beliefs, diversion and disposal control, and a conclusion lesson. Using a treatment-only design, pretest and posttest surveys were collected. A general linear repeated measures ANOVA was used to assess mean differences from pretest to posttest. Descriptive statistics were used to assess participant feedback about the intervention.

Results: After completing the intervention, participants showed significant mean changes with increases in knowledge of opioids (+0.459; \textit{P}<.001), less favorable attitudes toward opioids (–1.081; \textit{P}=.001), more positive beliefs about communication with providers (+0.205; \textit{P}=.01), more positive beliefs about pain management control (+0.969; \textit{P}<.001), and increased intentions to avoid opioid use (+0.212; \textit{P}=.03). Of the 33 practitioners who completed the program, most felt positive about the information presented, and almost 70% (23/33) agreed or strongly agreed that other workers in the industry should complete a program like this.

Conclusions: While attempts to address the opioid crisis have been made through public health policies and prescribing initiatives, opioid misuse continues to rise. Certain industries place workers at greater risk for injury and opioid dispensing, making interventions that target workers in these industries of particular importance. Results from this pilot study show positive impacts on knowledge, attitudes, and beliefs about communicating with providers and pain management control, as well as intentions to avoid opioid misuse. However, the dropout rate and small sample size are severe limitations, and the results lack generalizability. Results will be used to inform program revisions and future optimization trials, with the intention of providing insight for future intervention development and evaluation of mobile-friendly eHealth interventions for employees.
KEYWORDS

health care workers; opioid misuse; pain management; prescription opioids; prevention; substance abuse; substance use; workers

Introduction

There is an opioid crisis in the United States [1,2]. This includes not only the morbidity and mortality associated with the illegal sale and use of opioids [3,4] but also the misuse and abuse of prescribed opioids [5-8]. Prescription opioids are often prescribed to address injuries and pain. In the United States, at least 20% of adults reported experiencing pain every day or most days over the past 6 months [9], and 36% of those with chronic pain reported that their pain frequently limits their life and work activities [10]. The overprescription of opioid pain relievers, such as oxycodone and hydrocodone, has led to widespread addiction, and increased demand for illicit opioids like heroin and fentanyl. In 2020, there were an estimated 91,000+ overdose deaths, with a marked increase in deaths due to opioids since 2016 [11]. Interventions are needed to address the growing problems associated with opioid misuse.

A primary response to the opioid crisis has been to call for improved opioid prescribing practices, improved treatment retention, and increased involvement of psychiatrists [12]. Many educational programs related to prescription opioids are designed for physicians, nurses, and other health care providers and focus on the assessment and treatment of opioid use disorder (ie, buprenorphine) or aligning one’s prescription behaviors with the clinical guidelines for opioid use [13-15]. Proposed strategies for addressing misuse, abuse, and addiction often rely on health care providers for program or intervention delivery, often referred to as patient-centered education programs [16-19]. These strategies can benefit from considering individual risk factors such as a history of pain, previous substance use, or mental health challenges, which could inform delivery, implementation, or the focus of intervention [20]. Previous intervention studies on opioid misuse can also inform future prevention efforts.

In many studies that address education about opioids, knowledge of risks and safe practices is the primary outcomes, and increases in knowledge are frequently observed. For example, in a randomized pilot study, greater knowledge gains were found for those who received a 6-minute video on proper usage of opioids compared to those who received standard-care instructions [21]. An educational campaign using an advertising format focused on changing attitudes and beliefs of a young adult target population showed increased empathy and perceived risk of problems associated with opioid use [22]. A multidisciplinary clinic-based strategy that included the assessment and treatment of veterans diagnosed with unsafe opioid use patterns suggested better opioid management, but only for a small group of patients [23]. Another study that proposed adaptive prescription and consumption monitoring found significant declines in the risk of developing an opioid addiction associated with their smart prescription management intervention [24]. Among the key takeaways from these studies is the potential for affecting opioid use outcomes through increased provider-patient communication, a key component that other studies have also supported [25,26].

Although interventions for patients and providers may target the point of access and use, they may miss opportunities to establish supportive cultures and environments, such as those that could be implemented across worksites. Worksites represent a critical social determinant of prescription opioid misuse. Despite the risks associated with prescription opioid use, injured workers are frequently prescribed opioid pain medications at rates three times the national average [27,28] and almost one-third of injured workers continue to fill their prescriptions 90 days after their injury [29]. As many as 57% of accidental deaths due to opioid overdose occur following work injuries [30]. In addition, prescription opioid misuse costs US employers an estimated US $25-$26 billion each year due to lost productivity, turnover, and premature employee death [28,29]. Though 70% of employers report negative consequences of opioid misuse (eg, absenteeism, poor performance, and turnover), very few employers offer any type of prescription opioid prevention intervention [31] or feel prepared to address opioid-related concerns [32].

Intervening within worksites would allow for reaching a broader audience, implementing primary prevention before treatment of injury or pain, and addressing complex and unique social-behavioral motivators for behaviors related to health outcomes [33]. For example, coworker relationships can influence beliefs and attitudes about behaviors such as the misuse of prescription drugs [34]. In the case of construction workers, nurses, and nursing assistants, peer normalization of job strain, work-related injuries, and working through pain compounds the risk of musculoskeletal injury and the use of prescription opioids for pain [35,36]. While interventions aimed at preventing other substance use disorders have successfully targeted behavioral factors, particularly through web-based digital health technologies [37,38], a one-size-fits-all approach may ignore the unique predictors of prescription opioid misuse and the influence of work on behavior.

Among the strategies that have been proposed to increase the frequency and quality of communication is the development of smartphone apps [39,40]. Evidence-based mHealth interventions have become a popular and effective delivery method for health behavior [41,42] as 85% of Americans own a smartphone [43]. Further, mHealth interventions have demonstrated exceptional feasibility and effectiveness in the prevention of substance use among a variety of populations [44]. mHealth interventions can be more cost-effective (eg, professional staff do not need to be trained or hired to deliver the content and have a broader reach than facilitator-led interventions) [45].

This pilot study examines the use of an mHealth intervention for prescription opioid use prevention, referred to as “WorkWell,” which targets psychosocial outcomes that research suggests may influence intentions to use or misuse opioids, such as...
as knowledge, attitudes, perceived control of managing pain, and perceptions about communicating with providers [21-26,46,47]. WorkWell was designed to be delivered digitally because of the economy, privacy, scalability, and impact that mHealth interventions can have. In addition, brief mHealth interventions like WorkWell are ideal for workforces because (1) they can be accessed anytime, allowing self-paced learning; (2) they can be delivered outside of work hours and therefore do not disrupt safety and productivity during the workday; (3) workers can access them in any setting that provides access to wireless internet or a cellular network; and (4) they can deliver personalized feedback while being standardized to ensure fidelity. This paper presents information about the intervention and study protocols, as well as the preliminary results from the pilot data. This pilot study aimed to assess its short-term impact on targeted psychosocial outcomes, as described below.

Methods

Intervention

WorkWell presents information about pain management strategies, how opioids may impact acute and chronic pain, guidelines for use and disposal, the nature of the epidemic, tolerance and addiction, and the risks of overdose. In collaboration with an expert advisory panel with backgrounds in generalized pain and symptom management, prescription opioid use among young adults, and workplace culture, we scripted 10 brief lessons (referred to as touch points for participants). With input from the advisory panel, the following lessons targeted attitudes and beliefs: addiction beliefs, Centers for Disease Control and Prevention guidelines and recommendations, patient-provider relationships and communication, beliefs about self-monitoring pain and side effects, diversion and disposal, and the conclusion lesson. Lessons that targeted perceived control included addiction control, communicating with providers, control with self-monitoring pain and side effects, diversion and disposal, and the conclusion lesson. All lessons included knowledge components, and lessons 4 and 5 addressed aspects of patient-provider communication, a key point highlighted in previous research, as well as the input from our expert advisory panel.

We partnered with an educational design team to develop and narrate each lesson and the web-based platform within which it would be housed. Each lesson was expected to be completed in between 8 and 10 minutes. Lessons included text, animations, still images, and dynamic exercises in which participants indicated their experiences, preferences, or decisions when prompted with a scenario. For example, true or false questions were integrated into knowledge components within lessons, and attitudinal components included opportunities for participants to react or input their own beliefs to trigger tailored responses within the app. In addition to tailored feedback in response to their interactions with the material, participants received in-app industry-specific prescription opioid information and appropriately timed push notifications to prompt use and completion.

The prototype developed for pilot testing was tailored for delivery to health care workers, including images or examples specific to working in the health care industry. Nurses are among the most injured workers, just behind construction workers [9,48,49] and nursing assistants had the highest incidence rate of injury cases that required missed days of work [50]. Further, nurses, nursing assistants, and psychiatric aides are at high risk for musculoskeletal disorders and experience considerable work-related stress and job strain (ie, high psychological workload and low work-related decisional latitude) that challenge effective coping and lead to substance use and other behavioral, physical, and mental health concerns [51]. Prescription drug diversion has been cited as a specific problem among nurses, whose access to prescription opioids presents an additional challenge to misuse prevention [52,53]. Given the broad normalization of pain, injury, and musculoskeletal disorders and the distress associated with the perceived need to prematurely return from injury to work, it is clear that employment within this industry is an important risk factor for prescription opioid misuse.

Participants

Development of this intervention prototype took place in 2019, leaving recruitment for this pilot project to take place during the height of the COVID-19 pandemic. This not only limited in-person or higher-touch recruitment efforts but also created challenges for recruiting from an already overburdened workforce. We relied on recruitment emails sent out through relevant listserves by partnering with “INFOCUS Marketing,” the National Association of School Nurses, and nursing schools at 2 local universities. We were unable to track how many email addresses were valid across various listserves and were thus unaware of the total number of recruitment emails successfully sent. Although 856 health care workers fully completed the pretest survey, only 47 individuals continued to further participate in the intervention. Of those participating in the intervention, 33 (70%) individuals completed all aspects of the study, including the pretest, intervention, and posttest. Participants ranged in age from 26 to 73 years, with a mean age of 49 years. Most participants were female (30/33, 91%), White (31/33, 94%), and (30/33, 91%) current health care practitioners.

Measures

Pretest and posttest surveys assessed demographics (age, gender, ethnicity, race, and employment) and the following psychosocial measures: knowledge of opioids (eg, “I know how to safely use an opioid prescription for pain;” 7 items; pretest α=.873; posttest α=.935), attitudes toward opioids (eg, “rate from good to bad: using opioids to control pain is...”; 6 items; pretest α=.736; posttest α=.866), communicating with providers (eg, “rate from not important to very important: communicating with my provider about my opioid prescription is...”; 4 items; pretest α=.702; posttest α=.935), pain management control (eg, “rate from very difficult to very easy: refusing a prescription for opioids when I am in pain would be...”; 8 items; pretest α=.718; posttest α=.758), and intentions about opioid use (eg, “rate from strongly disagree to strongly agree: I do not intend to take prescription opioids for pain management;” 4 items; pretest α=.581; posttest α=.735).
At the posttest, participants provided ratings about the WorkWell program and its appropriateness as an educational intervention. Ratings on a 6-point scale ranging from “strongly disagree” to “strongly agree” were given to four prompts: (1) “The information presented in this program was too basic,” (2) “The information presented in this program was difficult to understand,” (3) “The information presented in this program was so boring that I became distracted,” and (4) “All workers in my industry should complete a program like this.”

**Design, Hypotheses, and Analysis Plan**

The pilot study was conducted as a pretest-treatment-posttest design without a control or comparison group. Pretest surveys were administered through Qualtrics (Silver Lake). Once participants had completed pretest surveys, the first touch point would be made accessible within the intervention platform. As each touch point was completed, the participant was given access to the next touch point. Therefore, all participants completed all touch points in the same order. Immediately following the completion of all 10 touch points, a posttest survey was made available, again administered through Qualtrics.

We hypothesized that there would be measurable improvements on all targeted psychosocial measures. We also hypothesized that ratings would reflect a positive evaluation of the intervention. Within-subject pretest and posttest psychosocial measures were compared using a general linear repeated measures analysis of variance. Descriptive statistics were provided for participants’ posttest ratings about the program.

**Ethical Considerations**

This study was reviewed and approved by the University of North Carolina Greensboro Institutional Review Board (#20-0090). Recruitment emails included information about the purpose of the study, what individuals are being asked to do, the incentive for participating, and instructions on how to learn more information about engaging in the study. Interested individuals were also shown a 2-minute animated video describing the study, intervention, and eligibility to receive an incentive for completion of all intervention and survey components in addition to viewing the web-based consent form. Informed consent documentation was provided to participants at the beginning of the pretest survey and included details about the study aims, how it relates to the person accessing the consent form, what they would be asked to do, the risks and benefits, confidentiality of the data, the right to refuse or withdraw from the study, and a US $25 gift card incentive provided to the first 200 people completing the pretest, intervention, and posttest. All participants voluntarily consented to participate in the study. All data are anonymous, with no personally identifiable information.

**Results**

The program resulted in significant changes in all 5 targeted psychosocial variables (Table 1). Knowledge of opioids, communicating with providers, pain management control, and intentions to avoid opioid use all improved pretest-to-posttest. That is, participants knew more about opioids, understood the importance of communicating with providers about opioids and pain, improved their perception that they could appropriately control their pain, and intended to not use opioids inappropriately. Attitudes toward opioids were reversed in that lower scores reflected a less favorable attitude. Participants expressed more negative attitudes toward opioids at the posttest.

Additionally, the 33 participants who completed both the pretest and posttest provided positive ratings for the program (Table 2). They judged the program to present information that was appropriate, easy to understand, and engaging. They generally agreed that other people in their industry should participate in this program or one like it.

**Table 1.** Generalized repeated measures ANOVA mean changes in psychosocial outcome variables from pretest to posttest for 33 health care practitioners who completed the intervention and surveys (N=33).

<table>
<thead>
<tr>
<th></th>
<th>F test (df)</th>
<th>P value</th>
<th>Pretest mean</th>
<th>Posttest mean</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge of opioids</td>
<td>21.812 (1, 32)</td>
<td>&lt;.001</td>
<td>3.524</td>
<td>3.983</td>
<td>0.459</td>
</tr>
<tr>
<td>Attitudes toward opioids</td>
<td>14.811 (1, 32)</td>
<td>.001</td>
<td>3.848</td>
<td>2.768</td>
<td>−1.081</td>
</tr>
<tr>
<td>Communicating with providers</td>
<td>7.609 (1, 32)</td>
<td>.01</td>
<td>5.174</td>
<td>5.379</td>
<td>0.205</td>
</tr>
<tr>
<td>Pain management control</td>
<td>16.991 (1, 32)</td>
<td>&lt;.001</td>
<td>5.242</td>
<td>6.211</td>
<td>0.969</td>
</tr>
<tr>
<td>Intentions to avoid opioid use</td>
<td>5.060 (1, 32)</td>
<td>.03</td>
<td>5.795</td>
<td>6.008</td>
<td>0.212</td>
</tr>
</tbody>
</table>
Discussion

Overview

WorkWell is an mHealth opioid misuse prevention intervention designed to integrate with the lifestyle of workers with higher risks of opioid dispensing and misuse, such as those in the health care industry. The goal of this pilot study was to be a proof-of-concept test of the WorkWell intervention that focused on improving psychosocial factors that underlie preventing opioid use and misuse. Previous research [46] has demonstrated strong relationships between opioid intentions and the other psychosocial variables we tested. The 10-lesson intervention resulted in improvements in participants' knowledge of opioids, attitudes toward opioids, communicating with providers, pain management control, and intentions to avoid opioid use. Furthermore, participants who completed all aspects—pretest, intervention, and posttest—reported favorably of the appropriateness, ease of use, and engagement of the intervention. However, 30% (14/47) of those who completed the pretest did not follow through with both intervention and posttest completion, biasing posttest results and missing the perspective of those who chose not to complete.

One major criticism of mHealth interventions is that participant use diminishes within a week or 2 of initiation if the information is static or relies on participants to initiate contact [54]. A hypothesized design strength of WorkWell is an intentional, scheduled approach to user engagement and retention, providing interactions and nudges based on participant response. However, questions remain about the effects of this design, as 30% (14/47) of the sample did not complete the full intervention and posttest. Although posttest data from health care workers who did complete WorkWell as a program generally agreed that it should be adopted as an industry standard for safety, these data exclude populations.

Table 2. Descriptive statistics for posttest feedback about the program from 33 health care practitioners who completed the intervention and surveys (N=33).

<table>
<thead>
<tr>
<th>The information presented in this program was too basic.</th>
<th>Strongly disagree, n (%)</th>
<th>Disagree, n (%)</th>
<th>Slightly disagree, n (%)</th>
<th>Slightly agree, n (%)</th>
<th>Agree, n (%)</th>
<th>Strongly agree, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (6)</td>
<td>10 (30)</td>
<td>12 (36)</td>
<td>8 (24)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>The information presented in this program was difficult to understand.</td>
<td>16 (49)</td>
<td>14 (42)</td>
<td>3 (9)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>The information presented in this program was so boring that I became distracted.</td>
<td>6 (18)</td>
<td>19 (58)</td>
<td>6 (18)</td>
<td>2 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>All workers in my industry should complete a program like this.</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>7 (21)</td>
<td>13 (39)</td>
<td>10 (30)</td>
</tr>
</tbody>
</table>

Limitations

There are several limitations to this pilot study. Results are based on a small sample size of only 33 participants who completed the pretest, intervention, and posttest. The posttest results leave out the perspectives of the 14 individuals who dropped out of the study before completing the posttest, further biasing the overall findings for the effects of the intervention as well as the participant feedback. The pilot testing took place during the COVID-19 pandemic, creating challenges for recruiting and retaining the targeted population of health care workers who experienced higher workloads and strain during the pandemic. We did not capture data that would inform whether COVID-19 was a part of the reason for choosing not to participate in the full pilot study. We also did not include a control group and are unable to conclude whether increases in psychosocial variables would have been present without exposure to the intervention. Also, this pilot study included only one immediate posttest, limiting the ability to determine whether the intervention has lasting effects on intentions to avoid opioid misuse. Lastly, the intervention prototype was tested only with health care workers and is thus not generalizable to at-risk workers in other industries.
Conclusions

While attempts to address the opioid crisis have been made through public health policy initiatives, the problem still permeates throughout our society as we can see opioid overdose deaths continue to rise. Prescription opioid misuse prevention efforts have focused heavily on restricting access to prescription opioids with drug monitoring programs, stricter prescribing guidelines, dose-limit laws, prescription take-back days, and increased law enforcement [61-64]. Interventions like WorkWell that focus on changing the individual and interpersonal motivational factors linked to opioid misuse could be an effective way to supplement current policy-level initiatives, though more research is needed to substantiate the effects of this particular program. Combining individual-level interventions with organizational or policy-level interventions provides a socioecological approach to preventing opioid misuse and overdose death. Additionally, because certain industries carry a greater risk for employee opioid use and misuse due to high rates of on-the-job injury [35,36], there is a specific need to provide easily accessible and scalable interventions for employees within these at-risk populations. This study presented an mHealth intervention that aims to mitigate opioid misuse and subsequent opioid morbidity and mortality among employees within an at-risk industry such as health care. Preliminary findings will inform the next phase of intervention development and testing, which will examine the effects of all combinations of intervention components against a control group to establish the most efficient intervention for use with at-risk workers using a MOST design [57-60].

Acknowledgments

This work was supported by the National Institute on Drug Abuse [1R43DA050404-01].

Data Availability

The data set generated during this study is available from the corresponding author on reasonable request.

Conflicts of Interest

This pilot project was funded as a Phase 1 Small Business Innovation Research grant with the intent to conduct optimized trials and eventually commercialize the intervention through Prevention Strategies. Authors working with Prevention Strategies have a vested interest in the success of the intervention.

References


Abbreviations

MOST: multiphase optimization strategy

©Stephen Hebard, GracieLee Weaver, William B Hansen, Scarlett Ruppert. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Lessons From the Field From a Volunteer Telehealth Ambassador Program to Enhance Video Visits Among Low-Income Patients: Qualitative Improvement Study

Delphine S Tuot¹,²,³, MD; Aarya Mukherjee⁴; Amanda Churape¹,², BA; Triveni DeFries¹, MD, MPH; George Su¹,², MD; Elaine C Khoong¹,1, MD, MS; Courtney Lyles¹,3,5,6,7, PhD

¹Department of Medicine, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States
²Center for Innovation in Access and Quality at Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States
³UCSF Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, San Francisco, CA, United States
⁴University of California, Berkeley, Berkeley, CA, United States
⁵Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States
⁶Center for Healthcare Policy and Research, University of California, Davis, Sacramento, CA, United States
⁷Department of Public Health Sciences, University of California Davis School of Medicine, Sacramento, CA, United States

Corresponding Author:
Delphine S Tuot, MD
Department of Medicine
Zuckerberg San Francisco General Hospital
University of California, San Francisco
1001 Potrero Ave
San Francisco, CA, 94110
United States
Phone: 1 6282068242
Email: Delphine.tuot@ucsf.edu

Abstract

Background: The prevalence of telehealth video use across the United States is uneven, with low uptake in safety-net health care delivery systems, which care for patient populations who face barriers to using digital technologies.

Objective: This study aimed to increase video visit use in an urban safety-net delivery system. We piloted a telehealth ambassador program, in which volunteers offered technical support to patients with access to digital technologies to convert primary care visits already scheduled as telehealth audio-only visits to telehealth video visits.

Methods: We used a descriptive approach to assess the feasibility, efficacy, and acceptability of the pilot telehealth ambassador program. Feasibility was quantified by the percentage of eligible patients who answered calls from telehealth ambassadors. Program efficacy was measured in two ways: (1) the percentage of patients with access to digital technology who interacted with the navigators and were successfully prepared for a telehealth video visit, and (2) the percentage of prepared patients who completed their scheduled video visits. Program acceptability was ascertained by a structured telephone survey.

Results: Telehealth ambassadors attempted to contact 776 eligible patients; 43.6% (338/776) were reached by phone, among whom 44.4% (150/338) were provided digital support between March and May 2021. The mean call duration was 8.8 (range 0-35) minutes. Overall, 67.3% (101/150) of patients who received support successfully completed a telehealth video visit with their provider. Among the 188 patients who were contacted but declined video visit digital support, 61% (114/188) provided a reason for their decline; 42% (48/114) did not see added value beyond a telehealth audio-only visit, 20% (23/114) had insufficient internet access, and 27% (31/114) declined learning about a new technology. The acceptability of the telehealth ambassador program was generally favorable, although some patients preferred having in-real-time technology support on the day of their telehealth video visit.

Conclusions: This high-touch program reached approximately one-half of eligible patients and helped two-thirds of interested patients with basic video visit capability successfully complete a video visit. Increasing the program’s reach will require outreach solutions that do not rely solely on phone calls. Routinely highlighting the benefits of video visits, partnering with community-based
organizations to overcome structural barriers to telehealth use, and offering in-real-time technology support will help increase the program’s efficacy.

(JMIR Form Res 2024;8:e49993) doi:10.2196/49993

KEYWORDS
digital barriers; digital support; digital technologies; equity; health care delivery; safety-net; telehealth

Introduction
The onset of the COVID-19 pandemic in March 2020 led to an unprecedented era of digital reliance resulting from quarantining and shelter-in-place precautions. The dependence on digital interactions as opposed to in-person communication disrupted health care delivery with the rapid rollout of telehealth services across the United States, including telehealth audio-only visits (ie, telephone visits) and telehealth video visits [1]. The prevalence of telehealth video visit use across different care delivery settings was uneven, however, with particularly low uptake in safety-net systems, which often consist of county hospitals, health clinics, and emergency departments that treat patients regardless of their ability to pay or their immigration status [2,3]. Reasons for low telehealth video uptake in safety-net systems are multifactorial, including suboptimal infrastructure for implementation as well as challenges faced by their patient populations due to structural and socioeconomic barriers and limited digital literacy [4,5]. There is a paucity of trial data that directly compares different telehealth visit modalities, and both likely have a role to play in primary care delivery [6]. However, retrospective studies examining data from electronic health records suggest that primary care telehealth video visits are associated with more clinical actions, including more medication prescriptions and diagnostic testing and fewer return in-person visits within 7 days compared to telehealth audio-only visits [7]. Additionally, telehealth video visits have been associated with less clinician concern for patient safety compared with telehealth audio [8].

Patient satisfaction (overall and within safety net settings) with telehealth services, including, but not limited to, telehealth video, has been generally favorable [9]. However, many patient groups have concomitantly voiced hesitation surrounding the shift from in-person to remote care delivery, including the use of patient portals and telehealth video visits [10]. These groups include individuals with low digital literacy, defined by a poor ability to use information technologies to find, evaluate, create, and communicate information, as well as those with low socioeconomic status. The presence of these conflicting messages is reinforced by data that suggest that socioeconomic disparities in health technology use are not primarily driven by low patient interest but rather by suboptimal knowledge or digital skills or both [11]. There is a need for robust training and support to bridge the gap between interest in telehealth and the actual use of digital technology.

During the early phase of the pandemic, the use of telehealth video visits in our urban safety-net health care delivery system required patients to download video platform software on their computer or mobile device. Efforts to support patients through this process revealed the need for substantial staff time to provide one-on-one, tailored technical support to match patients’ language, literacy, digital literacy level, and technical needs. Other safety-net health delivery systems had similar experiences, reinforcing the need for individualized counseling on how to use telehealth video services [12]. Our system focused its efforts on addressing one barrier in increasing uptake of telehealth video visits in our low-income population—that of low confidence in using digital tools to participate in telehealth video visits. Here, we describe the feasibility, efficacy, and acceptability of one pilot initiative—a telehealth ambassador program—that offered technical support to patients with access to digital technologies to convert primary care visits already scheduled as telehealth audio-only visits to telehealth video visits.

Methods

Study Setting and Patient Population
San Francisco Health Network (SFHN) is the integrated public health care delivery system that serves San Francisco’s low-income population. It consists of the Zuckerberg San Francisco General Hospital; a long-term care facility; and full-spectrum ambulatory care services delivered through 14 primary care clinics, jail health, specialty care clinics, and a variety of community-based programs. Its primary care clinics, in which this pilot program was conducted, serve 59,000 individual patients with 310,270 annual encounters. Patients are racially and ethnically diverse: 37% Latinx, 20% Asian, 18% White, 15% Black or African American, and 10% other or unknown, and one-third have limited English proficiency. Consistent with its safety-net health system designation, nearly 100% of SFHN patients are covered by government-sponsored insurance: 58% Medicaid, 32% Medicare, and 9% to 10% San Francisco County health access plans.

Telehealth Ambassador Program
Second-year preclinical medical students were given the opportunity to serve as volunteer telehealth ambassadors as part of a quality improvement course embedded in their medical school curriculum. Leveraging role-playing activities, students were trained by the quality improvement team to call patients, introduce themselves, offer technical support, and troubleshoot potential barriers to completing a telehealth video visit. Over the subsequent 10 weeks in early 2021, the telehealth ambassadors reached out to patients with telehealth audio-only appointments scheduled with a small number of primary care providers using a standardized script to introduce themselves as extensions of the primary care clinical team and confirm patients’ identities and upcoming scheduled telehealth audio-only appointments. Each eligible patient received at least 2 phone calls. The telehealth ambassadors then screened patients...
with standardized questions gauging: (1) willingness to participate in video visits; (2) access to a digital device; and (3) access to sufficient data for telehealth video visits, identified by asking about the use of web-based communication platforms (ie, Facebook, Facetime, or WhatsApp). In the absence of a gold standard screening tool for video visit use, the inclusion of these screening questions has been recommended as a best practice for enhancing patient engagement with digital technologies [13]. Among patients with access to digital technology and sufficient data to participate in a telehealth video visit, telehealth ambassadors helped patients download the videoconference app used by our health system on their internet-enabled device and offered a practice session to confirm their ability to use the software. Successful “on-boarding” of an individual patient included a complete download of the video visit app with audio and visual checks and participation in a practice session with the telehealth ambassador. Professional telephone interpreters were available to help patients with limited English proficiency. After successful onboarding, a member of the team with access to the electronic medical record then converted the patient’s upcoming appointment from a telehealth audio visit to a telehealth video appointment.

Ethical Considerations
This was a quality improvement project aimed at assessing the feasibility, efficacy, and acceptability of a new navigator program meant to increase the delivery system’s use of telehealth video. The quality improvement team did not collect or store patient-level, identifiable data. Telehealth ambassadors did not have access to patient electronic health records; they only had access to patient name, telephone number, and primary care clinician information. Such quality improvement activities do not require institutional review board approval at our institution (University of California, San Francisco), even if they include patient surveys, even if they include patient surveys, as long as the primary purpose of the survey is to gauge the opinions and perceptions of customers to improve the delivery of health care. As such, patients did not participate in a formal consent process to interact with the telehealth ambassadors, although they had the opportunity to decline telehealth ambassador services. Patients contacted after their scheduled video visit could also decline answering questions about their experience.

Analysis
We used a descriptive approach to assess the feasibility, efficacy, and acceptability of the pilot telehealth ambassador program. Feasibility was quantified by the percentage of eligible patients who answered calls from telehealth ambassadors. The efficacy of the program was measured in two ways: (1) the percentage of patients willing to participate in a telehealth video visit with access to digital technology who interacted with the navigators and were successfully “onboarded” and (2) the percentage of onboarded patients who completed their scheduled video visits. Telehealth ambassadors called patients within 2 weeks after their scheduled video visit, using a brief, standardized set of survey prompts (Textbox 1) that were developed by the team with input from operational leaders to ascertain the acceptability of the telehealth ambassador program. Feedback about the telehealth ambassador program was reviewed by the 2 members of the quality improvement team (DST and AC). Barriers to video visit completion were tabulated using descriptive statistics.

Textbox 1. Acceptability survey prompts.

1. Did you successfully complete the video visit (yes or no)?
2. Before having the setup call with a Telehealth Ambassador, how successful did you think you would be with completing the video visit (not at all, a little, somewhat, or very)?
3. Would you recommend a Telehealth Ambassador phone call to other patients to get set up for a video visit?
4. What would help you feel prepared to make the most out of a video visit?
5. What type of visit would you prefer for your next visit?

Results
All patients who had scheduled telehealth audio-only visits with a few primary care providers were eligible to be contacted by a telehealth ambassador (n=776). While we did not collect sociodemographic data for those 776 patients, SFHN primary care patients are racially, ethnically, and linguistically diverse (Table 1). During the study period, primary care patients with telehealth visits (n=12,851) compared to those with in-person visits (n=8345) were more likely to be aged between 45 and 64 years, female, of Asian race and non-Hispanic ethnicity, and speak a language other than English (Table 1).

Teledhealth ambassadors attempted to contact all 776 eligible patients with future telehealth audio visits scheduled with a participating primary care provider (Figure 1). Overall, 43.5% (338/776) of patients were reached by phone. Of those, 44.4% (150/338) voiced interest in participating, had access to digital technology and data, and were successfully onboarded after receiving ambassador support, which included a download of the video visit app and confirmation that the patient could use the software, most often with a practice video call. The mean call duration between telehealth ambassadors and the 150 patients was 8.75 (range 0-35) minutes.

Nearly two-thirds (101/150; 67.3%) of patients who were interested, received support, and were deemed ready for a video visit successfully completed their scheduled telehealth video visit with their provider. This translated into a 30% (101/338) video visit completion rate among eligible patients reached by phone. Brief phone surveys with patients who did not have a successful video visit revealed a variety of factors that impaired the video visit completion rate: patient nonattendance to their telehealth visit; internet or connection challenges on the day of the visit, which led to the provider recommending a switch to

https://formative.jmir.org/2024/1/e49993

Tuot et al

JMIR Form Res 2024 | vol. 8 | e49993 | p.3039

(page number not for citation purposes)
an audio-only visit; and a last-minute patient preference for a phone visit.

Table 1. Sociodemographic characteristics of San Francisco Health Network patients with primary care visits between January and February 2021.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>In-person (n=8345), n (%)</th>
<th>Telehealth (n=12,851), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range (years)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≤17</td>
<td>1528 (18.3)</td>
<td>1021 (8.9)</td>
<td></td>
</tr>
<tr>
<td>18-44</td>
<td>1856 (22.2)</td>
<td>2878 (22.4)</td>
<td></td>
</tr>
<tr>
<td>45-64</td>
<td>3015 (36.1)</td>
<td>5319 (41.4)</td>
<td></td>
</tr>
<tr>
<td>≥65</td>
<td>1946 (23.3)</td>
<td>3633 (28.3)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>Female</td>
<td>4551 (54.5)</td>
<td>7318 (56.9)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3793 (45.5)</td>
<td>5530 (43)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.01)</td>
<td>3 (0.02)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>White</td>
<td>1120 (13.4)</td>
<td>2002 (15.6)</td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>1434 (17.2)</td>
<td>1827 (14.2)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1520 (18.2)</td>
<td>3625 (28.2)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4213 (50.5)</td>
<td>5298 (41.2)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>58 (0.7)</td>
<td>100 (0.7)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3747 (44.9)</td>
<td>4681 (36.4)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>4540 (54.4)</td>
<td>8060 (62.7)</td>
<td></td>
</tr>
<tr>
<td>Unreported</td>
<td>58 (0.7)</td>
<td>100 (0.7)</td>
<td></td>
</tr>
<tr>
<td>Primary language</td>
<td></td>
<td></td>
<td>.006</td>
</tr>
<tr>
<td>English</td>
<td>4189 (50.2)</td>
<td>6164 (48)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4152 (49.8)</td>
<td>6679 (52)</td>
<td></td>
</tr>
<tr>
<td>Unreported</td>
<td>4 (0.05)</td>
<td>8 (0.06)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram of participants.

Figure 1: Consolidated Standards of Reporting Trials (CONSORT) diagram of participants.

A total of 55.6% (188/338) of individuals reached by telehealth ambassadors declined technical support and thus declined changing their audio-only visit to a telehealth video visit. A total of 60.6% (114/188) provided a reason for their decline. In all, 42.1% (48/114) believed video visits did not add value above a telehealth audio visit, 20.2% (23/114) had insufficient internet access or did not have an internet-enabled device, 27.2% (31/114) were hesitant to learn new technologies, 4.4% (5/114) reported time constraints limiting their ability to learn how to

https://formative.jmir.org/2024/1/e49993
download the videoconferencing app, and 6.1% (7/114) did not have a safe or private space to conduct video visits.

Nearly one-third (45/150, 30%) of individuals who received technical support responded to the brief acceptability survey. Most patients appreciated the telehealth ambassador phone call and preparatory support and recommended that the program continue. This was particularly true among those who successfully completed their video visits. Example quotes include “I would recommend it absolutely; the set-up call was very helpful” and “I would recommend a set-up call to others who are not familiar with using (the telehealth video software).” However, some patients still lacked confidence in participating in a video visit on the day of their appointment and relied on family members to successfully complete the video visit. A patient mentioned: “I still need help from family members [on the day of the visit]. [With the volunteer], it was set up successfully, but it didn’t work the day of the appointment.” Reactions were more mixed among patients who were not successful at completing their video visits despite receiving technical support. Some patients regarded the program favorably: “I was texted a link for the video visit, but I was not able to connect...The set-up call would likely help others have a successful video visit.” Others felt that the technical support was insufficient, particularly if they were alone, and stated a preference for technical support on the day of appointment: “It was difficult by myself. An IT call right before would be the best solution.”

Discussion

Overview

The SFHN telehealth ambassador volunteer program offered one-on-one remote technical support to patients to increase access to video visits during the COVID-19 pandemic. Overall, this high-touch program reached approximately one-half of eligible patients and helped two-thirds of interested patients with basic video visit capability successfully complete a video visit. Data offered insights into how we could revise the program to achieve greater success, which we define as an increase in the percentage of telehealth appointments that are video versus audio only among individuals equipped with the digital tools that allow them to participate in video visits.

The SFHN telehealth ambassador program relied on telephone communication to reach patients. Failure to make this initial connection represented a missed opportunity to offer digital support to patients. Increasing patient awareness of video visit availability through marketing campaigns and offering a centralized telehealth support desk for patients and families to call for support rather than waiting for a telehealth ambassador phone call could facilitate participation from a higher number of patients [14,15]. Automating the initial phone call or leveraging SMS text messaging might also help. Previous research has shown the practicality of automated interactive voice response apps to make large volumes of calls for patient outreach [16]. Importantly, these calls or texts would need to be multilingual and easily transferred to a telehealth ambassador with professional telephone interpreter support to successfully onboard patients and families who answered the automated phone call or reached out for additional help after receiving a SMS text message. These changes would increase the overall reach of the program, though they would not likely enhance efficacy.

Among those individuals who declined video visit support, many did not feel as though video visits added value to their health care compared to audio-only visits. This is consistent with data from the 2019-2020 nationally representative Medicare Current Beneficiary Survey, which suggested that 28.5% of patients with video visit experience prefer an audio-only visit over a telehealth video visit when clinically appropriate. Ambivalence about video visits has been overcome when routinely offered by health care teams, especially when they highlight provider preference for video visits over audio-only visits [17]. Similarly, frequent and repeated recommendations and offers to help patients use web-based portals have been associated with higher patient portal engagement [18,19]. Incorporating the telehealth ambassador program into clinical workflows will thus be key to increasing interest in video visits among eligible patients. In our next iteration, we are asking all primary care clinic personnel, including clinicians, nurses, medical assistants, front desk staff, nutritionists, etc to routinely recommend video visits over audio-only visits, even to patients who have previously declined a video visit, and simultaneously refer patients to the telehealth ambassador team for digital support. We are also recommending that telehealth ambassadors strongly encourage patients to have a telehealth video partner at home who can help them connect on the day of the visit. This stems from patient feedback among those who were and were not successful at completing their video visits. Such a partner could include a family member, or a community member, or a volunteer, from a community-based organization. Approximately 20% of patients reached by phone reported not having access to an internet-capable device or an unlimited data plan to facilitate video visits. While data from the San Francisco Department of Technology suggests that 87% of the San Francisco general population has high-speed internet access, this percentage is far lower among patients from racial or ethnic minority backgrounds and residents who are low-income, older, have limited English proficiency, or have a disability [20]. Empowering telehealth ambassadors to refer these individuals to partnering community-based organizations that can help them access low-cost devices will be a key feature of our next iteration of the telehealth ambassador program.

Conclusions

In summary, our first iteration of a stand-alone telehealth ambassador program to enhance patients’ digital literacy empowered a small group of patients to participate in video visits. As has been recommended by experts in the field, a multistakeholder and multipronged approach will be required to overcome inequities in telehealth video access [21]. The insights that we gleaned to enhance the reach and efficacy of the telehealth ambassador program must be considered alongside the limitations of our initial evaluation. Our small sample size limits the generalizability of our conclusions. Also, we did not collect patient-level data, so we were unable to discern whether
individual sociodemographic characteristics were associated with program engagement and efficacy. Nevertheless, the data suggest that the program is valuable to a certain group of patients. A more robust version of the telehealth ambassador program will play a key role in our future approach to increasing the percentage of telehealth visits that are video versus audio-only. Increasing the reach of our existing telehealth ambassador program will require creative solutions that enhance widespread awareness of the benefits of video visits compared to audio-only visits and outreach solutions that do not rely solely on person power. Increasing the efficacy of our program will require multilevel interventions, including integration of the telehealth ambassador program into clinical workflows with routine and consistent recommendations for the use of telehealth video over audio-only visits, offering in-real-time IT support, and partnering with community-based organizations to enhance patient access to broadband and internet-enabled devices [22, 23].

Acknowledgments
This work was supported by the Tides Foundation in collaboration with the Center for Care Innovations and the California Health Care Foundation (TF2007-093073). The research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the NIH (K23HL157750). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest
None declared.

References


Abbreviations

SFHN: San Francisco Health Network

©Delphine S Tuot, Aarya Mukherjee, Amanda Churape, Triveni DeFries, George Su, Elaine C Khoong, Courtney Lyles. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e49993 JMIR Form Res 2024 | vol. 8 | e49993 | p.3043 (page number not for citation purposes)
Original Paper

Mi Sleep Coach Mobile App to Address Insomnia Symptoms Among Cancer Survivors: Single-Arm Feasibility Study

Noel Arring¹, RN, DNP, PhD; Debra L Barton¹, RN, PhD; Carolyn Lafferty¹, PhD; Bryana Cox², BS; Deirdre A Conroy³, PhD; Lawrence An⁴, MD

¹College of Nursing, University of Tennessee, Knoxville, TN, United States
²School of Nursing, University of Michigan, Ann Arbor, MI, United States
³Behavioral Sleep Medicine Program, Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
⁴Division of General Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Noel Arring, RN, DNP, PhD
College of Nursing
University of Tennessee
1412 Circle Drive, Room 411
Knoxville, TN, 37966
United States
Phone: 1 8659741988
Email: narring@utk.edu

Abstract

Background: Rates of sleep disturbance among survivors of cancer are more than 3 times higher than the general population. Causes of sleep disturbance among survivors are many and multifaceted, including anxiety and fear related to cancer diagnosis and treatments. Cognitive behavioral therapy for insomnia (CBT-I) is considered a first-line treatment for insomnia; However, a lack of access to trained professionals and limited insurance coverage for CBT-I services has limited patient access to these effective treatments. Evidence supports digital delivery of CBT-I (dCBT-I), but there is only limited evidence to support its use among survivors of cancer. Broad adoption of smartphone technology provides a new channel to deliver dCBT-I, but no prior studies have evaluated mobile dCBT-I interventions for survivors. To address the need for accessible and efficacious CBT-I for survivors of cancer, the Mi Sleep Coach program was developed to adapt CBT-I for delivery to survivors of cancer as a self-directed mobile health app.

Objective: This single-arm feasibility study assessed the adherence, attrition, usefulness, and satisfaction of the Mi Sleep Coach app for insomnia.

Methods: A 7-week, single-arm study was conducted, enrolling adult survivors of breast, prostate, or colon cancer reporting sleep disturbances.

Results: In total, 30 participants were enrolled, with 100% completing the study and providing data through week 7. Further, 9 out of 10 app features were found to be useful by 80% (n=24) to 93% (n=28) of the 30 participants. Furthermore, 27 (90%) participants were satisfied with the Mi Sleep Coach app and 28 (93%) would recommend the use of the Mi Sleep Coach app for those with insomnia. The Insomnia Severity Index showed a decrease from baseline (18.5, SD 4.6) to week 7 (10.4, SD 4.2) of 8.1 (P<.001; Cohen d=1.5). At baseline, 25 (83%) participants scored in the moderate (n=19; 15-21) or severe (n=6; 22-28) insomnia range. At week 7, a total of 4 (13%) patients scored in the moderate (n=4) or severe (n=0) range. The number of patients taking prescription sleep medications decreased from 7 (23%) at baseline to 1 (3%; P<.001) at week 7. The number of patients taking over-the-counter sleep medications decreased from 14 (47%) at baseline to 9 (30%; P=.03) at week 7.

Conclusions: The Mi Sleep Coach app demonstrated high levels of program adherence and user satisfaction and had large effects on the severity of insomnia among survivors of cancer. The Mi Sleep Coach app is a promising intervention for cancer-related insomnia, and further clinical trials are warranted. If proven to significantly decrease insomnia in survivors of cancer in future randomized controlled clinical trials, this intervention would provide more survivors of cancer with easy access to evidence-based CBT-I treatment.

Trial Registration: ClinicalTrials.gov NCT04827459; https://clinicaltrials.gov/study/NCT04827459
cognitive behavioral therapy; insomnia; mobile health; breast cancer; prostate cancer; colon cancer; cancer survivor

Introduction

There are an estimated 18 million survivors of cancer in the United States [1], a number that is expected to increase to 22 million by 2030 [2]. As many as 20% to 75% of survivors of cancer report sleep disturbances depending on the type of cancer, which is more than 3 times higher than the general population [3-6]. Results from a recent systematic review suggest the overall prevalence of sleep disturbance in survivors of cancer is 61% [7]. Survivors of breast cancer are among the most likely to report insomnia at 60% [7]. Among survivors of prostate cancer, as many as 45% reported sleep disturbance [7]. In a study of survivors of colorectal cancer, those reporting sleep difficulty was as high as 75% (75% among patients of rectal cancer and 68% among patients of colon cancer) [6].

Problems with insomnia can persist long into the survivorship period. The causes of sleep disturbance among survivors are many and multifaceted, such as anxiety and fear related to cancer diagnosis and treatments [8,9], intrusive thoughts and faulty beliefs and attitudes about sleep [10,11], nausea and hot flashes from chemotherapy [8,9], or pain after surgery [8,9]. Symptom comorbidity creates a need for interventions that address multiple sequelae, as a lack of sleep can lead to a difficult cycle of poor sleep followed by aggravated daytime symptoms [12,13].

Cognitive behavioral therapy for insomnia (CBT-I) is considered a first-line treatment for insomnia [8,14-16]. Endorsed by the American Academy of Sleep Medicine [17], components of CBT-I include (1) stimulus control (ie, limiting the occurrence of nonsleep behaviors in the bedroom), (2) sleep restriction (ie, sleep schedule management), (3) sleep hygiene (ie, behaviors and practices that facilitate sleep), (4) cognitive restructuring (ie, addressing thoughts that interfere with sleep), and (5) relaxation (ie, use of guided imagery or meditation to reduce arousal). CBT-I is a well-established, evidence-based treatment for sleep that has shown improvement in various measures of sleep continuity without known side effects [18,19]. The findings from several systematic reviews and meta-analyses support the efficacy of CBT-I among survivors of cancer with sustained improvements in insomnia over time [16,20,21], as well as fatigue, depression, and quality of life [15,16,22,23].

A lack of access to trained professionals and limited insurance coverage for CBT-I services has limited patient access to these effective treatments [24]. In response, a growing body of research is exploring the role of internet-based or digital delivery of CBT-I (dCBT-I) and finding it effective, even comparable to in-person CBT-I at improving sleep [25-27] and comorbid symptoms such as fatigue, depression, and anxiety [25]. Further, 2 studies have examined internet-based dCBT-I for survivors of cancer using the Sleep Healthy Using the Internet program. In the Sleep Healthy Using the Internet study, 28 survivors were recruited and the participants in the intervention group showed significant improvements compared to controls for multiple sleep outcomes and general fatigue [28]. The second study was larger, included a national sample of 255 Danish survivors of breast cancer, and resulted in significant improvements in multiple sleep outcomes and reductions in fatigue [29]. To date, only these 2 studies have examined internet-based dCBT-I for survivors of cancer. While the results of these studies are encouraging, one of these studies was small (total n=28) and the other study enrolled only survivors of breast cancer [28,29].

The growth of mobile smartphone technology creates opportunities to deliver dCBT-I interventions [30]. In the past decade, we have seen rapid adoption of smartphones across diverse populations with fewer disparities compared to other technologies (eg, desktop or laptop computers, or broadband internet) [31-33]. Mobile phone apps can deliver highly usable interventions without the barrier of cost [34] and provide a realistic option to provide population access to CBT-I treatment [35]. They can also help overcome barriers such as cost [34], potentially reduce disparities in access to medical care, communication barriers with medical providers, and help with symptom monitoring and management [36]. There has in fact been a proliferation of commercial smartphone apps that address insomnia. Unfortunately, in a recent review, only 1 (CBTI Coach) of the 9 commercially available apps reviewed was found to be highly adherent to evidence-based CBT-I elements and had undergone rigorous evaluation for efficacy [34]. A small number of studies have demonstrated positive effects of mobile delivery of dCBT-I, but none of these studies involved survivors of cancer [35,37-41].

To address the critical need for accessible and efficacious CBT-I for survivors of cancer, the Mi Sleep Coach app was tested in a feasibility study that assessed adherence, attrition, usefulness, satisfaction, and effect size as an intervention for insomnia. The Mi Sleep Coach app provides access to all elements of the evidence-based CBT-I strategies. To support adoption and adherence to these strategies, the app engages users in a computerized conversational dialogue with a digital agent that is modeled on the interaction between clients and trained human CBT-I experts.

Methods

Study Design

This single-arm feasibility study aimed to assess adherence, usefulness, satisfaction, and effect size for the Mi Sleep Coach app as an intervention for insomnia.

Participants

Adult survivors of breast, prostate, and colon cancer who had completed curative treatment at for at least 3 months and not more than 5 years before study entry and reported sleep disturbances for at least 3 months (see Textbox 1 for full inclusion or exclusion criteria) were recruited from the University of...
Michigan Rogel Cancer Center and through social media between March 2021 and June 2022.

Textbox 1. Inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Aged 18 years or older</td>
</tr>
<tr>
<td>• Completed curative-intent treatment (chemotherapy, surgery, or radiation) at least 3 months and not more than 5 years before study entry</td>
</tr>
<tr>
<td>• Reported trouble falling asleep or staying asleep at least 3 nights per week (most weeks) for the last 3 months</td>
</tr>
<tr>
<td>• Owned an Android phone version 8 or higher or an iPhone running iOS 11 or higher</td>
</tr>
<tr>
<td>• Able to access the internet via smartphone</td>
</tr>
<tr>
<td>• Able to read and write English</td>
</tr>
<tr>
<td>• Able to provide informed written consent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Diagnosed with a sleep disorder other than insomnia (eg, sleep apnea, restless legs syndrome, or narcolepsy)</td>
</tr>
<tr>
<td>• Diagnosed with insomnia before cancer diagnosis</td>
</tr>
<tr>
<td>• Reported physical symptoms that interfered with sleep, such as shortness of breath, pain, hot flashes, or frequent urination</td>
</tr>
<tr>
<td>• Reported 3 or more of the following symptoms suggestive of sleep apnea: excessive daytime sleepiness; loud snoring; awakening with a dry mouth, sore throat, or morning headache; observed episodes of stopped breathing during the night; abrupt awakenings accompanied by gasping or choking</td>
</tr>
<tr>
<td>• Major psychiatric or medical condition other than cancer suspected to contribute to sleep disturbance</td>
</tr>
<tr>
<td>• Evidence of active cancer</td>
</tr>
<tr>
<td>• Currently or previously received cognitive behavioral therapy for insomnia</td>
</tr>
<tr>
<td>• Night shift workers or subject to other external restrictions on their opportunity to sleep at night</td>
</tr>
</tbody>
</table>

Intervention

Overview

The Mi Sleep Coach app was developed by the Center for Health Communications Research (CHCR), which is based at the University of Michigan. CHCR approached the development of the Mi Sleep Coach app guided by the principles of user-centered design [42] and included iterative engagement with members of the University of Michigan Rogel Cancer Center’s patient and family advisory council for feedback on program design and message content. The Mi Sleep Coach app adapted an evidence-based clinical CBT-I program for delivery as a self-directed mobile health app for survivors of cancer. The content for the Mi Sleep Coach app is based on core principles of CBT-I, which has empirical evidence developed in multiple clinical trials [43-45]. In adapting this clinical intervention, our sleep consultant (DAC), who is board certified in behavioral sleep medicine, imagined and described what else they might do or say to the patient if they were actually with the patient throughout the day as opposed to only seeing the patient during a weekly clinic visit. The Mi Sleep Coach program was developed for cross-platform deployment on both iOS and Android operating systems. Please see Multimedia Appendix 1. The intervention topic schedule and strategies are presented in Textbox 2.

To address issues of health and technology literacy, plain language guidelines [46] were applied to inform the design of several aspects of the Mi Sleep Coach app. Examples of these plain language principles and design features included serve your audience, address the user, keep it conversational, and organize information.
Textbox 2. Mi Sleep Coach app intervention schedule.

<table>
<thead>
<tr>
<th>Week 1</th>
<th>• Sleep hygiene education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2</td>
<td>• Personalized sleep plan</td>
</tr>
<tr>
<td></td>
<td>• Daily activity plan to promote sleep</td>
</tr>
<tr>
<td>Week 3</td>
<td>• Motivational interviewing (MI) for sleep plan adherence</td>
</tr>
<tr>
<td></td>
<td>• Updated sleep and daily activity plan</td>
</tr>
<tr>
<td>Week 4</td>
<td>• Relaxation techniques</td>
</tr>
<tr>
<td></td>
<td>• Updated sleep and daily activity plan with MI</td>
</tr>
<tr>
<td>Week 5</td>
<td>• Dysfunctional beliefs about sleep</td>
</tr>
<tr>
<td></td>
<td>• Updated sleep and daily activity plan with MI and relaxation techniques</td>
</tr>
<tr>
<td>Week 6</td>
<td>• Maintaining good sleep</td>
</tr>
<tr>
<td></td>
<td>• Updated sleep and activity plan with MI and relaxation</td>
</tr>
<tr>
<td></td>
<td>• Addressing dysfunctional beliefs</td>
</tr>
<tr>
<td>Week 7</td>
<td>• Program review</td>
</tr>
</tbody>
</table>

Serve Your Audience

The Mi Sleep Coach app provided users with a clear value proposition at initial program engagement (eg, “Think of this program as having personal Sleep Coach in your pocket to help retrain your body and mind to have better sleep.”). As part of this value proposition, the Mi Sleep Coach app explained how cancer and cancer treatments can disturb sleep, provided testimonials from survivors of cancer who had improved their sleep using CBT-I, and emphasized specific CBT-I strategies likely to address common causes of insomnia among survivors (eg, constructive worry to address anxiety about recurrence). The program tone consistently expressed empathy and encouraged patience and kindness to oneself when addressing sleep challenges.

Address the User

Rather than providing large volumes of information and asking users to find the specific pieces that are most personally relevant, the app directly assessed and addressed the individual’s specific needs through the application of personally tailored health messaging [47]. Examples of tailored communications (described in more detail below) include the creation of a personalized sleep and daily activity plan for each user, the application of motivational interviewing (MI) techniques to support sleep plan implementation, and the incorporation of users’ priorities and preferences in behavioral activation and cognitive reframing of dysfunctional beliefs about sleep.

Keep It Conversational

Prior work suggests that computer systems that engage in direct conversation-like interactions (ie, conversational or dialogue agents) can help address the needs of users with lower health literacy [48,49]. Informed by this work, the Mi Sleep Coach app was structured as a series of dialogues between the user and the program. This was accomplished by using a hybrid approach that combines sleep expert video messages with a Sleep Coach avatar (loosely modeled on the appearance of our sleep expert) that both addressed users directly. The Sleep Coach avatar functioned as a computerized dialogue agent and was able to have multiple turns of “conversation” with the user based on the user’s selection of specific fixed-choice response options. The several visual representations of the Sleep Coach avatar (ie, different facial expressions) enabled the character to exhibit different emotional states (eg, concern or caring, neutral thoughtfulness, or happiness or satisfaction) as part of interaction with users. The dialogue agent approach helped ensure that the presentation of program content was concise and easy to follow.

Organize Information

A series of lesson cards was integrated into the flow of conversation between the program and users. These lesson cards allowed the app to structure the presentation of information (eg,
main topic and section headings) and incorporate appropriate graphics and illustrations. Narrative analogies were developed to simplify complex concepts (eg, growing a daily “sleep garden” to explain aspects of sleep biology related to circadian rhythms and sleep drive).

**App Tailoring**

**Overview**

Tailored health messages in the Mi Sleep Coach app included personalized sleep plans, MI, addressing dysfunctional beliefs about sleep, and just-in-time behavioral activation.

**Personalized Sleep Plan**

The Mi Sleep Coach app asked users to complete a sleep log (items drawn from the Consensus Sleep Diary) [50] daily as well as items from the validated Sleep Needs Questionnaire (eg, Degree to which the individual is feeling tired or fatigued? Sleepy or drowsy? Taking naps or falling asleep during the day? Feeling that they are getting adequate sleep?) weekly [51,52]. The Mi Sleep Coach app then reviewed with users key metrics of their sleep from these assessments. This review focused on the user’s overall sleep efficiency (ie, time asleep or average time in bed) but also allowed the users to drill down to component and related metrics including average time asleep, average time in bed, average bedtime and consistency of bedtimes, average wake time and consistency of wake times, and average nighttime awakenings and time awake at night. At the start of each intervention week, this information was used to create a personalized sleep plan for each user. To create a personalized plan, users were prompted to pick a target wake time for the upcoming week. Based on sleep efficiency, average time in bed, and perceived sleep needs from the prior week, the Mi Sleep Coach app calculated and presented the user with a target bedtime and time in bed for the user for the upcoming week. The Mi Sleep Coach algorithm increased or decreased recommended time in bed (in 15-min intervals) to achieve a sleep efficiency (time asleep/total time in bed × 100) of 85% or more (80% for individuals aged 65 years or older) and a low reported unmet sleep need. Tailored messaging in the presentation of the sleep plan highlighted the importance of establishing a consistent bedtime and wake time and the recognition of sleep efficiency, rather than total time in bed, as a marker of healthy restful sleep. Throughout the program, users were able to review trends in program engagement (eg, number of consecutive days completing the sleep log) and progress through the program (eg, trends in sleep efficiency).

**About MI**

The Mi Sleep Coach app applied the principles of MI to support users in the implementation and adherence to their personal sleep plan [53]. Each week the program assessed the user’s motivation and confidence to adhere to the newly presented personalized sleep plan. The program reflected the user’s initial responses and then provided tailored encouragement matching the user’s level of motivation and confidence. MI Mi Sleep Coach–based activities included a review and selection (or creation) of personally relevant coping statements designed to build the user’s confidence in their natural ability to sleep (eg, Sleep is a natural process. Everybody, including me, has the natural ability to sleep. Taking control of my daily routines can improve my sleep.). The user’s preferred statements were then used as motivation- and confidence-building reminders throughout the remainder of the program.

**Dysfunctional Beliefs About Sleep**

The Mi Sleep Coach app asked users to complete the brief version of the validated Dysfunctional Beliefs and Attitudes About Sleep (DBAS) assessment [54]. User responses were used to identify that individual’s most prominent dysfunctional beliefs (eg, unrealistic expectations about sleep, catastrophizing after a night of poor sleep, misattribution of the causes of poor sleep, generalization of sleep difficulties to other areas of life). For each user’s top 3 most prominent dysfunctional beliefs, cognitive reframing was supported by providing users with a list of alternative thoughts addressing this belief. Users reviewed and selected specific alternate thoughts that were most personally relevant to them (or created their own alternative thoughts). These alternative thoughts were included as positive reminders to the patient throughout the remainder of the program. Dysfunctional beliefs were presented as “thought traps” and with the idea that specific alternative thoughts could be used to help “get out” of these traps. For example, an individual user who reported prominent unrealistic expectations regarding sleep needs (eg, I must have 8 hours of uninterrupted sleep every night in order to function) was asked to select from a list of alternative thoughts (eg, There is no one right amount of sleep that is right for everyone. In general, my body gets the sleep it needs to keep going. I sleep better when I am more relaxed about how much I sleep) or create their own personally relevant alternative. These alternative thoughts were then delivered back to them as positive reminders throughout the remainder of the program.

**Just-in-Time Behavioral Activation**

The Mi Sleep Coach app also took advantage of the just-in-time capabilities of mobile health interventions to support behavioral activation for improved sleep [55]. In addition to the personalized sleep plan (eg, target bedtime and wake time) described above, the Mi Sleep Coach app presented the users with a set of behavioral strategies that could be used at key portions of the day (eg, waking up and getting up, building sleep drive throughout the day, staying up to the target bedtime, winding down for bed, or awakening in the night). Users selected specific strategies that they would like to try at each of these times of day. The Mi Sleep Coach app then assessed experiences throughout the day and confidence to adhere to sleep-promoting practices throughout the day (eg, How easy or hard was it to get up this morning? How confident are you that you will stay up to your target bedtime?) and then delivered a push notification to users during specified portions of the day (eg, following their wake time, late morning, early afternoon, or 3 hours and 1 hour before their target bedtime) reminding them of their preferred sleep-building behavioral strategies. Users were encouraged to stop using screens and digital devices and employ relaxation strategies 1 hour before their target bedtime each day.
Measures
An adherence rate of at least 75% was considered acceptable. Participants who engaged with the app at least 4 days per week over the 7 weeks were considered adherent.

Attrition rates were assessed weekly and for the study overall. Anyone who did not complete data collection for the full 7 weeks of the study was considered to be prematurely withdrawn and included in our attrition count. An attrition rate of 25% or less was considered feasible.

User experience and satisfaction were assessed with a 40-item investigator-developed user experience survey based upon the Unified Theory of Acceptance and Use of Technology that assessed the perceived ease of use and usefulness of the Mi Sleep Coach app features [56]. Ease of use is assessed by asking participants to rate 10 statements on a scale of 1 (very hard) to 4 (very easy). Participants were asked to rate how useful they found 10 different features (chatting with the avatar, viewing video messages, daily sleep log, personal sleep plan, reading lessons, reading stories from other people with sleep problems, creating a personal list of daily activities, creating a list of coping strategies, constructive worry activities, or deep breathing activities) of the Mi Sleep Coach app on a scale of 1 (not at all useful) to 4 (very useful). Satisfaction was assessed with statements such as “overall, I am satisfied with the Mi Sleep Coach app” and “I would recommend the Mi Sleep Coach app to other people who have trouble sleeping” using a 1 (strongly disagree) to 5 (strongly agree) scale. They were also asked to answer open-ended statements such as “please tell us what you liked about the Mi Sleep Coach app.” This was collected at baseline and week 7 via QualtricsXM (Qualtrics), a survey software tool. We considered the app acceptable if (1) at least 75% (23/30) of participants considered at least 1 app feature at least a little useful (2 or greater on a scale of 1-4); or (2) at least 80% (24/30) of participants were satisfied (4 or greater, on a scale of 1-5) with the intervention and that at least 80% (24/30) of participants would recommend it to others (4 or greater) on a scale of 1 to 5.

The insomnia Severity Index (ISI) is a 7-item questionnaire that measures a patient’s perceptions of the nature, impact, and severity of insomnia. This scale rates items on a Likert scale ranging from 0 to 4 for a maximum score of 28. ISI has good internal consistency (Cronbach α 0.90) [57]. Higher scores indicate more severe insomnia. A 6- to 8.4-point reduction is associated with moderate improvement in insomnia after treatment [50,58]. This measure was collected via QualtricsXM at baseline and week 7. We estimated that the effect size of the intervention would be at least as high as other effect sizes reported in the literature (approximately Cohen d 0.6 to 0.7) for nontherapist delivered CBT-I.

Sleep medication use was assessed by asking participants to report both prescription and over-the-counter (OTC) medications used for sleep within the last 2 weeks at baseline and week 7 via QualtricsXM.

The brief DBAS-16 is a 16-item measure that assesses faulty beliefs and unrealistic expectations about sleep using statements such as “I am worried that I may lose control over my ability to sleep” [54]. The measure has a good internal consistency of 0.77 [54]. Higher scores indicate more DBAS. This measure was collected via QualtricsXM at baseline and week 7 and within the app during week 5.

Participants were asked to report any new symptoms, illnesses, or diagnosed health problems to study team members. These adverse events were categorized using the Common Terminology Criteria for Adverse Events [59].

Analysis
The data were summarized using descriptive statistics (mean, SD, median, percentage, and frequency). The demographic and clinical characteristics including ISI, DBAS-16, and adverse events were descriptively analyzed with effect size calculations done for the validated quantitative measure. The frequency of study participants taking prescription and OTC sleep medications at baseline and week 7 was compared using $\chi^2$ testing. A full missing-data analysis was conducted using IBM SPSS statistical software (version 27; IBM Corp).

Ethical Considerations
This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Institutional Review Board of the University of Michigan on March 18, 2021 (HUM00194610). Informed written consent was obtained from all individual participants included in the study. The study team follows best practices for protecting patient privacy, including the encryption of participant data when at rest or in motion and that final study data sets are deidentified [60]. No protected health information was collected through the app. CHCR is a trusted IT service provider for the University of Michigan Health System and all CHCR programs undergo extensive information assurance and cybersecurity review including active penetration testing.

Results
Overview
We recruited 30 participants diagnosed with breast (n=24, 80%), prostate (n=3, 10%), and colon (n=3, 10%) cancer between March 2021 and June 2022. The mean age of participants was 54 (SD 9.1) years. Further, 27 (90%) identified as female and 3 (10%) identified as male. In total, participants reported that they were White (n=22, 73%), Black (n=5, 17%), Asian (n=4, 13%), American Indian or Alaska Native (n=1, 3%), and Hispanic or Latino (n=1, 3%). Some participants reported being of more than 1 race or ethnicity. Most (n=25, 83%) had graduated from college or had an advanced degree after college. Furthermore, 5 (17%) reported a high school diploma or some college.

Adherence
Weekly adherence rates ranged from 96% to 100% (week 1: 97%, week 2: 97%, week 3: 100%, week 4: 100%, week 5: 100%, week 6: 100%, and week 7: 97%). The overall adherence rate for the study was 99%.
Attrition
All 30 (100%) participants completed the study and provided data through week 7.

Usefulness and Satisfaction
Every app feature, with the exception of reading stories from other people with sleep problems (9 out of 10), was found to be useful by 80% (24/30) to 93% (28/30) of participants. The highest-rated components were keeping a sleep log, obtaining a personal sleep plan, and reading lessons about sleep. Further, 27 (90%) of the 30 participants were satisfied with the Mi Sleep Coach app and 28 (93%) would recommend the use of the Mi Sleep Coach app for those with sleep issues.

Insomnia
The mean change score on the ISI of 8.1 was a significant decrease from baseline (18.5, SD 4.6) to week 7 (10.4, SD 4.2; \(P\leq.001\); Cohen \(d=1.5\); see Table 1). At baseline, all ISI scores ranged from 9 to 26, indicating mild sleep disturbance for 5 (17%) participants, moderate insomnia for 19 (63%), and severe insomnia for 6 (20%). At week 7, a total of 8 (27%) participants scored in the no clinical insomnia range (0-7) while 18 (60%) scored in the mild sleep disturbance range (8-14) and 4 (13%) scored in the moderate insomnia range (15-21). Further, 6 (20%) participants scored in the severe insomnia (22-28) range at baseline, and none scored in the severe insomnia range at 7 weeks (see Table 2).

Table 1. ISI\(^{a}\) and dysfunctional beliefs about sleep at baseline to week 7 among cancer survivors in the Mi Sleep Coach pilot study.

<table>
<thead>
<tr>
<th></th>
<th>Value, mean (SD)</th>
<th>Cohen (d)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISI</td>
<td></td>
<td>1.5</td>
<td>(\leq.001)</td>
</tr>
<tr>
<td>Baseline</td>
<td>18.5 (4.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>10.4 (4.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBAS(^{b})</td>
<td></td>
<td>1.3</td>
<td>(\leq.001)</td>
</tr>
<tr>
<td>Baseline</td>
<td>47 (10.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>36 (12.7)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\)ISI: Insomnia Severity Index.
\(^{b}\)DBAS: Dysfunctional Beliefs and Attitudes About Sleep.

Table 2. ISI\(^{a}\) clinical ranges at baseline and week 7 among cancer survivors in the Mi Sleep Coach pilot study (N=30).

<table>
<thead>
<tr>
<th>Clinical range</th>
<th>Baseline, n (%)</th>
<th>Week 7, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No clinical insomnia (ISI scores 0-7)</td>
<td>0 (0)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>Mild sleep disturbance (ISI scores 8-14)</td>
<td>5 (17)</td>
<td>18 (60)</td>
</tr>
<tr>
<td>Moderate insomnia (ISI scores 15-21)</td>
<td>19 (63)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Severe insomnia (ISI scores 22-28)</td>
<td>6 (20)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\(^{a}\)ISI: Insomnia Severity Index.

Sleep Medication
At baseline, 7 (23%) participants were taking prescription medications for sleep. At 7 weeks, only 1 (3%; \(\chi^2 P<.001\)) participant was still taking prescription sleep medications. Likewise, at baseline, 14 (47%) participants were taking OTC medications for sleep. At 7 weeks, this number deceased, so that 9 (30%; \(\chi^2 P=.03\)) participants were still taking OTC sleep medications.

About DBAS
The mean change score on the DBAS-16 of 11 was a significant decrease from baseline (47, SD 10.8) to week 7 (36, SD 12.7, \(P<.001\), Cohen \(d=1.3\); see Table 1). Beliefs and attitudes about sleep were moderately positively correlated with insomnia, as measured by the ISI, at baseline (\(R=0.574\)) but only mildly correlated with insomnia at 7 weeks.

Adverse Events
In total, 4 participants reported adverse events while in this study. All adverse events were not related to the Mi Sleep Coach intervention. Further, 2 were mild (1 pain in extremity and 1 stomach pain) and 2 were moderate and required medication (2 allergic rhinitis).

Discussion
Principal Findings
This study evaluated the innovative Mi Sleep Coach mobile dCBT-I program tailored for survivors of cancer. The Mi Sleep Coach program was found to be highly feasible with 100% (N=30) of participants completing all 7 weeks of the study and providing data through week 7. Additionally, the overall program adherence rate for the study was 98%. In total, 90% (n=27) of the Mi Sleep Coach app participants were satisfied with the program and 93% (n=28) would recommend the use of the Sleep Coach app for those with sleep issues. The Mi Sleep Coach...
Coach program also showed promise in addressing insomnia among survivors of cancer. Program users experienced a significant decrease in the mean change on the ISI of 8.1, from baseline (18.5, SD 4.6) to week 7 (10.4, SD 4.2; \(P\leq.001\)), which is also represented by shifts from moderate-to-severe to mild-to-no clinical insomnia. Program users also reported a spontaneous decrease in the use of both OTC and prescription sleep medications.

The high rates of engagement and adherence to the Mi Sleep Coach program are notable. Some caution that digital interventions are particularly susceptible to attrition because participants can easily access resources a few times before abandoning them [55]. However, the high usefulness and satisfaction scores attributed to the Mi Sleep Coach app might indicate a lower likelihood of dropout. Recent systematic reviews and meta-analyses of randomized controlled trials of internet-delivered CBT have found attrition rates in the range of 21.6% to 24.7% [27,61]. While only a few of the studies focus on survivors of cancer, similar or lower dropout rates have been found in other CBT-I studies in survivors of cancer [28,29,62]. More research is needed to understand how to ensure the retention of participants in digital interventions and the possible role of attrition as a moderating variable [63]. For multicomponent CBT-I interventions, adherence varies widely depending on the details of the study. Similar to our findings, a randomized controlled trial of 30 survivors of breast cancer found adherence to 2 eHealth CBT-I components ranged from 99% to 100% [64]. However, in that study, participants met weekly in a group video conference environment, making it less like the self-managed approach used by Mi Sleep Coach. Adherence to the Mi Sleep Coach program compares favorably to reports from a fully internet-delivered CBT-I intervention for survivors of breast cancer (adherence ranges from 59.7% to 82.1%) [29]. The use of engagement strategies in the Mi Sleep Coach, such as the digital health agent, may have resulted in higher adherence rates.

The large treatment effects of the Mi Sleep Coach program (eg, Cohen \(d\)=1.5 for ISI) are encouraging. This effect is consistent with findings from 2 meta-analyses of largely person-delivered CBT-I for survivors of cancer that found large treatment effects (pooled effect sizes 0.77-0.78) [16,22]. The Mi Sleep Coach effects also compare favorably to the effects of dCBT-I in general where 2 meta-analyses found moderate pooled effects (eg, Cohen \(d\)=0.40, decrease in insomnia severity by 4.3 points) [26,27]. The finding of reduced use of sleep medications among Mi Sleep Coach participants may be particularly important. A recent national study found that approximately half of survivors of cancer with poor sleep reported using sleep medications [65]. This is of concern because the use of OTC or prescription sleep medications is associated with an increased risk of falls and fractures among older adults and survivors of cancer [66-68].

The limitations of this study include a small sample size, lack of a comparator, and the use of self-report measures for sleep. However, this study was a feasibility study and as such was designed to assess adherence, attrition, and usefulness. Additionally, sleep diaries have been used to measure the primary outcome in numerous sleep studies and have been validated against both functional and physiologic measures of sleep [69].

The results of this pilot study support that the Mi Sleep Coach app is a promising intervention for cancer-related insomnia. Future research should be conducted to test the Mi Sleep Coach app in a randomized controlled trial with sufficient control to determine its efficacy. If proven to significantly decrease insomnia in survivors of cancer in future randomized controlled clinical trials, this intervention would provide more survivors of cancer with easy access to evidence-based CBT-I treatment.

Acknowledgments
This work was supported by the Breast Cancer Research Foundation and by the National Cancer Institute (P30CA046592) by the use of the Rogel Cancer Center Health Communications Shared Resource.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
LA, NA, DLB, and DAC contributed to the study conception and design. Material preparation and data collection and analysis were performed by LA, NA, DLB, DAC, CL, and BC. The first draft of this paper was written by NA, CL, and LA. All authors commented on previous versions of this paper. All authors read and approved the final paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Mi Sleep Coach feasibility presentation.
[PPTX File, 20159 KB - formative_v8i1e55402_app1.pptx ]

References
https://formative.jmir.org/2024/1/e55402


51. Spielman AJ. Sleep need questionnaire. URL: https://deploymentpsych.org/system/files/member_resource/Sleep%20Need%20Questionnaire%20fillable%20autoscore_0.pdf [accessed 2024-04-05]


Abbreviations

- CBT-I: cognitive behavioral therapy for insomnia
- CHCR: Center for Health Communications Research
- DBAS: Dysfunctional Beliefs and Attitudes About Sleep
- dCBT-I: digital delivery of cognitive behavioral therapy for insomnia
- ISI: Insomnia Severity Index
- MI: motivational interviewing
- OTC: over-the-counter
Integrating Virtual Mindfulness-Based Stress Reduction Into Inflammatory Bowel Disease Care: Mixed Methods Feasibility Trial

Kaitlyn Delaney Chappell¹, MSc; Diana Meakins², MD; Melanie Marsh-Joyal³, MD; Allison Bihari¹, BSc; Karen J Goodman¹, PhD; Jean-Michel Le Melledo², MSc, MD; Allen Lim¹, MD; Farhad Peerani¹, MD; Karen Ivy Kroeker¹, MSc, MD

¹Division of Gastroenterology, Department of Medicine, University of Alberta, Edmonton, AB, Canada
²Department of Psychiatry, University of Alberta, Edmonton, AB, Canada
³Department of Psychiatry, Royal Alexandra Hospital, Edmonton, AB, Canada

Corresponding Author:
Karen Ivy Kroeker, MSc, MD
Division of Gastroenterology
Department of Medicine
University of Alberta
130 University Campus NW
Edmonton, AB, T6G 2X8
Canada
Phone: 1 780 492 4873
Fax: 1 780 492 8121
Email: karen.kroeker@ualberta.ca

Abstract

Background: Individuals with inflammatory bowel disease (IBD) experience cycles of aggressive physical symptoms including abdominal pain, diarrhea, and fatigue. These acute symptoms regress and return, and chronic symptoms and complications often linger. The nature of the disease can also cause individuals to experience psychological distress including symptoms of anxiety and depression; however, unlike the physical symptoms of IBD, these psychological symptoms often remain untreated.

Objective: This study aims to evaluate the feasibility, acceptability, and effectiveness of virtual mindfulness-based stress reduction (v-MBSR) for adults with IBD.

Methods: IBD patients with self-reported anxiety or depression were recruited from clinics in Alberta, Canada to participate in an 8-week v-MSBR intervention. Eligible patients participated in v-MBSR delivered by psychiatrists using a videoconferencing platform. Primary feasibility outcomes included trial uptake, adherence, attendance, and attrition rates. Secondary effectiveness outcomes included measures of anxiety, depression, quality of life (QoL), and mindfulness. Effectiveness data were collected at 3 time points: baseline, at intervention completion, and 6 months after completion. To further assess feasibility and acceptability, participants were invited to participate in a semistructured interview after completing v-MBSR.

Results: A total of 16 of the 64 (25%) referred patients agreed to participate in v-MBSR with the most common reason for decline being a lack of time while 7 of the 16 (43.8%) participants completed the program and experienced encouraging effects including decreased anxiety and depression symptoms and increased health-related QoL, with both improvements persisting at 6-month follow-up. Participants described improved coping strategies and disease management techniques as benefits of v-MBSR.

Conclusions: Patients with IBD were interested in a psychiatrist-led virtual anxiety management intervention, but results demonstrate v-MBSR may be too time intensive for some patients with IBD patients. v-MBSR was acceptable to those who completed the intervention, and improvements to anxiety, depression, and QoL were promising and sustainable. Future studies should attempt to characterize the patients with IBD who may benefit most from interventions like v-MBSR.

(JMIR Form Res 2024;8:e53550) doi:10.2196/53550
KEYWORDS
inflammatory bowel disease; psychosocial care; multidisciplinary care; quality of care; quality of life; mental health; adult; adults; anxiety; depression; IBD; virtual mindfulness; feasibility trial; clinic; health facility; Canada; semistructured interview; psychiatrist; psychiatrists; videoconferencing; effectiveness; v-MBSR; coping; coping strategy

Introduction

Inflammatory bowel diseases (IBDs), including Crohn disease (CD) and ulcerative colitis, are incurable diseases of the digestive tract with debilitating symptoms that can impair many aspects of a person’s life [1]. There is a large body of evidence suggesting the existence of a bidirectional connection between the gut and brain [2], thus research has begun to investigate the reciprocal relationship between mental health and the pathogenesis and severity of IBD [3,4]. Many high-quality studies have observed significant relationships between symptoms of psychological distress, including anxiety and depression, and both the severity of disease symptoms [5] as well as the development of flares [6].

Psychiatric comorbidities are common in people with IBD. A recent study reported that compared with the general population, adults with IBD have 10 times the risk of suicide, anxiety diagnoses, and depression diagnoses [7]. Further, a meta-analysis of over 30,000 patients with IBD estimated the prevalence of anxiety and depression symptoms in patients with active IBD to be 57.6% and 38.9%, respectively [8]. On top of these comorbid psychiatric risks, approximately 25%-33% of patients with IBD experience symptoms of psychological distress and posttraumatic stress symptoms from medical procedures, hospital stays, disease symptoms, and pain [9]. Despite these associations, most IBD patients’ mental health concerns often remain untreated [1,9].

Increasing one’s ability to cultivate mindfulness is a skill shown to have a protective effect on mental health by mediating the negative effects of the symptoms of psychological distress described above [10,11]. One of the aims of mindfulness-based interventions (MBIs), such as mindfulness-based stress reduction (MBSR), is to help develop the trait of mindfulness that can aid participants in reducing their feelings of anxiety and allow these benefits to persist for many years after completing the intervention. Specifically, MBSR encourages the development of mindfulness by teaching participants to approach their thoughts with openness and acceptance and by helping them to focus on and be nonjudgmental of the present moment [10,12,13]. Within the medical setting, in-person MBSR is effective in improving health-related quality of life (HRQoL) and reducing fatigue, stress, anxiety, and depression [14,15] in a variety of patient populations; however, a limited number of studies have focused on the IBD population. Further, there is little evidence to support the feasibility and effectiveness of the virtual delivery of mindfulness interventions, including virtual mindfulness-based stress reduction (v-MBSR).

The objective of this study was to assess if v-MBSR was feasible, acceptable, and effective for patients with IBD using both quantitative and qualitative approaches.

Methods

Design
This study was a multicenter, single-arm, feasibility trial. Study staff recruited patients with IBD from academic and nonacademic outpatient clinics in Edmonton, Alberta, Canada. We invited participants who completed the intervention to participate in a nested qualitative study.

Ethical Considerations
The University of Alberta research ethics board granted ethics approval for the project as 2 separate studies (Pro00108955 and Pro00119852). Informed consent was obtained from all participants, and participants had the option to withdraw from the study at any time. All participant data were deidentified for analysis. Participants did not receive any compensation for their participation in the intervention or in the interviews.

Participants and Recruitment
The research team recruited people with IBD experiencing stress, anxiety, or depression symptoms from 4 hospital outpatient clinics and a variety of gastroenterology clinics in Edmonton, Alberta, Canada between January and August 2022. Patients could self-refer using the posted study advertisements or could be referred by their gastroenterologist. Inclusion criteria included a confirmed diagnosis of CD or ulcerative colitis, age between 18 and 65 years, ability to communicate in English, and stable IBD and psychotropic medications for 6 weeks prior to the start of intervention. Exclusion criteria included ongoing corticosteroid use, having undergone surgery within 6 weeks of v-MBSR start date, and a history of or existing psychiatric-specific symptoms including psychotic or dissociative symptoms, severe active untreated drug use, severe self-injurious behavior, suicidal ideation, or cognitive impairment. We chose to add criteria regarding medication changes and surgery within 6 weeks of the start of the intervention in order to limit additional factors that could influence psychosocial well-being and quality of life (QoL).

At the time of recruitment, many gastroenterologists in Edmonton were conducting appointments via telehealth, so the study coordinator (KDC) contacted referred patients over the phone. If patients were not reached after 3 call attempts, they were categorized as “never reached”. Patients interested in participating signed an electronic consent form [16] and began the screening process. As research consent was only obtained after the referral, individuals who declined study participation were given the option to provide a reason for declining. The first step of screening involved participants completing an evaluation of their current anxious and depressive symptoms by completing a Patient-Health Questionnaire–Somatic, Anxiety, and Depressive Symptoms Scale (PHQ-SADS). Participants with at least mild symptoms (defined as a PHQ-SADS score ≥5) were eligible to proceed to the next step of screening—a
60-minute semistructured assessment with a psychiatrist. One of the collaborating psychiatrists conducted the assessment digitally, discussed the expectations within v-MBSR, and ensured the intervention was safe for the participant. After the assessment, eligible participants had 1 week to consider participating before confirming their preferred start date.

**Intervention**

The MBSR intervention groups followed the standard protocol developed by Kabat-Zinn [10]. Participation involved 8 weekly sessions of 2.5 hours each, a 5-hour weekend session after week 6, and home practice lasting between 45 and 60 minutes per day. We provided participants with audio files and a program manual to assist in home practice. Phillips [17], a certified MBSR teacher, developed the protocol and audio recordings. The same 3 facilitators delivered the intervention to groups of 7-9 participants. All facilitators regularly deliver MBSR as part of their clinical practice and are qualified MBSR teachers. The facilitators delivered the weekly sessions in the evenings to accommodate participants attending school or work during the daytime, and the videoconferencing platform met legislative and provincial safety and privacy guidelines.

**Feasibility and Acceptability Outcome Measures**

The primary aim of this study was to assess the feasibility and acceptability of v-MBSR in adults with IBD. We selected 4 primary outcomes, adapted from the framework and recommendations for feasibility trials [18,19] to assess feasibility: recruitment success, adherence, attendance, and attrition. We defined recruitment success verbally as trial uptake and mathematically as the number of participants who enrolled divided by the total number of participants referred. We assessed adherence in the participants who completed the intervention. We defined adherence as the average number of minutes that participants who completed the intervention practiced per day divided by the average number of minutes they were asked to practice per day. As per the protocol of MBSR [10], the recommended home practice is 45 minutes each day. We assessed attendance and attrition together and verbally defined this combined measure as the number of participants who completed the intervention. As with adherence, we defined completion as recommended by the original MBSR protocol [10]: participants were considered to have completed v-MBSR if they attended 6 out of 8 weekly sessions as well as the weekend session. To assess this mathematically, we calculated the proportion of participants who we considered to have completed the intervention by dividing the number of enrolled participants who did not discontinue the intervention and met the defined completion criteria by the total number of participants who enrolled.

**Short-Term Effectiveness Outcome Measures**

The secondary outcomes of the trial were the short-term effectiveness of v-MBSR in the participants who completed it across the domain of psychosocial functioning. We defined the intervention effect as a change from baseline in psychosocial functioning following the intervention, with participants acting as their own controls. We also collected measures of disease activity to evaluate if patients experienced active disease during or after their participation in the group. The collection of short-term effectiveness data occurred at 3 time points: pregroup, postgroup, and 6-month postgroup. We collected baseline or pregroup data up to 2 weeks before the v-MBSR group started, postgroup data at 1 week following the completion of the group, and 6-month follow-up data 24-26 weeks after participants completed the group.

**Psychosocial Functioning Outcome Measures**

**Psychological Distress (Anxiety and Depression Symptoms)**

We measured psychological distress using assessment tools for somatic, anxiety, and depressive symptoms; HRQoL; dispositional mindfulness; and self-compassion.

The PHQ-SADS [20] is a 3-part scale used to assess anxiety, depression, somatization, and general stress symptoms. Scores range from 0 to 27 with higher scores indicating a more severe symptom burden. PHQ-SADS above 5 indicates the presence of at least mild symptoms and a score above 10 is considered clinically concerning [20].

**Health-Related Quality of Life (HRQoL)**

The Short Inflammatory Bowel Disease Questionnaire (SIBDQ) [21] is a tool used to assess disease-specific HRQoL in people with IBD. The 10-item questionnaire is adapted from the Inflammatory Bowel Disease Questionnaire and assesses QoL across 4 dimensions: bowel, systemic, social, and emotional. Scores range from 10 to 70 with lower scores indicating poorer QoL. Use of the SIBDQ, authored by Dr Jan Irvine et al, was made under license from McMaster University, Hamilton, Canada.

**Mindfulness**

The Mindful Attention and Awareness Scale [22] is a tool used to assess dispositional mindfulness, also referred to as the trait of mindfulness. The 15-item questionnaire focuses on respondents’ open and receptive awareness and their ability to focus their attention on the present. Scores range from 1 to 6 and higher scores indicate greater levels of dispositional mindfulness.

**Self-Compassion**

The Short Form Self-Compassion Scale [23] is a scale that assesses respondents’ ability to be kind and understanding to themselves in instances of pain and failure. The 12-item scale is adapted from the Self-Compassion Scale. Scores range from 1 to 5 with higher scores indicating greater levels of self-compassion.

**Disease Activity Outcome Measures**

The Harvey Bradshaw Index [24] and the Partial Mayo Score [25] were used to assess IBD clinical symptoms and disease activity. These clinical indicators of disease were supplemented with inflammatory markers of disease activity, including C-reactive protein and fecal calprotectin (FCP) if laboratory results were available within 2 months of the time point of interest.
Baseline Characteristics
We collected participant baseline and demographic information 0-2 weeks before the group start date. Information collected included age, self-identified gender, type of IBD, age at diagnosis, employment status, a list of medications, and the adverse childhood experience (ACE) questionnaire.

The ACE Scale [26] is a 10-item questionnaire that assesses childhood rearing and maltreatment contexts with scores ranging from 0 to 10. Previous studies have associated higher ACE scores with mental illness indicators including emotional distress and poorer general health outcomes [27]. ACE scores can be useful to identify patients who may have greater mental health needs now or in the future.

Statistical Analysis
We included all participants with baseline data in the pregroup analysis. Only participants who met the completion criteria were included in the postgroup and 6-month follow-up analysis. We performed the calculations of descriptive statistics and analyses using Stata (version 17.0; StataCorp) [28]. Our calculations included proportions as well as means with 95% CIs to demonstrate precision. We compared the participants’ preintervention and postintervention mean outcome scores with the preintervention score acting as the control.

Nested Qualitative Study
To understand participants’ experiences with v-MBSR, we invited participants who completed the intervention to participate in a semistructured interview. The interviews aimed to gather more data regarding the feasibility, acceptability, and effectiveness of v-MBSR and to identify contextual factors that may be relevant to implementing a mental health intervention such as v-MBSR in IBD care.

We designed the interviews from the philosophical perspective of interpretive description as described by Thorne et al [29,30]. The interpretive description focuses on gaining a deeper understanding of the participant’s experience and because of this, reaching saturation is not a priority of this type of inquiry. We invited all participants who completed v-MBSR to participate in the interviews. All videoconference interviews took place between November and December 2022. Subsequently, the first author (KDC) transcribed the video interviews and generated codes and then themes from the transcripts using latent thematic analysis. In keeping with best practices for qualitative inquiry, a critical friend with experience in qualitative methods (AB) evaluated and audited the data.

Results
Baseline Characteristics
Table 1 summarizes the characteristics of participants who enrolled in the feasibility trial. The average age was 36 (range 18-55) years and 62.5% (n=10) of participants were female, 50% (n=8) of participants had CD, and 68.8% (n=11) of the participants were employed or studying full-time. At the time of their participation, 87.5% (n=14) of participants were using a biologic to treat their disease and 56.3% (n=9) of participants were using at least 1 psychotropic medication with the most common (8/16, 50%) being antidepressants. The average ACE score was 3.0 (95% CI 1.6-4.5).
Table 1. Participant demographic information including self-identified sex, age, diagnosis information, education, and medication details (N=16).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10 (62.5)</td>
</tr>
<tr>
<td>Male</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td><strong>Type of IBD</strong></td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>8 (50)</td>
</tr>
<tr>
<td>UC</td>
<td>8 (50)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>3 (18.8)</td>
</tr>
<tr>
<td>25-34</td>
<td>3 (18.8)</td>
</tr>
<tr>
<td>35-44</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td>45-54</td>
<td>3 (18.8)</td>
</tr>
<tr>
<td>55-65</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td><strong>Age at diagnosis (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>4 (25)</td>
</tr>
<tr>
<td>18-29</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>5 (31.3)</td>
</tr>
<tr>
<td><strong>Years since diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>5 (31.3)</td>
</tr>
<tr>
<td>5-10</td>
<td>5 (31.3)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
</tr>
<tr>
<td>Employed full-time</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>Studying full-time</td>
<td>4 (25)</td>
</tr>
<tr>
<td>Unemployed/retired</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>Employed part-time</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td><strong>IBD medication</strong></td>
<td></td>
</tr>
<tr>
<td>Biologic therapy only</td>
<td>9 (56.3)</td>
</tr>
<tr>
<td>Combination therapy (biologic and immunosuppressant)</td>
<td>4 (25)</td>
</tr>
<tr>
<td>5-ASA&lt;sup&gt;d&lt;/sup&gt; only</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td>Biologic and 5-ASA</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td>None</td>
<td>1 (6.3)</td>
</tr>
<tr>
<td><strong>Psychotropic medication</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>9 (56.3)</td>
</tr>
<tr>
<td><strong>Amount of psychotropic medications given to people who opted “Yes”</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 (33.3)</td>
</tr>
<tr>
<td>2</td>
<td>2 (22.2)</td>
</tr>
<tr>
<td>3+</td>
<td>5 (55.6)</td>
</tr>
<tr>
<td><strong>Psychotropic medication by type</strong></td>
<td></td>
</tr>
</tbody>
</table>
### Feasibility and Acceptability Outcomes

#### Recruitment

During the 8-month recruitment period, 64 patients were referred to the trial. Nearly all patients (60/64, 93.8%) were referred by their gastroenterologist with 50% (n=32) of referrals coming from a gastroenterologist working at an academic center and 76.6% (n=49) of referrals being from male gastroenterologists. Overall, 7 individual gastroenterologists referred patients to the study.

Of the 64 patients referred, 25% (n=16) of patients enrolled in 1 of the 2 offered v-MBSR groups. Patient flow through the trial is shown in Figure 1. Other feasibility data including reasons for declining, attendance, adherence, and attrition are included in Table 2. Of note, the participants’ average adherence rate corresponded to 48% (21.7 of 45 minutes) of the recommended daily practice time. Further, no patients were excluded due to low PHQ-SADS scores or following the completion of the psychiatric assessment and all 7 participants who discontinued did so before the third weekly session.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antidepressant</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Sedative/hypnotic</td>
<td>5 (31.3)</td>
</tr>
<tr>
<td>Mood stabilizer/anticonvulsant</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>Antipsychotic</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>Stimulant</td>
<td>1 (6.3)</td>
</tr>
</tbody>
</table>

*a* IBD: inflammatory bowel disease.  
*b* CD: Crohn disease.  
*c* UC: ulcerative colitis.  
*d* 5-ASA: 5-aminosalicylic acid.
Figure 1. Flow diagram of participant progress through all stages of the study reasons for exclusion noted at each stage. MBSR: mindfulness-based stress reduction; PHQ-SADS: Patient-Health Questionnaire–Somatic, Anxiety, and Depressive Symptoms Scale; v-MBSR: virtual mindfulness-based stress reduction.

- Enrollment and consent
  - Participant expresses interest in study and self reports anxiety and/or depressive symptoms (n=4)
  - OR
  - Physician identifies patient as depressed or anxious and refers patient to study (n=60)
  - Total referrals: n=64

- Excluded (n=48), reasons:
  - Declined (n=26), reasons:
    - Too busy (n=21)
    - Not needed (n=3)
    - Group aspect unappealing (n=1)
    - Other (n=1)
  - Never reached/lost to follow-up (n=21)
  - Ineligible (n=1)

- Signs consent form and enrols in study (n=16)

- Initial screening by research coordinator
  - Screened for anxiety and depressive symptoms (PHQ-SADS≥5) and other inclusion criteria (n=16)

- Excluded (n=0)

- Final screening by psychiatrist
  - Clinical interview by psychiatrists, participant decides to participate in group (n=16)

- Excluded (n=0)

- v-MBSR starts (baseline or pre-MBSR)
  - Participants complete questionnaires up to 2 weeks before group starts
  - 8-week v-MBSR groups start (n=16)

- Excluded (n=9), reasons:
  - Did not complete v-MBSR (n=7)
  - Withdrawn due to time commitment (n=4)
  - Discontinued, no reason given (n=3)
  - Finished but did not meet completion criteria (n=2)

- v-MBSR complete (post-MBSR)
  - Participants repeat questionnaires 1 week after v-MBSR is complete (n=7)

- Excluded (n=1), reasons:
  - Lost to follow-up (n=1)

- 6 month follow-up
  - Participants repeat questionnaires 24-26 weeks after v-MBSR is complete (n=6)
Table 2. Recruitment, attrition, attendance, and adherence results.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recruitment (n=64), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Enrolled</td>
<td>16 (25)</td>
</tr>
<tr>
<td>Declined</td>
<td>26 (40.6)</td>
</tr>
<tr>
<td>Lost to follow-up/never reached</td>
<td>21 (32.8)</td>
</tr>
<tr>
<td>Ineligible</td>
<td>1 (1.56)</td>
</tr>
<tr>
<td><strong>Reasons for declining (n=26), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Too busy/lack of time</td>
<td>21 (80.8)</td>
</tr>
<tr>
<td>Help not necessary</td>
<td>3 (11.5)</td>
</tr>
<tr>
<td>Group aspect unappealing</td>
<td>1 (3.85)</td>
</tr>
<tr>
<td>Not enough help</td>
<td>1 (3.85)</td>
</tr>
<tr>
<td><strong>Attrition and attendance (n=16), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Completed</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>Finished, but did not complete</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>Discontinued</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td><strong>Reasons for discontinuing (n=7), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Lack of time</td>
<td>4 (57.1)</td>
</tr>
<tr>
<td>No reason noted</td>
<td>3 (42.8)</td>
</tr>
<tr>
<td><strong>Adherence (n=6; minutes), mean (95% CI)</strong></td>
<td></td>
</tr>
<tr>
<td>Average practice per day</td>
<td>21.7 (13.1-30.2)</td>
</tr>
</tbody>
</table>

**Participant Interviews: Barriers to v-MBSR**

We assessed barriers to the intervention in the nested qualitative study. When we asked participants to identify challenges to participating in the intervention, excessive time related to home practice was mentioned as the largest barrier by all 5 participants. One participant noted that the inflexibility of the weekly session (ie, having to attend the weekly session on the same day and at the same time every week) was also challenging. Quotes from interviews are presented in Figure 2.
Characterizing Those Who Completed the Intervention

We compared participants who completed the intervention to the remaining participants across a variety of characteristics including referral details, demographics, and baseline scores, as summarized in Tables 3 and 4. Compared with participants who did not complete v-MBSR, those who completed the intervention were all female and more likely to be older, receive care from a gastroenterologist at a nonacademic center, have CD, work or study full-time, and have higher average ACE scores. Notably, the group of participants who completed the intervention had higher average ACE scores despite having comparable psychosocial functioning scores at baseline.
Table 3. Baseline demographics and characteristics of participants organized by participant completion status.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All participants (n=16)</th>
<th>Complete participants (n=7)</th>
<th>Incomplete participants (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10 (62.5)</td>
<td>7 (100)</td>
<td>3 (33.3)</td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time work/study</td>
<td>11 (68.8)</td>
<td>4 (57.1)</td>
<td>7 (77.8)</td>
</tr>
<tr>
<td>Taking psychotropics, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>9 (56.3)</td>
<td>4 (57.1)</td>
<td>5 (55.6)</td>
</tr>
<tr>
<td>Disease type, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD a</td>
<td>8 (50)</td>
<td>5 (71.4)</td>
<td>3 (33.3)</td>
</tr>
<tr>
<td>Age (years), mean (95% CI)</td>
<td>36.4 (30.6-42.2)</td>
<td>38.1 (27.3-49.0)</td>
<td>35.0 (26.7-43.3)</td>
</tr>
<tr>
<td>Years since diagnosis, mean (95% CI)</td>
<td>10.4 (5.6-15.1)</td>
<td>9.7 (0.0-21.0) b</td>
<td>10.9 (6.2-15.6)</td>
</tr>
<tr>
<td>ACE c score, mean (95% CI)</td>
<td>3.06 (1.61-4.51)</td>
<td>3.43 (0.71-6.14)</td>
<td>2.78 (0.71-4.84)</td>
</tr>
<tr>
<td>Baseline scores, mean (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-SADS d</td>
<td>11.8 (9.2-14.3)</td>
<td>11.9 (7.5-16.3)</td>
<td>11.7 (7.8-15.6)</td>
</tr>
<tr>
<td>SIBDQ e</td>
<td>41.0 (35.7-46.2)</td>
<td>39.6 (29.0-50.2)</td>
<td>42.0 (35.0-49.0)</td>
</tr>
<tr>
<td>MAAS f</td>
<td>4.0 (3.5-4.6)</td>
<td>4.3 (3.3-5.3)</td>
<td>3.9 (3.0-4.8)</td>
</tr>
<tr>
<td>SCS g</td>
<td>2.3 (2.0-2.6)</td>
<td>2.2 (1.6-2.8)</td>
<td>2.2 (1.9-2.5)</td>
</tr>
</tbody>
</table>

aCD: Crohn disease.
b95% CI imprecise due to outliers, range 3-36.
cACE: adverse childhood experience.
dPHQ-SADS: Patient-Health Questionnaire–Somatic, Anxiety, and Depressive Symptoms Scale.
eSIBDQ: Short Inflammatory Bowel Disease Questionnaire.
fMAAS: Mindful Attention and Awareness Scale.
gSCS: Self-Compassion Scale.

Table 4. Characteristics of referring physicians organized by participant completion status.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All participants (n=16), n (%)</th>
<th>Complete participants (n=7), n (%)</th>
<th>Incomplete participants (n=9), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location of referring physician</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic center</td>
<td>9 (56.3)</td>
<td>2 (28.6)</td>
<td>7 (77.8)</td>
</tr>
<tr>
<td>Sex of referring physician</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3 (18.8)</td>
<td>1 (14.3)</td>
<td>2 (22.2)</td>
</tr>
</tbody>
</table>

Participant Interviews: Motivations, Expectations, and Past Experiences

We asked interview participants, all of whom completed v-MBSR, about their motivations to enroll in v-MBSR. All 5 noted they were motivated to engage in v-MBSR by at least 1 of the following factors: previous experience with mindfulness, a desire to improve their QoL and disease management through learning, their own research, specific intervention features such as the intervention sessions being led by a professional, and recommendations from others including family, friends, or health care providers. The 3 most common motivations were personal research, the desire to improve through learning, and recommendations from others.

We also asked interview participants to recall their expectations of v-MBSR before they participated in the intervention. Four of 5 interview participants had limited or negative expectations of the intervention. Specifically, these expectations included feelings of apprehension regarding their enjoyment of and success with the intervention as well as strong negative feelings regarding the length of the weekly meeting.

Additionally, we asked interview participants about their previous experience with trying to improve their well-being. All 5 participants described previously engaging in activities to improve their physical well-being including diet and exercise; 4 of 5 participants had tried mental health improvements including finding supportive social circles and seeing a mental health care practitioner such as a therapist or psychiatrist.
Participant Interviews: Positive Experiences and Recommendations

All 5 interview participants described their experience in v-MBSR positively and indicated that being able to participate in the group with other people with IBD contributed to their positive experience. Two of 5 participants also added that their participation led to changes in their attitudes and perceptions which contributed to the positive experience.

All 5 participants said they would recommend v-MBSR to others, and 3 participants emphasized that they would continue to practice mindfulness going forward and share their experience and tools with others.

Short-Term Effectiveness Outcomes

All 16 enrolled participants provided baseline data before the intervention began. Those who completed the intervention also provided data immediately after and 6 months after the intervention ended. One of the 7 participants who completed the intervention was lost to follow-up immediately after the intervention and did not provide postintervention data.

Psychosocial Functioning Outcomes

Figure 3, Figure 4, and Table 5 present psychosocial outcomes at all 3 time points, with results for individual participants as well as means at each time point. Participants’ mean PHQ-SADS score decreased from 11.2 at baseline to 7.8 immediately following the intervention and further to 6.4 at the 6-month follow-up. Participants’ SIBDQ scores showed a similar improvement. Their mean SIBDQ score increased from 40.7 at baseline to 48.1 postintervention. Participants’ mean SIBDQ score was 46.8 at the time of 6-month follow-up demonstrating sustained improvement.

Figure 3. Participant anxiety scores at baseline, postintervention, and 6 months postintervention organized by participant. PHQ-SADS: Patient-Health Questionnaire–Somatic, Anxiety, and Depressive Symptoms Scale; v-MBSR: virtual mindfulness-based stress reduction.
Figure 4. Participant quality of life scores baseline, postintervention, and 6 months postintervention organized by participant. SIBDQ: Short Inflammatory Bowel Disease Questionnaire; v-MBSR: virtual mindfulness-based stress reduction.
### Table 5. Participant psychosocial outcome scores at all 3 time points organized by participant.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Completion status</th>
<th>Pre&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Post&lt;sup&gt;f&lt;/sup&gt;</th>
<th>6M&lt;sup&gt;g&lt;/sup&gt;</th>
<th>Pre</th>
<th>Post</th>
<th>6M</th>
<th>Pre</th>
<th>Post</th>
<th>6M</th>
<th>Pre</th>
<th>Post</th>
<th>6M</th>
<th>Completion status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No&lt;sup&gt;h&lt;/sup&gt;</td>
<td>17.3</td>
<td>N/A</td>
<td>N/A</td>
<td>33</td>
<td>N/A</td>
<td>3.1</td>
<td>N/A</td>
<td>N/A</td>
<td>1.8</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>7.7</td>
<td>4.3</td>
<td>3.0</td>
<td>49</td>
<td>59</td>
<td>5.1</td>
<td>4.3</td>
<td>4.1</td>
<td>2.0</td>
<td>3.1</td>
<td>2.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>4.3</td>
<td>4.3</td>
<td>2.3</td>
<td>60</td>
<td>67</td>
<td>5.6</td>
<td>4.5</td>
<td>4.8</td>
<td>2.8</td>
<td>3.2</td>
<td>2.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>12.3</td>
<td>N/A</td>
<td>N/A</td>
<td>44</td>
<td>N/A</td>
<td>2.2</td>
<td>N/A</td>
<td>N/A</td>
<td>1.9</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>10.0</td>
<td>N/A</td>
<td>N/A</td>
<td>46</td>
<td>N/A</td>
<td>3.3</td>
<td>N/A</td>
<td>N/A</td>
<td>1.9</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>12.7</td>
<td>N/A</td>
<td>N/A</td>
<td>52</td>
<td>N/A</td>
<td>5.5</td>
<td>N/A</td>
<td>N/A</td>
<td>2.8</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>17.0</td>
<td>6.7</td>
<td>7.7</td>
<td>32</td>
<td>49</td>
<td>3.1</td>
<td>3.0</td>
<td>3.9</td>
<td>1.8</td>
<td>2.0</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Yes&lt;sup&gt;l&lt;/sup&gt;</td>
<td>15.7</td>
<td>N/A</td>
<td>N/A</td>
<td>33</td>
<td>N/A</td>
<td>3.5</td>
<td>N/A</td>
<td>N/A</td>
<td>3.3</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Yes</td>
<td>14.0</td>
<td>13.7</td>
<td>10.7</td>
<td>35</td>
<td>31</td>
<td>4.3</td>
<td>3.4</td>
<td>3.9</td>
<td>3.0</td>
<td>3.0</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td>9.3</td>
<td>7.0</td>
<td>8.0</td>
<td>41</td>
<td>40</td>
<td>4.1</td>
<td>3.2</td>
<td>2.3</td>
<td>1.8</td>
<td>3.1</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>No</td>
<td>21.3</td>
<td>N/A</td>
<td>N/A</td>
<td>28</td>
<td>N/A</td>
<td>2.8</td>
<td>N/A</td>
<td>N/A</td>
<td>2.1</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>No</td>
<td>10.0</td>
<td>N/A</td>
<td>N/A</td>
<td>47</td>
<td>N/A</td>
<td>4.3</td>
<td>N/A</td>
<td>N/A</td>
<td>2.7</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>No&lt;sup&gt;h&lt;/sup&gt;</td>
<td>10.3</td>
<td>N/A</td>
<td>N/A</td>
<td>30</td>
<td>N/A</td>
<td>4.3</td>
<td>N/A</td>
<td>N/A</td>
<td>2.2</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>No</td>
<td>4.7</td>
<td>N/A</td>
<td>N/A</td>
<td>50</td>
<td>N/A</td>
<td>3.7</td>
<td>N/A</td>
<td>N/A</td>
<td>2.0</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Yes</td>
<td>15.0</td>
<td>10.7</td>
<td>7.0</td>
<td>27</td>
<td>43</td>
<td>3.7</td>
<td>3.3</td>
<td>3.3</td>
<td>1.8</td>
<td>3.8</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>No</td>
<td>6.7</td>
<td>N/A</td>
<td>N/A</td>
<td>48</td>
<td>N/A</td>
<td>5.6</td>
<td>N/A</td>
<td>N/A</td>
<td>2.7</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean in all participants</td>
<td>11.8</td>
<td>N/A</td>
<td>N/A</td>
<td>41</td>
<td>N/A</td>
<td>4.0</td>
<td>N/A</td>
<td>N/A</td>
<td>2.3</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean in completed participants only (95% CI)</td>
<td>11.9 (6.1-16.3)</td>
<td>7.8 (3.9-11.7)</td>
<td>6.4 (3.1-9.8)</td>
<td>40 (28-53)</td>
<td>47 (34-62)</td>
<td>4.3 (3.3-5.3)</td>
<td>3.6 (3.0-4.2)</td>
<td>3.7 (2.8-4.6)</td>
<td>2.2 (1.6-2.8)</td>
<td>3.0 (2.4-3.6)</td>
<td>2.6 (2.0-3.3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>PHQ-SADS: Patient-Health Questionnaire–Somatic, Anxiety, and Depressive Symptoms Scale.

<sup>b</sup>SIBDQ: Short Inflammatory Bowel Disease Questionnaire.

<sup>c</sup>MAAS: Mindful Attention and Awareness Scale.

<sup>d</sup>SCS-SF: Short Form Self-Compassion Scale.

<sup>e</sup>Pre: score preintervention.

<sup>f</sup>Post: score postintervention.

<sup>g</sup>6M: score 6 months postintervention.

<sup>h</sup>Did not complete, but did finish.

<sup>i</sup>N/A: not applicable.

<sup>j</sup>Lost to follow-up.

### Participant Interviews: Perceived Benefits

When asked about the benefits of the program, all 5 interview participants noted the main benefits arose from the new techniques they were able to develop. All 5 indicated that they developed coping skills allowing them to better manage their anxiety and stress in addition to new focus and mindfulness tools. Four of 5 participants also noted they benefited from novel techniques that allowed them to manage disease symptoms and pain.

### Disease Activity Outcomes

As expected, given the study eligibility criteria, all participants started the intervention in inflammatory remission. One participant (Participant #15) experienced a flare after the intervention concluded, indicated by FCP levels greater than 250 µg/mg. All other participants remained in clinical remission throughout and in the 6 months following the intervention.

### Discussion

#### Principal Results

Together, the mix of quantitative and qualitative data from this study provides insight into the feasibility, acceptability, and preliminary effectiveness of v-MBSR for patients with IBD. To our knowledge, this is the first study to evaluate v-MBSR in this patient population and also the first study to use qualitative methods to explore experiences with MBSR among IBD patients.

Two other studies examined MBSR as a treatment for people with IBD: a 2014 randomized controlled trial conducted by...
Conclusions

Despite some of the struggles with recruitment and retention, there was substantial interest from patients with IBD and providers in a free, virtual, and psychiatrist-led stress reduction intervention, and participants who completed the intervention saw significant and sustained improvements to their psychosocial health. In view of these favorable results, v-MBSR may serve as a promising integrated treatment for a subset of patients with IBD who are able and willing to manage the intervention’s large time commitment and could be a beneficial tool to help patients with IBD regain a sense of control over their physical and mental health while they are in remission. Future studies should investigate if telehealth-based group interventions are noninferior to their face-to-face equivalents and characterize the patients with IBD who are able to commit to and benefit from intensive MBIs.
Acknowledgments

This work was supported by Pfizer Canada through a Quality Improvement Grant (76111355). The authors would like to thank the patients for their participation in the study. The authors thank Catherine Phillips for generously allowing us to use her mindfulness-based stress reduction course materials and for cofacilitating the intervention. The authors also appreciate the time Sarah MacArthur gave to cofacilitate the intervention and help with the psychiatric interviews. Further, the authors would like to thank Leo Dieleman, Lillian Du, Frank Hoentjen, and Karen Wong for referring their patients to the study.

Data Availability

All data generated during this study are available in this published thesis (KDC) [45].

Conflicts of Interest

None declared.

References


Abbreviations

ACE: adverse childhood experience
CD: Crohn disease
FCP: fecal calprotectin
HRQoL: health-related quality of life
IBD: inflammatory bowel disease
MBI: mindfulness-based intervention
MBSR: mindfulness-based stress reduction
PHQ-SADS: Patient-Health Questionnaire–Somatic, Anxiety, and Depressive Symptoms Scale
QoL: quality of life
SIBDQ: Short Inflammatory Bowel Disease Questionnaire
v-MBSR: virtual mindfulness-based stress reduction

©Kaitlyn Delaney Chappell, Diana Meakins, Melanie Marsh-Joyal, Allison Bihari, Karen J Goodman, Jean-Michel Le Mello, Allen Lim, Farhad Peerani, Karen Ivy Kroeker. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Impact of Electronic Patient-Reported Outcomes on Unplanned Consultations and Hospitalizations in Patients With Cancer Undergoing Systemic Therapy: Results of a Patient-Reported Outcome Study Compared With Matched Retrospective Data

Andreas Trojan¹,², MD; Christian Kühne², MD; Michael Kiessling², MD, PhD; Johannes Schumacher³, PhD; Stefan Dröse³, PhD; Christian Singer⁴, MD; Christian Jackisch⁵,⁶, MD; Christoph Thomssen⁷, MD; Gerd A Kullak-Ublick³, MD

¹Oncology, Breast Center Zürichsee, Horgen, Switzerland
²Department of Clinical Pharmacology and Toxicology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
³palleos healthcare, Wiesbaden, Germany
⁴Center for Breast Health and Female Medicine, University Hospital Vienna, Vienna, Austria
⁵Sana Clinic Offenbach, Offenbach, Germany
⁶Evangelische Kliniken Essen-Mitte GmbH, Essen, Germany
⁷Department of Gynecology, Martin Luther University Halle-Wittenberg, Halle, Germany

Corresponding Author:
Andreas Trojan, MD
Oncology
Breast Center Zürichsee
Seestrasse 88
Horgen, 8810
Switzerland
Phone: 41 76 34 30 200
Email: trojan@1st.ch

Abstract

Background: The evaluation of electronic patient-reported outcomes (ePROs) is increasingly being used in clinical studies of patients with cancer and enables structured and standardized data collection in patients’ everyday lives. So far, few studies or analyses have focused on the medical benefit of ePROs for patients.

Objective: The current exploratory analysis aimed to obtain an initial indication of whether the use of the Consilium Care app (recently renamed medidux; mobile Health AG) for structured and regular self-assessment of side effects by ePROs had a recognizable effect on incidences of unplanned consultations and hospitalizations of patients with cancer compared to a control group in a real-world care setting without app use. To analyze this, the incidences of unplanned consultations and hospitalizations of patients with cancer using the Consilium Care app that were recorded by the treating physicians as part of the patient reported outcome (PRO) study were compared retrospectively to corresponding data from a comparable population of patients with cancer collected at 2 Swiss oncology centers during standard-of-care treatment.

Methods: Patients with cancer in the PRO study (178 included in this analysis) receiving systemic therapy in a neoadjuvant or noncurative setting performed a self-assessment of side effects via the Consilium Care app over an observational period of 90 days. In this period, unplanned (emergency) consultations and hospitalizations were documented by the participating physicians. The incidence of these events was compared with retrospective data obtained from 2 Swiss tumor centers for a matched cohort of patients with cancer.

Results: Both patient groups were comparable in terms of age and gender ratio, as well as the distribution of cancer entities and Joint Committee on Cancer stages. In total, 139 patients from each group were treated with chemotherapy and 39 with other therapies. Looking at all patients, no significant difference in events per patient was found between the Consilium group and the control group (odds ratio 0.742, 90% CI 0.455-1.206). However, a multivariate regression model revealed that the interaction term between the Consilium group and the factor “chemotherapy” was significant at the 5% level (P=.048). This motivated a corresponding subgroup analysis that indicated a relevant reduction of the risk for the intervention group in the subgroup of
patients who underwent chemotherapy. The corresponding odds ratio of 0.53, 90% CI 0.288-0.957 is equivalent to a halving of the risk for patients in the Consilium group and suggests a clinically relevant effect that is significant at a 2-sided 10% level ($P=.08$, Fisher exact test).

Conclusions: A comparison of unplanned consultations and hospitalizations from the PRO study with retrospective data from a comparable cohort of patients with cancer suggests a positive effect of regular app-based ePROs for patients receiving chemotherapy. These data are to be verified in the ongoing randomized PRO2 study (registered on ClinicalTrials.gov; NCT05425550).

Trial Registration: ClinicalTrials.gov NCT03578731; https://www.clinicaltrials.gov/ct2/show/NCT03578731

International Registered Report Identifier (IRRID): RR2-10.2196/29271

(JMIR Form Res 2024;8:e55917) doi:10.2196/55917

KEYWORDS
systemic cancer therapy; electronic patient-reported outcome; ePRO; ePROs; Consilium Care; medidux; unplanned consultation; hospitalization; hospitalizations; hospitalized; cancer; oncology; side effect; side effects; adverse; chemotherapy; patient reported outcome; PRO; PROs; mobile health; mHealth; app; apps; application; applications; mobile phone

Introduction

An important component of optimal medical care is considered to be guaranteed when patients actively participate in their treatment and are involved in decisions about their treatment plan [1]. The use of smartphones during the outpatient treatment of patients with cancer can enable low-threshold contact between patients and their treatment centers and facilitate patient involvement. The data recorded by patients (patient reported outcomes; PROs) can be recorded electronically in real time (electronic patient reported outcomes; ePROs) using a smartphone app and made available to the doctor. The use of ePROs thus enables ongoing recording of patients’ daily well-being and state of health. The evaluation of ePROs is also used in clinical studies of patients with cancer [2] and enables structured and standardized data collection in patients’ everyday lives. By analyzing the information flows and anonymized data of many patients, as well as by networking with other (research) centers, a basis can be created that can promote the quality and efficiency of treatment.

However, the benefits of app-based ePROs for patients can go beyond improving communication. It is known that patients with cancer undergoing systemic therapy often obtain side effects, with fatigue (80%), pain (48%), and nausea or vomiting (48%) being the most common [3]. The type of chemotherapy and the patient’s “performance index” are associated with the hospitalization rate [4]. It has been reported that 35% of newly diagnosed patients with cancer experienced an unplanned hospitalization, and 67% of hospitalized patients had previously been to the emergency department [5]. In 154 patients with colorectal cancer, 28% of hospitalizations were due to complications of cancer treatment, of which 19% were identified as potentially preventable [6]. Another study of patients with colorectal cancer showed that the majority of unplanned consultations (72%) occurred within 30 days of their last chemotherapy treatment. Of these unplanned visits, 10% resulted in hospitalization [7]. Among the unplanned hospital admissions of outpatients summarized in a retrospective study, 74% had received chemotherapy in the previous 6 months. Further, 69.7% of these hospitalizations occurred within 4 weeks of receiving chemotherapy. It can be assumed that by structured recording of patient reported symptoms, adverse events of cancer therapy can be recognized at an early stage and higher degrees of severity can be avoided through timely action. This is supported by recent studies. It was shown that the use of a digital app or web-based system for symptom monitoring had reduced the number of emergency admissions and hospitalizations compared to a control group and had even extended (progression-free) survival times [8-11]. A reduction in the number of serious adverse events (SAEs) compared to the control group could also be attributed to the use of the digital application in 2 studies [9,12].

Consilium Care (recently renamed medidux) is a digital app for monitoring and alleviating symptoms during and after the treatment of patients with cancer. The app enables the standardized entry of symptoms according to the Common Terminology Criteria for Adverse Events (CTCAE). When entering symptoms, users receive tips on how to alleviate them and are prompted to contact their treating physician or clinic if defined severity threshold values are exceeded. By sharing data with the treating physician, treatment can be customized precisely to the patient’s individual needs. The aim of Consilium Care or medidux is to improve the quality of life of patients with breast cancer, enable early symptom monitoring, and establish a closer connection with their treatment teams. In an initial study (registered on ClinicalTrials.gov; NCT02004496), it was shown that patient well-being and awareness of adverse effects could be improved by using the Consilium Care app in collaboration with the treating physician [13]. With further developed versions of the app, the benefits of digital patient monitoring using Consilium Care have also been demonstrated during immunotherapy and targeted therapies for cancer in the form of more efficient symptom assessment and patient-physician communication as well as a reduced need for telephone consultations [14-16]. In the PRO study (registered on ClinicalTrials.gov; NCT03578731), the primary end point investigated the extent to which the self-assessment of the severity of undesirable outcomes between patients with cancer using the Consilium Care app for 90 days and the treating physicians was consistent [17,18]. The occurrence of unplanned (emergency) consultations and hospitalizations among patients with cancer was recorded by the investigators as a secondary
end point. As a control group without app usage was not included in the PRO study, retrospective patient data from a comparable cohort of patients with cancer were used for the present analysis to obtain initial indications as to whether a reduction in emergency consultations and hospitalizations can be observed when patients use the Consilium Care (medidux) app.

Methods

Data Collection of Unplanned Consultations and Hospitalizations Within the PRO Study’s Patient Population and Acquisition of Retrospective Data for the Control Group

For this retrospective comparison of the group of patients with cancer using the Consilium Care app with a control group that did not use the app, data were compiled from 2 different sources. For the Consilium group, respective data were taken from the PRO study that was conducted as a multicenter, observational, noninterventional study. Patients with breast, colon, prostate, or lung cancer, as well as those with hematological malignancies, aged 18 years and older, receiving systemic therapy in a neoadjuvant or noncurative setting were eligible to participate after providing written informed consent. In addition, participants had to speak German and own a smartphone. Eligible participants were recruited consecutively and without preselection according to the recommendation of the local tumor boards in centers in Switzerland (10 recruited patients in Switzerland), Germany (2 centers), and Austria (1 center). The results corresponding to the primary objective (assessment of the level of agreement, K, between symptom ratings by physicians at the time of the regular consultation and the ratings derived from the daily PRO between consultations) are published in Trojan et al [18]. Patients were assigned to medical oncology visits every 3 weeks and invited for shared reporting and intended symptom review, which were preferably scheduled on days of therapeutic intervention. The observation period covered a total of 90 days (for further details see Trojan et al [18]). Relevant to the present analysis is the recording of a secondary study end point by the participating physicians during the 3-weekly oncology visits: the number and reasons for unplanned consultations were surveyed and recorded, and unplanned (emergency) hospitalizations and their duration were documented. The latter were divided into ≤2 days and >2 days. In addition to the criteria defined in the analysis of the primary end point [18], only patients who completed the 90-day observation period according to the protocol were included in the present analysis, so that patients who died during the observation period or withdrew from this study for other reasons were not included. The group that used the app (Consilium group) finally consisted of 178 patients.

For the control group, which did not use the Consilium Care app, retrospective patient data were compiled from the databases of 2 oncology centers in Zurich. The patients were selected systematically in a 2-stage process whereby attention was initially only paid to a comparable composition of the patient collective concerning the most comparable distribution of therapy types, tumor entities, and tumor stages as well as age and gender. First, patients who had started and completed systemic cancer therapy between January 2016 and October 2020 were selected consecutively in reverse order to the completion date from the Zurich Oncology Center database, that is, starting with the patient who was the last to complete 90 days of cancer treatment and then continuing with the patient with the next completion date, etc. As the proportion of patients with breast cancer was still too small, a further 43 patient data sets from the Breast Center Zurich were selected also in reverse consecutive order. Patient data were recorded anonymously, with only the respective (psynonymized) patient ID of the 2 centers serving as proof of identity during the compilation of the control group. The retrospective data of the selected patients were then searched for documented unplanned consultations and hospitalizations by a second person who was not involved in their selection. Patient demographics and relevant systemic cancer treatment data were also collected for a period of 90 days from the start of treatment.

Ethical Considerations

The PRO study was approved by the responsible ethics committees in Switzerland (Lead EC: KEK-ZH: 2017-02028), Germany, and Austria and conducted per the principles of the Declaration of Helsinki and registered on ClinicalTrials.gov (NCT03578731). Patients with breast, colon, prostate, or lung cancer, as well as those with hematological malignancies, aged 18 years and older, and initiating adjuvant or neoadjuvant systemic therapy were eligible to participate after providing written informed consent. In addition, participants had to speak German and own a smartphone. Eligible participants were recruited consecutively and without preselection according to the recommendation of the local tumor boards in centers in Switzerland, Germany, and Austria.

All study documents were de-identified by assigning a unique ID to each patient. Functional data security was ensured by identification being made only possible via the patient’s ID. The data on the patient’s device were encapsulated in the app and the data exchange was encrypted with the patient’s ID. There was no compensation provided to participants.

Informed consents for the control group were present for the included patients of the 2 respective cancer centers. These were obtained as part of normal patient care based on the internal processes at the cancer centers. These generally provide a declaration of consent for patients upon admission, which allows for the anonymized use of collected data for research purposes. Patients can refuse consent without affecting further treatment and can also withdraw it at any time.

Division of Patients Into Chemotherapy and Nonchemotherapy Subgroups

All participating patients in the PRO study as well as in the retrospective control group received systemic cancer therapy according to the local standard of care. Due to the large number of different systemic cancer therapies used as standard treatment for the cancer indications investigated, it was not possible to create a control group with exactly the same distribution of systemic drugs. Since the therapy forms’ influence on the number of unplanned consultations and hospitalizations was to
be evaluated, the therapies were recorded, and the systemic therapies were divided into 2 groups: chemotherapy and nonchemotherapy. The chemotherapy group included therapies with classic chemotherapeutic agents such as alkaloids, alkylating agents, antitumor antibiotics, etc; the nonchemotherapy group included therapeutic agents such as antihormones, aromatase inhibitors, antibodies, checkpoint inhibitors, or cyclin-dependent kinase 4/6 inhibitors. If a patient’s treatment consisted of nonchemotherapeutic agents, such as antibodies, in addition to classic chemotherapeutic agents, the patient was assigned to the chemotherapy group according to the higher expected toxicity.

**Objective of Analysis**

The objective of this retrospective analysis was to evaluate whether Consilium Care guided ePROs for symptom monitoring resulted in fewer unplanned consultations and hospitalizations (summarized as events) in patients with cancer compared to patients who did not use the app.

**Mobile App**

In the PRO study, the Consilium Care app (version 2.0) was used. In brief, the app facilitated the selection of well-being, symptoms, medication, and private notes. Symptoms, which were structured in groups according to organ systems, could be selected. The symptom entry display (52 distinct symptoms were available for which severity, onset, and duration could be indicated) was equipped with date and time stamps. Symptom severity, with descriptions based on the CTCAE, could be selected via a slider. The symptom history was displayed on a timeline with individual colors for each symptom. In addition, diary entries and information on diagnosis and therapy were indicated separately. Patients were encouraged to capture data on well-being and symptoms daily. Recording usually started on the day of the therapy’s initiation or of a change in therapy and continued throughout an observational period of 90 days. The app allowed the continuous recording of well-being and symptoms based on the CTCAE through the use of virtual analogue scales. Information for self-care (derived from the Swiss Cancer League) was provided to them via the app depending on the severity of symptoms upon data entry. In the case of severe symptoms, patients were encouraged by push notifications to seek medical advice. The history of recorded data was displayed and visualized in the form of a symptom progression chart. For further information, refer to Trojan et al [18].

**Statistical Analyses**

The question of whether the number of recorded unplanned consultations and hospitalizations per patient can be reduced by using the Consilium Care app was addressed by considering the incidence proportions (proportion of patients with at least one event) in both patient groups. Exploratory analysis was performed using the Fisher exact test and multivariate logistic regression. Effect sizes are reported as odds ratios (OR; Consilium group vs retrospective control group) with a 90% CI. The 2-sided 10% significance threshold was chosen to demonstrate initial trends in this exploratory setting that would warrant further investigation in a fully prospective trial. The results are reported for the overall group, as well as for the subgroup of patients who underwent chemotherapy. Analyses were performed with R (version 4.0.2; June 22, 2020; R Core Team).

**Results**

**Baseline Characteristics**

Both patient groups were comparable in terms of age (age range 23-83 years; mean age 54.2, SD 12.0 years in the Consilium group vs age range 27-85 years; mean age 56.4, SD 13.8 years in the control group), gender ratio (155 women and 23 men vs 157 women and 21 men), cancer entity (breast cancer n=139 vs n=133; lung cancer n=13 vs n=11, colorectal cancer n=11 vs n=15, hematological malignancies n=9 vs n=14, prostate cancer n=6 vs n=5) and American Joint Committee on Cancer stages (Table 1). Further, 139 patients from each group were treated with chemotherapy and 39 with other therapies (nonchemotherapy).
Table 1. Baseline characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Consilium group (n=178)</th>
<th>Control group (n=178)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>23-83</td>
<td>27-85</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>54.2 (12.0)</td>
<td>56.4 (13.8)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>155 (87)</td>
<td>157 (88)</td>
</tr>
<tr>
<td>Male</td>
<td>23 (13)</td>
<td>21 (12)</td>
</tr>
<tr>
<td><strong>Cancer entity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>139 (78)</td>
<td>133 (75)</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>13 (7)</td>
<td>11 (6)</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>6 (3)</td>
<td>5 (3)</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>11 (6)</td>
<td>15 (8)</td>
</tr>
<tr>
<td>Hematological malignancies</td>
<td>9 (5)</td>
<td>14 (8)</td>
</tr>
<tr>
<td><strong>AJCC stage, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>22 (12)</td>
<td>25 (14)</td>
</tr>
<tr>
<td>Stage II</td>
<td>68 (38)</td>
<td>51 (29)</td>
</tr>
<tr>
<td>Stage III</td>
<td>37 (21)</td>
<td>29 (16)</td>
</tr>
<tr>
<td>Stage IV</td>
<td>51 (29)</td>
<td>73 (41)</td>
</tr>
<tr>
<td><strong>Treatment, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>139 (78)</td>
<td>139 (78)</td>
</tr>
<tr>
<td>Nonchemotherapy</td>
<td>39 (22)</td>
<td>39 (22)</td>
</tr>
</tbody>
</table>

aAJCC: American Joint Committee on Cancer.

Incidence of Unplanned Consultations and Hospitalizations

In the Consilium group, a total of 36 unplanned consultations and hospitalizations occurred during the observation period (Table 2). These 36 events occurred in 29 patients, with 1 event documented for 25 patients and 2 or 3 events for 2 patients each. The events could be divided into 23 unplanned consultations, 5 unplanned hospitalizations lasting up to 2 days, and 8 unplanned hospitalizations lasting longer than 2 days (Table 2). In comparison, there were a total of 38 documented unplanned consultations and hospitalizations in the control group, which were recorded in 37 patients. Further, 1 event was documented for 36 patients and 2 events for 1 patient. These were divided into 29 unplanned consultations, 4 hospitalizations up to a maximum of 2 days, and 5 unplanned hospitalizations of more than 2 days. Per the nominally largest proportion of patients with breast cancer in both groups, most events also occurred within this subgroup, with the remainder distributed among the 4 other cancer entities (Table 2).

Looking at the sheer numbers of events, no difference can be found between the Consilium and the retrospective control group, as already mentioned in Trojan et al [18]. Since multiple occurrences of events in a single patient cannot be assumed to be independent, the proportion of patients with at least one event (29 for the Consilium group and 37 for the control group) was used for the statistical analysis. Looking at all patients, there was also no statistically robust positive effect of app use on unplanned consultations and hospitalizations (OR 0.742, 90% CI 0.455-1.206; Fisher exact test P=.34 at a 2-sided 10% significance level).

To take into account the heterogeneity of the overall collective, a multivariate regression model was evaluated for the binary end point of the occurrence of at least one unplanned event (consultation or hospitalization). The model analyzed the factor of the intervention group (Consilium versus control group) and adjusted for treatment type and tumor entity. Due to the small subgroup sizes, the tumor entity variable was dichotomized into the values “breast cancer” and “other.” Furthermore, the model contains interaction terms between intervention and chemotherapy or breast cancer subgroups. Table 3 summarizes the results and reports the corresponding OR for all coefficients to interpret the effect size, as well as a P value, which indicates the statistical significance of the respective coefficient. Looking at the heterogeneous overall collective, no statistically robust positive effect of app use on unplanned consultations and hospitalizations can be detected in this model (Table 3; model term for Consilium group, $P>0.10$). However, the interaction term between the Consilium group and the factor “chemotherapy” was significant at the 5% level ($P=.048$), indicating a relevant reduction of risk for the intervention group in the collective of patients who underwent chemotherapy. This motivated a further subgroup analysis within the collective of patients who underwent chemotherapy.
In each group, that is, the Consilium group and the control group, 139 patients had received chemotherapy. In the Consilium group, 17 (12.2%) patients with unplanned consultations and hospitalizations were recorded; in the control group, there were 29 (20.9%) patients. In the chemotherapy subgroup, a clinically relevant effect of app use on these events was observed, with an OR of 0.53, 90% CI 0.288-0.957 (Fisher exact test; \(P=0.08\)) which is significant at a 2-sided 10% significance level. The observed OR is equivalent to a halving of the risk for patients in the intervention group.

Table 2. Occurrence of at least one event in patients.

<table>
<thead>
<tr>
<th>Events</th>
<th>Consilium group (n=178)</th>
<th>Control group (n=178)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with (\geq 1) event, n (%)</td>
<td>29 (16.3)</td>
<td>37 (20.8)</td>
</tr>
<tr>
<td><strong>Total events, n</strong></td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>Unplanned (emergency) consultations</td>
<td>23</td>
<td>29</td>
</tr>
<tr>
<td>Hospitalizations (\leq 2) days</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Hospitalizations &gt;2 days</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td><strong>Subgroups for tumor stages, (events, n/patients, n)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AJCC(^a) stage I</td>
<td>1/22</td>
<td>0/25</td>
</tr>
<tr>
<td>AJCC stage II</td>
<td>9/68</td>
<td>9/51</td>
</tr>
<tr>
<td>AJCC stage III</td>
<td>4/37</td>
<td>9/16</td>
</tr>
<tr>
<td>AJCC stage IV</td>
<td>15/51</td>
<td>19/41</td>
</tr>
<tr>
<td><strong>Subgroups for therapies, (events, n/patients, n)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>17/139</td>
<td>29/139</td>
</tr>
<tr>
<td>Nonchemotherapy</td>
<td>12/39</td>
<td>8/39</td>
</tr>
<tr>
<td><strong>Subgroups for cancer entities, n</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Hematological malignancies</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

\(^a\)AJCC: American Joint Committee on Cancer.

Table 3. Multivariate logistic regression for “unplanned consultation or hospitalization.”

<table>
<thead>
<tr>
<th>Model term</th>
<th>Odds ratio (90% CI)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consilium group</td>
<td>1.16 (0.370-3.6220)</td>
<td>.83</td>
</tr>
<tr>
<td>Type of therapy: chemotherapy</td>
<td>1.09 (0.527-2.380)</td>
<td>.86</td>
</tr>
<tr>
<td>Cancer entity: breast cancer</td>
<td>0.46 (0.241-0.892)</td>
<td>.051(^a)</td>
</tr>
<tr>
<td>Interaction Consilium group: chemotherapy</td>
<td>0.29 (0.101-0.805)</td>
<td>.048(^b)</td>
</tr>
<tr>
<td>Interaction Consilium group: breast cancer</td>
<td>1.89 (0.681-5.455)</td>
<td>.31</td>
</tr>
</tbody>
</table>

\(^a\)Above \(P<0.05\) significance.

\(^b\)Below \(P<0.05\) significance.

**Discussion**

**Principal Findings**

In our comparative exploratory analysis of data from the PRO study and a matched retrospective control group, the effect of Consilium Care, an app for structured and regular self-assessment of side effects by ePROs, on unplanned consultations and hospitalizations of patients with cancer was analyzed. When considering all included patients, no statistically robust difference in unplanned consultations and hospitalizations between the groups could be demonstrated. In the Consilium group, there were a total of 36 unplanned consultations and hospitalizations, which occurred in 29 (16.3%) different patients. In comparison, the control group had 38 consultations and hospitalizations in 37 (20.8%) different patients. This simple comparison of the event numbers led Trojan et al [18] to conclude that a positive effect of the app was not demonstrable—but without showing the data and the analyses used. However, a multivariate regression model revealed that
the interaction term between the Consilium group and the factor chemotherapy was significant at the 5% level \((P=.048)\) and indicated a relevant reduction in the risk in the intervention group in the collective of patients who underwent chemotherapy. Within the subgroup of patients who underwent chemotherapy (139 in each group), 7 events were documented in the Consilium group, while 29 were recorded for the control group, which corresponds to a halving of the risk (OR 0.53, 90% CI 0.288-0.957) at a 2-sided 10% level. This indicates a relevant reduction in the risk in the Consilium group in the collective of patients who underwent chemotherapy and provides initial indications that the concomitant use of the Consilium Care app could have a positive, clinically relevant effect on patients with cancer receiving chemotherapy.

It should be noted that although care was taken to ensure a comparable composition of the 2 groups, they were still heterogeneous in terms of cancer entities, the American Joint Committee on Cancer stages, cancer therapies, and age structure. In addition, the data of the Consilium group in the PRO study were collected at 14 different centers—mainly in Switzerland, but also in Germany and Austria—while the retrospective data of the control group only came from 2 Swiss oncology centers. It should also be noted that within the PRO study, the participating physicians were explicitly asked in this study’s protocol to inquire about unplanned consultations and hospitalizations during standard patient visits. In contrast, the unplanned consultations and hospitalizations of the control group were drawn retrospectively from the patient records, which may have had a lower level of documentation in this respect. That a reduction in the events studied was only observed in the chemotherapy subgroup may be related to the fact that the majority of unplanned consultations and hospitalizations occur within 30 days of the last chemotherapy, a period that was covered by the chosen observation period [3,6]. Accordingly, any effects arising from app usage should have been able to manifest themselves during this period. Since the type of chemotherapy and the patient’s “performance index” are also associated with the hospitalization rate [3], it can be assumed that early detection of side effects supported by regular documentation of the patient in the form of app-based ePROs prevents higher severity and thus also reduces the hospitalization rate.

In other studies, with heterogeneous cancer populations, a positive clinically relevant effect of web- and smartphone-based apps that were used in a comparable way for monitoring and self-assessment of health status was demonstrated. Basch et al [7,8] were able to show that patients with metastatic solid tumors (metastatic breast, urogenital, gynecological, or lung cancers) receiving chemotherapy according to the standard of care were less likely to be admitted to the emergency department (34% vs 41% in the control group; \(P=.02\)) or hospitalized (45% vs 49%; \(P=.08\)) when regularly carrying out health self-assessment. In another study with patients with cancer (various tumor entities) who received approved oral cancer drugs, days of hospitalization (2.82 days vs 4.44 days, \(P=.02\)), and treatment-related toxicity \(\geq3\) CTCAE (27.6% vs 36.9%, \(P=.02\)) were reduced in the Consilium group compared to a control group without the app [9]. The recently published results of the PreCycle study with a well-defined patient population (hormone receptor-positive, human epidermal growth factor receptor 2–negative locally advanced, or metastatic breast cancer) also show that not only was self-assessment-based quality of life greatly improved in a group of patients using a full version of a physician-supported app for regular self-assessment compared to a control group using an app with limited functionality [10], but also that patients using the full version were significantly less likely overall to have an SAE [11].

Conclusions

The results of the analysis presented here clearly indicate a positive effect on the incidence of unplanned consultations and hospitalizations of an app that is used alongside cancer therapy to document side effects and support communication with the treating physician. This is also supported by studies with comparable apps, which also demonstrate a direct positive effect on the incidence of SAEs. Due to the exploratory nature of this study, the randomized PRO2 study (NCT05425550) is currently being conducted with the medidux app, the further-developed successor to the Consilium Care app, in which the incidence of high-grade adverse events (CTCAE \(\geq2\)) is being investigated in a better-defined patient population (patients with human epidermal growth factor receptor 2–positive breast cancer) under controlled conditions to support the results of this exploratory analysis.

Acknowledgments

The authors thank all patients who participated in the PRO study and the investigators and their teams.

Conflicts of Interest

AT reported being chief medical officer, being cofounder, and owning stocks (as a majority shareholder) of mobile Health AG. GAK-U reported equity in Novartis. JS and SD reported payment by palleos healthcare GmbH. CT received honoraria for lectures and served on the advisory boards of AMGEN, Astra-Zeneca, Aurikamed, ConEvent Gesellschaft zur Entwicklung und Förderung von Kommunikationsprozessen mbH, DaiichiSankyo, EKH Wittenberg, ESMO, Gilead, J
cor Eickeler, KH Chemnitz, Lilly, med update GmbH, mediculut Kasseler Institut für Frauengesundheit und Weiterbildung, MSD, Onkowissen.de, Pfizer, Roche, Roland Berger, Seagen, streamedup GmbH. All other authors have declared no conflicts of interest.

References


Abbreviations

CTCAE: Common Terminology Criteria for Adverse Events
ePRO: electronic patient-reported outcome
PRO: patient-reported outcome
OR: odds ratio
SAE: serious adverse event
Feasibility of Deploying Home-Based Digital Technology, Environmental Sensors, and Web-Based Surveys for Assessing Behavioral Symptoms and Identifying Their Precipitants in Older Adults: Longitudinal, Observational Study

Wan-Tai M Au-Yeung1, BSME, MS, PhD; Yan Liu1, MS; Remonda Hanna2, BS; Sarah Gothard1, BS; Nathaniel Rodrigues1, BS; Cierra Leon Guerrero1, BS; Zachary Beattie1, PhD; Jeffrey Kaye1, MD

1Oregon Health & Science University, Portland, OR, United States
2Fariborz Maseeh Department of Mathematics + Statistics, Portland State University, Portland, OR, United States

Corresponding Author:
Wan-Tai M Au-Yeung, BSME, MS, PhD
Oregon Health & Science University
3181 SW Sam Jackson Park Road
Mailcode: UHN13
Portland, OR, 97239
United States
Phone: 1 503 494 6695
Email: auyeungm@ohsu.edu

Abstract

Background: Apathy, depression, and anxiety are prevalent neuropsychiatric symptoms experienced by older adults. Early detection, prevention, and intervention may improve outcomes.

Objective: We aim to demonstrate the feasibility of deploying web-based weekly questionnaires inquiring about the behavioral symptoms of older adults with normal cognition, mild cognitive impairment, or early-stage dementia and to demonstrate the feasibility of deploying an in-home technology platform for measuring participant behaviors and their environment.

Methods: The target population of this study is older adults with normal cognition, mild cognitive impairment, or early-stage dementia. This is an observational, longitudinal study with a study period of up to 9 months. The severity of participant behavioral symptoms (apathy, depression, and anxiety) was self-reported weekly through web-based surveys. Participants’ digital biomarkers were continuously collected at their personal residences and through wearables throughout the duration of the study. The indoor physical environment at each residence, such as light level, noise level, temperature, humidity, or air quality, was also measured using indoor environmental sensors. Feasibility was examined, and preliminary correlation analysis between the level of symptoms and the digital biomarkers and between the level of symptoms and the indoor environment was performed.

Results: At 13 months after recruitment began, a total of 9 participants had enrolled into this study. The participants showed high adherence rates in completing the weekly questionnaires (response rate: 275/278, 98.9%), and data collection using the digital technology appeared feasible and acceptable to the participants with few exceptions. Participants’ severity of behavioral symptoms fluctuated from week to week. Preliminary results show that the duration of sleep onset and noise level are positively correlated with the anxiety level in a subset of our participants.

Conclusions: This study is a step toward more frequent assessment of older adults’ behavioral symptoms and holistic in situ monitoring of older adults’ behaviors and their living environment. The goal of this study is to facilitate the development of objective digital biomarkers of neuropsychiatric symptoms and to identify in-home environmental factors that contribute to these symptoms.

(JMIR Form Res 2024;8:e53192) doi:10.2196/53192

KEYWORDS
neuropsychiatric symptoms; mild cognitive impairment; dementia; unobtrusive monitoring; digital biomarkers; environmental precipitants; mobile phone
Introduction

Neuropsychiatric symptoms (NPSs) are experienced by most people living with mild cognitive impairment (MCI) and people living with dementia [1,2]. NPSs include apathy, anxiety, depression, agitation, wandering, psychosis, and sleep disturbance [1]. It is estimated that NPSs affect up to 90% of all people living with dementia over the course of their illness. Even in the early stages of cognitive impairment, NPSs are frequent, with estimated rates of 35% to 85% in people living with MCI [1] and are independently associated with poor outcomes including distress among patients and caregivers, long-term hospitalization, misuse of medication, and increased health care costs [3-5]. For people living with MCI, apathy, depression, and anxiety are the most commonly observed NPSs, each of which have been linked to cognitive and functional decline in daily activities and disease progression [6].

Currently, a major challenge in the assessment of NPSs is that they are assessed subjectively, briefly, and episodically by people living with MCI, people living with dementia, or caring partners [7-12]. As subjective data can be prone to bias [13] and these symptoms may wax and wane in severity, current methods of self-reports or infrequent queries may misidentify or miss the emergence or changes in these symptoms over time. Using digital biomarkers (defined as “objective, quantifiable, physiological, and behavioral data that are collected and measured by means of digital devices, such as embedded environmental sensors, portables, wearables, implantables, or digestibles” [14]), one may begin to address these limitations by providing more continuous and objective measurements of behaviors that are signatures of these NPSs in real-life settings. Within this context, the overall goal of this study is to develop objective, largely passive, and ecologically valid digital biomarkers of the NPSs among older adults, focusing in particular on prevalent apathy, depression, and anxiety. The findings reported in this paper address the feasibility of capturing digital biomarkers and examining how they relate to the reported measures of apathy, depression, and anxiety. The development of these behavioral, digital biomarkers may reduce subjectivity in the assessment of these symptoms; enable continuous monitoring of these symptoms; and reduce the survey burden of people living with MCI, people living with dementia, or their caregivers. As NPSs such as apathy, depression, and anxiety often precede the diagnosis of MCI or dementia [15-17], effective and continuous monitoring of such symptoms can enable the early detection of MCI or dementia, so that timely treatments may be initiated or maintained. Behavioral digital biomarkers may also reduce the sample size needed in clinical trials as shown in previous studies [18,19].

In addition to more objectively and continuously assessing NPSs, digital technologies may also have an important role in identifying the precipitants of NPSs. In particular, environmental conditions (e.g., light, noise, and temperature) may be stimulating or mediating factors in the emergence of NPSs [20-22]. Therefore, including environmental sensing technology in the home technology assessment suite may provide additional, important, linked information in the eventual assessment and treatment of NPSs, because the environment may be readily modified if any of these factors are found to play a role in stimulating or sustaining NPSs. Importantly, intelligent environmental manipulation or modification that mediates NPSs can be achieved at low cost and could be implemented widely. This can potentially decrease the amount of pharmacological intervention administered to patients, which are frequently associated with adverse side effects. Thus, one of the major aims of this study is to develop and integrate environmental sensing into the digital biomarker assessment suite and identify the possible associations of NPSs with environmental conditions at home or in the community.

Currently, most of the literature is focused on developing digital biomarkers and finding environmental precipitants of NPSs in people living with moderate- to later-stage dementia, whereas those for people living with MCI or early-stage dementia have been insufficiently examined. At the same time, to date, most studies that measured the physical environment were conducted in long-term, assisted living units. The few studies that were conducted at homes neither deployed objective, physical environmental sensors [23] nor focused on apathy, depression, or anxiety [24]. This study addresses a critical need as addressing these NPSs early on may slow disease progression and help people live independently for a longer period in their home. This study examines the feasibility of deploying a weekly web-based survey in addition to deploying an in-home behavioral and environmental assessment suite in the homes of older adults who may have these NPSs.

Methods

Ethical Considerations

Ethics approval was obtained from Oregon Health and Science University (OHSU) institutional review board (IRB; approvals 24210, 2765, and 20236).

To protect participants’ privacy during recruitment, consent, and study procedures, participants were assigned a code that were used instead of their name or any personally identifying information. Electronic files for data analysis only contained the participant code. The key associating the codes and the participants personally identifying information was restricted to the PI and study staff. The key was kept secure on a restricted OHSU network drive in a limited access folder. Any documents that contained personally identifying information were handled according to standard institutional practices at OHSU to maintain the confidentiality and security of data. The study database was password-protected. Data were encrypted and personal identifying information was excluded from data uploads to prevent interception of data during broadband transmission. Firewall protection and password-restricted access protected against unauthorized access to the central server.

Study Design and Overview

This is an observational, longitudinal study (OHSU IRB 24210) with a duration of 9 months. Study participants are men or women recruited from the Digital Technology Core (DTC) of the Oregon Alzheimer’s Disease Research Center (OADRC; OHSU IRB 2765) and the Oregon Royal Center for Care Support Translational Research Advantaged by Integrating...
Technology (ORCASTRAIT; OHSU IRB 20236). When participants are first enrolled into the DTC or ORCASTRAIT, they are assessed in person using standardized health, behavioral, and cognitive tests (Table 1).
Table 1. List of standardized health, behavioral, and cognitive tests that are used to assess participants in the Digital Technology Core (DTC) or Oregon Roybal Center for Care Support Translational Research Advantaged by Integrating Technology (ORCASTRAIT) when they enroll.

<table>
<thead>
<tr>
<th>Tests</th>
<th>DTC</th>
<th>ORCASTRAIT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UDS</strong>—required forms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1—Subject Demographics</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>A2—Informant Demographics</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>A3—Subject Family History</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>A4—Subject Medications</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>A5—Subject Health History</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B4—Global Staging CDR</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B5—Neuropsychiatric Inventory Questionnaire</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B6—Geriatric Depression Scale</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B7—Functional Assessment Questionnaire</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B8—Neurological Examination Findings</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B9—Clinician Judgement of Symptoms</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C2—Neuropsychological Battery Scores</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MoCA or Blind MoCA version</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Craft Story 21 (immediate)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Benson Complex Figure Copy (immediate)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Number Span Forward</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Number Span Backward</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Category Fluency (animals)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Category Fluency (vegetables)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Trail Making Test (parts A and B) or Oral Trail (parts A and B)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Craft Story 21 (delayed)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Benson Complex Figure Recall</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multilingual Naming Test</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Letter Fluency (F, L, and C)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>D1—Clinician Diagnosis</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>D2—Clinician-assessed Medical Conditions</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CERAD—Word List</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Visual Reproduction 1 and 2 (WMS-R)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Stroop Test</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Block Design (WAIS-R)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Oregon Gait and Balance Inventory</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>OARS ADL or IADL</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Collateral Clinical Dementia Rating Interview Questions</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Collateral Clinical Dementia Rating Scale</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>QOL-AD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Lubben Social Network Scale</td>
<td>Web based</td>
<td></td>
</tr>
<tr>
<td>GAD-7</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Adverse Childhood Experience</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
### Weekly NPS Assessment

Participants are sent an email every Monday morning at 9:00 AM, which contains a link to a web-based survey (Qualtrics) that assesses their level of behavioral symptoms including apathy, depression, and anxiety in the past 7 days. Each symptom is assessed using a distinct assessment survey that is placed on a separate web page, and the participants can leave any questions unanswered. Apathy is assessed using the Withdrawal-Apathy-Vigor (WAV) subscale from the 30-item Geriatric Depression Scale [30,31]. WAV scores range from 0 to 6, with 0 indicating no apathy and 6 indicating the most severe apathy. WAV has been shown to be congruent with disengagement or depletion [30]. Depression and anxiety were assessed using the 4-item depression and anxiety subscales from the 29-item Patient-Reported Outcomes Measurement Information System [32]. The possible scores for the weekly assessment of each symptom subscale range from 4 to 20, with a score of 4 being the mildest and a score of 20 being the most severe. These two 4-item subscales have been shown to have good internal reliability and convergent validity [33]. In addition, all 3 subscales have the advantage of being brief, so they should not be very burdensome to participants. In addition to the weekly web-based assessment of the 3 NPSs, a 13-item questionnaire is also sent weekly to assess other relevant health-related and activity-related events and experiences including pain, loneliness, illness, falls, visits to the emergency department or hospital, medication changes, and the need for additional assistance at home [25]. The participants receive no training for completing the surveys.

### Digital Technology

As part of the OADRC DTC or ORCASTRAIT cohort participation, all participants have the ORCATECH home assessment platform deployed in their residences. The platform includes the following: activity sensing watches (Steel; Nokia)

---

### Table 1: Tests Used

<table>
<thead>
<tr>
<th>Tests</th>
<th>DTC</th>
<th>ORCASTRAIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caregiving Annual Questions (CP only)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Dyadic Relationship Scale (CP only)</td>
<td>Web based</td>
<td></td>
</tr>
<tr>
<td>Zarit Burden Interview (CP only)</td>
<td>Web based</td>
<td></td>
</tr>
<tr>
<td>Weekly questionnaire</td>
<td>Web based</td>
<td></td>
</tr>
</tbody>
</table>

---

aUDS: Uniform Data Set.  
bCDR: Clinical Dementia Rating.  
cMoCA: Montreal Cognitive Assessment.  
dCERAD: Consortium to Establish a Registry for Alzheimer’s Disease.  
eWMS-R: Wechsler Memory Scale-Revised.  
fWAIS-R: Wechsler Adult Intelligence Scale-Revised.  
‡OARS ADL or IADL: Older Americans Resources and Services Activities of Daily Living or Instrumental Activities of Daily Living.  
‡QOL-AD: Quality of Life-Alzheimer’s Disease.  
‡GAD-7: General Anxiety Disorder-7.  
‡CP: Care partner.

All participants are sent a weekly web-based survey querying about their health and activity measures [25] and have a technology platform deployed in their homes that was developed by the Oregon Center for Aging and Technology (ORCATECH) [26,27]. When participants consent to participate in the study described in this paper (OHSU IRB 24210), they also agree to complete a weekly web-based survey for assessing their level of apathy, depression, and anxiety in the past week. The weekly queries are web-based, allowing for responding through any internet-connected device (eg, smartphone, tablet, laptop, or PC). Participants are not provided with a device but use their own internet-connected devices to complete the surveys. In addition, participants of this study had an environmental sensor added to the ORCATECH home assessment platform to measure the indoor environment of their residence.

### Participants

Study coordinators reach out to potential participants in the OADRC DTC and ORCASTRAIT study cohorts through phone calls or emails to gauge their interest and eventually obtain informed consent for participation in this study. Participants have the option of consenting electronically (e-consent) or using paper consent forms. An individual is considered capable of informed consent if they have a score ≤0.5 on the Clinical Dementia Rating scale [28]. If not capable, the assent of the participant is required and is obtained by explaining the study to the participant and having the participant sign the consent form along with their legally authorized representative. The target population is older adults aged ≥60 years without major comorbidity. People with stable, common, age-associated medical conditions (eg, hypertension, osteoporosis, and osteoarthritis) are not excluded. Participants may have normal cognition, MCI, or early-stage dementia diagnosed using National Alzheimer’s Coordinating Center criteria [29]. They may or may not have symptoms of apathy, depression, or anxiety as assessed using the Uniform Data Set Form B5: Neuropsychiatric Inventory Questionnaire or Uniform Data Set Form B9: Clinician Judgment of Symptoms [29]. The goal is to recruit 6 participants with normal cognition and 6 participants with either MCI or early-stage dementia. As apathy, depression, and anxiety are highly prevalent in people living with MCI or early-stage dementia and less common in people with normal cognition, the expectation is to recruit a sample of participants with and those without the NPSs of interest.
or Activité; Withings), sleep and nighttime activity sensing bed mats (ENM-9360-0656-30-P; Emfit), electronic pillboxes (TimerCap), passive infrared (PIR) motion sensors (NYCE), and contact sensors (NYCE). Participants are instructed to wear the activity sensing watch on their wrist to measure their number of steps. The bed mat is placed underneath the mattress of each participant and provides sleep measurements such as total sleep duration, sleep latency, or total duration that the participant is awake after having initially fallen asleep. The bed mat also measures physiological metrics such as heart rate and respiratory rate. The motion sensors are installed in each living space (eg, kitchen, living room, bedrooms, and bathrooms) to detect the participant’s presence in each living space. Contact sensors are placed on egress doors of a participant’s home to detect door openings and closings. More details regarding the ORCATECH technology platform can be found in the paper by Beattie et al [27]. All the digital technologies were designed to be as unobtrusive as possible and require minimal input from a participant. The ORCATECH home assessment platform is deployed in participant homes for the duration of the study.

A previous study showed that apathy was associated with physical inactivity in community-dwelling older adults [34]. Moreover, the literature suggests that sleep disturbances are associated with anxiety and depression in older adults [35]. With the digital technology in the ORCATECH platform, activity level and sleep can be measured objectively, and they can be compared against the self-reported severity of behavioral symptoms to identify correlation.

Environmental Sensing

After consenting to participate in this study (OHSU IRB 24210), an environmental condition sensing device (Awair Omni; Awair) is added to the ORCATECH home assessment platform during a scheduled visit from a study technician to the participant’s home. The environmental sensor is placed at shoulder height on the wall of the room where the participant typically sleeps. The Awair Omni is deployed for the entire duration of the study and is connected to the internet via the ORCATECH platform hub computer in the home. The Awair Omni measures ambient light level (lux), noise level (dB), temperature (°C), relative humidity (%), carbon dioxide level (parts per million), total volatile organic compounds (parts per billion), and particulate matter–2.5 (PM2.5; µg/m³) every 5 minutes. The Awair Omni is an indoor RESET Air [36] Accredited-Grade B monitor for carbon dioxide, total volatile organic compounds, relative humidity, temperature, and PM2.5. The Awair Omni also generates a metric (Score), which quantifies the air quality and ranges from 0 to 100, with a higher number indicating better air quality. Participants’ indoor environmental data are downloaded from the Awair company website. The Awair Omni device was chosen because it measures multiple physical environmental variables that may be associated with apathy, depression, and anxiety in older adults. Light may reduce depressive symptoms in older adults and improve their quality of life [37]. Noise is associated with more severe anxiety [38]. Air quality has also been suggested to be associated with depressive and anxiety symptoms in older adults [39]. The Awair Omni device also collects data passively and requires no interaction from the participant.

Data Analysis

We addressed the feasibility of deploying the web-based weekly surveys, the digital technology, and the environmental sensing tool by examining protocol adherence and the amount of missing data. Answers to weekly surveys were summarized through descriptive statistics. In addition, we explored how digital biomarkers from the technology platform related to the reported measures of NPSs and identified the possible associations between NPSs and environmental conditions at homes. Weekly means were calculated for the digital biomarkers and environmental data. These statistics were compared to the self-reported level of NPSs using Spearman correlation coefficients intraindividually.

Results

Clinical Characteristics of Participants

At 13 months after recruitment began, a total of 9 participants had enrolled into the study (Table 2). At that time, participant 1 had withdrawn after 7.5 months, participants 3 and 5 have completed the 9-month study, and the remaining 67% (6/9) of the participants were actively participating in the study. The average age of participants at enrollment was 76.3 (SD 5.10) years, 56% (5/9) are men, and 44% (4/9) are women. Participants had 17.6 (SD 3.5) years of education and, to date, have 219 (SD 46.3) days of follow-up. In terms of cognition, of the 9 participants, 5 (56%) have normal cognition, 3 (33%) have MCI, and 1 (11%) has early-stage Alzheimer disease. The Montreal Cognitive Assessment–BLIND score for the 9 participants at baseline was 19 (SD 1.80). Of the 9 participants, 1 (11%) had all 3 NPSs at baseline, 3 (33%) had none, and the other 5 (56%) had either 1 or 2 of the NPSs being studied.
Table 2. Summary of participants’ demographics and the presence of behavioral symptoms during their baseline clinical assessments.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Cognition</th>
<th>Sex</th>
<th>Age (y)</th>
<th>Race</th>
<th>Years of education</th>
<th>Behavioral symptoms reported during baseline assessment (UDS$^a$–B5: Neuropsychiatry Inventory Questionnaire or UDS–B9: Clinician Judgment of Symptoms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Early Alzheimer disease</td>
<td>Male</td>
<td>80</td>
<td>White</td>
<td>18</td>
<td>Y$^b$</td>
</tr>
<tr>
<td>2</td>
<td>Normal cognition</td>
<td>Female</td>
<td>80</td>
<td>White</td>
<td>16</td>
<td>N$^c$</td>
</tr>
<tr>
<td>3</td>
<td>MCI$^d$</td>
<td>Male</td>
<td>80</td>
<td>White</td>
<td>18</td>
<td>N, N, Y</td>
</tr>
<tr>
<td>4</td>
<td>Normal cognition</td>
<td>Female</td>
<td>71</td>
<td>White</td>
<td>16</td>
<td>N, N, Y</td>
</tr>
<tr>
<td>5</td>
<td>MCI</td>
<td>Male</td>
<td>69</td>
<td>White</td>
<td>25</td>
<td>Y, N</td>
</tr>
<tr>
<td>6</td>
<td>MCI</td>
<td>Male</td>
<td>74</td>
<td>White</td>
<td>19</td>
<td>Y, Y</td>
</tr>
<tr>
<td>7</td>
<td>Normal cognition</td>
<td>Male</td>
<td>71</td>
<td>White</td>
<td>16</td>
<td>N, Y, N</td>
</tr>
<tr>
<td>8</td>
<td>Normal cognition</td>
<td>Female</td>
<td>83</td>
<td>Black</td>
<td>12</td>
<td>N, N, N</td>
</tr>
<tr>
<td>9</td>
<td>Normal cognition</td>
<td>Female</td>
<td>79</td>
<td>White</td>
<td>18</td>
<td>N, N, Y</td>
</tr>
</tbody>
</table>

$^a$UDS: Uniform Data Set.  
$^b$Y: symptoms reported at baseline.  
$^c$N: no such symptom reported at baseline.  
$^d$MCI: mild cognitive impairment.

Web-Based Weekly Surveys

Table 3 summarizes the weekly surveys, both at the individual and group level. Among a total of 278 web-based questionnaires sent, only 4 (1.4%) were not even started. The remaining were either fully completed (241/278, 86.7%) or partially completed (33/278, 11.9%).

Table 3. Weekly survey adherence and neuropsychiatric symptom score statistics.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Weekly surveys sent, n$^a$</th>
<th>Responses received, n$^b$</th>
<th>Apathy score, mean (SD)$^c$</th>
<th>Depression score, mean (SD)$^d$</th>
<th>Anxiety score, mean (SD)$^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
<td>32</td>
<td>1.13 (0.942)</td>
<td>5 (1.04)</td>
<td>5.29 (1.65)</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>23</td>
<td>0 (0)</td>
<td>4.24 (0.526)</td>
<td>4.14 (0.343)</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>40</td>
<td>2.03 (0.357)</td>
<td>4 (0)</td>
<td>4.03 (0.156)</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>32</td>
<td>2.06 (1.01)</td>
<td>4.25 (0.612)</td>
<td>4.22 (0.544)</td>
</tr>
<tr>
<td>5</td>
<td>39</td>
<td>39</td>
<td>5.00 (0.392)</td>
<td>10.6 (0.718)</td>
<td>10.5 (0.873)</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>32</td>
<td>5.97 (0.174)</td>
<td>14.3 (1.46)</td>
<td>13.5 (1.60)</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>24</td>
<td>0.625 (1.03)</td>
<td>5.92 (0.400)</td>
<td>4 (0)</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>30</td>
<td>1.47 (0.806)</td>
<td>5.45 (0.723)</td>
<td>5.97 (1.02)</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>23</td>
<td>0 (0)</td>
<td>6.74 (1.36)</td>
<td>5.23 (0.849)</td>
</tr>
</tbody>
</table>

$^a$Group mean 30.9 (SD 6.53).  
$^b$Group mean 30.6 (SD 6.37).  
$^c$Group mean 2.47 (SD 2.09).  
$^d$Group mean 6.82 (SD 3.50).  
$^e$Group mean 6.56 (SD 3.45).

Responses to the weekly queries regarding apathy, depression, and anxiety were highly variable both across and within individuals over time (Figure 1). Of the 9 participants, 2 (22%; participant 5 and participant 6) were characterized as having higher apathy, depression, and anxiety scores over time compared to the rest of the cohort. Both these individuals had MCI at baseline. Participant 5 had symptoms of apathy at baseline, and participant 6 had symptoms of apathy and depression at baseline.
Adherence to and Feasibility of Data Collection Using Digital Technology and Environmental Sensor

All participants had PIR motion and contact sensors successfully deployed in their homes. All participants except participant 1 (8/9, 89%) were given an electronic pillbox. Participant 1’s home had connectivity issues with the electronic pillbox device and thus did not have the device when enrolled in this study. Of the 9 participants, 7 (78%) wore the activity sensing watches. Participants 1 and 2 did not wear them. Participant 1 had their own commercial wearable and preferred that to the study’s activity sensing device. Participant 2 found wearing the activity sensing device uncomfortable. All participants except participant 6 (8/9, 89%) had bed pressure mats installed. Participant 6 reported the bed pressure mat as being very intrusive.

Table 4 shows the number and percentages of days for which data were collected from the respective device and each participant’s home over their follow-up period. At the sample level, over the mean follow-up of 219 days, the PIR motion and contact sensors captured data for an average of 196 (89.5%) days, the bed mats captured data for 157 (71.7%) days, the activity sensing watch captured data for 164 (74.9%) days, and the electronic pillbox captured data for 165 (75.3%) days. The Raspberry Pi (hub computer in each home) occasionally went offline, requiring a restart that accounted for most data outages. Participant 3 moved during the study, which accounted for some of the data outages for their home. On other occasions, the devices themselves were disconnected from the Wi-Fi or lost communication with the hub computer. Other times, data were not collected because participants did not use the devices. For example, when participants went out of town or did not sleep on their beds, then no data are collected from the bed mat. In addition, participants would occasionally forget to wear the activity sensing watches in the morning, so no data were collected throughout that day. It is not possible to identify the reason for every data loss. For devices that participants interact with, only the medication adherence tracking devices would automatically confirm that they were functioning every day regardless of whether the participants used them. If the medication adherence tracking devices confirmed that it was functioning, but no data were collected regarding participants’ uses, then it can be assumed that the participants did not use the devices. However, if there is no activity sensing watch or bed mat data, it can be difficult to know whether they are not being used or whether there was a technical issue. As the motion and contact sensors would also check in every day, it was possible to know whether they were working.
The Awair Omni environmental sensors were deployed in all homes (9/9, 100%). The number of days the Awair Omnis were deployed were shorter than the study duration because they were added after participants had consented to participate in this study and it took time for the study staff to schedule home visits to install these sensors. An average of 90.4% (SD 10%) of environmental data were successfully collected from all homes during the monitoring periods when the Awair Omnis were marked as active. There were 2 occasions when an Awair Omni was marked as inactive. One occasion was due to construction at the participant’s home, and the other was due to the participant moving. The Awair Omni would become disconnected from the internet occasionally. This issue was often resolved by power cycling the device. However, this did not always lead to data loss because the environmental data can be stored locally on the device for up to approximately 10 days even when it is disconnected from Wi-Fi.

Figure 2 shows the environmental data from each participant’s home. There was time-dependent heterogeneity across homes in all the environmental domains measured. The PM2.5 levels were highest in the homes of participants 5 (purple) and 6 (brown) in October 2022, which were time aligned with the wildfires occurring during that time in Oregon. Moreover, there was seasonality in the relative humidity data. As participants 5 and 6 had the longest follow-up periods for their in-home environment, the humidity level was lower during winter compared to summer and autumn, which is contrary to the outdoor environment. This was likely due to participants switching on their heaters during winter. In addition, by examining the data shown in Figure 2, it was noticeable that the environment data for participant 4’s home changed considerably, starting from the beginning of 2023. It is speculated that the environmental sensor at this home had not been working properly since 2023 as the daily PM2.5 level dropped to approximately 0 µg/m³, which is extremely low.

### Table 4. Number and percentage of days with data for each digital technology.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Follow-up period (days), N</th>
<th>Bed mat, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Motion and contact sensor, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Activity sensing watch, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Medication adherence tracking device, n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Days Awair Omni was deployed, n</th>
<th>Awair Omni, n (%)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>225</td>
<td>195 (86.7)</td>
<td>N/A</td>
<td>N/A</td>
<td>186</td>
<td>182</td>
<td>(97.8)</td>
</tr>
<tr>
<td>2</td>
<td>161</td>
<td>146 (90.7)</td>
<td>N/A</td>
<td>155 (96.3)</td>
<td>160</td>
<td>159</td>
<td>(99.4)</td>
</tr>
<tr>
<td>3</td>
<td>274</td>
<td>193 (70.4)</td>
<td>228 (83.2)</td>
<td>232 (84.7)</td>
<td>80 (29.2)</td>
<td>197</td>
<td>174 (88.3)</td>
</tr>
<tr>
<td>4</td>
<td>224</td>
<td>203 (90.6)</td>
<td>223 (99.6)</td>
<td>217 (96.9)</td>
<td>204 (91.1)</td>
<td>210</td>
<td>208 (99)</td>
</tr>
<tr>
<td>5</td>
<td>275</td>
<td>243 (88.4)</td>
<td>244 (88.7)</td>
<td>258 (93.8)</td>
<td>255 (92.7)</td>
<td>254</td>
<td>248 (97.6)</td>
</tr>
<tr>
<td>6</td>
<td>264</td>
<td>N/A</td>
<td>206 (78)</td>
<td>264 (100)</td>
<td>262 (99.2)</td>
<td>243</td>
<td>196 (80.7)</td>
</tr>
<tr>
<td>7</td>
<td>162</td>
<td>141 (87)</td>
<td>162 (100)</td>
<td>162 (100)</td>
<td>161 (99.4)</td>
<td>159</td>
<td>158 (99.4)</td>
</tr>
<tr>
<td>8</td>
<td>210</td>
<td>185 (88.1)</td>
<td>201 (95.7)</td>
<td>191 (90.9)</td>
<td>190 (90.5)</td>
<td>202</td>
<td>201 (99.5)</td>
</tr>
<tr>
<td>9</td>
<td>172</td>
<td>106 (61.6)</td>
<td>172 (100)</td>
<td>150 (87.2)</td>
<td>172 (100)</td>
<td>128</td>
<td>111 (86.7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Denotes the number of days with data and the percentage of days with data over the follow-up period.

<sup>b</sup>Denotes the number of days with data over the number of days deployed and the percentage of days with data over the number of days deployed.

<sup>c</sup>N/A: not applicable.
Figure 2. Weekly mean environmental data along with error bars showing the SDs at participants’ homes throughout their follow-up periods. ppm: parts per million; ppb: parts per billion; PM2.5: particulate matter–2.5; VOC: volatile organic compound.

Exploratory Data Visualization and Analysis
In general, there were large variations in NPSs and the related digital biomarker signals (eg, sleep measures and step counts). For example, the within-individual correlation between duration of sleep onset and self-reported level of anxiety is especially noteworthy. The mean duration of sleep onset for the 89% (8/9) of the participants who had bed pressure mats was 22.3 (SD 6.84) minutes. The mean duration of sleep onset for each individual participant ranged from 20 to 27.7 minutes. Figure 3 shows the scatter plots of duration of sleep onset from the bed mats versus self-reported level of anxiety grouped based on participant, along with the corresponding Spearman correlation coefficients ($\rho$) and statistical significance. As shown by the correlation coefficients, durations of sleep onset were positively correlated with the self-reported levels of anxiety for all participants, with one of them being statistically significant ($P<.001$).

Figure 3. Scatter plots of duration of sleep onset in minutes versus self-reported level of anxiety in weekly surveys for (A) Participant 1, (B) Participant 2, (C) Participant 3, (D) Participant 4, (E) Participant 5, (F) Participant 7, (G) Participant 8 and (H) Participant 9. Some individuals had little or no change in the levels of anxiety (participants 2, 3, and 7), whereas others had a wide range of change in anxiety or duration of sleep onset. ns: not significant.
Figure 4 presents an example of how an environmental condition (noise level per participant home) may relate to an NPS (anxiety). Average weekly noise levels were typically >50 dB, which is above the level of a typical quiet room [40]. Anxiety was significantly and positively correlated with the noise level for 33% (3/9) of the participants ($P = .02$ for participant 5, and $P = .005$ for participants 4 and 9). Due to the abovementioned suspected technical difficulties with the Awair Omni in the home of participant 4, the observed noise levels <50 dB and the observed correlation with anxiety should be considered with a degree of uncertainty. For the other 44% (4/9) of the participants who had variation in their weekly levels of anxiety, the weekly indoor noise levels were negatively but nonsignificantly correlated with their weekly levels of anxiety.

Figure 4. Scatter plots of self-reported level of anxiety from weekly web-based surveys versus weekly mean noise level (dB) at home for (A) Participant 1, (B) Participant 2, (C) Participant 3, (D) Participant 4, (E) Participant 5, (F) Participant 6, (G) Participant 7, (H) Participant 8 and (I) Participant 9. The corresponding Spearman correlation coefficients ($\rho$) and $P$ values are provided. ns: not significant.

Discussion

Principal Findings

In general, the sensing platform including the environmental sensing device were successfully deployed to homes despite occasional nonadherence and data loss. NPS levels of participants show week-to-week fluctuations even though the within-person variability of level of symptoms is small compared to the possible range of the level of symptoms. The study and its recruitment are ongoing, with the current duration of follow-up relatively limited. The planned collection of up to 40 weeks of data from each participant will assist in providing a wider spectrum of possible NPS changes over time and will guide the study durations for future studies. Nevertheless, the weekly collection of data about the participants’ levels of NPSs, digital biomarkers, and indoor home environments allows us
to examine their changes over time and correlations among them both intraindividually and interindividually.

Participants’ levels of NPSs may be correlated with their adherence rate of completing the weekly survey as evidenced by participant 6 who had the highest level of all 3 symptoms and the lowest response rate among the 9 participants. When a person has these NPSs, they may consider completing surveys a nuisance and burdensome. This emphasizes the importance of being able to measure their level of NPSs objectively and unobtrusively.

The in-home environmental data collected in this study emphasize the importance of collecting such data because the in-home environment often does not align with the outdoor environment. The temperature and relative humidity data are such examples. With older adults spending most of their time at home, indoor environment monitoring will provide a more accurate estimate about their environmental exposure. However, indoor environment sensing is a relatively new feature added to the ORCATECH platform, and one of the lessons learned while conducting this study was that even high-quality environmental sensor may not provide reliable and valid measurements after some time, as evidenced in the data of participant 4’s home. Parts may fail, and there is variability in the quality of sensors and their life span even when they were manufactured by the same company. Frequent monitoring of the environmental data by the study personnel will help in detecting unreliable data, so that a malfunctioning sensor may be replaced. At the same time, this may not always be achievable if the changes caused by the deterioration of sensors are small.

Adherence to using the technology is an issue for the activity sensing watch, for which 2 (22%) of 9 participants did not opt for the device provided by the study. This points to the importance of having technology that is unobtrusive as some participants may find using a wearable uncomfortable. At the same time, some participants may already have their own wearable that they prefer. One future direction in research is to let participants “bring your own device.” However, bringing their own device has its own challenges as each device will have its own operating system and may not be compatible with the ORCATECH platform.

The ORCATECH platform was designed to be minimally intrusive to collect ecologically valid data from older adults as they live their daily lives. The platform does not record video nor audio, which preserves the privacy of our participants. However, some participants may still find some devices, such as bed mat, to be intrusive. As bed mat can collect valuable information about participants’ sleep and physiology, including an extensive educational session regarding study technology may increase participant comfort level with study devices, leading to higher acceptance and compliance.

Missing data are inevitable when data are collected in free-living environments as several problems can arise, ranging from power issues and internet issues to participants not using the devices. When such data are being analyzed, analytical methods that can account for missing data, such as mixed-effects models, will be especially applicable.

Cognitive impairment is strongly correlated with treatment incompliance and nonadherence as shown in a previous study on older adult patients with hypertension [43]. Cognitive impairment may contribute to device incompliance as memory problems may cause participants to forget to wear the activity sensing watches. These are issues that merit further investigation in future studies with a larger sample.

Several digital biomarkers and highly intensive environmental data were collected in this study with which multiple tests can be performed to examine the associations between digital biomarkers and level of NPSs and associations between the environment and the level of NPSs. This pilot study data can be used to generate hypotheses and to calculate the sample size for more definitive studies in the future. In this study, the preliminary data for duration of sleep onset collected using bed pressure mats are positively correlated with self-reported level of anxiety in all participants who had fluctuations in their level of anxiety (7/7, 100%), even though only 1 of them was statistically significant at the P<.05 level. Furthermore, the levels of anxiety for 22% (2/9) of the participants were statistically significantly and positively correlated with the noise level at their home (excluding participant 4). These preliminary and individual results suggest feasibility and agree with existing literature [38,44].

Limitations
This study had limitations. One limitation of this observational study is that we did not record residential environmental factors that may affect the measured environmental variables experienced by our participants. Our participants likely used heaters, air conditioners, or other devices to keep the indoor environmental conditions within small ranges. Small variability in environmental conditions make it difficult to find associations between the level of NPSs and the environment. Moreover, as indoor environmental sensors are still costly, deploying them in the bedrooms only will not provide a good estimate of people’s environmental exposures when they are outside or in other parts of the home. Another limitation of this study is that despite the use of subscales from well-validated scales, the validity of them being used weekly for a maximum period of 9 months has not been proven. It is unknown whether participants would take the surveys less seriously as time passes. While recruitment is still ongoing, the sample size presented in this paper is small. Therefore, one future direction is to conduct a similar study but with a larger sample size, so that estimates of effect size can be derived. This study provides an examination of feasibility, an exploration of possible pitfalls, and pilot data that can help guide future power and sample size calculations.

Conclusions
To conclude, this study demonstrates the feasibility of assessing the severity of behavioral symptoms weekly through web-based surveys and deploying in-home, digital assessment technologies among older adult individuals. It provides a guide for holistic, in situ, longitudinal monitoring of levels of NPSs and about how these behaviors may respond to changes in the living environment. The implications of deploying this approach include enabling early detection of these NPSs, characterizing the responses to treatment or intervention more effectively with...
Acknowledgments

The authors would like to thank the participants for their participation. This study was supported by the National Institute on Aging of the National Institutes of Health under awards P30AG066518 and P30AG024978. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

Oregon Health and Science University, ZB, and JK have financial interest in Life Analytics Inc, a company that may have commercial interest in the results of this study and technology. This potential conflict of interest has been reviewed and managed by Oregon Health and Science University. All other authors report no other conflicts of interest.

References


36. RESET® Air. RESET. URL: https://www.reset.build/standard/air [accessed 2024-04-05]


Abbreviations

DTC: Digital Technology Core
IRB: institutional review board
MCI: mild cognitive impairment
NPS: neuropsychiatric symptom
OADRC: Oregon Alzheimer’s Disease Research Center
OHSU: Oregon Health and Science University
ORCASTRAIT: Oregon Roybal Center for Care Support Translational Research Advantaged by Integrating Technology
ORCATECH: Oregon Center for Aging and Technology
PIR: passive infrared
PM2.5: particulate matter--2.5
WAV: Withdrawal-Apathy-Vigor

©Wan-Tai M Au-Yeung, Yan Liu, Remonda Hanna, Sarah Gothard, Nathaniel Rodrigues, Cierra Leon Guerrero, Zachary Beattie, Jeffrey Kaye. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
The Primary Care and Environmental Health e-Learning Course to Integrate Environmental Health in General Practice: Before-and-After Feasibility Study

Jean-Baptiste Tostain, MD; Marina Mathieu, MD; Agnès Oude Engberink, MSc, MD; Bernard Clary, MD; Michel Amouyal, MD; Béatrice Lognos, MD, PhD; Pascal Demoly, MD, PhD; Isabella Annesi-Maesano, MD, PhD, DSC; Grégory Ninot, PhD; Nicolas Molinari, MD, PhD; Arnaud Richard, PhD; Maha Badreddine, MSc; Claire Duflos, MD, PhD; Francois Carbonnel, MD, PhD

1Desbret Institute of Epidemiology and Public Health, University of Montpellier, National Institute for Health and Medical Research, Montpellier, France
2Department of General Practice, University of Montpellier, Montpellier, France
3Clinical Research and Epidemiology Unit, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
4Service de Santé Pluriprofessionnelle Universitaire Avicenne, Cabestany, France
5Service de Pneumologie, Allergologie et Oncologie Thoracique, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
6University of Toulon, Toulon, France
7Department of Pedagogical Engineering and Audiovisual Production, Faculty of Medicine, University of Montpellier, Montpellier, France

Corresponding Author:
Francois Carbonnel, MD, PhD
Desbret Institute of Epidemiology and Public Health
University of Montpellier
National Institute for Health and Medical Research
Campus Sante, IURC
641 Avenue du Doyen Gaston Giraud
Montpellier, 34095
France
Phone: 33 684014834
Email: francois.carbonnel@umontpellier.fr

Abstract

Background: Environmental and behavioral factors are responsible for 12.6 million deaths annually and contribute to 25% of deaths and chronic diseases worldwide. Through the One Health initiative, the World Health Organization and other international health organizations plan to improve these indicators to create healthier environments by 2030. To meet this challenge, training primary care professionals should be the priority of national policies. General practitioners (GPs) are ready to become involved but need in-depth training to gain and apply environmental health (EH) knowledge to their practice. In response, we designed the Primary Care Environment and Health (PCEH) online course in partnership with the Occitanie Regional Health Agency in France. This course was used to train GP residents from the Montpellier-Nimes Faculty of Medicine in EH knowledge. The course was organized in 2 successive parts: (1) an asynchronous e-learning modular course focusing on EH knowledge and tools and (2) 1 day of face-to-face sessions.

Objective: This study assessed the impact of the e-learning component of the PCEH course on participants’ satisfaction, knowledge, and behavior changes toward EH.

Methods: This was a pilot before-and-after study. Four modules were available in the 6-hour e-learning course: introduction to EH, population-based approach (mapping tools and resources), clinical cases, and communication tools. From August to September 2021, we recruited first-year GP residents from the University of Montpellier (N=130). Participants’ satisfaction, knowledge improvements for 19 EH risks, procedure to report EH risks to health authorities online, and behavior change (to consider the possible effects of the environment on their own and their patients’ health) were assessed using self-reported questionnaires on a Likert scale (1-5). Paired Student t tests and the McNemar χ² test were used to compare quantitative and qualitative variables, respectively, before and after the course.
Results: A total of 74 GP residents completed the e-learning and answered the pre- and posttest questionnaires. The mean satisfaction score was 4.0 (SD 0.9) out of 5. Knowledge scores of EH risks increased significantly after the e-learning course, with a mean difference of 30% (P<.001) for all items. Behavioral scores improved significantly by 18% for the participant’s health and by 26% for patients’ health (P<.001). These improvements did not vary significantly according to participant characteristics (eg, sex, children, place of work).

Conclusions: The e-learning course improved knowledge and behavior related to EH. Further studies are needed to assess the impact of the PCEH course on clinical practice and potential benefits for patients. This course was designed to serve as a knowledge base that could be reused each year with a view toward sustainability. This course will integrate new modules and will be adapted to the evolution of EH status indicators and target population needs.

(JMIR Form Res 2024;8:e56130) doi:10.2196/56130

KEYWORDS
environmental health; medical education; One Health; environment; environmental; eLearning; e-learning; remote; learning; online learning; primary care; satisfaction; awareness; behavioral; behavior change; questionnaire; survey; course; educational; teaching; GP; general practice; general practitioner

Introduction

The World Health Organization (WHO) defines environmental health (EH) as “those aspects of human health, including quality of life, that are determined by the physical, chemical, biological, social, psychosocial and aesthetic factors in our environment” [1]. According to the WHO, environmental exposures are responsible for 12.6 million deaths each year worldwide and contribute to cancer, cardiovascular, respiratory, neurological, and infectious diseases [2]. Environmental and behavioral factors are estimated to be the cause of approximately one-quarter of deaths and chronic diseases (ie, 20% of cancers, 30% of cardiovascular diseases, and 35% of respiratory diseases). In Europe, over 1.4 million premature deaths are caused by polluted environments [2]. The Food and Agriculture Organization of the United Nations, United Nations Environment Programme, WHO, and World Organisation for Animal Health have set a goal of creating healthier environments by 2030 in working together through the One Health initiative [3]. The French High Council for Public Health reiterated that achieving this goal was a political, financial, and scientific priority. However, inequalities in health are widening in France, mainly because of underresourced and poorly organized prevention and primary care systems [4].

Involving primary care in EH is an important support in the training of health care professionals. This training is a priority issue and has been reiterated as such in the fourth French national EH plan [5]. General practitioners (GPs) are clearly ready to become involved in this initiative; 93% of GPs believe that they should play an important role in informing their patients about EH risks [6], whereas only 55% feel they can respond to this topic, and 75% of them have not been trained on the topic, an observation shared by other disciplines [6-9]. The French national agency for continuing professional development has set understanding EH issues as one of its priorities for 2023-2025, which represented less than 1% of its approved courses in 2022 [10]. Training GPs while they are still at university could be a worthwhile investment and help to overcome these difficulties, as 84% of French health students express a desire to be trained in this area [11]. In the United Kingdom, the integration of an EH course into medical school curricula has improved the ability of GPs to discuss EH with their patients [12]. The COVID-19 pandemic has encouraged the development of online (e-learning) courses [13]. An e-learning course in EH aimed at pediatric health professionals has already shown an average 30% increase in general knowledge, with a greater increase in subjects whose level of knowledge was initially low [14]. GPs appreciate this format for their continuing professional training, especially when combined with face-to-face teaching [15,16].

Methods

Study Design

This was an interventional study based on a before-and-after, open, non-randomized, monocentric design. The study is reported per the guidelines for nonrandomized pilot and feasibility studies [17] and in line with the CONSORT (Consolidated Standards of Reporting Trials) 2010 checklist for reporting a pilot or feasibility trial (items pertinent to randomization were considered to be not applicable) [18].

To structure our inquiry, we relied on the Kirkpatrick model [19], which is a commonly used framework for evaluating learning. This model frames training on four levels: (1) reaction, (2) learning, (3) behavior, and (4) results. Reaction is the degree to which participants find the training favorable, engaging, and
relevant to their jobs; learning is the degree to which participants acquire the intended knowledge, skills, attitude, confidence, and commitment based on their participation in the training; behavior is the degree to which participants apply what they learned during training when they are back on the job; and results is the degree to which targeted outcomes occur as a result of the training and the support and accountability package. A more recent version of this model added that the behavior level refers to “required drivers” or factors that increase the likelihood that people will retain and apply what they have learned in a given setting [20].

Participants
The study population included first-year GP residents at the Montpellier-Nimes Faculty of Medicine. An email was sent to the potential participants, inviting them to identify themselves on Moodle, the learning management system of the University of Montpellier, to access the online training course. Only participants who completed the pre- and posttest questionnaires were included in the final analysis.

e-Learning Intervention Content
The general objective of the e-learning course was to enable GPs to integrate EH risks for a given region into their practice. The course consisted of 4 modules, which are described in Table 1.

Table 1. Content of the e-learning Primary Care Environment and Health (PCEH) course.

<table>
<thead>
<tr>
<th>Modules and learning objectives</th>
<th>Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction to EH*: Raising awareness of the impact of the environment on health</td>
<td></td>
</tr>
<tr>
<td>Define EH and the exposome</td>
<td>Commented slideshow</td>
</tr>
<tr>
<td>Define an EH risk</td>
<td>8 videos (outdoor air, indoor environment, water, noise, UV, classified facilities and emitters, endocrine disrupters, vector-borne pathogens)</td>
</tr>
<tr>
<td>Describe current regulations on environmental risks</td>
<td>Interactive texts: modular courses with interactive sorting, labeled graphics, process, timeline, accordion blocks, and responsive quiz to enhance student feedback</td>
</tr>
<tr>
<td>Define a territorial EH diagnosis</td>
<td>Text document, illustrated with figures</td>
</tr>
<tr>
<td>Describe the fourth national EH plan</td>
<td>Commented slide show and, when relevant, examples of potential medical thesis topics</td>
</tr>
<tr>
<td>2. Population-based approach: Identifying the environmental risks of a population</td>
<td></td>
</tr>
<tr>
<td>Differentiate population-based approach, individual approach, and situational analysis</td>
<td>Interactive texts</td>
</tr>
<tr>
<td>Identify an EH risk in an area using mapping tools (GEODES, Atlasanté, and SIRSE)</td>
<td>Interactive texts and videos</td>
</tr>
<tr>
<td>List the EH resources available to GPs b</td>
<td>Interactive texts</td>
</tr>
<tr>
<td>List the tools (institutional and noninstitutional) for assessing and analyzing the various risks associated with an area and a population (Occitanie territorial EH diagnosis)</td>
<td>Interactive texts and commented slideshows</td>
</tr>
<tr>
<td>Organize monitoring methods for the practice’s patients (Recosanté, ATMO, Pollen.fr, DGS-urgent)</td>
<td>Interactive texts</td>
</tr>
<tr>
<td>Describe how to report an EH risk event to health authorities</td>
<td>Interactive texts</td>
</tr>
<tr>
<td>3. Integrating EH into my practice for my patients: Identify the environmental risks to which an individual is exposed in a clinical case, use tools to assess the various risks associated with an area and a population in an educational situation, include the patient’s opinion in defining an appropriate response to their EH risk</td>
<td>Three clinical cases based on 5-6 multiple-choice questions (exposure to lead, atmospheric pollution, and UV radiation) with immediate debriefing of the answers to help participants realize how far they had progressed and the practical usefulness of the tools proposed</td>
</tr>
<tr>
<td>4. Communication: Promoting an EH intervention designed as part of PCEH training</td>
<td></td>
</tr>
<tr>
<td>List the internal, institutional, and noninstitutional means of communication for promoting an EH intervention</td>
<td>Interactive texts</td>
</tr>
<tr>
<td>Analyze EH communication using the example of the Occitanie website of the Agence Régionale de Santé [21].</td>
<td>Video</td>
</tr>
</tbody>
</table>

aEH: environmental health.
bGP: general practitioner.
Each module contained a hypertext link to the Rise360 software, which was interfaced with Moodle to generate interactive online courses. This platform can be used to display alternating slideshows, videos, audio, and interactive text in a fluid manner, which facilitates acquisition, creates an environment that involves the participants, and enhances attention span throughout the course [22]. The course lasted between 4 and 6 hours and served as a prerequisite for the compulsory face-to-face training course on September 9, 2021, which validated their participation for the GP curriculum.

**Data Collection**

The assessment was carried out directly on Moodle using Microsoft Forms. The pre- and posttest questionnaires contained 34 questions (see Multimedia Appendix 1). These questionnaires were created by the author team given the lack of an existing validated questionnaire for this population in this study context. The posttest questionnaire was available if all 3 first learning modules were completed, as the fourth module (Communication) had been supplied late and was not required to answer the questions. There were 6 questions on participants’ characteristics, including sex, county of residence, time spent in the region, projected county for professional settlement, number of children, and sources of EH information. The question of age was not asked, as the participants in this year’s class were all aged between 24 and 26 years. The participants’ data were collected and selected to be comparable with a barometric survey evaluating EH perceptions, knowledge, and behaviors in the regional population [23]. In addition, there were 2 EH knowledge questions to identify which of the suggested organizations is responsible for EH at the regional level and one Boolean question to indicate if they are familiar with the following mapping tools available to gather EH information on French territories: Rezone, Geodes, Sirsé, Atlasanté, Recosanté, ATMO-France, and RNSA [24-30]. These questions were asked only in the pretest questionnaire.

There were 20 questions related to EH to rate the participants’ knowledge for each of the 19 following EH risks from 1 (poor) to 5 (excellent): outdoor air quality, indoor air quality in buildings, noise, soil quality, radon, carbon monoxide, bathing water quality, tap water quality, Legionnaire disease, endocrine disruptors, lead, other heavy metals (cadmium, aluminum), electromagnetic waves, pesticides, nanomaterials, allergenic plant pollens, vector-borne diseases (eg, Chikungunya, Zika, yellow fever, malaria), substandard housing, and the first 1000 days of life concept. Finally, there was 1 question on the web-based reporting of EH risks to the health authorities.

There were 2 questions on behavior, as defined by the more recent version of the Kirkpatrick model [20]. Participants were asked to select how much they consider the possible effects of the environment on their health and on their patients’ health (consumption, protection, vigilance) on a scale from 1 (not at all) to 5 (totally).

There were 2 questions about their medical thesis: one asking if they would like to work on an EH thesis and one to invite them to write down their ideas for a thesis about EH.

There was 1 question about their general satisfaction level with the e-learning (completely satisfied, somewhat satisfied, don’t know, somewhat not satisfied, not at all satisfied) and 1 question to rate their level of satisfaction for each module. These questions were only available in the posttest questionnaire. Finally, there was 1 general-comments section.

The data from each of the pre- and posttest questionnaires were collected from the university’s secure website, reported into a spreadsheet, matched for each participant, and anonymized for statistical analysis. Participants who did not complete the posttest questionnaire were excluded from the final analysis.

**Data Analysis**

The data were analyzed by a statistician from the research department of the Montpellier University Hospital Center. The responses given on Likert scales were analyzed as quantitative data by calculating a before versus after difference (Δ) for each question; these Δ values are described using the mean with SD and range (minimum to maximum values). Quantitative variables are described using mean and SD. For the qualitative questions, we calculated the numbers and percentages of responses.

We used paired statistical tests to compare the responses before and after the e-learning and to compare the Δ value of one question with those of other questions; the paired Student t test was used for comparisons of quantitative variables and the McNemar χ² test was used for comparisons of qualitative variables. We controlled the α risk at .05 for each family of tests (knowledge, behavior) using the Hochberg method. There was no imputation of missing data. The data were analyzed using SAS Enterprise software Guide 8.2.

**Ethical Considerations**

The need for informed consent from participants was deemed unnecessary according to national regulations; evaluation of changes in practices brought about by training medical or paramedical staff for research purposes is considered noninterventional in France and therefore does not require an ethical review or approval [31].

Ethics committee approval was not required for this study according to General Data Protection Regulation recommendations. All data for this study are stored on a hard disk protected by secure authentication. Participant data were anonymized. Participants were not compensated to take part in this program, which was valued as part of their general medicine curriculum.

**Results**

**Participants**

Of the 94 participants who completed the pretest questionnaire, 20 did not complete the posttest questionnaire; thus, 74 participants completed both the pre- and posttest questionnaires (Figure 1). The study was carried out from July 21, 2021, to September 5, 2021 (47 days) to complete the training course and answer the questionnaires. The sociodemographic characteristics of the participants are presented in Table 2. The
CONSORT 2010 checklist of information to include when reporting a pilot or feasibility trial is provided in Multimedia Appendix 2.

The mean initial EH knowledge score of the participants was 2.0 (SD 0.8) out of 5. Participants initially rated their awareness of the effects of the environment on their own health at an average of 2.9 (SD 0.9) and their awareness of the effects of the environment on their patients’ health at an average of 2.6 (SD 0.9), which are both also based on a maximum total score of 5.

Figure 1. Flowchart of participation.
Table 2. Participants’ sociodemographic characteristics (N=74).

<table>
<thead>
<tr>
<th>Sociodemographic characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female sex</strong></td>
<td>50 (67.5)</td>
</tr>
<tr>
<td><strong>County of residence</strong></td>
<td></td>
</tr>
<tr>
<td>Hérault</td>
<td>34 (46)</td>
</tr>
<tr>
<td>Gard</td>
<td>19 (25.7)</td>
</tr>
<tr>
<td>Pyrénées-Orientales</td>
<td>12 (16.2)</td>
</tr>
<tr>
<td>Aveyron</td>
<td>4 (5.4)</td>
</tr>
<tr>
<td>Aude</td>
<td>3 (4.0)</td>
</tr>
<tr>
<td>Lozère</td>
<td>2 (2.7)</td>
</tr>
<tr>
<td><strong>Projected county for professional settlement</strong></td>
<td></td>
</tr>
<tr>
<td>I don’t know yet</td>
<td>34 (46)</td>
</tr>
<tr>
<td>Hérault</td>
<td>17 (23)</td>
</tr>
<tr>
<td>Gard</td>
<td>4 (5.4)</td>
</tr>
<tr>
<td>Pyrénées-Orientales</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Another county</td>
<td>5 (6.8)</td>
</tr>
<tr>
<td>Aude</td>
<td>5 (6.8)</td>
</tr>
<tr>
<td>Lozère</td>
<td>2 (2.7)</td>
</tr>
<tr>
<td><strong>Time lived in Occitanie, France</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;1 year</td>
<td>25 (33.8)</td>
</tr>
<tr>
<td>&gt;1 year</td>
<td>49 (66.2)</td>
</tr>
<tr>
<td><strong>Number of children</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>73 (98.6)</td>
</tr>
<tr>
<td>1 or more</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td><strong>EH³ information sources</strong></td>
<td></td>
</tr>
<tr>
<td>Press/radio/TV</td>
<td>55 (74.3)</td>
</tr>
<tr>
<td>Official website (Ministry of Health/regional health authority)</td>
<td>34 (45.9)</td>
</tr>
<tr>
<td>Relatives (friends/family)</td>
<td>36 (48.6)</td>
</tr>
<tr>
<td>Health professionals</td>
<td>31 (41.9)</td>
</tr>
<tr>
<td>I don’t know where to begin</td>
<td>18 (24.3)</td>
</tr>
<tr>
<td>University</td>
<td>24 (32.4)</td>
</tr>
<tr>
<td>Encyclopedic website</td>
<td>17 (23)</td>
</tr>
<tr>
<td>Website or public forum</td>
<td>9 (12.2)</td>
</tr>
<tr>
<td>Environmental protection association</td>
<td>7 (9.5)</td>
</tr>
<tr>
<td>I am not interested</td>
<td>2 (2.7)</td>
</tr>
<tr>
<td><strong>Knowledge of mapping tools</strong></td>
<td></td>
</tr>
<tr>
<td>REZONE</td>
<td>4 (5.4)</td>
</tr>
<tr>
<td>RecoSanté</td>
<td>6 (8.1)</td>
</tr>
<tr>
<td>RNSA³</td>
<td>4 (5.4)</td>
</tr>
<tr>
<td>Géodes</td>
<td>4.1 (3)</td>
</tr>
<tr>
<td>ATMO³</td>
<td>6.8 (5)</td>
</tr>
<tr>
<td>AtlaSanté</td>
<td>2 (2.7)</td>
</tr>
<tr>
<td>SIRSe³</td>
<td>1 (1.4)</td>
</tr>
</tbody>
</table>
Satisfaction
The overall average satisfaction score was 4.0 (SD 0.9) out of 5. The average satisfaction score for each of the modules 1 to 4 was 4.0 (SD 0.9), 4.0 (SD 0.7), 4.5 (SD 0.6), and 3.5 (SD 1.3), respectively. For the qualitative assessment of posttest satisfaction, 94.6% (70/74) of participants indicated being satisfied, including 35.1% (26/74) completely satisfied and 59.5% (44/74) fairly satisfied; 5.4% (4/74) were fairly dissatisfied; and none declared being completely dissatisfied. However, 23% (17/74) of the participants declared that they had not been able to access module 4.

Knowledge
There was a statistically significant increase in the knowledge of participants for all the environmental items studied (Figure 2), with an average increase of 30% (+1.5/5, SD 0.9) points for all questions (raw \(P<.001\); all questions were considered to show a statistically significant difference at \(\alpha=.05\) after applying the Hochberg method).

Knowledge of 3 environmental items (lead, the first 1000 days of life, and web reporting) improved significantly more than the others (\(P=.005\)). The question on web reporting showed the greatest improvement, with an average improvement of 2.7/5 (SD 1.1) points (\(P<.001\)), representing an increase of 1.1/5 (SD 0.2) points on average compared to the increase in knowledge for the other questions.

No difference in the improvement of knowledge scores was found among participants according to their sociodemographic characteristics such as sex, children, and place of working or living.

The detailed pre- and posttest responses to Likert-scale questions on knowledge are presented in Multimedia Appendix 3.

Behavior
There was a statistically significant increase in EH behaviors with an increase of 18% (+0.9/5 points, SD 0.4; \(P<.001\)) for participants considering EH-related factors in their own health and of 26% (+1.3/5 points, SD 0.5; \(P<.001\)) for considering EH...
with respect to their patients’ health (Figure 3). There was no significant increase in the number of residents declaring that they would write a thesis on EH ($P=0.07$). After the training, 68.9% (n=51) of the 74 participants stated that they did not know if they wished to focus their thesis on EH, 8.1% (n=6) had answered “yes,” and 22.9% (n=17) answered “no.” Among those who answered “no” or “I don’t know,” 29.4% (20/68) described ideas for a thesis related to EH but did not know how to start. No differences in improvement were found according to participants’ sociodemographic characteristics.

Figure 3. Differences in self-assessed behavior scores before and after the course. The diamonds correspond to the means, the boxes indicate the IQR, and the vertical bar indicates the median. The thin lines correspond to the extent of the difference ($\Delta$ values). A paired Student $t$ test was performed for each of the questions; the results were only considered statistically significant if they met the Hochberg procedure for $\alpha$ risk control. The results were all significant (family $P<0.05$).

### Discussion

#### Main Findings

The participants’ knowledge improved significantly overall and for each of the 21 environmental items studied. These items were chosen because they are easily identifiable in primary care and have a negative impact on population health [32]. The level of EH knowledge of practicing GPs is also known to be low due to their lack of training and information on the topic [33,34]. In our study, 74.3% of participants obtained information on this subject via the mainstream media, which is similar to the rate of 83.1% of citizens in Occitanie [35]. However, 45.9% of the participants of this study also used official websites to obtain this information, compared with only 18% of citizens. Only 32.4% of the participants indicated that their knowledge on this topic came from the university and 24.3% stated that they “didn’t know where to start” obtaining such knowledge. Acting on these proportions would require EH to be taught as part of the initial training. A course designed for students in their second or third year of medicine by the French Conference of Deans was deployed in medical faculties starting in 2023.

The average progress made by our participants was greater for items where their initial level of knowledge was low compared to that of topics for which the initial knowledge level was high. This expected result suggests that an initial assessment of the participants’ knowledge could facilitate providing a personalized course, prioritizing the areas of knowledge that are the least mastered, such as lead poisoning, the concept of the first 1000 days, and the web reporting of environmental pathologies to health authorities [6,14].

Self-reported behaviors showed significant improvement after the course in terms of considering the impact of the environment on their own health and the health of their patients. This suggests that the PCEH course has the potential to engage GP residents in actions on the environment, which is a promising finding given that EH will be a compulsory part of the additional year of the French GP internship [36]. Writing a medical thesis in EH would be further proof of the residents’ interest in this area. However, 68.9% of the residents included in this study stated that they did not know whether they wanted to do a thesis in EH and there was no significant increase in this proportion after the course. This question may have been asked at too early a
stage in their studies. The participants were only in their first year of medical school and still had another 5 years until they need to defend their medical thesis. This will change with the introduction of junior doctors in general practice, which will reduce this period to 3 years instead of 6 years.

Further studies will be also needed to assess whether the change in behavior reported by the participants is taking place in their personal and professional lives (stage 4—results of the Kirkpatrick model). Achieving this level remains a challenge for health educators [37]. Indeed, this was the main motivation for designing this pilot study (ie, to prepare for the implementation of the junior doctoral year integrating EH). These findings will therefore help us to develop a higher-quality study to evaluate the impact of the PCEH course.

Limitations
The satisfaction rate of the e-learning among participants was 94.6%. Satisfaction rates of training courses delivered in the same format often exceed 70% [16]. Overall satisfaction with the course (e-learning and face-to-face training) was not assessed. Our priority in this study was the e-learning, as it is intended to be duplicated every year. Participants were less satisfied with module 4, Communication. This may be linked to the fact that 23% of the participants declared that they had not been able to access this module, as it took longer to produce and was published after the other modules.

The consistent improvement for all environmental knowledge items and related behaviors validates the two corresponding levels of the Kirkpatrick model.

The main limitation of this study concerned the self-scaling nature of the knowledge and behaviors reported by the participants. Selection and social desirability biases may have influenced the responses. As stated in the Study Design subsection of the Methods, a recent version of the Kirkpatrick model has added that the behavior level refers to the “processes and systems that reinforce, encourage, and reward performance of critical behaviors on the job” [20]. We assumed that taking greater account of EH for their own health and for the health of their patients could be seen as a catalyst for applying what had been learned in the PCEH course. This was also why we asked participants about their desire to write their medical thesis on EH.

A hetero-assessment would not have made it possible to simply assess all of the knowledge areas of this detailed training course. This was also a pilot study designed to test the training and guide future updates to improve e-learning.

A control group could have increased the internal validity of the results and limited possible bias. This could have been achieved by comparing two classes from different universities. This could be considered in the future with a view to the junior doctor year, as many faculties have not yet designed an EH course for GPs.

A more balanced and larger sample size would have been required if we wanted to show any influence of certain characteristics of our participants on their commitment to EH. The barometer carried out on the regional population identified the number of children as a factor favoring their commitment to EH [23]. Our sample was not superimposable with this population because 81% of our participants had no children, which is consistent for first-year residents, who are usually aged between 24 and 26 years and the average age at which French female doctors have their first child is 30 years [38].

The number of participants lost to follow-up was high (32/94, 28%) and may have induced a motivation bias. This can be explained by the short duration of the entire training course (47 days) over a summer holiday period. We tried to limit this by using an asynchronous and anonymous self-assessment questionnaire.

Finally, we noted a regression in scores after training for a minority of participants, which was likely due to a Dunning-Kruger effect [39]. For the training, some participants recognized their ignorance, causing their score to drop in the posttest as a loss-of-confidence effect.

Conclusions and Perspectives
The results from this pilot study will be used to improve the existing e-learning itself as well as the study design of the planned EH modules to be created. In 2026, the French general medicine curriculum will include an additional year of junior doctor status. EH will be one of the new themes taught during this year and we plan to disseminate this training across other medical universities of France. This study was an opportunity to test this training with a view to more advanced studies (clinical impact) in pedagogy.

Developing a training program in EH is in line with the new training curriculum for GPs, but also with the evolution of primary care in France that is largely reorganizing itself toward coordinated practice. By 2023, such coordinated practices represented 2251 multiprofessional health houses, 455 multiprofessional health centers, and 389 territorial professional health communities, although they did not exist before 2007 [40,41]. These types of coordinated practices must be structured around a health project. Incorporating an EH diagnosis could enable coherent action to be taken on the health determinants of the populations in the areas in which they operate as part of a population-based approach [4,42]. This is in line with the “scale up actions on environment, climate change and health three-step framework”: data-based situation assessment, definition of targets and selection of actions, and implementation monitoring [43]. The PCEH course could then be assessed at level 4 of the Kirkpatrick model based on concrete actions implemented in the field for the prevention and care of patients.

Acknowledgments
We thank the students who took part in this training course. We are grateful to the administrative, audiovisual, and communication services of the Montpellier-Nîmes Faculty of Medicine for their essential help. Our respectful thanks go to Professors Michel
Mondain and Isabelle Laffont, Deans of the Montpellier-Nîmes Faculty of Medicine, for their support of this training program. Special thanks to Occitanie Health Regional Agency, directed by Pierre Ricordeau and Didier Jaffre, for funding the creation and implementation of this course. Our special thanks go to the public health department of the Occitanie Health Regional Agency directed by Catherine Choma and its environmental health division led by Dr Marie-Pierre Allie for their trust and commitment to this program. The Primary Care Environment and Health course was financed by the third regional Environmental Health Plan run by the Occitanie Regional Health Agency, as part of the French government’s national environmental health plan. The university multiprofessional health house Avicenne was a winner of the related call for projects.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Study questionnaire.
[DOCX File, 40 KB - formative_v81e56130_app1.docx ]

Multimedia Appendix 2
CONSORT (Consolidated Standards of Reporting Trials) 2010 checklist of information to include when reporting a pilot or feasibility trial.
[DOC File, 228 KB - formative_v81e56130_app2.doc ]

Multimedia Appendix 3
Pre- and posttest responses to Likert-scale questions on knowledge and behavior.
[PNG File, 181 KB - formative_v81e56130_app3.png ]

References


26. Système d'information inter-régional en santé (Sirsé) [Inter-regional health information system]. Atlasanté. URL: https://sirsatlasante.fr [accessed 2024-01-04]


29. Fédération des associations agrées de surveillance de la qualité de l'air [French federation of air quality monitoring associations]. ATMO-France. URL: https://www.atmo-france.org/ [accessed 2024-01-04]

30. Réseau National de Surveillance Aérobiologique (RNSA) [National Aerobiological Monitoring Network]. URL: https://pollens.fr/ [accessed 2024-01-04]


40. Les maisons de santé [Health houses]. Ministère du Travail et de la Santé et des Solidarités. URL: https://sante.gouv.fr/systeme-de-sante/structures-de-soins/article/les-maisons-de-sante-300889 [accessed 2024-01-04]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
EH: environmental health
GP: general practitioner
PCEH: Primary Care Environment and Health
SPES: Soins Primaires Environnement et Santé
WHO: World Health Organization

©Jean-Baptiste Tostain, Marina Mathieu, Agnès Oude Engberink, Bernard Clary, Michel Amouyal, Béatrice Lognos, Pascal Demoly, Isabella Annesi-Maesano, Grégoire Ninot, Nicolas Molinari, Arnaud Richard, Maha Badreddine, Claire Duflos, Francois Carbonnel. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e56130

Edited by A Mavragani; submitted 07.01.24; peer-reviewed by D Cole, C Perdrix; comments to author 30.01.24; revised version received 13.02.24; accepted 16.03.24; published 09.05.24.

Please cite as:
URL: https://formative.jmir.org/2024/1/e56130
doi:10.2196/56130
PMID:
Changepoint Detection in Heart Rate Variability Indices in Older Patients Without Cancer at End of Life Using Ballistocardiography Signals: Preliminary Retrospective Study

Naotake Yanagisawa¹, PhD; Yuji Nishizaki², MD, MPH, PhD; Bingwei Yao³, BSE; Jianting Zhang⁴, MS; Takatoshi Kasai⁵, MD, PhD

¹Medical Technology Innovation Center, Juntendo University, Tokyo, Japan
²Division of Medical Education, Juntendo University School of Medicine, Tokyo, Japan
³E3 Enterprise, Tokyo, Japan
⁴Zhejiang Huiyang Technology, Huzhou, China
⁵Department of Cardiovascular Biology and Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan

Corresponding Author:
Takatoshi Kasai, MD, PhD
Department of Cardiovascular Biology and Medicine
Juntendo University Graduate School of Medicine
3-1-3 Hongo
Bunkyo-ku
Tokyo, 113-8421
Japan
Phone: 81 338133111
Email: kasai-t@mx6.nisiq.net

Abstract

Background: In an aging society such as Japan, where the number of older people continues to increase, providing in-hospital end-of-life care for all deaths, and end-of-life care outside of hospitals, such as at home or in nursing homes, will be difficult. In end-of-life care, monitoring patients is important to understand their condition and predict survival time; this information gives family members and caregivers time to prepare for the end of life. However, with no clear indicators, health care providers must subjectively decide if an older patient is in the end-of-life stage, considering factors such as condition changes and decreased food intake. This complicates decisions for family members, especially during home-based care.

Objective: The purpose of this preliminary retrospective study was to determine whether and how changes in heart rate variability (HRV) indices estimated from ballistocardiography (BCG) occur before the date of death in terminally ill older patients, and ultimately to predict the date of death from the changepoint.

Methods: This retrospective pilot study assessed the medical records of 15 older patients admitted to a special nursing home between August 2019 and December 2021. Patient characteristics and time-domain HRV indices such as the average normal-to-normal (ANN) interval, SD of the normal-to-normal (SDNN) interval, and root mean square of successive differences (RMSSD) from at least 2 months before the date of death were collected. Overall trends of indices were examined by drawing a restricted cubic spline curve. A repeated measures ANOVA was performed to evaluate changes in the indices over the observation period. To explore more detailed changes in HRV, a piecewise regression analysis was conducted to estimate the changepoint of HRV indices.

Results: The 15 patients included 8 men and 7 women with a median age of 93 (IQR 91-96) years. The cubic spline curve showed a gradual decline of indices from approximately 30 days before the patients' deaths. The repeated measures ANOVA showed that when compared with 8 weeks before death, the ratio of the geometric mean of ANN (0.90, 95% CI 0.84-0.98; P=.005) and RMSSD (0.83, 95% CI 0.70-0.99; P=.03) began to decrease 3 weeks before death. The piecewise regression analysis estimated the changepoints for ANN, SDNN, and RMSSD at −34.5 (95% CI −42.5 to −26.5; P<.001), −33.0 (95% CI −40.9 to −25.1; P<.001), and −35.0 (95% CI −42.3 to −27.7; P<.001) days, respectively, before death.

Conclusions: This preliminary study identified the changepoint of HRV indices before death in older patients at end of life. Although few data were examined, our findings indicated that HRV indices from BCG can be useful for monitoring and predicting...
survival time in older patients at end of life. The study and results suggest the potential for more objective and accurate prognostic tools in predicting end-of-life outcomes.

(JMIR Form Res 2024;8:e53453) doi:10.2196/53453

KEYWORDS
ballistocardiography; BCG; noninvasive monitoring; heart rate variability; end-of-life care; prognosis prediction

Introduction

Background
The population of Japan was estimated to be 125 million as of October 1, 2020, and has been declining for 11 consecutive years since 2011 [1]. Meanwhile, the aging rate is increasing, and the number of people aged 65 years and older is estimated at 36 million, accounting for 28.8% of the total population. As the total population declines further, this percentage is expected to continue to increase, reaching approximately 33.3% in 2036 and 38.4% in 2065 [2].

The number of deaths has also increased over the years, exceeding 1 million in 2003 and 1.37 million in 2020. Malignant neoplasms are the leading cause of death, accounting for 27.6% of all deaths, followed by heart disease (15%) and senility (9.6%) [3]. Notably, in 2018, deaths due to senility surpassed cerebrovascular disease and became the third leading cause of death [4], possibly due to the growing number of people aged 90 years and older. The increasing number of deaths among older people creates difficulty in caring for all deaths in hospitals. Therefore, it is expected that end-of-life care outside hospitals will increasingly be provided in nursing homes and at home. In end-of-life care, it is important to monitor patients to understand their condition and predict survival time; this information helps patients anticipate what is to come and gives family members and caregivers time to prepare for the end of life [5,6].

However, this process may be very difficult under conditions when resources are limited, such that medical staff may not be available immediately at home. Furthermore, the physical, spiritual, and psychosocial conditions of older patients at end of life vary [7]. Unnecessary monitoring is also burdensome to patients and should be discouraged. To reduce the burden on patients and families, noninvasive and nonintrusive monitoring systems have been increasingly important, and several studies have demonstrated their usefulness in palliative care [8-10].

While tools to predict the survival of patients with terminal cancer have been developed and many studies have shown their usefulness [11-13], the reliability of these tools in predicting survival time for end-of-life patients without cancer has been questionable, and the tools are considered difficult to use [14,15]. Therefore, we focused on continuous and unobtrusive monitoring of patients using heart rate variability (HRV) indices estimated from ballistocardiography (BCG) obtained from a sheet-type device as a means of monitoring their condition. BCG is a measurement of the vibration signals generated by the ejection of the blood at each cardiac cycle [16], which has been used frequently in recent years to acquire biological signals [17-20]. The advantages of this method are that BCG signals can be obtained by placing a device under a patient’s mattress, making long-term, around-the-clock monitoring possible without attaching electrodes to the patient. The device could also reduce the burden on nurses by allowing them to monitor patients remotely. For example, many recent applied studies have been conducted using BCG to detect hypertension and apnea using HRV indices estimated from BCG signals [21-26]. HRV is a measure of the variation in time between heartbeats, usually recorded by an electrocardiogram (ECG). This variation is related to the autonomic nervous system and reflects a person’s health status [27]. A decrease in HRV generally results in a higher mortality rate [28,29] in patients with myocardial infarction, and HRV has also been demonstrated to be associated with a variety of diseases [30-33]. If the prognosis, including the patient’s survival, can be accurately estimated by HRV indices through noninvasive and noncontact monitoring, this would be beneficial for end-of-life patients, as well as their clinicians, caregivers, and family members, to plan future care and prepare for the end of life.

Objectives
The aim of this preliminary retrospective study was to investigate whether there is a changepoint in HRV indices obtained from BCG signals prior to death to predict the survival time in older patients without cancer at end of life.

Methods

Study Design and Patients
This was a single-center, retrospective pilot study in which all medical records of end-of-life patients (median age 93, IQR 91-96 years) were extracted from one nursing hospital in China. The study included data from 15 patients who were admitted to the facility and died between August 1, 2019, and December 31, 2021. All patients were without cancer. Patients with implanted pacemakers and patients with chronic atrial fibrillation or frequent extrasystoles were excluded. In addition, patients with less than 2 months of data from the date of death were excluded.

Measurements
Age and sex were collected as patient characteristics, and the following classical time-domain HRV indices were obtained from a medical database and used for analyses: average normal-to-normal (ANN) interval for each 5-minute segment of HRV recording; SD of all normal-to-normal (SDNN) intervals; and root mean square of successive differences (RMSSD). These HRV indices were collected up to 2 months before the patients’ deaths and used for data analyses.
BCG Sensing Device

The BCG device used in this study is made by Zhejiang Huiyang Technology Co, Ltd; it is a sheet-type device equipped with a highly sensitive motion detection sensor that is placed under pillows or mattresses to continuously measure the patient's body movements. In addition to HRV indices, this device can measure arousal, sleep, respiratory rate, and bed-withdrawal time. HRV parameters were processed from continuously recorded raw BCG signals at 133 Hz. The BCG signal was preprocessed through filtering and detection of the BCG wave (the J-wave) based on proprietary algorithms. The typical wave form of BCG signals is shown in Figure 1. After detection of J-J wave intervals, ectopic intervals were excluded in time series and formed normal-to-normal intervals. These intervals were used as an alternative to the RR interval (RRI) obtained from ordinary ECGs.

Figure 1. Typical wave form of ballistocardiography signal.

Statistical Analysis

Data are summarized as mean (SD) for continuous variables if the data followed normal distribution and as median (IQR) otherwise. Categorical variables are expressed as numbers and percentages. The time-course changes of HRV indices were examined by drawing a spline chart with a restricted cubic spline model and 3 internal knots fitting each HRV index as a function of time. A repeated measures ANOVA model was conducted to evaluate the within-subject effect, that is, the average of an individual trend of HRV indices over the period for which data were obtained. Since multiple analyses were performed with the reference and the values at each time point, multiplicity was adjusted using the Dunnett-Hsu method. Before analysis, data were natural-log transformed to reduce the right-skewness of the original data, and then daily or weekly averages were calculated for each patient. After analysis, estimates were back-transformed. The results at each time point represent the ratio of the geometric mean of values to reference values. To explore the location of the changepoint of HRV indices over the observation period, a piecewise regression model was used. The regression coefficients were back-transformed; the regression coefficients represent the fold-change for every 10-unit change in the independent variables. The statistical 2-sided significance level was set at .05, and P<.05 was considered statistically significant. All statistical analyses were conducted using SAS (version 9.4; SAS Institute).

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki (revised in 2013) and the Ethical Guidelines for Medical and Health Research Involving Human Subjects in Japan (revised in 2017). The study received approval from the Research Ethics Committee, Faculty of Medicine, Juntendo University (M19-0287). The committee waived the requirement for informed consent for this retrospective study, as it involved the analysis of existing data. All study data were anonymized, and any identifying information was removed to ensure the complete confidentiality of patients. Only authorized research personnel had access to the collected data, and any publication or presentation of the results strictly maintained the anonymity of the study. Patients were not offered any compensation for their participation in this study.

Results

Patient Characteristics

A summary of patient demographic characteristics is presented in Table 1. The data were obtained from 15 patients aged 90 to 100 years, including 8 men and 7 women. The median age for the entire group was 93 (IQR 91-96) years, and by sex, the median age was 94 (IQR 90-97) years for women and 93 (IQR 91.5-94.5) years for men. The other underlying cardiac diseases of the patients were not a concern; however, only cases with no change in medication during the period were included, resulting in a final study total of 15 patients.
Table 1. Characteristics of end-of-life older patients (N=15).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), median (IQR)</td>
<td>93 (91-96)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7 (47)</td>
</tr>
<tr>
<td>Male</td>
<td>8 (53)</td>
</tr>
</tbody>
</table>

**Time-Course Changes of HRV Indices**

Time-course changes in HRV indices were plotted by spline charts (Figure 2). The x-axis in this figure represents the date of the event as 0 and proceeds backwards from there. As shown in Figure 2, the value for ANN gradually decreased from approximately –60 days to approximately –30 days, but the degree of the trend slightly increased after approximately –30 days until the patients’ deaths. A similar trend was observed in RMSSD and SDNN; however, for SDNN, the values seemed to be constant between –60 days to approximately –30 days. After –30 days, a decreasing trend was observed.
Repeated Measures ANOVA

The results of the repeated measures ANOVA are shown in Table 2. In this analysis, the average of the log-transformed data at 8 weeks before the patients’ deaths was used as a reference and compared to each averaged log-transformed result from 7 weeks to 1 week before death. Then, the estimates were back-transformed in order to represent the results as the ratios to reference. At 3, 2, and 1 week before the patients’ deaths, there were significant decreases in ANN (3 w: 0.90, 95% CI 0.84-0.98; \( P = .005 \); 2 w: 0.90, 95% CI 0.84-0.98; \( P = .004 \); 1 w: 0.86, 95% CI 0.80-0.93; \( P < .001 \)) and RMSSD (3 w: 0.83, 95% CI 0.70-0.99; \( P = .03 \); 2 w: 0.83, 95% CI 0.70-1.00; \( P = .04 \); 1 w: 0.78, 95% CI 0.66-0.93; \( P = .002 \)) compared to the reference. Although not statistically significant, SDNN began to decrease from 4 weeks, and a statistically significant decrease was observed at 1 week (0.81, 95% CI 0.67-0.97; \( P = .02 \)) compared to the reference.
Table 2. Repeated measures ANOVA for heart rate variability indices. The values represent ratios of each week’s value to the reference and their corresponding 95% CIs.

<table>
<thead>
<tr>
<th>Weeks before the patients’ deaths</th>
<th>Average normal-to-normal interval (95% CI)</th>
<th>SD of all normal-to-normal intervals (95% CI)</th>
<th>Root mean square of successive differences (95% CI)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Reference (0.93-1.08)</td>
<td>0.99 (0.83-1.19)</td>
<td>0.99 (0.83-1.18)</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>7</td>
<td>1.00 (0.92-1.08)</td>
<td>1.00 (0.83-1.21)</td>
<td>0.99 (0.83-1.18)</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>6</td>
<td>0.98 (0.91-1.06)</td>
<td>1.01 (0.84-1.22)</td>
<td>0.90 (0.76-1.07)</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>5</td>
<td>0.94 (0.87-1.02)</td>
<td>0.96 (0.80-1.16)</td>
<td>0.90 (0.76-1.07)</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>4</td>
<td>0.90 (0.84-0.98)</td>
<td>0.88 (0.73-1.05)</td>
<td>0.83 (0.70-0.99)</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>3</td>
<td>0.90 (0.84-0.98)</td>
<td>0.85 (0.70-1.02)</td>
<td>0.83 (0.70-1.00)</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>2</td>
<td>0.86 (0.80-0.93)</td>
<td>0.81 (0.67-0.97)</td>
<td>0.78 (0.66-0.93)</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>These P values represent the results at each time point relative to the reference, that is, the Dunnett-Hsu–adjusted P value.

<sup>b</sup>These P values represent the results for the change in the heart rate variability indices over the entire observation period.

Changepoint Estimation

The changepoints in HRV indices were explored by using a piecewise regression model. Results are shown in Table 3. The changepoints for ANN, SDNN, and RMSSD were estimated as occurring at –34.5 (95% CI –42.5 to –26.5), –33.0 (95% CI –40.9 to –25.1), and –35.0 (95% CI –42.3 to –27.7) days, respectively, before the date of the patient’s death at 0. The regression coefficients of the slope for the HRV indices before the changepoint were not statistically significantly different from 0; however, after the changepoint, the coefficients indicated that HRV indices were significantly decreased by a factor of the coefficients for every 10 units of increase in days (ANN: 0.959, 95% CI 0.950-0.968; P<.001; SDNN: 0.930, 95% CI 0.908-0.953; P<.001; RMSSD: 0.925, 95% CI 0.905-0.946; P<.001).
Table 3. Changepoint analysis in heart rate variability indices.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimates (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average normal-to-normal interval</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept before changepoint</td>
<td>793.9 (717.7 to 878.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope before changepoint</td>
<td>0.993 (0.980 to 1.006)</td>
<td>.25</td>
</tr>
<tr>
<td>Intercept after changepoint</td>
<td>705.1 (650.9 to 763.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope after changepoint</td>
<td>0.959 (0.950 to 0.968)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Changepoint (days)</td>
<td>−34.5 (−42.5 to −26.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>SD of all normal-to-normal intervals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept before changepoint</td>
<td>51.8 (37.4 to 71.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope before changepoint</td>
<td>1.020 (0.986 to 1.055)</td>
<td>.23</td>
</tr>
<tr>
<td>Intercept after changepoint</td>
<td>38.2 (28.8 to 50.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope after changepoint</td>
<td>0.930 (0.908 to 0.953)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Changepoint (days)</td>
<td>−33.0 (−40.9 to −25.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Root mean square of successive differences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept before changepoint</td>
<td>51.4 (37.9 to 69.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope before changepoint</td>
<td>1.011 (0.980 to 1.043)</td>
<td>.45</td>
</tr>
<tr>
<td>Intercept after changepoint</td>
<td>37.6 (28.9 to 49.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope after changepoint</td>
<td>0.925 (0.905 to 0.946)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Changepoint (days)</td>
<td>−35.0 (−42.3 to −27.7)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*a*The regression coefficient of slope indicates that for every 10-unit increase in days, heart rate variability indices increase or decrease by the value multiplied by the regression coefficient.

*b*Values for changepoint represent days before the date of the patient’s death at 0.

**Discussion**

**Principal Findings**

This preliminary retrospective study aimed to explore the changepoint of HRV indices based on estimates from BCG signals measured over a period of time in older patients at end of life who were residents in a nursing home. Our analysis, conducted by plotting a spline curve, revealed a gradual decline in HRV indices from approximately 30 days before the patients’ deaths. To assess whether this change could be detected statistically, we compared the HRV indices 8 weeks before the date of death with the HRV indices from each week, starting from 7 weeks to 1 week before death, with a repeated measures ANOVA. This analysis found that the ANN and RMSSD values decreased from 3 weeks before the date of death. Furthermore, for a more precise estimation of when the changepoint occurred, we conducted a piecewise regression analysis. This analysis revealed the changepoints for the indices in days before the patients’ deaths for ANN (~34.5, 95% CI ~42.5 to ~26.5 days), SDNN (~33.0, 95% CI ~40.9 to ~25.1 days), and RMSSD (~35.0, 95% CI ~42.3 to ~27.7 days). The results indicate that if the changepoint can be detected during BCG monitoring, it may be possible to predict the survival time from that changepoint: approximately 1 month later. Accurate estimation of the time of death approximately 1 month in advance by continuously monitoring HRV indices obtained from BCG could provide families or health care providers with more time to prepare for end-of-life care. Although our results are from a retrospective study and the sample size was small, to our knowledge, this is the first study to analyze data from long-term monitoring of older patients without cancer at end of life using HRV indices estimated from BCG.

HRV measures the time interval between adjacent heartbeats, that is, the variation in the time interval from R wave to R wave (ie, RRI) on the ECG. HRV indices are a useful noninvasive means of assessing autonomic function. The gold standard for evaluation of HRV indices is to evaluate the RRI in an ECG. However, in recent years, other systems have been developed to acquire BCG signals, most of which come in the form of a mattress or chair [34], and BCG is considered to be a potential substitute for HRV indices [35,36]. Martín-Yebra et al [36] evaluated the JJ, II, KK, and HH intervals of the BCG signal as an alternative to the ECG’s RRI and showed that the JJ interval was almost consistent with the RRI measured simultaneously with BCG. We also analyzed the HRV indices estimated from the JJ intervals of the BCG signal in our study.

The rapid aging of the world’s population and the widespread use of HRV have increased interest in the prognostic value of HRV in older patients, even outside the specific field of cardiology [37]. Kurita et al [38] examined the survival of
patients based on HRV and blood tests such as serum albumin and C-reactive protein levels at admission among older Japanese individuals in special nursing homes. They reported that there was no statistically significant difference between the death and survival groups in terms of blood test values. However, the SDNN and coefficient of variables of RRI values, which have been suggested to be related to parasympathetic activity, were statistically higher in the survivor group. In another study, which used HRV to predict survival time for terminally ill patients with cancer [39], the length of survival time was shorter in groups with lower SDNN (21.3 ms or less) or higher heart rates (100 or more beats per minute) measured at baseline. Another prospective study using HRV to evaluate the prognostic values for discharge from the hospital in cancer patients compared the high-frequency (HF) and low-frequency components of HRV at admission between patients who were discharged and those who died. In the study, none of the HRV indices were statistically different, but the estimated value of the HF components of HRV as an index reflecting parasympathetic nerve activity (vagal activity), tended to be higher in the discharged group [40]. Since our study also showed a decrease in ANN, which represents an increase in heart rate, and a decrease in SDNN and RMSSD, which reflect parasympathetic function, it may be appropriate to use HRV indices related to parasympathetic function to make prognoses at end of life in older patients.

Our results showed a change in HRV indices approximately 35 days before patients’ deaths. A previous study of signs of mortality and the timing of their appearance in end-of-life patients in Japan [41] was based on the subjective perceptions of nurses working at a special nursing home for older people; it reported that the signs of mortality were divided into 19 categories and classified them according to the time they appeared. The study found that some signs appeared approximately 1 month before death and others appeared 2 days before. Signs that appeared 1 month prior to death included lack of eyesight, paleness, lack of vitality, somnolence, and decreased food intake compared to previous stable daily activities. Notably, the changepoint of the HRV indices was observed approximately 1 month before death in our study, and there may be some relationship between the signs recognized by the nurses and the change in HRV indices, that is, they may run in parallel. Therefore, in the future, it may be possible to construct a more accurate model that can predict the survival time of patients by combining more objective changes in HRV with the subjective symptoms identified by nurses.

**Strength**

The strength of this study lies in the analysis of data collected continuously from patients for more than 2 months. For example, the Palliative Performance Scale (PPS) and Palliative Prognostic Index (PPI) scores at admission are commonly used as prognostic tools for terminal cancer patients [42]. However, many studies rely solely on values at admission to assess prognoses. In contrast, Kao et al. [43] postulated that using only the PPI score on the first day may be limited because it does not consider subsequent changes in the patient’s condition. They showed that combining PPI scores from day 1 and 1 week later improved prognostic accuracy. In addition, Chan et al. [44] examined changes in PPS scores at admission, week 1, and week 2 in terminally ill patients to predict patient survival. They found that all of these changes were independent predictors, with the greater the change, the higher the hazard ratio. These studies underscore the importance of analyzing changes over time. Although it is relatively convenient to measure the items needed to calculate PPS and PPI in a clinical setting, it is burdensome to observe these items daily. Therefore, if the sheet-type device used in this study can be used to measure HRV indices without inconvenience to the individual and facilitate making prognoses in end-of-life patients, it could be a very useful approach.

**Limitations**

There are several limitations in this study. First, it was retrospective and used existing data from a single institution; therefore, selection bias was unavoidable. Second, the study used a small amount of data from 15 patients, which could have affected the statistical validity. Third, the patients’ detailed background information was not available, and the study included patients with a variety of underlying diseases. Therefore, the results of this study may have been influenced by chance or by confounding factors, such as the patients’ medical histories and medical conditions while they still survived.

**Conclusions**

To our knowledge, this is the first study to assess HRV changepoints estimated from BCG signals in older patients at end of life. We found changepoints occurred approximately 1 month before a patient’s death, indicating that these changes could be used to predict patients’ deaths. This study also suggests the possibility of developing more objective and accurate predictive tools and offers valuable insights for future research. Such tools could be created by integrating BCG data with the subjective judgments of nurses, caregivers, and other health professionals, as well as established tools like the PPS, to make prognoses in cancer patients. However, since the findings were obtained in a retrospective study, future research is needed to determine how to detect these changes when observed prospectively.

**Acknowledgments**

This work was supported by a grant from JSPS KAKENHI Grants-in-Aid (JP21K12800), partly supported by other grants from JSPS KAKENHI Grants-in-Aid (JP17K09527, JP18K15904, JP21K16034), and supported by the Intractable Respiratory Diseases and Pulmonary Hypertension Research Group of the Ministry of Health, Labor and Welfare of Japan (20FC1027) and a research grant from the Japanese Center for Research on Women in Sport at Juntendo University. The funding agency had no role in the study design, any aspects of data analysis or interpretation, or manuscript publication.
Data Availability
The data sets generated during and/or analyzed during this study are not publicly available due to the confidential nature of the information, and consent to share data with outside parties has not been obtained from the patients.

Conflicts of Interest
TK is affiliated with departments endowed by Philips, ResMed, and Fukuda Denshi, as well as with Paramount Bed. The other authors have no conflicts of interest to declare.

References


Abbreviations

- **ANN**: average of normal-to-normal interval
- **BCG**: ballistocardiography
- **ECG**: electrocardiogram
- **HF**: high frequency
- **HRV**: heart rate variability
- **PPS**: Palliative Performance Scale
- **PPI**: Palliative Prognostic Index
- **RMSSD**: root mean square of successive differences
- **RRI**: RR interval
- **SDNN**: SD of normal-to-normal interval

©Naotake Yanagisawa, Yuji Nishizaki, Bingwei Yao, Jianting Zhang, Takatoshi Kasai. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Engaging Cancer Care Physicians in Off-Label Drug Clinical Trials: Human-Centered Design Approach

Maren C Parsell¹,², MBA; Morgan N Greenleaf¹,²,³, MS; Greeshma G Kombara²,⁴, MPH; Vikas P Sukhatme²,⁴,⁵, MD, SCD; Wilbur A Lam¹,²,³,⁶, MD, PhD

¹Georgia Clinical and Translational Science Alliance, Emory University School of Medicine, Emory University, Atlanta, GA, United States
²Emory School of Medicine, Emory University, Atlanta, GA, United States
³Center for the Advancement of Diagnostics for a Just Society (ADJUST), Emory University, Atlanta, GA, United States
⁴The Morningside Center for Innovative and Affordable Medicine, Emory University School of Medicine, Emory University, Atlanta, GA, United States
⁵Departments of Medicine and Hematology and Medical Oncology, Emory University, Atlanta, GA, United States
⁶Aflac Cancer and Blood Disorders Center of Children’s Healthcare of Atlanta and Department of Pediatrics, Emory University, Atlanta, GA, United States

Corresponding Author:
Maren C Parsell, MBA
Georgia Clinical and Translational Science Alliance
Emory University School of Medicine
Emory University
1440 Clifton Road
Atlanta, GA, 30322
United States
Phone: 1 4048896149
Email: maren.parsell@emory.edu

Abstract

Background: Using a human-centered design (HCD) approach can provide clinical trial design teams with a better understanding of the needs, preferences, and attitudes of clinical trial stakeholders. It can also be used to understand the challenges and barriers physician stakeholders face in initiating and completing clinical trials, especially for using off-label drugs (OLDs) to treat unmet clinical needs in cancer treatment. However, the HCD approach is not commonly taught in the context of clinical trial design, and few step-by-step guides similar to this study are available to demonstrate its application.

Objective: This study aims to demonstrate the feasibility and process of applying an HCD approach to creating clinical trial support resources for physician stakeholders to overcome barriers to pursuing clinical trials for OLDs to treat cancer.

Methods: An HCD approach was used to develop OLD clinical trial support concepts. In total, 45 cancer care physicians were contacted, of which 15 participated in semistructured interviews to identify barriers to prescribing OLDs or participating in cancer OLD clinical trials. Design research is qualitative—it seeks to answer “why” and “how” questions; thus, a sample size of 15 was sufficient to provide insight saturation to address the design problem. The team used affinity mapping and thematic analysis of qualitative data gathered from the interviews to inform subsequent web-based co-design sessions, which included creative matrix exercises and voting to refine and prioritize the ideas used in the final 3 recommended concepts.

Results: The findings demonstrate the potential of HCD methods to uncover important insights into the barriers physicians face in participating in OLD clinical trials or prescribing OLDs, such as recruitment challenges, low willingness to prescribe without clinical data, and stigma. Notably, only palliative care participants self-identified as “frequent prescribers” of OLDs, despite high national OLD prescription rates among patients with cancer. Participants found the HCD approach engaging, with 60% (9/15) completing this study; scheduling conflicts caused most of the dropouts. Over 150 ideas were generated in 3 co-design sessions, with the groups voting on 15 priority ideas that the design team then refined into 3 final recommendations, especially focused on increasing the participation of physicians in OLD clinical trials.

Conclusions: Using participatory HCD methods, we delivered 3 concepts for clinical trial support resources to help physician stakeholders overcome barriers to pursuing clinical trials for OLDs to treat cancer. Overall, integrating the HCD approach can aid in identifying important stakeholders, such as prescribing physicians; facilitating their engagement; and incorporating their
Introduction

Project Goals

Cancer clinical trials involve the coordination of many stakeholders but frequently fail to meet enrollment goals, prespecified end points, and timelines [1,2]. Effective stakeholder engagement can be essential to cancer clinical trial design, conduct, and reporting of clinical research [3]. However, traditional design approaches often lack consideration of stakeholders’ needs, preferences, and experiences, which can lead to recruitment and retention challenges or to study results that are less generalizable or applicable to real-world settings. Trial stakeholders include physicians who face unique barriers, especially when designing trials for off-label drugs (OLDs) to treat unmet clinical needs in cancer treatment [4]. Assumptions that stakeholders will be willing to participate in a trial may not adequately factor in the unique challenges and barriers that patients with cancer and stakeholders face. In addition, underrepresented and marginalized populations’ needs can also be missed because their perspective is rarely represented by clinical trial designers, which can lead to a lack of diversity in clinical trial participants and potential disparities in treatment outcomes [1].

To overcome these limitations, a human-centered design (HCD) approach offers both a process to learn about the needs of the trial stakeholders as well as flexible tools to test assumptions, uncover barriers, and work collaboratively with study stakeholders to optimize the clinical trial design. An emerging body of literature on HCD for health care innovation yields many HCD methods to choose from [5], making it difficult for novice innovators to know where to start and which methods to use. The purpose of this paper is to demonstrate HCD methods through a case study, where cancer care physicians interviewed, 8 were oncologists, 3 were palliative care specialists, 2 were urologists, 1 was an anesthesiologist, and 1 was a family medicine physician (Multimedia Appendix 1). Although this sample size is small, it was adequate to achieve insight saturation, or the point where participants were providing similarly themed insights.

Background

HCD is a problem-solving methodology that focuses on discovering needs and developing solutions for individuals within a system; it is increasingly being used in cancer care settings [6-8]. HCD is particularly well-suited for uncovering and understanding the complex attitudes and barriers contributing to physician off-label prescribing behavior for a variety of Food and Drug Administration–approved drugs with strong safety profiles such as aspirin that may have anticancer properties [9]. These medications have shown early evidence when prescribed off-label, in reducing the risk of developing certain types of cancer such as colorectal cancer, or in improving patient outcomes. However, due to low financial incentives and resources, and the lack of further clinical testing to provide definitive prospective data, physician awareness is low, and prescription adoption has been minimal [10]. To help build incentives and engage physicians to consider repurposed OLD options, the Morningside Center for Innovative and Affordable Medicine (Morningside Center) [11] based at Emory University worked with staff trained in HCD methods from the Innovation Catalyst program within the Georgia Clinical and Translational Science Alliance (CTSA) [12]. The Georgia CTSA team set out to first learn more about the barriers to these types of prescriptions and then work together with physicians to propose support programs to increase the consideration of repurposed OLD treatment options for their patients. We explored the implications of these findings for physician receptiveness to conducting clinical trials with repurposed OLDS.

Methods

Study Design

HCD is an iterative and flexible process for generating solutions to problems that typically includes 3 phases: learning about the humans involved in the situation, coming up with and refining concepts, and implementation. Additional steps can be added to address the specific needs of the project, and the CTSA research team emphasized stakeholder analysis because many stakeholders influence prescription decisions, so understanding who had the greatest influence was important to forming the guiding questions for the ideation activity. Thus, the team followed four steps with each step matched with an appropriate HCD method: (1) collect, sort, and analyze insights using semistructured interviews and affinity mapping; (2) identify influences using a stakeholder power-interest grid; (3) generate and prioritize ideas using a co-design activity called a “creative matrix” [13]; and (4) produce concepts to prototype by refining and ranking ideas generated from the co-design activity (Figure 1). The team identified 45 physicians who met the cancer care inclusion criteria within the Emory academic medical center ecosystem or community practice setting. From this pool, 15 agreed to participate in semistructured interviews. Of the 15 physicians interviewed, 8 were oncologists, 3 were palliative care specialists, 2 were urologists, 1 was an anesthesiologist, and 1 was a family medicine physician (Multimedia Appendix 1). Although this sample size is small, it was adequate to achieve insight saturation, or the point where participants were providing similarly themed insights.
Collect Insights

Within the HCD approach, qualitative research methods such as interviews, focus groups, or ethnography are typically used to gain insights about the situation, influencing factors, as well as the needs and wants of the stakeholders involved. The Georgia CTSA research team conducted semistructured interviews, which involve asking open-ended questions that allow for more free-flowing conversation than in structured interviews. A set of predetermined questions or topics guided the conversation, but there was also flexibility to follow-up on interesting or unexpected responses and to explore topics in more detail. Some of the key questions the team sought to answer were:

- Have you prescribed OLDS? Why or why not?
- What level of information or data do you need to prescribe?
- Where do you find information on OLDS?
- Who influences your OLDS prescription decisions?
- What influences your decision to participate in OLDS clinical trials?

The interviews produced dozens of pages of transcribed data, which the team reduced to key points on Post-it notes in preparation for a thematic analysis process called affinity mapping. Concepts that are related to each other were grouped, and a theme was coded for each group (Figure 2). In addition, stakeholders who influenced prescription decisions were also noted.
Identify Influencers

To validate assumptions about who ultimately makes OLD prescribing decisions and to verify decision-making impact, the Georgia CTSA research team identified stakeholders based on the insights collected from the interviews and affinity mapping process (Multimedia Appendix 2). This list of stakeholders was then mapped to a power-interest grid, which involves plotting stakeholders on a 3D grid based on their level of power or influence over the situation and their level of interest in outcomes (Figure 3). Stakeholders with high levels of power and high levels of interest are considered key players and require the most engagement while high-interest and lower-power stakeholders should be kept updated.
Figure 3. Screenshot of power-interest grid method used for stakeholder analysis in this study. FDA: Food and Drug Administration; IRB: institutional review board; MD: Doctor of Medicine; mgmt: management; OLD: off-label drug; PCP: phencyclidine; PI: principal investigator.

Generate Ideas
Co-design activities include the end users of a solution in the ideation of that solution. Co-design is an effective way to increase stakeholder engagement, which is essential to project success, particularly in cancer care delivery research [3]. There are many methods to collaboratively come up with ideas, such as rapid ideation, brainwriting, mind mapping, thumbnail sketching, and many others. Which method to select depends on the goals of the session, the time allotted, and who is participating. The Georgia CTSA team selected the creative matrix method (Figure 4), which concentrates ideation on a particular topic using a visual grid with 2 questions and up to 4 categories. This method can generate a high volume of ideas in a short time and within a specific context. The research team selected salient topics that arose from the affinity mapping exercise to create the context for 2 creative matrix exercises used in three 1-hour co-design sessions with 9 physicians. Participants were allotted 6 minutes to generate ideas and 3 minutes to select 2 concepts that they believed had the most valuable to help support the consideration of repurposed OLDs for prescription. These “best ideas” were subsequently mapped on a grid for impact versus effort to identify which ideas would have the most impact and lowest effort.
**Vote and Produce Prototypes**

Involving co-design participants in the idea evaluation process is important for getting feedback on the feasibility, desirability, and viability of the ideas [14]. It can increase participants’ sense of ownership and investment in the final solution, potentially leading to greater adoption and success. While co-design sessions can generate a high volume of ideas quickly, further refinement to create workable prototypes for testing is sometimes necessary, and this study’s team distilled concepts that were well-received by physicians during the co-design sessions into 3 prototype-ready directions.

**Ethical Considerations**

This study was reviewed by an ethics committee to ensure the protection of research participants, and it was approved by Emory University Institutional Review Board (STUDY00004025). Each participant provided consent to be interviewed and that their responses could be used in our design study without additional consent. The privacy and confidentiality of participants and their responses were ensured by removing names and any other fact that might point to participants’ identity in this study; also, research records are kept private to the extent required by law. The compensation for participation in this study was a US $50 gift card.

---

**Table: Guided Ideation Framework**

<table>
<thead>
<tr>
<th>Context</th>
<th>Question 1</th>
<th>Question 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do we support Emory physicians for them to consider OLDs as an option for concurrent treatment?</td>
<td>How might we offer access to the right information to support prescription decisions?</td>
<td>How might we provide support to increase OLD prescription comfort?</td>
</tr>
</tbody>
</table>

**Diagrams**

- **Context:** How do we support Emory physicians for them to consider OLDs as an option for concurrent treatment?
- **Question 1:** How might we offer access to the right information to support prescription decisions?
- **Question 2:** How might we provide support to increase OLD prescription comfort?

**Figure 4:** Screenshot of creative matrix method of guided ideation used in this study. AI: artificial intelligence; AMC: academic medical center; EMR: electronic medical record; FDA: Food and Drug Administration; OLD: off-label drug; WG: working group.
Results

Thematic Analysis
Of the 45 physicians we contacted, 15 agreed to be interviewed and share their experiences and opinions on OLD prescribing for cancer care. Key themes that emerged from the interviews showed that, in contrast to documented high off-label prescription rates in cancer [15-17], the interviewed physicians were less likely to self-identify as “frequent prescribers,” citing a range of barriers. A lack of clinical trial evidence to support the prescription was most frequently expressed as a barrier to prescription, as well as cost and reimbursement concerns and the lack of knowledge of off-label options. A notable exception to off-label prescription hesitancy was seen in palliative care specialists who expressed few barriers to prescribing off-label but emphasized that their focus was on easing symptoms and not providing treatment. To learn about OLDS, physicians described a variety of sources but most often relied on professional social networks to discover and deepen knowledge; however, they struggled to find information on specific OLDS when they needed it.

On the power-influence grid, cancer care physicians were identified as the stakeholders having the highest power and interest in patient prescriptions, validating the assumption that they are the primary prescription decision makers who require the most engagement. Patients and their families were mapped to the high-interest but less power quadrant and were sometimes the source of OLD prescription requests that did not always lead to prescriptions.

Concept Generation
The Georgia CTSA research team synthesized insights from the interviews into 2 brainstorming themes used during 3 co-design sessions with 9 physician participants. The first context explored types of support needed for physicians to consider OLDS as concurrent treatment, which generated 70 concepts. The second context focused on what support was needed to make it easier to start and complete an OLD clinical trial, which produced 95 concepts. Of the 165 total concepts, participants voted for 14 of the ideas that they felt best supported their needs, which the physicians mapped to identify which were high impact or lower effort. The selected concepts included providing dedicated clinical research coordinators, protected research time, electronic medical record prompts or reminders, incentive programs, an OLD reference database or app, and endorsements for OLD trials at team meetings to support recruiting efforts (Multimedia Appendix 3).

The research team further synthesized these concepts into 3 prototype-ready directions listed in Textbox 1.

Textbox 1. Recommended concepts delivered to the Morningside Center after this study.

<table>
<thead>
<tr>
<th>Concept and description</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Off-label drug (OLD) fellowship program</td>
</tr>
<tr>
<td>• An invitation-only program for physicians initiating OLD clinical trials that emphasizes promotional communications for the principal investigator (PI) to promote the study and offers protected research time.</td>
</tr>
<tr>
<td>• OLD clinical trial support package</td>
</tr>
<tr>
<td>• Sponsored package of select administrative resources including a clinical resource coordinator to facilitate OLD trials.</td>
</tr>
<tr>
<td>• OLD searchable database</td>
</tr>
<tr>
<td>• On-demand digital tool to provide national repurposed OLD updates, identify center-wide OLD trials and PIs, and look up available clinical trial resources.</td>
</tr>
</tbody>
</table>

In a web-based survey, all 15 physician participants were asked to force rank the concepts that they felt was the most impactful to supporting repurposed OLD prescription decisions; we received 7 responses (Multimedia Appendix 4). The fellowship program concept surfaced as the leading idea with a mean of 2.0 (SD 0.93; Figure 5). Overall, concepts that resonated most either supported clinical trial development or bolstered social interfaces from which they could learn about OLDS from other physicians.
Discussion

Principal Findings

Using this case study, we have demonstrated the flexible nature of HCD as it relates to accommodating unique characteristics of the project, selecting from a wide range of ideation techniques, and adapting to participant ideas and preferences. The findings also demonstrate the potential of HCD methods to uncover important insights into physician decision-making processes that can lead to improving OLD prescription and clinical trial practices. Clinical trial designers can similarly use these methods to gain a deeper understanding of the target audience’s needs, preferences, and attitudes toward a potential trial. Insights can help overcome barriers to participation, such as language or cultural differences, and inform more effective trial protocols that lead to greater participant engagement and retention. Without using the grounded approach of first synthesizing stakeholder insights, ideation sessions may not be as productive.

A notable feature of HCD is high engagement among participants, also demonstrated in this study with 60% (9/15) of physicians participating from the initial interviews to the ideation and final voting exercises. Further, 1 participant said, “I really appreciate being listened to and would be willing to volunteer to test these ideas.” HCD can be used to increase patient engagement by incorporating the needs, preferences, and attitudes of potential trial participants [18]. Participants who did not participate in all activities of this study either did not respond to the invitations or declined because of scheduling conflicts.

HCD’s participatory nature contrasts with the traditional clinical trial design, which often does not engage the participants for which the trial is designed. This can lead to important perspectives or data being missed and result in low patient engagement and incomplete or biased research outcomes. The high failure rate of clinical trials has led to increasing interest in using stakeholder and patient engagement to support clinical trial design [3,19]. In the context of clinical trial design, a co-design activity could involve collaborating with patients, health care providers, and other stakeholders to co-design trial materials, such as patient education materials or data collection forms to ensure that they are acceptable and feasible.

While HCD offers a flexible process and many methods to choose from, it is very different from the linear thinking and hierarchical norms associated with traditional scientific approaches [20], which may thwart the adoption or effective implementation of the HCD process. Additional challenges can include the potential for interviewer bias and the time- and resource-intensive nature of conducting interviews and co-design workshops. While HCD is becoming an increasingly familiar term, access to training remains low as few institutions have internal design teams, usually found in innovation centers [21], to consult with or to run projects. Thus, acquiring and building HCD expertise can be expensive, and it remains to be seen how future clinical innovation funding or such services will be allocated.

Conclusions

Although a relatively new problem-solving approach in health care, HCD can provide tangible, flexible, and reproducible methods that include stakeholders in the ideation and problem-solving process [22]. HCD can help identify and engage

![Figure 5. Force ranked physician votes for 3 prototype-ready concepts to support repurposed OLD prescription decisions. OLD: off-label drug.](image-url)
important stakeholders such as physicians and patients in the design process and then integrate their perspectives and needs into the subsequent solutions, which is increasingly seen as an important contributor to cancer clinical trial success. This case study provides a step-by-step guide on how to apply the HCD process and selected methods to generate stakeholder-centered solutions. HCD has the potential to be an important tool to increase success rates of clinical trials, but increased institutional support and researcher training will be needed for HCD to provide its fullest benefit.

Acknowledgments
This work was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR002378) and the Emory Morningside Center for Innovative and Affordable Medicine.

Data Availability
All data generated or analyzed during this study are included in this published paper and its supplementary information files.

Conflicts of Interest
None declared.

Multimedia Appendix 1
List of physicians who participated in the study, their specialties, and co-design session attendance status.
[PDF File (Adobe PDF File), 36 KB - formativew818e51604_app1.pdf ]

Multimedia Appendix 2
Brainstorm activity to record all stakeholders who influence the prescription of off-label drugs, used as the first step in the stakeholder mapping exercise.
[PDF File (Adobe PDF File), 243 KB - formativew818e51604_app2.pdf ]

Multimedia Appendix 3
Co-design activity concepts organized by innovation question.
[PDF File (Adobe PDF File), 277 KB - formativew818e51604_app3.pdf ]

Multimedia Appendix 4
Descriptions of prototype concepts with survey results from physician participants.
[PDF File (Adobe PDF File), 378 KB - formativew818e51604_app4.pdf ]

References
5. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR mHealth uHealth 2021;9(12):e0238099 [FREE Full text] [doi: 10.1371/journal.pone.0238099] [Medline: 32833974]


Abbreviations

CTSA: Clinical and Translational Science Alliance
HCD: human-centered design
OLD: off-label drug
the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Implementation Documentation and Process Assessment of the PharmNet Intervention: Observational Report

Lori Ann Eldridge¹, PhD; Beth E Meyerson²,³, PhD; Jon Agley⁴, MPH, PhD

¹Department of Health Education and Promotion, College of Health and Human Performance, East Carolina University, Greenville, NC, United States
²Harm Reduction Research Lab, Family and Community Medicine, College of Medicine-Tucson, University of Arizona, Tucson, AZ, United States
³Comprehensive Center for Pain and Addiction, University of Arizona Health Sciences, University of Arizona, Tucson, AZ, United States
⁴Prevention Insights, Department of Applied Health Science, School of Public Health Bloomington, Indiana University Bloomington, Bloomington, IN, United States

Corresponding Author:
Jon Agley, MPH, PhD
Prevention Insights, Department of Applied Health Science
School of Public Health Bloomington
Indiana University Bloomington
809 E 9th Street
Bloomington, IN, 47405
United States
Phone: 1 812 855 3123
Email: jagley@indiana.edu

Related Article:
This is a corrected version. See correction statement: https://formati

Abstract

Background: The number of overdose deaths in the United States involving opioids continues to exceed 100,000 per year. This has precipitated ongoing declarations of a public health emergency. Harm reduction approaches, such as promoting awareness of, ensuring access to, and fostering willingness to use naloxone to reverse opioid overdose, are a key component of a larger national strategy to address the crisis. In addition, overdose reversal with naloxone directly and immediately saves lives. Because of pharmacies’ ubiquity and pharmacists’ extensive clinical training, community pharmacies are well-positioned, in principle, to facilitate naloxone access and education.

Objective: In 2022, a single-site pilot study of PharmNet, a community pharmacy intervention incorporating naloxone distribution, awareness building, and referral, showed promising outcomes for both naloxone and resource distribution in the community. As a next step, this study was intended to be a pilot randomized controlled trial of PharmNet in 7 pharmacies. However, due to circumstances outside of the study team’s control, data collection was unable to be fully completed as planned. In keeping with open research standards, we transparently report all available data from the study and discuss trial barriers and processes. We do so both to provide insights that may inform similar studies and to avoid the “file-drawer” (publication bias) problem, which can skew the aggregated scholarly literature through nonpublication of registered trial results or selective publication of findings affirming authors’ hypotheses.

Methods: This paper reports an in-depth implementation study assessment, provides the available observational data, and discusses implementation considerations for similar studies in independent (eg, nonchain) community pharmacies.

Results: Retrospective assessment of study outcomes and fidelity data provided for robust discussion around how resource differences in independent community pharmacies (vs well-resourced chain pharmacies), as well as high demands on staff, can affect intervention implementation, even when leadership is highly supportive.

Conclusions: Community pharmacies, particularly independent community pharmacies, may require more support than anticipated to be successful when implementing a new intervention into practice, even if it might affect estimates of real-world effectiveness. Further implementation science research is needed specific to independent community pharmacies. All study elements are outlined in the International Registered Report Identifier (IRRID) PRR1-10.2196/42373. Although this paper reports results associated with that registration, results and conclusions should not be given the weight assigned to findings from a preregistered study.
Introduction

Background
In the last decade, hundreds of thousands of lives have been lost in the United States due to fatal opioid-involved overdoses [1], and annual overdose death rates (computed as 12-month year-over-year incidence) have remained over 100,000 in recent reports from the Centers for Disease Control and Prevention [2]. Additionally, overdose deaths represent over one-third of all accidental deaths in the United States [3], with the majority involving illicitly manufactured fentanyl, a synthetic opioid [4]. In the state of Indiana, where this study was conducted, preliminary data indicate 2250 overdose deaths in 2022 [5], around 71% of which involved fentanyl or other synthetic opioids [5].

Naloxone, an opioid antagonist, works by blocking the effects of opioids and can allow persons who have overdosed to resume normal breathing [6]. It is a safe and effective way to rapidly reverse the effects of opioid overdose and can prevent death [6,7]. Research has shown that overdose deaths decline when overdose education and naloxone distribution (OEND) efforts are undertaken in communities [8], and that OEND programs are both an effective [9,10] and cost-effective [11] strategy to address overdose fatality.

In 2018, the US Surgeon General issued an official recommendation that all US citizens carry naloxone [12]. By 2021, most US states, including Indiana, had begun to address the overdose epidemic using combinations of policies that generally included “Good Samaritan” laws and standing orders for naloxone at pharmacies [13]. More recently, the declaration of an opioid public health emergency was renewed in September 2023 [14], approximately a year after the release of the US National Drug Control Strategy, which outlines goals that include the expansion of access to evidence-based harm reduction practices, including naloxone [15].

Beyond the impact in terms of lives saved, individuals who are revived with naloxone and who also have opioid use disorder (OUD) have an opportunity to engage with an “Opioid Cascade of Care” [16], such as through treatment initiation with medications for opioid use disorder (MOUD) in the emergency department [17]. For these benefits to accrue, however, naloxone must be easily accessible by those who can use it to reverse an opioid overdose. In the United States, nasal formulations of naloxone were recently approved for over-the-counter sale [18], following a unanimous vote by Food and Drug Administration advisors [19]. While this change might in principle increase the number of people carrying naloxone, data from Australia suggest that it may be a helpful but not sufficient step to increase access [20].

Even as additional means of obtaining naloxone are developed (eg, naloxone vending machines [21]), community pharmacies will likely remain an important component of supporting naloxone access and education. In some populations, patients interact with their pharmacists more often than with their primary care physicians [22], and pharmacists have been identified as some of the most accessible health care providers in the United States [23,24]. Studies show that community pharmacists generally acknowledge their role in assisting in harm reduction [25-27], and an Arizona study found that community pharmacists were typically comfortable providing consultation around and dispensing naloxone, even when they were uncomfortable with other harm reduction approaches such as syringe dispensing without a prescription [28]. As such, community pharmacists are theoretically well-positioned to improve harm reduction education and accessibility of related supplies. At the same time, recent studies also continue to suggest that naloxone remains unavailable for purchase in some community pharmacies and may be less likely to be available in independent pharmacies compared to chain pharmacies [29].

The PharmNet Intervention
To better understand the dynamics and potential of community pharmacy harm reduction practice, we completed multiple preliminary studies [26,28,30-35]. Our work examined naloxone access, pharmacist comfort with harm reduction, pharmacy-based research approaches, and the “fit” of the pharmacy environment for harm reduction–related services. The studies were conducted from 2016 through 2022, and many of them engaged community and academic pharmacists as part of the research team.

By the end of the process, we had determined that our initial conceptualization of a “short” or “minimally demanding” intervention (eg, the first draft of PharmNet, a 10- to 15-minute screening and brief intervention-style intervention) would likely be perceived as useful but not necessarily feasible without a substantive added support system (eg, financial and personnel). Further, there was variability in the a priori intervention components that would be available at different pharmacy types. For example, a systematic review [36] in addition to our own earlier research in Indiana [33] and a recent study in North Carolina [29] identified independent pharmacies as being significantly less likely to stock or dispense naloxone compared to chain pharmacies (eg, ability to dispense naloxone, even at cost, could not be assumed). While independent pharmacies are, in theory, more likely to be able to participate in a small intervention implementation program (eg, they do not need to seek approval from a large bureaucracy associated with a national or international chain), they also lack the financial and systemic resources that are available to chain pharmacies, which can have implications for stocking and dispensing of products, as well as pharmacists’ availability to undertake additional work.
Thus, the PharmNet intervention was modified with the goal “to study procedures that have as minimal an impact as possible on pharmacy costs and operational functioning while maximally facilitating harm reduction from opioid overdose—in other words, to find an optimal intersection point of those concerns” [37]. The overall goals of the finalized PharmNet intervention are outlined in a previous paper, which reported the results of our initial single-site pilot study [37]. For clarity, we excerpt them here verbatim:

1. Building awareness of naloxone availability at the site among patients (eg, the use of yard signs and scrolling messages on television screens).
2. Supporting awareness among pharmacists and pharmacy technicians about proactively offering naloxone (eg, customized post-it notes).
3. Facilitating service provision by conducting a priori negotiations and establishing written agreements with local nonprofits to facilitate a pipeline of no-cost naloxone for the pharmacy.
4. Emphasizing bidirectional naloxone provision (eg, pharmacist- and pharmacy technician-initiated offers in addition to patient-initiated requests), and
5. Facilitating referral by providing a physical, durable, and curated list of community resources that can be used by pharmacists, handed to patients, and placed into pharmaceutical bags.

Interestingly, these goals are generally aligned with the intervention approaches described in a scoping review of opioid counseling and naloxone services published after our study concluded [38]. The review found that “all studies incorporated naloxone recommendations into their pharmacy-based opioid counseling and naloxone services,” and that most programs focused mostly on OEND [38]. Further, most of our awareness and communication strategies were similar to those in the review, though only one study in the review used outdoor yard signs, and none appear to have established pathways for pharmacies to receive no-cost naloxone [38].

Our first implementation of PharmNet was in a single independent pharmacy in the Midwestern United States. In that 3-month single-site pilot study, naloxone dispensing by any means (paid or no cost) nearly doubled, increasing by 96.48% (+9.33 doses per month); this included both distribution of no-cost doses (+6 doses per month) and an increase in the monthly rate of naloxone sold (+3.33 doses per month) compared to the 3 months before the intervention [37]. An average of 2.85 referral and resource cards (1/4 page in size) customized to the local area were distributed to patients each day. The initial pilot study also allowed us to further adjust the procedures based on feedback from community pharmacists [37]. The most substantive change was a drop from 3 sets of different post-it reminders to one, because only one type was regularly used by the pharmacy staff members in practice.

To better understand the results, and to begin assessing the intervention’s scalability, we then developed and published a study protocol for a pilot randomized controlled trial (RCT) of PharmNet [39]. Our stated goal was “to facilitate an improved understanding of whether it is reasonable to believe that the PharmNet intervention causes increased dispensing of naloxone...” and we proposed to test the hypothesis that, “Monthly naloxone dispensing (combined sales and no-cost distribution) will be significantly increased in the pharmacies implementing PharmNet compared to those in the control arm” [39].

**This Report**

This paper was originally planned to describe the results from a small RCT of PharmNet based on our protocol paper that also served as a registered report (International Registered Report Identifier: PRR1-10.2196/42373) [39]. However, several events and circumstances beyond the control of the study team (eg, a COVID-19 outbreak among pharmacists) resulted in substantial deviations from the proposed approach and outcome measures. The conceptual magnitude of these differences and the resultant effects on data and procedures mean that it would be inappropriate to attempt to assert that the information reported in this paper serves as answers to our original research questions.

At the same time, it is important for us to provide a complete description of the study and any resultant observations and lessons learned. In doing so, we avoid contributing to the “file drawer” (publication bias) problem, wherein studies that are perceived to be unsuccessful or that have nonsignificant results are never published to the detriment of the scientific enterprise and those affected by it [40-42]. Publication bias has been described as being conceptually similar to the practice of p-hacking [43]; it can skew the perception of a field of study in a particular direction through suppression (either deliberate or inadvertent) of results that do not align with the scientists’ goals. This can happen in at least 2 ways. First, results from registered trials may never be published at all (a recent study of RCTs in New Zealand found that 27% of results had not been published more than a decade after the study registration date) [44]. This can produce the mistaken impression that not much is known about a particular approach. Second, a “failed” study may be repeated with some minor modifications, and the resultant data reported as though it was the originally registered trial. This can lead to overestimation of the intervention’s likely outcomes as well as reducing the likelihood that procedural and implementation considerations—some of which may be important—are fully and transparently addressed.

Therefore, we share the results of this study in the form of an observational report that includes (1) the findings that we were able to collect, (2) the methods to collect that information, as well as (3) specific information related to deviations from planned data collection and analysis. Then, we provide a detailed written record of fidelity or process checks and the results of those checks. Finally, the implications of the external issues leading to these events, including discourse around the role that implementation fidelity checks may play in the conduct of intervention studies, and of the unique nature of independent pharmacies, are discussed.

**Methods**

**Study Design**

We planned to conduct a pilot cluster RCT with 7 independent community pharmacies consisting of 2 parallel groups. In total,
4 pharmacies were allocated to the intervention arm, and 3 pharmacies were allocated to the control arm. The original project flow diagram is provided as Figure 1 [39,45], along with colored marks added after the study to indicate the degree of completion for that step.

Detailed implementation notes were collected from the 4 pharmacies randomly assigned to the intervention arm. We also obtained data on naloxone sales and no-cost distribution for the 3 months prior to the study period as well as for a period of 3 or more months after the intervention began (more details on this timeframe are provided subsequently). Some preintervention staff data are provided and discussed.

No data were collected from control pharmacies, no posttest data were able to be collected from pharmacy staff members, and no counts of referral cards were able to be obtained from any pharmacy.

Figure 1. Original study protocol diagram for the PharmNet pilot trial, modified with indicators of completion status (adapted from Eldridge et al [39], which is published under the terms of the Creative Commons Attribution License 4.0 [45]).

Sources of the Data

There were 2 types of participants in this study, pharmacies and pharmacy staff members.

- The 7 participating pharmacies were part of a single independent pharmacy organization. They were allocated simultaneously to a study arm using a web-based random sequence generator with a 4:3 ratio.
- These pharmacies all operate in the Midwestern United States and have a shared leadership structure (eg, there are individuals who oversee operations at all 7 pharmacies). In addition, each pharmacy has a managing pharmacist (which we call “lead pharmacists” in this paper to avoid conflating their management role with those of pharmacists overseeing the whole organization).
- Lead pharmacists manage day-to-day pharmacy operations at single sites.
- Pharmacists and pharmacy staff members at each of the 7 sites (both the intervention and control arms) were asked to participate in the study by providing individual-level data at pretest. Posttest data collection was planned but not completed (as described in the Results section).

Intervention Description

The development of this intervention was described in detail in prior papers [37,39], and a structured table outlining the sequential steps of PharmNet was included in the published protocol paper [39]. The specific steps involved in the intervention are included in Tables 1 and 2, which describe each planned step (excerpted verbatim or paraphrased from the protocol paper, as appropriate, for clarity) as well as the actual outcome.
Table 1. Planned implementation steps and observed outcomes prior to the PharmNet study start date.

<table>
<thead>
<tr>
<th>Intervention stage and planned steps</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction to the study (prior to the start date)</strong></td>
<td></td>
</tr>
<tr>
<td>1. Digital video (.mp4) describing the intervention is distributed to the pharmacy manager.</td>
<td>An email was sent to the pharmacy organization management team, and they responded with an indication that they had received the video.</td>
</tr>
<tr>
<td>2. Pharmacy manager disseminates the video.</td>
<td>The managing pharmacist reported they would show it to all 4 intervention pharmacies prior to the start date at their monthly employee meetings.</td>
</tr>
<tr>
<td>3. All staff members view the video.</td>
<td>As far as we can determine, staff members viewed the video (based on phone conversations).</td>
</tr>
<tr>
<td>4. Lead pharmacists are asked by the managing pharmacist to include the video in the orientation package for new pharmacists and pharmacy technicians who missed the initial meeting.</td>
<td>We were unable to verify whether this occurred.</td>
</tr>
<tr>
<td><strong>Early implementation (prior to the start date)</strong></td>
<td></td>
</tr>
<tr>
<td>1. Confirm with the managing pharmacist that all intervention pharmacies have viewed the video at the “all hands” staff meeting.</td>
<td>LAE spoke with the managing pharmacist on the phone on the start date to confirm this.</td>
</tr>
<tr>
<td>2. Deliver the following supplies for each participating pharmacy:</td>
<td>JA verified all materials and personally delivered them to organizational management approximately 1 month prior to the start date. On the start date, LAE verified with the managing pharmacist that all 4 intervention pharmacies had received the supplies and that the intervention pharmacies were “live” with the intervention as of November 1, 2022.</td>
</tr>
<tr>
<td>- A set of 30 doses of naloxone (individually packaged Narcan nasal spray) to each pharmacy in numerically marked bags imprinted with “Not for Sale.”</td>
<td></td>
</tr>
<tr>
<td>- Deliver a set of 500 consecutively numbered harm reduction referral slips to each pharmacy (customized for each pharmacy location).</td>
<td></td>
</tr>
<tr>
<td>- Deliver sets of reminder post-it notes to pharmacies.</td>
<td></td>
</tr>
<tr>
<td>- Deliver the requested number of yard signs to each pharmacy.</td>
<td></td>
</tr>
<tr>
<td>3. Ensure the pharmacy’s in-store television has access to the scrolling reminder text.</td>
<td>In principle, this was made available. As we note below, we cannot verify that it was used.</td>
</tr>
</tbody>
</table>
Table 2. Planned implementation steps and observed outcomes concurrent with the PharmNet study start date.

<table>
<thead>
<tr>
<th>Intervention stage and planned steps</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Launch (on the start date)</strong></td>
<td></td>
</tr>
<tr>
<td>1. Verify the intervention has launched.</td>
<td>Launch verification calls were made by LAE to the intervention pharmacies on November 2, 2022 (the day after launch).</td>
</tr>
<tr>
<td>• Reminder post-its are placed in visible areas on the interior of the dispensing location (eg, staff side rather than patient side).</td>
<td>Pharmacy 1: The pharmacist who was working that day was a floating pharmacist and transferred LAE to the managing pharmacist. The managing pharmacist reported that the pharmacy “has the supplies and is ready to go.”</td>
</tr>
<tr>
<td>• Harm reduction referral cards are placed at each checkout for easy access.</td>
<td>Pharmacy 2: The pharmacist who was working that day said they had just returned from leave and did not know where the PharmNet intervention materials were in the pharmacy. They reported they would contact the managing pharmacist to find out where the materials were. The pharmacist stated they had watched the training video and were aware of the PharmNet intervention.</td>
</tr>
<tr>
<td>• Yard signs are placed outside the building in visible locations where traffic (vehicle or foot) is common.</td>
<td>Pharmacy 3: The pharmacist who was working that day reported they were not aware of receiving the PharmNet materials. They put LAE on hold to check if the supplies were in the pharmacy. They reported to LAE that they could not locate any supplies.</td>
</tr>
<tr>
<td>• Scrolling television banner will be enabled to run continuously through the intervention, reading “76.7% of overdoses happen at home. Bystanders who witness an overdose can be effective in reducing overdose mortality. Ask us about how you can save a life with naloxone today!”</td>
<td>Pharmacy 4: The pharmacist who was working that day reported that they had watched the training video and were aware of the PharmNet intervention. They reported that the pharmacy had not received any PharmNet supplies.</td>
</tr>
<tr>
<td>2. Obtain additional details about the launch.</td>
<td>On day 4 of the intervention, LAE called the organizational management team. The primary purpose of the call was to request that they contact intervention pharmacies 2-4 to inform the pharmacists where the PharmNet materials were. LAE asked the managing pharmacist to inform the lead pharmacists at the intervention pharmacies to begin PharmNet if they had not already done so.</td>
</tr>
<tr>
<td></td>
<td>The managing pharmacist informed the researcher that they would contact the pharmacists over the next few days. Unfortunately, soon after that call, we learned that the managing pharmacist had COVID-19 and was on sick leave. In addition, at least 1 additional lead pharmacist was also on sick leave for COVID-19, and pharmacists from other sites (and when possible) the managing pharmacist, were often required to staff those roles when no additional staff members were available to do so.</td>
</tr>
</tbody>
</table>

Data

The primary data points that we report include preimplementation naloxone sale data from intervention pharmacies (August 1, 2022, through October 31, 2022), postimplementation naloxone sale data from the same pharmacies (November 1, 2022, through January 31, 2023), and doses of no-cost naloxone distributed between the receipt of the doses at the end of May 2023. These data were obtained through objective pharmacy record extraction and physical (manual) counts of remaining naloxone doses by the lead pharmacist at each intervention pharmacy.

Data from individual staff members were collected via a survey administered prior to the implementation start date for PharmNet. The questionnaire was built in Qualtrics (Qualtrics), a secure web-based survey and data collection tool. The organizational pharmacy manager was provided with a link to distribute to pharmacists and pharmacy technicians at all independent community pharmacies in the study (both the intervention and control pharmacies). Measures included sociodemographic information, questions about pharmacy practice, comfort with harm reduction practices, and belief-based measures about people who use drugs. The wording for each item is provided in Multimedia Appendix 1. Completion of the survey was specifically not required for any individual pharmacist or staff member, but they were encouraged to participate. A copy of the full survey instrument is available as an appendix to the published protocol [39].

Finally, we also originally intended to have a single in-person fidelity check for each intervention pharmacy around 1 month into the project’s launch. For the reasons we describe in the Results section, multiple additional checks were added, and the modality was switched to telephonic and email correspondence to facilitate the increased frequency.

Ethical Considerations

This study received ethics reviews prior to being initiated. Participation by pharmacies was designated as not human subjects research by the Indiana University institutional review board (document 12339). Participating pharmacy sites in the intervention arm were each offered an institutional incentive of US $1000 for their participation. Pharmacy employees completed a study information sheet prior to participating in the individual-level data collection and were compensated with a US $5 digital gift card for completing the pretest survey. Individual-level procedures were reviewed separately by the
Indiana University institutional review board and received a designation of “exempt” (document 13956). Due to the small number of individual participants, data are reported only in aggregate and without cross-sectional tabulation.

Results

Pharmacy-Level Outcomes

On February 1, 2023, the intervention was concluded. From February through the first week of April 2023, we attempted to obtain pharmacy-level outcome data via phone calls, voicemails, in-person visits, and physical notes. However, these processes were unsuccessful. Our subjective observation is that a combination of illness and staffing complexity made it extremely difficult to collect the outcome data as originally planned in the protocol.

Although they were not originally involved in data collection, we were able to collaborate with the lead pharmacists at intervention pharmacies, who had participated in fidelity checking processes and were familiar with our team, to obtain the limited data set described in the Data section in May 2023. Importantly, we do not know the exact date when each pharmacy began offering no-cost naloxone doses, and the numbers of doses of no-cost naloxone dispensed were obtained several months after the intervention technically ended. As a result, we cannot compare naloxone distribution prior to and during the intervention period. However, we can still provide raw numbers (Table 3).

Table 3. Pharmacy-level naloxone sale and dispensing data, August 2022 through May 2023.

<table>
<thead>
<tr>
<th>Pharmacy</th>
<th>Preimplementation naloxone sale data (August 1, 2022-October 31, 2022)</th>
<th>Implementation naloxone sale data (November 1, 2022-January 31, 2023)</th>
<th>Doses of no-cost naloxone dispensed (of 30) as of May 2023</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmacy 1</td>
<td>19</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>Pharmacy 2</td>
<td>45</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>Pharmacy 3</td>
<td>9</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>Pharmacy 4</td>
<td>11</td>
<td>10</td>
<td>24</td>
</tr>
</tbody>
</table>

Pharmacy Staff Preintervention Data

A total of 8 staff members at least partly completed the study pretest during the period prior to the intervention, including 3 pharmacists and 5 pharmacy technicians. Pharmacy 4 was well-represented with 5 respondents, and there were 2 from pharmacy 3 and 1 from pharmacy 1. No staff members from pharmacy 2 or the control pharmacies participated.

All but 1 respondent who provided demographics (n=6) were female, and all were White, non-Hispanic, and heterosexual. Staff ages ranged from 32 to 52 years (n=5). Staff members who responded were comfortable with most harm reduction approaches, but fewer were comfortable with consulting with patients about pre-exposure prophylaxis for HIV prevention or dispensing syringes for nonprescription drug use. No participants indicated they were uncomfortable dispensing naloxone for overdose reversal or making referrals to community services.

Further, staff member attitudes toward patients who use drugs or have OUD were not typically stigmatizing, though they expected that others (“most people”) might blame persons with OUD or believe that they are dangerous. A table of these data is provided in Multimedia Appendix 1.

Fidelity Data

We also collected a full written accounting of the 9 distinct fidelity checkpoints during the intervention period. This information helped to contextualize some of the implementation issues around which we provide interpretation and discussion in the next section. Although not all study processes were in place at the time of the final fidelity check at week 10 of the 12-week intervention, by the time we collected outcome data, the intervention was apparently functioning at all sites (even though the technical “intervention period” was over). All documentation is available in Tables 4 and 5.
Table 4. Implementation fidelity checkpoint documentation for the PharmNet study through day 22.

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Outcome</th>
</tr>
</thead>
</table>
| Day 10—Calls to all intervention pharmacies  | • Pharmacy 1: The pharmacist who answered the phone reported that they were not the usual pharmacist and only work 1 day a week, but that the pharmacy did have the yard signs and the no-cost naloxone. However, they reported that the yard signs were not placed outside the store. They also indicated that the managing pharmacist was on leave and that they recommend calling when they returned the following week to follow-up about the yard signs.  
  • Pharmacy 2: The pharmacist reported they were not the usual pharmacist and only work at the pharmacy 2 times a month, and that the regular pharmacist who works at that store would be back the following week. They did not know about PharmNet nor about the no-cost naloxone, yard signs, or harm reduction referral sheets.  
  • Pharmacy 3: LAE spoke with the same pharmacist as previously (the startup verification). The pharmacist reported that they did not think they had received PharmNet supplies, but that they were ready to start the intervention as soon as they received the supplies.  
  • Pharmacy 4: We spoke with the same pharmacist as previously (the startup verification). The pharmacist reported that they did not think they had received PharmNet supplies.                                                                                                                                                                                                                      |
| Day 14—Attempt to coordinate with the pharmacies | • LAE attempted to reach the managing pharmacist at 2 separate pharmacies. A pharmacy technician called an additional 5 pharmacies but was unable to locate them. We left a voicemail at their office. We later learned that they were on leave during this period.                                                                                                                                                                                                                                           |
| Days 16, 17, and 21—Follow-up calls, after which we learned about illnesses or leaves among the pharmacists | • We called the original pilot pharmacy attempting to contact the leadership team. LAE left a voicemail and sent them an email on each of the 3 days.                                                                                                                                                                                                                                                                                         |
| Day 22—Calls to all intervention pharmacies  | • LAE learned that the managing pharmacist had returned to the office, but they were in a meeting when we called. A pharmacy technician took a message. Calls were also made to each of the pharmacies in the intervention arm.  
  • Pharmacy 1: When LAE called and requested to speak to a pharmacist, the pharmacy technician transferred her to the original pilot pharmacy location (which was not part of the study). LAE spoke with a pharmacist at this location, and this pharmacist reported they were unaware of the status of PharmNet at pharmacy 1 as they never worked at that site. They transferred the researcher back to pharmacy 1, where the pharmacist technician forwarded the researcher to voicemail. LAE left a message.  
  • Pharmacy 2: We spoke with a pharmacist who confirmed that the supplies had been received and that the intervention had begun.  
  • Pharmacy 3: The pharmacist reported that they still were not aware of having received the PharmNet supplies.  
  • Pharmacy 4: The pharmacist reported that they still had not received the PharmNet supplies. The pharmacist reported they had reached out to their leadership about the supplies and were informed that their supplies were located at pharmacy 1 and would be delivered after Thanksgiving.                                                                                                                                                                                             |
Barriers Specific to Pharmacy Naloxone Distribution

The research literature has identified several barriers that can occur when implementing pharmacy-based OEND programs, but interestingly, this study of the PharmNet intervention never encountered most of these barriers (though it encountered others). In some cases, this may have been due to the intervention’s design, as we spent considerable time in the development phase determining how we might obviate frequently discussed implementation issues. For example, multiple publications have identified that reimbursement for products and services can be an implementation barrier (eg, for the costs of naloxone, pharmacist time, or other program features) [38,47]. In certain program structures (eg, where free or reduced-cost naloxone doses are unavailable), patients’ out-of-pocket costs to purchase naloxone can also be a concern [48]. However, PharmNet included a planning step where the research team negotiated with nonprofit organizations to procure sources of no-cost naloxone. This meant that pharmacies could continue selling naloxone where appropriate, but that they also had the ability to provide free doses when indicated (eg, to patients who could not afford it) without needing to find a source of reimbursement.

Discussion

Context of the Report

PharmNet was developed with independent community pharmacies in mind as a core implementation venue. This was partly because of trends around naloxone stocking and dispensing, as described in the Background section, which suggested that independent pharmacies might especially benefit from intervention support. In addition, there is a large segment of the US population, typically older, rural, and with lower income levels, for whom independent pharmacies are the primary point of pharmacy access (recent estimates place the number of individuals at over 15 million) [46]. Thus, it is particularly important to know whether interventions can operate effectively in independent pharmacy venues.

As we have previously indicated, our single-site pilot, which preceded this study, was successful [37], whereas this study encountered multiple barriers, even though they were conducted within the same small, independent pharmacy group. Our subjective qualitative assessment is that these were meta-structural barriers—perhaps common to independent pharmacies or related to the implementation decisions—and not a function of the individuals in the independent pharmacy group. Further, many of the barriers we encountered did not appear to be especially related to the program’s focal area (ie, naloxone distribution).
Similarly, while pharmacy-based counseling is sometimes included in pharmacy OEND interventions, we were concerned about its feasibility in low-resource pharmacies, especially after COVID-19. Thus, for PharmNet, we opted to provide referral and resource cards in prescription bags that were tailored to each unique geographic location. In doing so, we attempted to avoid common barriers reported by similar studies such as lack of time or space for counseling [48]. Another larger study in 2 pharmacy chains also reported adaptations resulting from COVID-19, such as “streamlined counseling and drive-thru provision adaptations ...” [49].

Further, while a study of opioid overdose prevention programs in pharmacies identified barriers related to pharmacy leadership support [48], our perception is that the leadership team and pharmacy management in place at this independent pharmacy group were incredibly motivated, competent, and willing to support this intervention to the best of their ability—and it appeared that they were primarily limited by factors of raw capacity. In addition, though our pharmacist and pharmacy technician data were not collected from all intervention pharmacies, the limited data set that we procured did not identify any substantive concerns around comfort with dispensing naloxone.

Barriers Related to Independent Pharmacies as Scalable Intervention Sites

We intentionally designed this pilot trial to be pragmatic, meaning that the implementation environment would mirror real-world practice [39]. Our primary concern was that the resources provided as part of an intensive trial support system would introduce a kind of heterogeneity of treatment effect [50], where pharmacies working to conduct the intervention outside of the context of the RCT would be systematically different by virtue of not having the same startup support. This would potentially inflate the effect size of the intervention within the study relative to its true effectiveness. Therefore, for individuals working in single pharmacies (eg, lead pharmacists, pharmacy technicians, and other staff members), we did not plan any direct face-to-face interaction with our team. Instead, the intervention materials themselves—introductory videos, instructions to the overall leadership of the organization, and materials, along with a planned fidelity check—were the primary means of dissemination.

However, the pharmacy practice environment is also experiencing a variety of substantive external pressures. For example, a 2022 survey from the National Community Pharmacists Association found nearly one-quarter of responding pharmacy owners or managers had difficulty filling pharmacist roles, and 88.8% had similar difficulty filling pharmacy technician positions [51]. A 2019 survey of full-time pharmacists found that even though independent pharmacists had higher job satisfaction, organizational commitment, and perceived control in the work environment than chain, supermarket, or mass-merchandiser pharmacies, those numbers had still declined over the past 5 years [52]. Strains on the broader pharmacy working environment have recently appeared in media reports as well [53]. Understanding these pressures and accounting for them with pragmatic intervention design will likely be important to sustainable implementation success.

We suspect that these or similar external stressors played a meaningful role in the outcomes of this trial. Lessons from remote decentralized clinical trials, for example, highlight how reducing the research burden on some elements of a team or partnership may inadvertently shift the burden onto others working in the field [54]. In other words, eventually, someone must do the work, and when staffing levels are low and exogenous stressors (eg, a COVID-19 outbreak) occur, an unreasonably large volume of responsibility may accrue to a small number of people. This is a particular concern due to the prevalence of pharmacist burnout, which a systematic review recently estimated to exceed 50% [55], though a slightly older review reported a lower mean percentage [56] (both reviews indicated heterogeneity among included studies). In this case, while we do not have (nor are intending to imply) any direct evidence of burnout, per se, we noted that there was not a “deep bench” of staff and that when illnesses or similar issues occurred in these independent pharmacies, the work was often covered by others who already had full-time (or greater) workloads.

In contrast, in one well-conducted, harm reduction-focused, pragmatic RCT of 175 community chain pharmacies, the researchers encountered both COVID-19 and wildfires [57]. Their study team conducted the majority of their fidelity checks face-to-face and had regular phone contact with the pharmacies. Their published protocol states the research team maintained a “close collaboration and regular meetings with the pharmacy corporate leadership.” Unexpected circumstances that arose were sometimes able to be mitigated by pharmacy corporate leadership (eg, the provision of additional pharmacists to relieve pharmacists from their regular work duties so that the onsite “academic detailing” could be conducted). These types of resources and supports may not be accessible to most independent pharmacies.

Do Fidelity Checks Contribute to Intervention Effectiveness?

The major differences between our single-site pilot study (which was completed successfully and which produced positive outcome data) [37] and this follow-up study remain of interest to our team. One of our working theories relates to the nature of implementation fidelity checking that occurred. Implementation fidelity checking has long been understood as an important component of implementation science [58], though community pharmacy intervention studies do not always report implementation fidelity tools or processes [59].

Our single-site pilot study incorporated face-to-face fidelity checks on a regular basis [37]. This involved a study investigator regularly visiting the pharmacy in person, speaking with staff members and pharmacists, obtaining interim information, and identifying necessary points of action. In designing this follow-up study, we did not consider the effects that regular in-person fidelity checking might have had on intervention effectiveness in our single-site pilot. In other words, we do not know how much of the implementation effectiveness in the original study was a direct or indirect result of the routine physical presence of a researcher in the pharmacy. It is not...
necessarily the case that a single-site pilot is an insufficient first step to prepare for a larger-scale study, but rather, it seems that researchers might benefit from carefully examining the ways in which intensive fidelity procedures may actively contribute to successful implementation rather than merely documenting what happened.

Limitations

As noted throughout this paper, there were numerous limitations that resulted in our reporting of fidelity encounter data and limited outcome measures. Even though this study was preregistered, due to the differences in the data collected from what was planned, we could not test the planned hypotheses. This study should therefore be considered exploratory. In addition, while we carefully consider the context of the study in our Discussion section, the ideas presented therein are hypotheses only and should not be interpreted as definitive statements of fact.

Conclusions

Independent pharmacies, especially in rural areas, will likely continue to be key parts of US health care teams. For example, in April 2023, the Express Scripts business of Evernorth, a subsidiary of The Cigna Group, announced that they are extending their efforts to expand health care in rural communities via independent pharmacies [60]. As the prevalence of fatal opioid overdose remains high, multiple avenues to prevent fatality are warranted. Even though nasal formulations of naloxone are now available over the counter, pharmacies have a role in creating awareness and fostering accessibility for patients for whom the cost is prohibitive. As one peer reviewer (Steven H Linder MD) aptly noted, “the optimal means of naloxone community distribution remains to be determined.”

In terms of broader program implementation considerations for pharmacies, it may be the case—especially for independent pharmacies—that the start of a new initiative needs to be accompanied by more substantive support in the current pharmacy practice environment. If a trial is conducted over a longer period, support could in theory be withdrawn, and effectiveness measured longitudinally. However, we could not, with the data obtained from this study, disentangle issues around fidelity checking and externalities from those related to the overall pharmacist workload and burden and the intervention mechanics.

The strength of this paper is its detailed process description and transparent disclosure of the issues that affected the intervention. Given the likely importance of independent pharmacies to the national overdose response effort, we encourage further implementation science research that is specific to independent community pharmacies, which are often collapsed with other types of pharmacies but that have unique facilitators and barriers to intervention uptake.

Acknowledgments

This paper was funded by the Indiana University Grand Challenge: Responding to the Addictions Crisis. The funder did not have a role in the preparation of this paper, and the opinions and information expressed herein do not necessarily reflect the views or opinions of the funder.

Data Availability

All data generated or analyzed during this study are included in this published paper (Multimedia Appendix 1), except for the raw data for the small sample (N=8) of staff members who completed the pretest, which are reported in aggregate but not at the individual level due to privacy concerns.

Authors' Contributions

All authors conceptualized the study, conducted the investigation, worked on the methodology, administered the project, contributed resources toward the project, and validated the information presented herein. LAE and JA provided data curation and wrote the original draft. BEM and JA obtained funding for the study. JA managed visualization. All authors reviewed the final paper.

Conflicts of Interest

The authors have no conflicts of interest related to the content of this paper but note that they participate in ongoing research (as cited herein) related to this project and similar projects through grants awarded to their institutions and have served as grant peer reviewers on harm reduction funding streams for federal agencies and foundations.

Multimedia Appendix 1

Staff-level preintervention data.

[DOCX File , 16 KB - formative_v8i1e54077_app1.docx ]

References


45. Attribution 4.0 International (CC BY 4.0 Deed). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-02-23]


Abbreviations

MOUD: medications for opioid use disorder
OEND: opioid education and naloxone distribution
OUD: opioid use disorder
RCT: randomized controlled trial
Machine Learning–Based Prediction of Changes in the Clinical Condition of Patients With Complex Chronic Diseases: 2-Phase Pilot Prospective Single-Center Observational Study

Celia Alvarez-Romero¹, MSc; Alejandro Polo-Molina², MSc; Eugenio Francisco Sánchez-Úbeda², PhD; Carlos Jimenez-De-Juan³, MSc; Maria Pastora Cuadri-Benítez³, MSc; Jose Antonio Rivas-Gonzalez¹, BSc; Jose Portela³, PhD; Rafael Palacios², PhD; Carlos Rodriguez-Morcillo², PhD; Antonio Muñoz², PhD; Carlos Luis Parra-Calderon¹, MSc; Maria Dolores Nieto-Martin³, PhD; Manuel Oller-Baturone³, PhD; Carlos Hernández-Quiles³, PhD

¹Computational Health Informatics Group, Institute of Biomedicine of Seville, Virgen del Rocio University Hospital, Consejo Superior de Investigaciones Científicas, University of, Seville, Spain
²Institute for Research in Technology (IIT), ICAI School of Engineering, Comillas Pontifical University, Madrid, Spain
³Internal Medicine Department, Virgen del Rocio University Hospital, Sevilla, Spain

Corresponding Author:
Carlos Hernández-Quiles, PhD
Internal Medicine Department
Virgen del Rocio University Hospital
Av Manuel Siurot s/n
Sevilla, 41013
Spain
Phone: 34 697950012
Email: quiles_es@yahoo.es

Abstract

Background: Functional impairment is one of the most decisive prognostic factors in patients with complex chronic diseases. A more significant functional impairment indicates that the disease is progressing, which requires implementing diagnostic and therapeutic actions that stop the exacerbation of the disease.

Objective: This study aimed to predict alterations in the clinical condition of patients with complex chronic diseases by predicting the Barthel Index (BI), to assess their clinical and functional status using an artificial intelligence model and data collected through an internet of things mobility device.

Methods: A 2-phase pilot prospective single-center observational study was designed. During both phases, patients were recruited, and a wearable activity tracker was allocated to gather physical activity data. Patients were categorized into class A (BI ≤ 20; total dependence), class B (20<BI ≤ 60; severe dependence), and class C (BI>60; moderate or mild dependence, or independent). Data preprocessing and machine learning techniques were used to analyze mobility data. A decision tree was used to achieve a robust and interpretable model. To assess the quality of the predictions, several metrics including the mean absolute error, median absolute error, and root mean squared error were considered. Statistical analysis was performed using SPSS and Python for the machine learning modeling.

Results: Overall, 90 patients with complex chronic diseases were included: 50 during phase 1 (class A: n=10; class B: n=20; and class C: n=20) and 40 during phase 2 (class B: n=20 and class C: n=20). Most patients (n=85, 94%) had a caregiver. The mean value of the BI was 58.31 (SD 24.5). Concerning mobility aids, 60% (n=52) of patients required no aids, whereas the others required walkers (n=18, 20%), wheelchairs (n=15, 17%), canes (n=4, 7%), and crutches (n=1, 1%). Regarding clinical complexity, 85% (n=76) met patient with polyopathy criteria with a mean of 2.7 (SD 1.25) categories, 69% (n=61) met the frailty criteria, and 21% (n=19) met the patients with complex chronic diseases criteria. The most characteristic symptoms were dyspnea (n=73, 82%), chronic pain (n=63, 70%), asthenia (n=62, 68%), and anxiety (n=41, 46%). Polypharmacy was presented in 87% (n=78) of patients. The most important variables for predicting the BI were identified as the maximum step count during evening and morning periods and the absence of a mobility device. The model exhibited consistency in the median prediction error with a median absolute error close to 5 in the training, validation, and production-like test sets. The model accuracy for identifying the BI class was 91%, 88%, and 90% in the training, validation, and test sets, respectively.
Conclusions: Using commercially available mobility recording devices makes it possible to identify different mobility patterns and relate them to functional capacity in patients with polyopathy according to the BI without using clinical parameters.

(JMIR Form Res 2024;8:e52344) doi:10.2196/52344

KEYWORDS
patients with complex chronic diseases; functional impairment; Barthel Index; artificial intelligence; machine learning; prediction model; pilot study; chronic patients; chronic; development study; prognostic; diagnostic; therapeutic; wearable; wearables; wearable activity tracker; mobility device; device; physical activity; caregiver

Introduction
The Spanish strategy for the approach to chronicity in the National Health System defines patients with complex chronic diseases as patients with 1 or more chronic diseases that present greater complexity in their management due to changing needs that force continuous evaluations and make necessary the coordinated use of various care levels and, in some cases, health and social [1]. Social changes and health advances mean that we are living longer and better and that most diseases affecting us are becoming chronic. Several of them are accumulating, which causes the growing phenomenon of people living with polyopathy or complex chronic diseases. This concept includes not only people with the primary disease that triggers other secondary conditions but also those people where 2 or more chronic diseases coexist. It is a population characterized by frailty, polymedication, old age, hyperfrequent use of emergency services, and frequent re-entering. It is estimated that 70% to 95% of the older people in our environment have 1.2 to 4.2 chronic diseases, which constitute the leading death cause in the world (60% of the total) [2]. These patients generate a greater demand for attention in different care settings and use a more significant number of health and social resources. It is predominantly seen in older patients presenting with limiting and progressive diseases (eg, renal or cardiac insufficiency), polypharmacy, and some degree of functional impairment [3].

Functional impairment is one of the most decisive prognostic factors in patients with complex chronic diseases. A more significant functional impairment indicates that the disease is progressing, which requires implementing diagnostic and therapeutic actions that stop the exacerbation of the disease. The functional assessment of patients with complex chronic diseases can be performed using tools such as the Barthel Index (BI) [4], mobility tests, the 4-meter gait test [5], the balance test [6], and the timed “up and go” test [7].

The BI has excellent predictive value for variables such as mortality, hospital admission, and stay length in rehabilitation departments. In addition, it is an indicator to assess the functional and prognostic capacities of patients with complex chronic diseases [8,9]. The BI is a simple measure developed on empirical bases in obtaining and interpreting it. It is about assigning, to each patient, a score based on their degree of dependence to perform a series of basic activities related mainly to the individual’s mobility (eg, moving between the chair and the bed, moving, going up and down stairs, or showering). The total score can vary between 0 (fully dependent) and 100 points (completely independent) [10].

Concerning functional capacity, physical inactivity is defined as the spectrum of any decrease in body movement that reduces energy expenditure toward the baseline level. Physical inactivity affects many aspects of a person, such as respiratory capacity, bones, or the central nervous system, among others, and can even lead to various diseases [11]. In addition, physical inactivity itself decreases the physical fitness of the person, the duration of good health, and the age of onset of his or her first chronic illness. Relative to this, there are several parameters to assess the physical inactivity of the person, such as the number of daily steps, the time spent sitting, or the immobilization of the limbs, among others.

On the other hand, the possible causal relationship between sedentary behavior and mortality due to various causes has been studied. Various studies used accelerometers on the thigh to control the body’s position, and the chances of experiencing illnesses increased for every additional hour of sitting. Regarding limb immobilization in older people, one of the main concerns is the inability to recover the loss of bone strength and muscle mass [12].

Recent technological advances allow mobility monitoring through smartphones or wrist devices, which are widely distributed throughout the population. These devices provide information on the paths, the number of steps, the speed of the march, and the periods of falls, among others. Specific initiatives have tried to apply this information to the health sector. For instance, a multiagent system equipped with sensors has been developed to collect vital signs from patients. This system is intended to facilitate various tasks within the residences of older or disabled people [13]. Additionally, mobility monitoring by sensors in different rooms of the house has been considered to study translations between rooms and measure the length of stay in each room for older patients living alone [14].

Furthermore, individual physical activity can be monitored using accelerometers placed on the patient’s trunk and thigh [15]. At the same time, smartwatches have been used to evaluate movement and gait patterns in patients with Parkinson disease and essential tremor [16]. These advancements are driving the development of more hardware devices to enhance health care delivery and turn the concept of “a doctor in your pocket” into a reality for patients.

We would like to emphasize that using sensors to obtain health information currently has a specific trajectory [17-19]. Mobility has long presented prominent importance when dealing with diseases whose onset and symptomatic progression affect the functional capacity of the subject [20]. Recently, machine learning (ML) techniques are increasingly being considered to
characterize the movement, or some particularities of the movement, which can provide relevant information about the patient’s clinical status [21,22]. Some works investigate the relationships between movement and specific clinical pathologies [23,24].

Despite these initiatives, the evaluation of the mobility of patients with complex chronic diseases and their relationship with the functional capacity measured by the BI has yet to be explored [25]. For all these reasons, and with this background, this study aims to develop and validate mobility patterns based on artificial intelligence and the internet of things (IoT) environment, aiming to predict changes in the clinical condition of patients with complex chronic diseases through the prediction of the BI to know the clinical and functional status of the patients.

Methods

Study Design and Recruitment

This 2-phase pilot observational study has been designed to analyze how mobility deterioration can reflect changes in the patient’s clinical condition and possible degeneration in the integrated care of patients with complex chronic diseases. To this end, a prospective, single-center, descriptive study was carried out.

Eligible patients met the criteria of chronic patients with complex health needs defined according to the Integrated Patient Care Process of the Andalusian Ministry of Health [26]. Concretely, the study population included patients older than 65 years of age with multimorbidity (ie, diagnosed with at least 2 chronic diseases), and the recruitment took place at the Virgen del Rocio University Hospital of Seville, Spain. In addition, those patients in a situation of agony or those whose vital prognosis was limited, patients with psychiatric disease, and patients or caregivers unable to use mobility devices were excluded from the study. The study subjects were patients of the Internal Medicine Department of the Virgen del Rocio University Hospital of Seville, as part of the Andalusian Health Service, Spain.

The research was conducted in 2 phases. In the initial phase (January to November 2022), a cohort of 50 patients was enrolled, and their BI was measured before the allocation of the wearable activity trackers (WATs), during routine doctor appointments after providing informed consent. Approximately 1 month after the first assessment (encounter 1), the BI was measured again (encounter 2) to evaluate any changes in their functional status (Figure 1). The recruitment was conducted according to different degrees of patients with complex chronic diseases dependence based on the BI measured during the first assessment. In particular, the enrolled patient’s group was classified into 3 groups based on their BI scores: class A included patients with BI ≤20 (total dependence), class B comprised patients with 20<BI ≤60 (severe dependence), and class C consisted of patients with BI>60 (moderate or mild dependence, or independent) [27].

In the second phase (July 2022 to May 2023), 40 patients were recruited. Similar to phase 1, patients were recruited during doctor appointments, and their BI was measured 3 months after encounter 1 and encounter 2. Therefore, for phase 2 patients, there was an additional encounter 3.

An IoT framework was deployed to gather patient mobility data after analyzing the existing devices and applications in the market. The IoT-based infrastructure consisted of using mobile devices and WAT to measure the mobility activities of patients, considering the no or minimal invasion in the development of the daily tasks for the patients under study. The WAT used in this study recorded the step count, the cardiac activity, and the sleep duration from which both the step count and the heart rate were analyzed (Figure 2).
Once the 90 patients were included in the study, they were assigned a WAT, and different mobility and functional status tests were conducted. An information system for data storage (Analytics Datastore) was developed. This database allowed both the dumping of the information collected through the WAT and the storage of the relevant clinical information of the patients extracted from the electronic health records (Figure 3). For this purpose, confidentiality protocols of information and the security of the center’s systems were followed and in compliance with the ethical approval obtained by the hospital’s ethics committee.

The study patients’ exposure and clinical variables of interest were analyzed to characterize patient groups regarding mobility, using mobility measurement devices and clinical conditions. Demographic and clinical variables such as diseases, fragility, and polypathology criteria; pharmacological variables; and functional tests such as the BI, balance test, and timed “up and go” test were collected. Statistical analysis was performed using SPSS Statistics software (version 25; IBM Corp) and Python (version 3.10.9; Python Software Foundation) for the ML modeling.

**Data Preparation for the ML Model**

The WAT automatically gathered continuous and noninvasive data on a range of parameters, encompassing heart rates, step counts, and sleep duration. However, these raw data must be processed to apply ML techniques. Furthermore, given the potential influence of the walking aids on mobility patterns, patients were classified into 3 distinct groups. The first group encompassed patients using wheelchairs; the second comprised individuals using canes, walkers, or receiving aid from a caregiver during ambulation; and the third consisted of those with no reliance on assistance.

To ensure high data quality, instances where the median heart rate is missing are identified as null, along with the corresponding count of steps. The steps taken within 1-hour intervals are aggregated, and the median heart rate for these hourly intervals is computed. The resulting time series data of hourly step counts are then smoothed by applying a centered rolling window with a window size of 3. Subsequently, the data are grouped by the specific hour, resulting in an average representation of each patient’s activity throughout a 24-hour period. The data used to generate the mean activity profile consist of the information recorded during the 30-day period.
before encounter 2 requiring at least 14 days’ worth of data to consider the patient in the data set. This approach aims to develop a methodology that allows the estimation of the BI at any given moment using the information collected by the WAT over the last 30 days. Therefore, it holds the potential to provide a more dynamic and real-time assessment of the BI based on continuous monitoring through WAT. Additionally, mobility profiles from encounter 3 served as a production-like test set and were excluded from the model’s training. This strategy evaluated the model’s real-world performance and generalization on unseen data.

The 24-hour mean activity profiles were partitioned into 4-time segments: morning (7 AM-1 PM), afternoon (2 PM-7 PM), evening (8 PM-11 PM), and overnight (midnight-6 AM). This methodology aimed to reduce the dimensionality of the data inputted into the model. Various approaches were considered to reduce dimensionality, including summing the steps within each interval, calculating the mean, and determining the maximum value. The 4-time segments were selected based on the findings of Polo-Molina et al [25], where it was demonstrated that mobility patterns can be categorized into distinct clusters. The study highlighted that the maximum value of steps within each interval aligns with the suggested division, regardless of individual variations in the mobility patterns.

Once the data set was generated, it was divided into training and validation sets, with 70% (n=63) of the records allocated for training and 30% (n=27) for validation. To ensure that the proportions of each group of walking aids were maintained at this ratio, the division was performed within each group, and then the data were combined to create the final training and validation data sets. This approach aimed to ensure representative and well-balanced distributions of walking aids in both sets, allowing for robust evaluation of the model’s performance across different modes of mobility.

**Explainable ML Model**

A decision tree regressor has been considered to predict the BI. Decision trees iteratively select variables to maximize information gain or minimize impurity at each decision node, creating a hierarchical structure. Therefore, starting from the whole set of variables, at each split, the training algorithm selects the variable that generates the best split [28].

Moreover, to optimize the performance of the regression model, the hyperparameters were fine-tuned using a cross-validation approach with 7 folds. This technique ensures robustness and selects the optimal settings that yield the best predictive accuracy for the BI. The cross-validation optimization considered the hyperparameters “min_impurity_decrease” (ranging from 0.0 to 1.0 in increments of 0.01), “min_samples_leaf” (from 1 to 10 in steps of 1), and “min_samples_split” (spanning 1 to 10 with an interval of 1).

In addition, to assess the quality of the predictions, several metrics were considered, including the mean absolute error, median absolute error (MAD), and root mean squared error.

Finally, the permutation importance from explanatory variables was computed by permuting individual feature values while measuring the subsequent decline in model performance [28]. This iterative process assigned a score to each feature based on the decrease in predictive power caused by permutations, with elevated scores indicating significant contributions to accurate predictions.

**Ethical Considerations**

First, ethical approval was obtained in the health organization based on the regional regulations before involving it in the study execution. Likewise, informed consent procedures were defined, including informed consent and information sheets for the patients who were included in the study. Before starting this study, and based on the ethical and legal regulations, ethical approval was requested from the Ethics Committee of the Virgen del Rocio University Hospital of Seville, Spain. The study protocol, informed consent documents, and information sheets were submitted, and approval from the ethics committee was received. The study began, and patients who met the inclusion criteria were invited to participate after explaining the study procedures. Those who accepted and signed the informed consent and information sheets were included in the clinical study.

In addition, to ensure the protection of the privacy and confidentiality of the study participants, sensible data were anonymized and deidentified. Likewise, confidentiality protocols of information and the security of the center’s systems were followed and in compliance with the ethical approval obtained by the hospital’s ethics committee.

**Results**

**Statistical Analysis**

A total of 90 patients were included in the study and were classified into 3 categories according to their BI. Concretely, 50 patients were enrolled in the first phase (10 in the BI class A, 20 in BI class B, and 20 in BI class C), and in the second phase, 40 patients were included (20 patients in BI class B and 20 patients in BI class C).

Of the patients, 94% (n=84) had a caregiver, of which 40% (n=34) had a son or a daughter, 32% (n=27) had a spouse, 17% (n=14) had other relatives, and 11% (n=9) had a professional caregiver. The mean value of functional capacity measured by the BI was 58.31 (SD 24.5). Concerning mobility aids, 58% (n=52) of patients did not require it, 20% (n=18) required a walker, 17% (n=15) a wheelchair, 4% (n=4) required a cane, and 1% (n=1) required crutches. The clinical complexity was high with 76 (85%) patients meeting the criteria for patients with polyopathy, with a mean of 2.7 (SD 1.25), and 19 (21%) patients met the criteria for patients with complex chronic diseases. A total of 61 (69%) patients met the frailty criteria.

The most characteristic symptoms of this population were dyspnea (n=73, 82%), with 47% (n=42) of patients requiring home oxygen therapy; chronic pain (n=63, 70%); asthenia (n=61, 68%); and anxiety (n=41, 46%; Table 1). The mean number of drugs taken chronically was 12.19 (SD 11.88), with 87% (n=78) meeting the polypharmacy criteria and 70% (n=63) meeting the extreme polypharmacy criteria. Psychotropic drugs were the most consumed pharmacological group (n=29, 33%). Five patients died during the study.
Table 1. Description of the baseline characteristics of the study in line with the mobility classification.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mobility classification</th>
<th>20 BI≤60 (severe dependence)</th>
<th>20 BI&lt;60 (total dependence)</th>
<th>P value</th>
<th>20 BI&gt;60 (moderate or mild dependence, or independent)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase 1 (n=20)</td>
<td>Phase 2 (n=20)</td>
<td></td>
<td></td>
<td>Phase 1 (n=20)</td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>72.56 (14.5)</td>
<td>79.6 (8)</td>
<td>75.35 (8.9)</td>
<td>.49</td>
<td>78.65 (6.5)</td>
<td>.90</td>
</tr>
<tr>
<td>Gender (women), n (%)</td>
<td>8 (80)</td>
<td>11 (55)</td>
<td>9 (45)</td>
<td>.40</td>
<td>8 (40)</td>
<td>.40</td>
</tr>
<tr>
<td>Caregiver, n (%)</td>
<td>10 (100)</td>
<td>20 (100)</td>
<td>20 (100)</td>
<td>.12</td>
<td>16 (80)</td>
<td>.12</td>
</tr>
<tr>
<td>BI, mean (SD)</td>
<td>13.89 (4.8)</td>
<td>42.25 (10.12)</td>
<td>50.25 (10.44)</td>
<td>.89</td>
<td>80 (12.1)</td>
<td>.89</td>
</tr>
<tr>
<td>Balance test, mean (SD)</td>
<td>0 (0)</td>
<td>1.4 (2.1)</td>
<td>1.6 (1.46)</td>
<td>.74</td>
<td>4.6 (1.39)</td>
<td>.74</td>
</tr>
<tr>
<td>30-second chair stand and go, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 points</td>
<td>10 (100)</td>
<td>16 (80)</td>
<td>16 (80)</td>
<td>.99</td>
<td>16 (80)</td>
<td>.99</td>
</tr>
<tr>
<td>1 point</td>
<td>0 (0)</td>
<td>4 (20)</td>
<td>4 (20)</td>
<td></td>
<td>4 (20)</td>
<td></td>
</tr>
<tr>
<td>Timed “up and go” test, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.47</td>
<td></td>
<td>.06</td>
</tr>
<tr>
<td>No frailty</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (5)</td>
<td></td>
<td>15 (75)</td>
<td></td>
</tr>
<tr>
<td>HRF(^b)</td>
<td>8 (80)</td>
<td>11 (55)</td>
<td>10 (50)</td>
<td></td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>2 (20)</td>
<td>9 (45)</td>
<td>9 (45)</td>
<td></td>
<td>5 (25)</td>
<td></td>
</tr>
<tr>
<td>Frailty criteria, n (%)</td>
<td>7 (70)</td>
<td>14 (70)</td>
<td>16 (80)</td>
<td>.46</td>
<td>12 (60)</td>
<td>.74</td>
</tr>
<tr>
<td>PP(^c) criteria, n (%)</td>
<td>10 (100)</td>
<td>19 (95)</td>
<td>18 (90)</td>
<td>.32</td>
<td>14 (70)</td>
<td>.64</td>
</tr>
<tr>
<td>Patient with polypathology categories, mean (SD)</td>
<td>3.3 (1)</td>
<td>2.89 (1.15)</td>
<td>3 (1.48)</td>
<td>.23</td>
<td>2.3 (1.08)</td>
<td></td>
</tr>
<tr>
<td>PCCD(^d), n (%)</td>
<td>4 (40)</td>
<td>4 (20)</td>
<td>2 (10)</td>
<td>.08</td>
<td>17 (85)</td>
<td></td>
</tr>
<tr>
<td>Dyspnea, n (%)</td>
<td>5 (50)</td>
<td>17 (85)</td>
<td>18 (90)</td>
<td>.78</td>
<td>18 (90)</td>
<td>.33</td>
</tr>
<tr>
<td>Home oxygen, n (%)</td>
<td>4 (40)</td>
<td>7 (35)</td>
<td>7 (35)</td>
<td>.45</td>
<td>9 (45)</td>
<td>.26</td>
</tr>
<tr>
<td>Chronic pain, n (%)</td>
<td>8 (80)</td>
<td>9 (45)</td>
<td>3 (15)</td>
<td>.001</td>
<td>11 (55)</td>
<td>.2</td>
</tr>
<tr>
<td>Pressure ulcer, n (%)</td>
<td>6 (60)</td>
<td>2 (10)</td>
<td>1 (5)</td>
<td>.54</td>
<td>0 (0)</td>
<td>.15</td>
</tr>
<tr>
<td>Insomnia, n (%)</td>
<td>5 (50)</td>
<td>10 (50)</td>
<td>7 (35)</td>
<td>.17</td>
<td>10 (50)</td>
<td>.37</td>
</tr>
<tr>
<td>Anxiety, n (%)</td>
<td>8 (80)</td>
<td>9 (45)</td>
<td>10 (50)</td>
<td>.66</td>
<td>7 (35)</td>
<td>&lt;.99</td>
</tr>
<tr>
<td>Asthenia, n (%)</td>
<td>8 (80)</td>
<td>14 (70)</td>
<td>15 (75)</td>
<td>.49</td>
<td>12 (60)</td>
<td>&lt;.99</td>
</tr>
<tr>
<td>Anorexia, n (%)</td>
<td>40 (40)</td>
<td>8 (40)</td>
<td>6 (30)</td>
<td>.21</td>
<td>3 (15)</td>
<td>.12</td>
</tr>
<tr>
<td>Nausea and vomiting, n (%)</td>
<td>3 (30)</td>
<td>5 (25)</td>
<td>2 (10)</td>
<td>.12</td>
<td>2 (10)</td>
<td>.29</td>
</tr>
<tr>
<td>Diarrhea, n (%)</td>
<td>4 (40)</td>
<td>2 (10)</td>
<td>1 (5)</td>
<td>.23</td>
<td>1 (5)</td>
<td>&lt;.99</td>
</tr>
<tr>
<td>Number of drugs, mean (SD)</td>
<td>13.11 (3.6)</td>
<td>11.85 (3.6)</td>
<td>12.5 (3.2)</td>
<td>.86</td>
<td>11.8 (3.3)</td>
<td>.44</td>
</tr>
<tr>
<td>Polypharmacy, n (%)</td>
<td>10 (100)</td>
<td>20 (100)</td>
<td>20 (100)</td>
<td>&lt;.99</td>
<td>19 (95)</td>
<td>&lt;.99</td>
</tr>
<tr>
<td>Extreme polypharmacy, n (%)</td>
<td>8 (80)</td>
<td>15 (75)</td>
<td>16 (80)</td>
<td>.46</td>
<td>15 (75)</td>
<td>.43</td>
</tr>
<tr>
<td>Death, n (%)</td>
<td>2 (20)</td>
<td>2 (10)</td>
<td>0 (0)</td>
<td>.14</td>
<td>1 (5)</td>
<td>.31</td>
</tr>
</tbody>
</table>

\(^a\)BI: Barthel Index.

\(^b\)HRF: high risk of fall.

\(^c\)PP: polypathological patient.

\(^d\)PCCDs: patients with complex chronic diseases.

Regarding the baseline characteristics, comparing the different phases 1 and 2 categories, significant differences were only found in the presence of pain in the classification BI class B (Table 1). In that category, the mean BI at the beginning of the study was 50.25 (SD 10.44), with an increase at the end of the study to 63.53 (SD 28.92). In BI class C, an initial BI of 80.75 (SD 11.27) and a final BI of 86.76 (SD 15.806) were observed. In the initial balance test for BI class B, the value was 1.6 (SD 1.46) points and 4.8 (SD 1.65) points at the end of the study.
while in the BI class C, the initial value was 4 (SD 1.6) and the final value was 5.5 (SD 2.1).

Data for the ML Model

Following the aforementioned methodology, the patient cohort was reduced to 54 patients, taking into account only those who had at least 14 days’ worth of records before the doctor’s appointment. The principal factors contributing to this lack of data were predominantly attributed to mortality or patients becoming bedridden, subsequently ceasing to use the wristband.

Table 2. Description of the candidate features used to train the regression tree model.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>morning_max</td>
<td>Maximum number of steps recorded during the morning period</td>
</tr>
<tr>
<td>afternoon_max</td>
<td>Maximum number of steps recorded during the afternoon period</td>
</tr>
<tr>
<td>evening_max</td>
<td>Maximum number of steps recorded during the evening period</td>
</tr>
<tr>
<td>overnight_max</td>
<td>Maximum number of steps recorded during the overnight period</td>
</tr>
<tr>
<td>no_walking_aid</td>
<td>Avoidance of any type of walking aid</td>
</tr>
<tr>
<td>cane_or_walker</td>
<td>Use of either a cane, walker, or caregiver’s help for walking</td>
</tr>
<tr>
<td>wheelchair</td>
<td>Use of wheelchair</td>
</tr>
</tbody>
</table>

The fitted model, whose parameters were selected through cross-validation, is a decision tree regressor with a depth of 3, minimum impurity decrease of 0.0, minimum samples in a leaf node of 2 and minimum samples in a split of 9 (Figure 4). Among the features considered, the most important variables for predicting the BI were identified as the maximum step count during the evening and morning periods, and the absence of a mobility device. These key predictors were determined based on their significant impact on the functional status of the patients (Figure 5 [28]).

Based on the results in Table 3, the model exhibits consistency in MAD with a value close to 5 in the training, validation, and test sets. Furthermore, according to Figure 6, when observing the predicted values compared to the real ones, the model does not present a significant difference between the predicted and the real BI.

Once the BI prediction was performed, the intervals defining each BI class were further considered. Subsequently, a classification prediction is carried out by converting the predicted value into its corresponding class label, thereby assigning the appropriate class to the given BI prediction. As observed in Table 3 and Figure 7, in the training set, the model achieved precision, recall, and $F_1$-scores of 0.88, 0.93, and 0.90 for class B, respectively. For class C, the model obtained precision, recall, and $F_1$-scores of 0.94 for all 3 measures. However, for class A, all the metrics were 0.00 due to the limited support for that class (only 1 instance). In the validation set, the model demonstrated consistent performance with precision, recall, and $F_1$-scores of 0.88 for both class B and class C. On the other hand, in the data coming from the test set, the model achieved precision, recall, and $F_1$-scores of 0.5, 1, and 0.67 for class B, respectively. For class C, the model obtained precision, recall, and $F_1$-scores of 1, 0.94, and 0.97, respectively. Finally, only 1 member from class A was predicted as class B.
**Figure 5.** Descending variable importance ranking for the fitted model. Each violin plot represents the distribution of the decrease in R2 score when a single feature value, represented in the y-axis, is randomly shuffled. The importances of the permutations are labeled A (Train set), B (Validation set) and C (Test set).

**Table 3.** Training, validation, and test errors and accuracy of the model when transforming the predicted values to class A, B, and C.

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>91%</td>
<td>88%</td>
<td>90%</td>
</tr>
<tr>
<td>MAE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.30</td>
<td>8.75</td>
<td>10.10</td>
</tr>
<tr>
<td>MAD&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.42</td>
<td>4.50</td>
<td>6.00</td>
</tr>
<tr>
<td>RMSE&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.13</td>
<td>12.99</td>
<td>14.05</td>
</tr>
</tbody>
</table>

<sup>a</sup>MAE: mean absolute error.
<sup>b</sup>MAD: median absolute error.
<sup>c</sup>RMSE: root mean squared error.
Discussion

The use of mobility recording devices identifies different mobility patterns and relates them to functional capacity in patients with polyopathy. One of this study’s findings is improving functional capacity measured by the BI in patients after using the mobility devices for 6 months in a real environment. This improvement is slight, 13 points in the case of moderate dependence and 6 points in mild dependence. We believe that it reflects an effect of the empowerment experienced...
by patients with the use of mobility devices. Therefore, the use of such mobility monitoring devices may have a potential impact on the management of complex chronic patients and could be included as part of clinical follow-up practices. Specifically, during the study execution in a real environment, patient empowerment was related to patient participation in decision-making, gaining control, and learning about their health.

In addition, patients’ sense of empowerment was related to less frustration with the technology [29]. This effect is well-known in the literature. A systematic review of 71 articles analyzing patients’ expectations of digital tools found that mobile apps increase patient engagement and motivation, especially when they can visualize parameters graphically and thus monitor their outcomes over time [30]. The evidence is scarce in patients with polyopathy; but in another study of our group, we found similar results with slight improvements in functional capacity [31]. This empowerment is going to help patients in the self-management of their diseases. Health status has shown that incorporating digital technology into patients’ lives increases their awareness of lifestyle behaviors, which has helped them understand how to manage their health better and promote autonomy [32]. Longer term studies are needed to confirm this benefit, although it could be an alternative to integrate into the clinical practice of these patients to minimize their functional impairment.

Another possible beneficial effect of continuous monitoring of the functional capacity of patients with complex chronic diseases is the early detection of functional deterioration that may be the beginning of exacerbations of their diseases. If these data are integrated into the health care computer system, alarm situations could be determined that would allow early reaction by health staff to treat such exacerbation, prevent its progression, and minimize the functional deterioration that could be caused to the patient.

It should be noted that commercially available mobility monitoring devices have been used for this study and devices specifically designed for the study were not required. This favors cost reduction when considering the implementation of activity monitoring in patients with polyopathy in real-world settings. Since the population with mobility devices is growing, with 515 million units sold in 2022, and patients with polyopathy are a population that continue to increase and probably have their own mobility measurement device [33], the costs are thus reduced by integrating data in the informatics systems of the different health care organizations.

Another contribution of this study was to determine that mobility devices do not accurately recognize patients’ steps when using walking aids. For that reason, and to avoid this possible bias, patients were categorized into 3 groups depending on the walking aids. Additionally, caregivers assisting patients during physical activity have been classified similarly to canes or walkers due to their similarity in providing walking support. After the inclusion of an extra variable with the group to which the patient belongs, the ML model has managed to alleviate these limitations, achieving a good performance.

Concerning the data for the ML model, the use of the maximum value of steps taken in each of the 4 intervals defined (morning-afternoon-evening-overnight) yielded the most promising outcomes. This finding can be attributed to the limited and typically short-lived movements observed in patients with complex chronic conditions, which rarely extend beyond an hour. By leveraging the maximum step count within each time segment, we effectively capture the most significant and representative activity level during that period, thus optimizing the model’s performance. The selected intervals concur with the typical Spanish timetable for meals. Furthermore, the mean and SD data values found in the training and validation sets were similar, suggesting consistent levels of variability in both data sets. Moreover, including the heart rate information, measured by the WAT, and its relationship with the step count is proposed as a future study. Therefore, it could help to distinguish the requirement of the physical activity considering the cross-correlation or the cosine similarity between the step count and the heart rate.

Partitioning the data set into 7 equally sized subsets, with each fold serving as a validation set while the remaining folds are used for training, ensures robustness in selecting the optimal settings that yield the best predictive accuracy for the BI [34]. The hyperparameters play a crucial role in controlling the complexity and generalization of the decision tree model. By tuning these hyperparameters, the cross-validation process aims to find the optimal combination that balances model complexity and performance, resulting in a decision tree model with improved predictive capabilities [28].

Using a decision tree as a regression model holds paramount importance in the biomedicine field, particularly due to the necessity of using a highly interpretable model that can be effectively used and comprehended by the medical team [35,36]. The interpretability of the model enables medical professionals to understand the underlying decision-making process and gain insights into the factors influencing the predictions. This transparency fosters trust and facilitates collaboration between the model and the clinicians. Although more complex approaches exist, such as random forest or extreme gradient boosting, the ability to provide better results than decision trees in terms of accuracy most of the time, their lack of interpretability, and the limited sample size of this study advise against its use. Under these circumstances, a valid alternative to regression trees is multiple linear regression. However, a linear regression model based on the same variables as the decision tree has been performed, yielded inferior results.

Regarding the model performance, upon comparing the results obtained from the model’s predictions with the ground truth, the decision tree model generates accurate predictions. Therefore, the decision tree model can assess the functional capacity of patients based on data collected from the WAT. As observed, the errors remain similar among the training, validation, and test sets. Hence, this confirms that the model can generalize to unseen cases.

There is an imbalanced distribution of classes in the production-like test set, as shown in Figure 7. This discrepancy arises from the natural transition of patients between classes B

https://formative.jmir.org/2024/1/e52344
to C and A, due to changes in the BI, coupled with the criterion of minimum data required for data set inclusion. It is noteworthy that the test set could have been randomized to achieve an even distribution of class numbers, akin to the training and validation sets. However, given the primary goal of evaluating the model’s performance in a realistic production setting, this randomization was deliberately omitted.

In addition, it is worth noting that the MAD is the most suitable performance measure in this case. This choice is justified by the variability in step measurements captured by the WAT and the relatively small data set. It is possible that a patient with inaccurately measured data could significantly influence the error measure, particularly in terms of absolute or squared errors. By considering the MAD as the primary metric, we mitigate the impact of outliers or measurement inconsistencies, ensuring a more robust evaluation of the model’s performance in predicting the BI.

On the other hand, when considering the performance obtained in the classification problem, the accuracy remains consistent in training, validation, and test sets. The main objective of performing regression followed by classification into classes A, B, and C is due to the continuous nature of the BI variable. When aggregated into these 3 intervals, estimating BI solely through classification becomes complex, as small differences in BI values may result in a class label change. Therefore, regression allows the model to capture the underlying continuous relationship within the BI data, enhancing its ability to make more accurate and robust predictions while assigning the appropriate class labels based on the predicted values.

Furthermore, the model does not need to include clinical information such as specific disease, number of comorbidities, severity of disease, and so forth. Therefore, it can be regarded as a general model for patients with complex chronic diseases without specific clinical data, facilitating the development of a methodology that allows estimating the BI at any given moment using the information collected by the WAT over the last 30 days.

This study has several limitations. Since this was a pilot study with a small number of patients, the results should be confirmed by studies with a larger population. Prospective studies are needed to analyze whether identifying mobility changes and their transfer to health care systems can have care implications and improve the health status of patients with multiple pathologies. A notable element is that the bracelet does not register well the physical activity of patients who use a cane or crutches (or a wheelchair) since it cannot measure steps. A priori, this could be a limitation of the study. Still, adjusting the model by identifying walking aid devices and evaluating other parameters makes it possible to identify and predict mobility patterns in these patients.

In conclusion, using commercially available WATs makes it possible to identify different mobility patterns and relate them to functional capacity in patients with polypathology according to the BI without using clinical parameters.

Acknowledgments
This work was supported by the chronic-IoT project (Agencia Estatal de Investigación, PID2019-110747RB-C21/AEI/10.13039/501100011033), which has received funding from the Ministry of Science, Innovation and Universities of the Government of Spain and the State Research Agency. Also, this research has been cosupported by the Carlos III National Institute of Health, through the IMPaCT-Data program (code IMP/00019), and through the Platform for Dynamization and Innovation of the Spanish National Health System industrial capacities and their effective transfer to the productive sector (code PT20/00088), both co-funded by European Regional Development Fund (FEDER) “A way of making Europe.”

Data Availability
The data sets generated during and analyzed during this study are not publicly available because there is no Ethics Committee approval for this purpose, since the received Ethics Committee’s approval was related to carrying out this study. However, they could be available from the corresponding author on reasonable request and this authorization shall then be requested.

Conflicts of Interest
None declared.

References


Abbreviations

BI: Barthel Index
IoT: internet of things
MAD: median absolute error
ML: machine learning
WAT: wearable activity tracker

©Celia Alvarez-Romero, Alejandro Polo-Molina, Eugenio Francisco Sánchez-Úbeda, Carlos Jimenez-De-Juan, Maria Pastora Cuadri-Benitez, Jose Antonio Rivas-Gonzalez, Jose Portela, Rafael Palacios, Carlos Rodriguez-Morcillo, Antonio Muñoz, Carlos Luis Parra-Caldron, Maria Dolores Nieto-Martin, Manuel Ollerro-Baturone, Carlos Hernández-Quiles. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Privacy Concerns About Sharing General and Specific Health Information on Twitter: Quantitative Study

Pouyan Esmaeilzadeh, PhD
Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States

Corresponding Author:
Pouyan Esmaeilzadeh, PhD
Department of Information Systems and Business Analytics, College of Business
Florida International University
Modesto A. Maidique Campus 11200 S.W. 8th St
RB 261 B
Miami, FL, 33199
United States
Phone: 1 3053483302
Email: pesmaeil@fiu.edu

Abstract

Background: Twitter is a common platform for people to share opinions, discuss health-related topics, and engage in conversations with a wide audience. Twitter users frequently share health information related to chronic diseases, mental health, and general wellness topics. However, sharing health information on Twitter raises privacy concerns as it involves sharing personal and sensitive data on a web-based platform.

Objective: This study aims to adopt an interactive approach and develop a model consisting of privacy concerns related to web-based vendors and web-based peers. The research model integrates the 4 dimensions of concern for information privacy that express concerns related to the practices of companies and the 4 dimensions of peer privacy concern that reflect concerns related to web-based interactions with peers. This study examined how this interaction may affect individuals’ information-sharing behavior on Twitter.

Methods: Data were collected from 329 Twitter users in the United States using a web-based survey.

Results: Results suggest that privacy concerns related to company practices might not significantly influence the sharing of general health information, such as details about hospitals and medications. However, privacy concerns related to companies and third parties can negatively shape the disclosure of specific health information, such as personal medical issues ($\beta=-.43; P<.001$). Findings show that peer-related privacy concerns significantly predict sharing patterns associated with general ($\beta=-.38; P<.001$) and specific health information ($\beta=-.72; P<.001$). In addition, results suggest that people may disclose more general health information than specific health information owing to peer-related privacy concerns ($t_{165}=4.72; P<.001$). The model explains 41% of the variance in general health information disclosure and 67% in specific health information sharing on Twitter.

Conclusions: The results can contribute to privacy research and propose some practical implications. The findings provide insights for developers, policy makers, and health communication professionals about mitigating privacy concerns in web-based health information sharing. It particularly underlines the importance of addressing peer-related privacy concerns. The study underscores the need to build a secure and trustworthy web-based environment, emphasizing the significance of peer interactions and highlighting the need for improved regulations, clear data handling policies, and users’ control over their own data.

(JMIR Form Res 2024;8:e45573) doi: 10.2196/45573

KEYWORDS

concern for information privacy; CFIP; peer privacy concern; PrPC; health information disclosure; Twitter; empirical study

https://formative.jmir.org/2024/1/e45573
**Introduction**

**Background**

Reports and analyses highlight that approximately 60% of health-related tweets contain links to health-related websites [1]. Twitter users frequently share health information related to chronic diseases, mental health, and general wellness topics [2]. Twitter is a popular platform for health-related conversations because it allows users to share their thoughts, experiences, and information in real time [3]. This platform can be particularly useful for sharing information about health events, such as disease outbreaks or public health campaigns. In addition, Twitter can be used to connect with others with similar health concerns or interests and to access information from health care professionals and organizations [4]. Twitter is a social media platform with a character limit, which makes sharing detailed information about health issues difficult. Moreover, the information shared on Twitter may not always be accurate or reliable, as it is not always fact-checked or verified. However, Twitter is a common platform for people to share opinions, discuss health-related topics, and converse with a wide audience. According to a survey conducted by the Pew Research Center, 21% of Twitter users have used the platform to share information about a health condition [5]. The survey also found that 20% of Twitter users have followed a health organization or medical professional on the platform, and 15% have searched for information about a health condition on Twitter. Users often express their views about health policies, medical breakthroughs, health care services, and public health issues [6]. Health professionals, researchers, advocacy groups, and patients actively participate in these discussions, contributing diverse perspectives and sharing evidence-based information. This open and rapid exchange of ideas allows health information dissemination and facilitates conversations that can influence public opinion and policy decisions [7].

Sharing health information on Twitter raises privacy concerns as it involves sharing personal and sensitive data on a public platform [8]. Information privacy refers to individuals' control over collecting, using, and disclosing their personal information. Regarding sharing personal health information, web-based information privacy refers to protecting sensitive health data from unauthorized access, secondary use, or disclosure [9]. It involves ensuring that individuals can make informed decisions about how their health information is shared and used and that appropriate safeguards are in place to protect the confidentiality and security of this information. The potential privacy risks associated with sharing health information on Twitter can be grouped into 3 reasons. First, potential identification and disclosure of personal information—sharing health information on Twitter can inadvertently lead to disclosing personally identifiable information. A study found that anonymized data from health-related tweets could be reidentified to reveal the identity of users [10]. Researchers were able to reconstruct personal health stories and connect them to specific individuals, highlighting the potential privacy risks involved. Second, data mining and analytics—third parties can analyze and use health-related tweets for various purposes, including targeted advertising or creating consumer profiles. Researchers analyzed tweets related to mental health and found that the content could be used to predict users' self-reported diagnoses, medication use, and other personal information [11]. This demonstrates the potential for extracting sensitive health-related data from Twitter. In addition, a study used deep neural networks to identify personal health experience tweets, highlighting the potential for using Twitter as a data source for health surveillance studies [12]. Third, public disclosure of sensitive health information—sharing health information on Twitter might inadvertently expose individuals to public scrutiny and judgment. A study examined tweets related to mental health and found that users often disclosed personal experiences, symptoms, and treatments [13]. Although this sharing can provide support, it can also expose individuals to potential stigma, discrimination, or unwanted attention.

Previous studies suggest that individuals may be comfortable with sharing general information that is not sensitive on social media [14]. However, people may not be likely to share personal information, especially health-related data, owing to privacy concerns [15]. According to previous studies, privacy concerns can arise from companies' information collection and use policies in the age of medical big data [16] and web-based social interactions that may threaten information privacy [17]. Twitter is reported as an important data set for vendors, researchers, and medical companies to collect health-related information [18]. Many medical companies collect health information and patient experiences from Twitter for big data analysis to find patterns for public health management [19]. Although big data collection and data mining techniques could help generate intelligence for monitoring public health issues, they can cause privacy concerns. Reports highlight that many Twitter users have experienced invasion of privacy owing to companies' collection, sharing, and analytics practices that use information from their tweets, including private health information [20].

Although there are various studies of vendor-related privacy concerns [21] and peer-related privacy concerns [22], little is known about whether these 2 aspects of privacy concerns may collectively influence information-sharing behaviors. As privacy violations can be related to peers (such as inappropriate comments and unauthorized retweeting) and companies (sharing personal information with third parties), more studies are required to examine whether information-sharing disclosure can be affected equally by vendor-related and peer-related privacy concerns. In this study, we aimed to determine whether both aspects of privacy concerns (ie, concern for information privacy [CFIP] and peer privacy concern [PrPC]) can mutually influence information-sharing behaviors. As privacy violations can be related to peers (such as inappropriate comments and unauthorized retweeting) and companies (sharing personal information with third parties), more studies are required to examine whether information-sharing disclosure can be affected equally by vendor-related and peer-related privacy concerns. In this study, we aimed to determine whether both aspects of privacy concerns (ie, concern for information privacy [CFIP] and peer privacy concern [PrPC]) can mutually influence information-sharing decisions or whether one's effects can dominate or overshadow the impact of the other. For instance, whether the nature of the relationships and contexts with 2 different information trustees (ie, vendors and peers) can influence information dissemination behavior. Thus, we argue that both aspects of privacy concerns should be considered in a model to better characterize information privacy on social media. Investigating the importance of privacy concerns related to companies and vendors (such as Twitter analytics) and web-based peers (such as retweeting a focal user's health information without permission) in case of disclosing...
public and private health information on Twitter would be the main contribution that makes this study different from others. This argument is built based on 4 reasons. First, although the main interactions on social media are mainly peer oriented, vendors can still collect a lot of personal data (such as health information) without authorization and use it for unconssented purposes [23]. There have been instances of social media platforms, including Twitter, being used by organizations for health-related data mining and analysis [24]. Twitter data can provide valuable insights into health-related trends for health care organizations through analytics [25]. Researchers and companies (such as pharmaceutical manufacturers) have used Twitter data to track and analyze health-related trends, including disease outbreaks, medication use, and public health concerns [26]. For example, a study found that Twitter data could be used to track the spread of influenza and predict outbreaks [27]. Another study uncovered that Twitter data could be used to monitor adverse drug reactions and identify potential safety concerns [28]. Pharmaceutical companies have also used Twitter data to monitor medication use and patient experiences. For example, a study reported that Twitter data could be used to monitor patient experiences with antidepressant medications [29]. Moreover, it is common for organizations and vendors, including those in the health care industry, to monitor social media platforms to gather insights about consumer opinions, preferences, and trends [30]. Twitter, as a popular social media platform, has been used for these purposes [31]. Health care organizations and vendors may collect health-related information, such as discussions about medical conditions, treatment experiences, and patient preferences, from public Twitter profiles [19]. These insights can be valuable for marketing and market research purposes.

Second, although the primary purpose of peer-to-peer (P2P) interactions on social media is to maintain social connections with peers and there are no explicit business-to-consumer interactions, companies can still use social media analytics to investigate published health information. Companies can leverage various analytics tools to gather and find meaning and patterns in data collected from social channels to support business decisions (such as predicting the risk factors to manage public health) [32]. Third, individuals share a variety of information (such as about lifestyle, health status, chronic issues, and medication) on social media, which can be more sensitive than conventional e-commerce information (such as transaction records). Disclosing a wide range of information across social media platforms could raise concerns about whether companies and peers misuse the shared data (eg, health information). Fourth, Web 2.0, a fundamental technology supporting social media, mainly focuses on bilateral relationships between peers. However, it does not remove traffic between companies and social media users. Thus, peers can comment on conversations about a user’s health condition and share others’ personal health information on their own channels. In contrast, companies can use big data analytical tools to collect, use, or share users’ personal health data with third parties.

**Study Objectives**

The main objective of this study was to investigate the concept of information privacy concerns in the context of social media based on both vendor-related and peer-related aspects. To do so, we used the survey research methodology and Twitter as the empirical context. We also relied on theories discussing 2 dimensions of information privacy (ie, CFIP and PrPC) as the theoretical foundation of our proposed model. In this study, CFIP, emphasizing both consumer perspectives and company responsibilities, represented privacy concerns related to web-based practices of companies and vendors (such as collection and sharing of self-shared information), and PrPC referred to privacy concerns about losing control over digital communications and web-based interactions with peers. Thus, we suggest that information privacy on social media can be multidimensional, focusing on privacy violations associated with companies’ (vendors’) practices and sharing behaviors of peers. The integration of CFIP and PrPC can comprehensively present the entirety of privacy concerns about web-based health information. This study contributes to both theory and practice. We shed more light on information privacy conceptualization in the context of social media. This study also provides an interactive outlook and practical recommendations for handling privacy issues by explaining how web-based vendors and peers may cause privacy violations when dealing with health information (general and specific) shared over the web.

**Variable Conceptualization, Theoretical Foundation, and Research Hypotheses**

**General and Specific Health Information**

Individuals can use web-based channels to share general health information, such as information about treatments, medications, side effects, hospitals, medical costs, and healthy behaviors [33]. For instance, people are likely to tweet about general obesity-related topics, such as the relationship between fast food and weight gain [34]. Another study identifies general tobacco-related tweets (such as information about smoking, cigarette risks, and quitting) as the primary conversational data sets for health-related topics on Twitter [35]. Moreover, people can use tools such as Twitter to share specific health-related information, including past medical history, allergies, personal medications, private health issues, and signs and symptoms. For example, a study indicates that people disseminate information about diagnoses, advice based on personal experience, use of specific medications, side effects, negative reactions, and treatments on Twitter [31]. Another study highlights that people use Twitter to share their COVID-19–related symptoms and personal health issues during the early stages of the pandemic [36].

Sharing public and private health information can be valuable for web-based peers and affect their health-related decisions. General information can enable web-based users to find some facts about hospitals, physicians, and diseases. Disseminating specific information can share important insights and advice based on personal health conditions, medical treatments, care planning, and medical experiences with chronic diseases. General health information can be publicly available regardless of personal experiences. However, specific health information
can be unpleasant to share because it may contain more private information. As health information dissemination has 2 sides, questions still remain as to what dimensions of information privacy may strongly affect sharing behaviors on Twitter.

**CFIP Constructs**

There is evidence suggesting that companies use tweets to collect health information. For example, reports show that public health researchers use Twitter data to study the world’s health. A recent study indicates that the amount of textual health-related data, which could be personal, collected by various organizations is growing (especially during the COVID-19 pandemic) [37]. Another study argues that health care researchers and research companies have used social media data sources such as Twitter to study public health [19]. Owing to the importance of the Twitter database, the Centers for Disease Control and Prevention (CDC) designed a document to guide employees and contractors on using Twitter to disseminate health information and engage with individuals and partners [38]. A study indicates that companies increasingly use Twitter to share public health information and collect real-time health data using crowdsourcing methods [39]. Information privacy, which refers to people’s ability to control their information, is essential in e-commerce and social media [40]. Several studies explain the privacy concerns specific to the mobility data collection context [41]. Thanks to emerging technology (such as Web 2.0), protecting personal information has become a growing concern for web-based users. CFIP is a general concern about how organizations can use and protect consumers’ information [21]. CFIP explains concerns about organizations’ information collection practices, use policies, and access to consumers’ personal information [42]. Previous studies indicate that examining consumers’ concerns about how companies (vendors) may use their personal information significantly affects their willingness to engage in web-based transactions actively [43].

In this study, following most previous studies, CFIP is posited as a multidimensional construct with 4 dimensions to measure individuals’ concerns about organizations’ information privacy practices [44]. Collection pertains to individuals’ concerns about what web-based information is collected and whether such information is stored properly. Unauthorized secondary use explains individuals’ concerns about whether the information collected for a consented purpose may be unethically and illegally used for other purposes without obtaining authorization. Improper access implies individuals’ concerns about whether unauthorized people (entities) can access, view, and share their information. Finally, concerns about errors reflect whether individuals’ information is appropriately protected to minimize accidental or intentional errors [44]. Therefore, the multidimensional scale of CFIP reflects the complexity of individuals’ privacy concerns [21]. According to Stewart and Segars [40], CFIP is developed as a second-order construct with 4 reflective first-order factors. In this study, we also considered CFIP as a high-order construct with reflective factors. The logic behind conceptualizing this construct as reflective was that the privacy concerns related to companies are reflective of the 4 dimensions (ie, collection, unauthorized access, errors, and secondary use) and the expected interactions among them.

Therefore, these dimensions can reflect the same theme and may covary.

Although sharing information on Twitter is more oriented toward interactions with web-based peers, privacy concerns about the collection and misuse of digitized health information by vendors and companies still remain significant. Previous studies provide strong evidence suggesting that web-based users of Twitter are concerned about several aspects of their information privacy, from collection of a lot of data to misuse [45]. Our study focused on individuals’ perceptions about general CFIP owing to policies and practices of vendors and organizations that may collect, access, and use health information shared on Twitter rather than concerns about a particular vendor. According to the four dimensions of the CFIP construct, individuals who demonstrate high privacy concerns believe that (1) a lot of health information is collected by organizations from users’ Twitter accounts, (2) such health information is not appropriately protected against possible errors, (3) various organizations may use health-related information on Twitter for other purposes without authorization (such as data mining, surveillance, research, and business intelligence), and (4) there is lack of visibility into accurate security measures to control who can access and use health information from tweets.

Thus, the CFIP construct can be extended to privacy concerns about a wide range of vendors and companies accessing and using tweets containing health information. This concern is not the same as privacy issues owing to interactions with a specific vendor in the context of e-commerce (such as retail platforms). In these conventional interactions, privacy concerns may focus on personal, factual information shared in web-based transactions and services (such as demographic information). However, CFIP in the social media domain deals with concerns associated with the following uncertainty: which organizations collect personal posts, which unauthorized entities can view and share information, why and how the information is used (for instance, data mining), and how the information is protected from internal and external errors and misuse. Therefore, we argue that CFIP cannot be ignored in examining information privacy in social media because users may not have direct relationships with organizations on these digital platforms, but they are still concerned about how their posts can be collected and misused by various companies.

Sharing general health information could indicate a user’s rich medical information and wealth of medical knowledge. In contrast, sharing specific health information can show that the user may want to contribute or seek informational and emotional support by disseminating personal experiences and medical history. However, when privacy concerns about the collection and misuse of shared data by organizations are not addressed, users are not likely to disseminate general or specific health information on Twitter. Moreover, we can expect that because specific health information is more sensitive and private, web-based users may generally become more cautious about sharing it. Therefore, we hypothesized the following:

- Hypothesis 1A (H1A): CFIP negatively influences general health information dissemination on Twitter.
• Hypothesis 1B (H1B): CFIP negatively influences specific health information dissemination on Twitter.
• Hypothesis 1C (H1C): CFIP has a more negative effect on specific health information dissemination than on general health information sharing on Twitter.

**PrPC Constructs**

Owing to the nature of Web 2.0, users can communicate, create content, and share it via communities, social networks, and virtual worlds [46]. Web-based users can share a wide range of information and experiences on social media. The information can be objective (based on factual data) or subjective (based on personal interpretation, feelings, tastes, or opinions) [47]. The range can start with demographic information (eg, age, gender, and race); continue with political views, humanitarian opinions, and health information; and end with comments on others’ posts [48]. People can use different formats, such as text, pictures, and videos, to disseminate information. Peers are important components of social networks; however, they can threaten information privacy through inappropriate sharing behaviors and unintended consequences of web-based interactions [49].

Web-based transactions with peers on social media affect users’ decisions about whether they want to reveal their personal information (such as feelings and likes) and create an image consistent with their personal identity [50]. In this study, peers could be web-based friends who may have long-lasting and affect-laden connections with a user and any web-based users who interact through social media channels. Previous studies highlight the importance of PrPCs in the context of web-based interpersonal relationships where other peers can access and view a user’s web-based information [51]. Peer-related privacy refers to possible risks of privacy invasion because of direct and indirect web-based interactions with peers [17]. Social bots and fake and spam accounts can also raise privacy violation risks by potentially exposing several peers to a focal user’s posts using machine learning algorithms [52]. Previous studies indicate the threat of using social bots on social networks, increasing the likelihood of privacy breaches where even more private user data are exposed [53]. Understanding who can access web-based information (such as a post related to signs and symptoms of depression) and with whom such information is shared can significantly raise privacy concerns. For example, a study shows that sharing information with only selected friends in social networking services perceived higher control than sharing information with all friends [54].

Thus, information-sharing behaviors on social media may erode the ability of users to control their virtual space and personal boundaries. Leaving an inappropriate comment for a user who posted about seeking ways to lose weight, can increase privacy concerns about lack of control to maintain the privacy of their Twitter space. A study posits that managing the privacy of virtual territory refers to defining the level of access to and interaction others can have within a user’s territory (eg, allowing peers to see or comment on the post) [55]. Peers can also play a bilateral role in web-based social interactions. They can intentionally or unintentionally share a user’s personal health information with others and expose the user to others’ personal information that they might not like to view. The user may think that if others’ personal health information has been shared with me, my posts can also be revealed to others. Thus, communication privacy can significantly affect how individuals and relational parties share private information on social media [56].

A recent study defines PrPC as the sense of inability to control personal boundaries in web-based interactions owing to web-based peers’ behaviors [22]. They describe this term using 4 reflective dimensions: peer-related information privacy, psychological privacy, virtual territory privacy, and communication privacy. Peer-related information privacy denotes concerns about who can see what type of information and when and how such information is disclosed to other web-based peers. For posts shared by a user, the main concern is unauthorized access and secondary use of data by other peers. On Twitter, this can happen through retweeting and commenting. Peers can also initiate posts or conversation threads to disclose a user’s personal information without authorization. A privacy concern is about the accuracy of personal information shared by peers. Thus, peers’ sharing can be a source of private information leaks.

Psychological privacy explains the control over input information coming from others to shape feelings, opinions, and beliefs. Information sharing is 2-way traffic in social media (ie, from a user to peers and from peers to a user) [57]. As people are exposed to posts shared by celebrities, business magnates, politicians, and other web-based users, their behaviors and opinions are increasingly affected by input information from peers. Peers on social media can influence users’ behavior by applying social influence through public comments on posts [58]. Privacy concerns become more intense when users’ opinions and psychological independence are intentionally manipulated by social bots [59]. In this situation, users are not able to make a decision independent of other web-based peers’ ideas. Moreover, receiving a lot of unwanted information from peers may influence value systems, attitudes, identities, and choices.

Virtual territory privacy represents concerns about an individual’s inability to achieve control over other peers’ interactions with their virtual properties (such as Twitter accounts) and shared conversations (postings). Previous studies suggest that the sense of ownership and emotional attachment to personal territory can be generalized to the social media domain [60]. Similar to other personal belongings, virtual properties are seen as private. Thus, any unwanted addition to or revision of personal information can be considered as an intrusion, which may increase privacy violation risks [45]. Finally, communication privacy reflects an individual’s lack of control over how and when other peers can make direct web-based conversations. For example, peers may use various communication tools to engage individuals in a group conversation about potentially embarrassing or stigmatic health-related topics. Then, users may feel pressured by being involved in such undesirable conversations with unfamiliar people.

Individuals may become more likely to share general or specific health information on Twitter when they think it is useful for
other web-based peers (eg, they can make better medical decisions). However, peer-related concerns may prevent them from disseminating such information. Peers are participants in social media and can freely collect and share information that is sometimes considered as unwanted interference. For instance, if peers retweet a post containing personal information about postsurgery recovery plans without authorization or tag a user who posted general educational content about HIV, these web-based interactions may violate privacy needs and raise privacy concerns. In return, users may change the pattern of health information dissemination and become more cautious in sharing medical facts or personal experiences. Thus, we formulated the following hypotheses:

- **Hypothesis 2A (H2A):** PrPC negatively influences general health information dissemination on Twitter.
- **Hypothesis 2B (H2B):** PrPC negatively influences specific health information dissemination on Twitter.
- **Hypothesis 2C (H2C):** PrPC has a more negative effect on specific health information dissemination than on general health information sharing on Twitter.

### CFIP and PrPC Are Privacy Concerns for Twitter Users

Although tweets are publicly accessible by default, users likely expect some degree of privacy and control over their personal health information shared on the platform. Previous literature has found that even when posting content publicly on social media, individuals still have privacy interests and concerns about how their data might be used or accessed [61]. General health information shared publicly on Twitter, such as mentions of hospitals, physicians, and common diseases, is not considered protected or private. However, more specific personal health details, such as past medical history, allergies, medications, and current symptoms, could reveal private information about an individual’s health status. Although these details may be shared publicly by default on Twitter, users likely still have privacy concerns about this content being widely disseminated or used without their consent.

The concepts of CFIP and PrPC capture these types of privacy concerns. Although users are voluntarily sharing health information publicly on Twitter, they may still desire control over how these data are accessed and used. CFIP reflects concerns about using or sharing personal health data by third parties such as researchers or companies without the user’s knowledge or permission. Even if users willingly post health information publicly, they may still desire control over how that data are collected, analyzed, or shared by entities such as researchers, pharmacies, insurance companies, and so on. PrPC represents concerns about controlling boundaries around health disclosures and limiting exposure to certain audiences, such as employers or insurers, who could misuse the information. Users must balance sharing personal details with managing social risks if the information reaches unintended viewers such as employers, family members, or friends. Thus, although Twitter data are technically public, users are likely to have nuanced privacy interests surrounding their health disclosures. Therefore, concepts such as CFIP and PrPC are useful for quantifying expectations regarding control, anonymity, and audience boundaries that persist even when posting health care–related content openly over the web.

### Difference Between the Conceptualization of CFIP and PrPC

#### Overview

We used an interactive approach to provide a holistic view of information privacy in the context of sharing health information on Twitter. Using this approach, this study actively engaged with the 2 aspects of privacy concerns (CFIP and PrPC) in a dynamic way, considering the interplay between them, as opposed to treating them as isolated, independent factors. Therefore, we examined how these 2 aspects of privacy concerns interact with each other and how this interaction affects individuals’ behavior on Twitter. It should be mentioned that the dimensions used for CFIP and PrPC may differ because of the different nature of the relationships and contexts involved. Although the underlying concept of privacy concerns remains the same, the specific dimensions or factors that contribute to CFIP and PrPC may vary owing to the distinct characteristics of vendors and peers as information trustees.

#### Role and Control

Vendors typically have a professional or business relationship with individuals, where they are entrusted with handling personal information for specific purposes (eg, health care providers and web-based retailers). In this context, individuals may be concerned about vendors’ control over their information; how it is collected, used, and shared; and the potential for data breaches or unauthorized access.

#### Trust and Reputation

CFIP dimensions often include factors related to trust and reputation, such as trustworthiness, perceived reliability, and credibility of vendors. As individuals rely on vendors to handle their personal information responsibly, dimensions related to trust and reputation become important for CFIP measurement.

#### Legal and Ethical Considerations

CFIP dimensions may also include factors related to legal and ethical considerations, such as compliance with privacy laws, informed consent, and transparency in data practices. Individuals may be concerned about whether vendors meet the legal requirements and ethical standards in protecting their health information.

In contrast, peers, who are individuals within an individual’s social network or community, may have different dimensions of privacy concerns. Social interactions, trust, reciprocity, and the potential for social consequences typically characterize peer relationship dynamics. Some factors that could influence PrPC dimensions include the following.

#### Social Norms and Expectations

PrPC dimensions may reflect concerns about social norms and expectations related to privacy within the peer group. Individuals may worry about how their health information might be perceived, shared, or used by their peers and the potential impact on their social relationships or reputation.
Social Influence and Peer Pressure
PrPC dimensions may capture the influence of peer pressure or the fear of negative social consequences. Individuals may be concerned about potential judgment, stigma, or discrimination based on their health information within their peer group.

Personal Boundaries and Intimacy
PrPC dimensions may include factors related to personal boundaries and the level of intimacy within peer relationships. Individuals may be concerned about the extent to which personal health information should be shared with peers and the potential impact on their privacy, autonomy, and self-disclosure.

Although the underlying concept of privacy concerns is present in both CFIP and PrPC, the dimensions may differ owing to the distinct characteristics and dynamics of the relationships involved. Thus, considering these differences when developing measurement instruments is important to accurately capture individuals’ concerns regarding privacy in different trust relationships.

Research Model
The model focuses on health information and Twitter (as the research context). There are a few critical differences in the privacy concerns around health information compared with other types of information. First, health information is considered to be very sensitive and private. It can reveal details about medical conditions, treatments, prescriptions, family history, and so on. Other types of information, such as social media posts or shopping habits, are generally not as sensitive.

Second, health information has strict legal protections such as Health Insurance Portability and Accountability Act in the United States and General Data Protection Regulation in the European Union. These laws place restrictions on how health data can be collected, shared, and used. Other information does not have the same level of legal safeguards. Third, health information could potentially be used to discriminate against people in areas such as employment, insurance, and so on. This type of discrimination is legally prohibited, but the risk remains owing to the sensitive nature of the data. Other data, such as social media posts, have less potential for this type of discrimination. Finally, breach of health information is considered very serious, given the sensitivity of the data. Strong security protections are needed, and breaches can carry heavy penalties. Breaches of other types of data may not have the same level of severity.

Regarding privacy on social media, there are some key characteristics of the concerns around Twitter compared with other platforms. First, most Twitter content is public by default, whereas other platforms such as Facebook allow more privacy controls. This can raise concerns about a lack of control over dissemination. Second, tweets are often archived and searchable indefinitely; therefore, there are concerns about permanent availability even for “deleted” content. Other platforms may have more ephemeral sharing. Third, the open nature of Twitter makes it easy for tweets to spread rapidly and become viral compared with platforms such as Instagram, where sharing can be more controlled. This raises concerns about loss of context and lack of containment. Finally, the ability to create anonymous accounts on Twitter is greater than that on platforms such as Facebook that require real identities. This raises concerns about harmful speech, misinformation, and so on.

We proposed the following research framework for disclosing general and specific health information on Twitter by integrating 2 aspects of information privacy concerns (Figure 1). As several studies may have found empirical evidence for the hypotheses proposed in this study, we need to clarify what is new in our study. First, this study integrated both aspects of privacy concerns for the first time in a model. Previous studies examined either privacy concerns related to companies’ practices with web-based information (CFIP) [62] or concerns related to the web-based behaviors of peers (PrPC) [63]. However, as mentioned in the previous section, individuals may be concerned about disseminating their health information on Twitter because companies’ collection practices and web-based peers’ behaviors could violate their privacy. In this study, we wanted to examine whether both aspects of privacy concerns (ie, CFIP and PrPC) can collectively change health-related information-sharing decisions or whether one can dominate the other. For instance, whether the nature of the relationships and contexts with 2 different information trustees (ie, vendors and peers) can shape information dissemination behavior. Second, as Twitter is considered as a rich database for collecting individual health-related information to examine sentiments and manage public health [64] and reports highlight that individuals may be concerned about web-based interactions with peers [65], Twitter would be the best research context to meet the goals of this study. Third, this study distinguished between general and specific health information. Thus, we could offer more insights about privacy concern levels and disclosure behaviors related to the 2 types of health information on Twitter. These 3 reasons can make our study different from previous studies in the privacy literature.

In addition, we controlled for several variables such as age, gender, education, Twitter experience, privacy violation experience, and misrepresentation of identity on Twitter. According to previous studies in the privacy concern domain, some demographics, such as age [66], gender [67], and education level [68], can affect people’s intention to disclose information on social media. Moreover, the impacts of these variables have been examined in previous studies investigating individuals’ perceptions about sharing eHealth-related information [69,70]. The effects of these variables are often controlled in previous studies in the field of information privacy threats [71]. Thus, we assumed that individuals of different ages, genders, and educational levels engage in various disclosure behaviors because they have diverse backgrounds, individual characteristics, and personal differences. Therefore, we considered these demographics to be control variables in the proposed research model.

Moreover, the effects of misrepresentation of identity, experience with technology, and privacy violation experiences are controlled in previous studies examining relationships between privacy concerns and self-disclosure [22,42]. Thus, it is believed that individuals with different privacy violation experiences, previous identity misrepresentation, and experiences with Twitter are more likely to demonstrate various
disclosure behaviors. Therefore, we treated these experience-related variables as control variables in our model.

**Figure 1.** Research model. 1A: hypothesis 1A; 1B: hypothesis 1B; 2A: hypothesis 2A; 2B: hypothesis 2B.

**Methods**

**Research Approach and Survey Development**

We administered a web-based survey questionnaire to achieve the defined objectives and test the proposed model and research hypotheses. The survey consists of 4 sections. In the first part, the purpose of the study is described clearly, and a qualifying question is used to select respondents. The question for filtering respondents attempts to screen individuals with a Twitter account. Thus, individuals without a Twitter account are excluded from data collection and analysis. In the second section, respondents are asked to express their perceptions about privacy concerns associated with companies and third parties, peer-related privacy concerns, and health information dissemination behaviors. In the third section, demographic questions (ie, age, gender, education, income, and race) are asked. Finally, the last section focuses on personal privacy experiences (ie, Twitter experience, privacy violation experience, and misrepresentation of identity).

Questions to measure each construct were adapted from validated instruments available in the existing literature. Slight changes in the wording were made to fit the context of this study. We adapted items to measure CFIP (as a second-order construct with 4 dimensions) from the study by Stewart and Segars [40]. Following Zhang et al [22], we also conceptualized and measured PrPC as a second-order reflective construct with 4 dimensions. Previously defined scales to measure general and specific health information disclosure were adapted from the study by Hsu et al [72]. Respondents rated all the measuring items included in the survey using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Multimedia Appendix 1 shows the questions used in the web-based survey.

**Data Collection and Data Analysis**

Data were collected in April 2022 by uploading the questionnaire to Amazon’s Mechanical Turk (MTurk). MTurk is a crowdsourcing platform that enables researchers to access data from potential target samples to conduct a study. MTurk has been recognized as an acceptable web-based means for collecting individual-level data. Literature about health care analytics shows a growing number of studies using MTurk for health-related research [73]. Previous studies highlight that MTurk can measure individual perceptions in various domains, such as social media [74]. As the target population of this study was US citizens who use Twitter for web-based interactions, we limited the respondents’ location to the United States. Moreover, 2 attention-check questions were used to remove participants who chose answers without correctly replying to reverse-coded filler items [75]. The filtering questions were as
follows: (1) It does not bother me that my peers may try to influence me through comments on my health-related postings on Twitter and (2) I am not concerned that I have little control over who can start a health-related conversation with me on Twitter. We received 364 questionnaires and excluded 35 (9.6%) that were either incomplete or failed the response quality questions, resulting in 329 (90.4%) valid and usable responses. The average response time to complete the questionnaire was 12 minutes. The descriptive statistics for demographics were performed using SPSS (version 26; IBM). The research model was tested using AMOS (version 26; IBM) within the structural equation model framework.

**Ethical Considerations**

The institutional review board of Florida International University reviewed and approved the study (approval 112755). According to the institutional review board approval, written informed consent to participate in the study was obtained from all participants. Moreover, the data collected in this study were anonymous. We considered US $1 as an incentive for each respondent to participate in the study.

**Results**

**Instrument Validation**

We used confirmatory factor analysis to assess convergent and discriminant validity. Table 1 shows the results of the convergent validity test. The standardized factor loadings for all constructs exceeded 0.7, which is the acceptable range for factor loadings [76]. The composite reliability values and Cronbach α values were above the recommended value of .7, demonstrating the adequate reliability of the constructs [77]. All the values of average variance extracted (AVE) exceeded 0.5, which is the cutoff value [78]. These measures indicated the acceptability of the measurement model’s convergent validity.

Table 2 shows the discriminant validity of the constructs. All diagonal values (square roots of the AVEs) were >0.7 and greater than off-diagonal values (correlations) between any pair of constructs [79]. Thus, the discriminant validity requirements were satisfied for the research model. Moreover, we checked the convergent and discriminant validity of the second-order constructs. The composite reliability, Cronbach α, and AVE values for CFIP were 0.91, .88, and 0.64, respectively, and these measures for PrPC were 0.94, .89, and 0.72, respectively. The correlation between the second-order variables (eg, CFIP and PrPC) was 0.58. Finally, the square roots of the AVEs for both constructs were >0.7 and higher than the correlations between the constructs. These results confirm an acceptable convergent and discriminant validity for both second-order constructs in the model.
Table 1. Results of the convergent validity test.

<table>
<thead>
<tr>
<th>Constructs, subdimensions, and items</th>
<th>Standardized factor loading (&gt;0.7)</th>
<th>Composite reliability (&gt;0.7)</th>
<th>Cronbach α (&gt;0.7)</th>
<th>AVE³ (&gt;0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concern for information privacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL b</td>
<td>0.89</td>
<td>.75</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>COLL1</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL2</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL3</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL4</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU c</td>
<td>0.90</td>
<td>.79</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>USU1</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU2</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU3</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU4</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAC d</td>
<td>0.86</td>
<td>.79</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>IAC1</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAC2</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAC3</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR e</td>
<td>0.89</td>
<td>.84</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>ERR1</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR2</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR3</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR4</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer privacy concern</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC f</td>
<td>0.88</td>
<td>.84</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>PPC1</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC2</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC3</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC4</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC g</td>
<td>0.89</td>
<td>.88</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>CPC1</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC2</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC3</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC4</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC h</td>
<td>0.89</td>
<td>.87</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>VTPC1</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC2</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC3</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC4</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer-related information privacy concern–SSIPC i</td>
<td>0.87</td>
<td>.82</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>SSIPC1</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSIPC2</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSIPC3</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constructs, subdimensions, and items</td>
<td>Standardized factor loading (&gt;0.7)</td>
<td>Composite reliability (&gt;0.7)</td>
<td>Cronbach α (&gt;0.7)</td>
<td>AVEa (&gt;0.5)</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SSIPC4</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Peer-related information privacy concern–PSIPC</strong> j</td>
<td>0.89</td>
<td>.87</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>PSIPC5</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSIPC6</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSIPC7</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSIPC8</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GHID</strong> k</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/Ai</td>
<td>0.87</td>
<td>.84</td>
<td></td>
<td>0.64</td>
</tr>
<tr>
<td>GHID1</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHID2</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHID3</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHID4</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SHID</strong> m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/Ai</td>
<td>0.90</td>
<td>.88</td>
<td></td>
<td>0.69</td>
</tr>
<tr>
<td>SHID1</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHID2</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHID3</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHID4</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aAVE: average variance extracted.
bCOLL: collection.
cUSU: unauthorized secondary use.
dIAC: improper access.
eERR: error.
fPPC: psychological privacy concern.
gCPC: communication privacy concern.
hVTPC: virtual territory privacy concern.
iSSIPC: self-shared information privacy concern.
jPSIPC: peer-shared information privacy concern.
kGHID: general health information disclosure.
lN/A: not applicable.
mSHID: specific health information disclosure.
Table 2. Results of the discriminant validity test.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Score, mean (SD)</th>
<th>CFIP-COLL&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CFIP-USU&lt;sup&gt;b&lt;/sup&gt;</th>
<th>CFIP-IAC&lt;sup&gt;c&lt;/sup&gt;</th>
<th>CFIP-ERR&lt;sup&gt;d&lt;/sup&gt;</th>
<th>PrPC-PPC&lt;sup&gt;e&lt;/sup&gt;</th>
<th>PrPC-CPC&lt;sup&gt;f&lt;/sup&gt;</th>
<th>PrPC-VTPC&lt;sup&gt;g&lt;/sup&gt;</th>
<th>PrPC-PRIPC&lt;sup&gt;h&lt;/sup&gt;</th>
<th>GHID&lt;sup&gt;i&lt;/sup&gt;</th>
<th>SHID&lt;sup&gt;j&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFIP-COLL</td>
<td>4.03 (0.66)</td>
<td>0.81&lt;sup&gt;k&lt;/sup&gt;</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>CFIP-USU</td>
<td>4.01 (0.67)</td>
<td>0.71</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>CFIP-IAC</td>
<td>3.94 (0.77)</td>
<td>0.62</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>CFIP-ERR</td>
<td>3.85 (0.82)</td>
<td>0.67</td>
<td>0.69</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>PrPC-PPC</td>
<td>3.92 (0.78)</td>
<td>0.32</td>
<td>0.35</td>
<td>0.17</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
</tr>
<tr>
<td>PrPC-CPC</td>
<td>3.86 (0.88)</td>
<td>0.24</td>
<td>0.39</td>
<td>0.33</td>
<td>0.36</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>PrPC-VTPC</td>
<td>3.83 (0.86)</td>
<td>0.30</td>
<td>0.28</td>
<td>0.32</td>
<td>0.31</td>
<td>0.74</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>PrPC-PRIPC</td>
<td>3.87 (0.81)</td>
<td>0.19</td>
<td>0.36</td>
<td>0.31</td>
<td>0.33</td>
<td>0.69</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>GHID</td>
<td>3.84 (0.85)</td>
<td>0.34</td>
<td>0.32</td>
<td>0.42</td>
<td>0.36</td>
<td>0.35</td>
<td>0.42</td>
<td>0.30</td>
<td>0.49</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>SHID</td>
<td>3.81 (0.91)</td>
<td>0.30</td>
<td>0.30</td>
<td>0.38</td>
<td>0.31</td>
<td>0.37</td>
<td>0.38</td>
<td>0.41</td>
<td>0.41</td>
<td>0.53</td>
<td>0.83</td>
</tr>
</tbody>
</table>

<sup>a</sup>CFIP-COLL: concern for information privacy–collection.
<sup>b</sup>CFIP-USU: concern for information privacy–unauthorized secondary use.
<sup>c</sup>CFIP-IAC: concern for information privacy–improper access.
<sup>d</sup>CFIP-ERR: concern for information privacy–error.
<sup>e</sup>PrPC-PPC: peer privacy concern–psychological privacy concern.
<sup>f</sup>PrPC-CPC: peer privacy concern–communication privacy concern.
<sup>g</sup>PrPC-VTPC: peer privacy concern–virtual territory privacy concern.
<sup>h</sup>PrPC-PRIPC: peer privacy concern–peer-related information privacy concern.
<sup>i</sup>GHID: general health information disclosure.
<sup>j</sup>SHID: specific health information disclosure.
<sup>k</sup>Italicization represents the square roots of the average variance extracted.
<sup>l</sup>Not applicable.

Respondents’ Characteristics

Table 3 shows the participants’ characteristics. The descriptive statistics demonstrate that respondents were fairly distributed across gender, where 52.9% (174/329) were men and 47.1% (155/329) were women. The age range was positively skewed, indicating that most participants were young, with a range between 25 and 34 years (155/329, 47.1%) being high, followed by the range between 35 and 44 years (102/329, 31%). Approximately half (178/329, 54.1%) of the respondents had undergraduate or graduate education levels, which aligns with previous studies highlighting that people with high education levels tend to search more often for web-based health information [80]. The annual income was fairly distributed, with income between US $60,000 and US $79,999 showing a high range (135/329, 41%) among the provided categories. Most respondents were White (174/329, 52.9%), followed by Hispanic and African American individuals. Approximately half (174/329, 52.9%) of the respondents reported using Twitter for 4 to 6 years. Overall, 52% (171/329) of the respondents indicated that they had a privacy violation experience at least once (for instance, their account was hacked), and 38.9% (128/329) mentioned that they tried to use a fake account on Twitter (at least once).
Table 3. Descriptive statistics (N=329).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>174 (52.9)</td>
</tr>
<tr>
<td>Women</td>
<td>155 (47.1)</td>
</tr>
<tr>
<td><strong>Age group (y)</strong></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>13 (4)</td>
</tr>
<tr>
<td>25-34</td>
<td>155 (47.1)</td>
</tr>
<tr>
<td>35-44</td>
<td>102 (31)</td>
</tr>
<tr>
<td>45-54</td>
<td>33 (10)</td>
</tr>
<tr>
<td>55-64</td>
<td>20 (6.1)</td>
</tr>
<tr>
<td>≥65</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>10 (3)</td>
</tr>
<tr>
<td>High school</td>
<td>66 (20.1)</td>
</tr>
<tr>
<td>College</td>
<td>76 (23.1)</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>105 (31.9)</td>
</tr>
<tr>
<td>Graduate</td>
<td>72 (21.9)</td>
</tr>
<tr>
<td><strong>Annual income (US $)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;20,000</td>
<td>23 (7)</td>
</tr>
<tr>
<td>20,000-39,999</td>
<td>59 (17.9)</td>
</tr>
<tr>
<td>40,000-59,999</td>
<td>49 (14.9)</td>
</tr>
<tr>
<td>60,000-79,999</td>
<td>135 (41)</td>
</tr>
<tr>
<td>80,000-99,999</td>
<td>46 (14)</td>
</tr>
<tr>
<td>≥100,000</td>
<td>16 (4.9)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>36 (10.9)</td>
</tr>
<tr>
<td>Asian</td>
<td>23 (7)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>66 (20.1)</td>
</tr>
<tr>
<td>Native American</td>
<td>13 (4)</td>
</tr>
<tr>
<td>White</td>
<td>174 (52.9)</td>
</tr>
<tr>
<td>Mixed</td>
<td>16 (4.9)</td>
</tr>
<tr>
<td><strong>Twitter experience (y)</strong></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>63 (19.1)</td>
</tr>
<tr>
<td>4-6</td>
<td>174 (52.9)</td>
</tr>
<tr>
<td>7-9</td>
<td>66 (20.1)</td>
</tr>
<tr>
<td>10-12</td>
<td>20 (6.1)</td>
</tr>
<tr>
<td>&gt;12</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td><strong>Privacy violation experience (eg, being hacked)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>171 (52)</td>
</tr>
<tr>
<td>No</td>
<td>158 (48)</td>
</tr>
<tr>
<td><strong>Identity misrepresentation (eg, using fake accounts)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>201 (61.1)</td>
</tr>
<tr>
<td>No</td>
<td>128 (38.9)</td>
</tr>
</tbody>
</table>
Analysis of the Dimensions

When implementing second-order variables in a measurement model, there are 2 common approaches: the repeated items approach and the 2-step approach [81]. This study used a repeated items approach to measure reflective second-order constructs. In the repeated items approach, the indicators used to measure the second-order construct are included in the measurement model twice: once as indicators of the second-order construct and once as indicators of the corresponding first-order constructs [82]. This approach allows for a direct assessment of both the second-order and underlying first-order constructs in a single measurement model. The repeated items approach provides a holistic view of the measurement model by simultaneously assessing the second-order construct and its underlying dimensions [83]. Using the repeated items approach provides a more integrated perspective about how CFIP and PrPC are influenced by their respective first-order constructs. It allows for a direct examination of the relationships between the second-order construct and its underlying factors.

Both CFIP and PrPC are conceptualized as second-order reflective constructs, consistent with existing literature. A reflectively measured construct shares a common theme across subdimensions; the dimensions are expected to be highly correlated [84]. Table 2 shows that, consistent with reflective measurement, the 4 dimensions of CFIP (ie, collection, unauthorized secondary use, improper access, and errors) are highly correlated with each other. As expected, the 4 dimensions of PrPC (ie, psychological, communication, virtual territory, and peer-related information privacy concerns) are also highly correlated. Results show that the 4 dimensions of CFIP, as first-order factors, load significantly on the second-order construct. The loadings were 0.91 for collection, 0.80 for unauthorized secondary use, 0.95 for improper access, and 0.88 for errors. Therefore, the interaction among 4 dimensions reflects CFIP, which shares a common theme of losing control over information privacy owing to companies’ sharing behaviors. Furthermore, the 4 dimensions of PrPC also load significantly on the second-order construct. The loadings were 0.90 for psychological privacy concerns, 0.92 for communication privacy concerns, 0.86 for virtual territory privacy concerns, and 0.95 for peer-related information privacy concerns. Thus, interactions among these 4 dimensions represent PrPC, which exhibits a shared theme of privacy concerns about losing personal control owing to web-based peer behaviors.

Structural Model and Path Analysis

Consistent with privacy literature, we controlled variables such as age, gender, education, years of experience, privacy violation experience, and misrepresentation of identity in the structural model [42]. Findings demonstrate that when the control variables are present, the coefficients and $R^2$ change significantly. Specifically, when age ($\beta=-.12; P=.01$), education ($\beta=.19; P=.003$), and privacy violation experience ($\beta=-.58; P=.008$) are present in the model, they significantly influence health information disclosure. Thus, the findings confirm that young people with high education levels who have not experienced privacy violations are more likely to disclose health information on Twitter. However, no effects of gender, years of experience, and misrepresentation of identity were found on health information–sharing behaviors. We used the structural equation model technique to analyze the factors affecting health information disclosure on Twitter. The results of model fit indexes exhibit a good fit with the goodness-of-fit indexes ($\chi^2_{353}=2.22$; goodness-of-fit index=0.84; adjusted goodness-of-fit index=0.81; comparative fit index=0.90; normed fit index=0.91; incremental fit index=0.90; standardized root mean square residual=0.03; and root mean square error of approximation=0.04) where all indexes meet their recommended cutoff values [85]. Table 4 depicts the summary of path analysis for 4 hypotheses (ie, H1A, H1B, H2A, and H2B).

Figure 2 shows the standardized path coefficients of the structural model. Support is not found for H1A, which proposes that CFIP significantly influences general health information disclosure on Twitter ($\beta=-0.07; P=0.16$). In contrast, the findings support H1B by confirming that CFIP significantly attenuates sharing behaviors when disclosing specific health information on Twitter ($\beta=-.43; P<.001$). H2A, which posits that PrPC would directly affect the disclosure of general health information on Twitter, is supported ($\beta=-.38; P<.001$). The analysis also exhibits that PrPC negatively shapes the sharing of specific health information on Twitter ($\beta=-.72; P<.001$), and this significant relationship supports H2B.

Regarding H1C and H2C, an alternative model was created for each hypothesis, and the 2 relationships in that hypothesis were constrained [86]. Next, a 2-tailed $t$ test was used to compare the difference between the alternative and the original model. H1C posits a significant difference between the impact of CFIP on general and specific information–sharing behaviors. As the $t$ value was significant ($t_{165}=3.45; P<.001$), we confirm that CFIP imposes a more negative effect on specific health information dissemination than on sharing general health information on Twitter. In addition, H2C proposes that people may disclose more general health information than specific health information owing to peer-related privacy concerns. The $t$ value was significant ($t_{165}=4.72; P<.001$); thus, the effect of PrPC was more prominent in specific health information sharing than in disclosing general health information on Twitter.

Finally, the model explains 41% of the variance in general health information disclosure and 67% in specific health information sharing on Twitter. The $R^2$ scores suggest that the 2 aspects of information privacy concerns (ie, concerns about the web-based practices of companies and peers’ behaviors) can provide reliable explanatory power to predict the variance in sharing general and specific health information.
### Discussion

#### Principal Findings

Information sharing is one of the most important objectives of social media. People use Twitter for conversation, and information sharing can initiate a web-based exchange of ideas about an issue. As health information is more sensitive than other types of personal information, disclosing such data can raise privacy concerns. Most previous studies have mainly focused on privacy concerns related to companies and vendors as they may collect and use individuals’ personal information for other purposes or may not properly protect the collected information [87]. Few studies have also explained privacy concerns related to the web-based behaviors of peers [22]. However, previous literature did not consider both sides of information privacy concerns in a model in the context of social media. Moreover, disclosure behaviors on social media can be contingent upon the type of health information owing to sensitivity levels. Few studies have examined the sharing

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Path</th>
<th>Standardized coefficient, β</th>
<th>SE</th>
<th>Critical ratio</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>CFIP–general health information disclosure</td>
<td>−.07</td>
<td>0.04</td>
<td>1.54</td>
<td>Not supported</td>
</tr>
<tr>
<td>1B</td>
<td>CFIP–specific health information disclosure</td>
<td>−.43</td>
<td>0.03</td>
<td>4.21</td>
<td>Supported</td>
</tr>
<tr>
<td>2A</td>
<td>PrPC–general health information disclosure</td>
<td>−.38</td>
<td>0.05</td>
<td>5.37</td>
<td>Supported</td>
</tr>
<tr>
<td>2B</td>
<td>PrPC–specific health information disclosure</td>
<td>−.72</td>
<td>0.05</td>
<td>7.12</td>
<td>Supported</td>
</tr>
</tbody>
</table>

aCFIP: concern for information privacy.  
bSignificance level, P<.001.  
cPrPC: peer privacy concern.

**Figure 2.** Model paths. H1A: hypothesis 1A; H1B: hypothesis 1B; H2A: hypothesis 2A; H2B: hypothesis 2B; *P<.001.
behaviors based on the unique characteristics of general and specific health information [86].

Although both antecedents (CFIP and PrPC) have been examined separately in previous studies, this study's findings could propose scientific novelty. The study differs from previous research in this field because we integrated 2 aspects of privacy concerns (eg, related to companies and peers) to investigate the disclosure of general and specific health information on Twitter. In this study, we examined whether both aspects of information privacy concerns can jointly influence sharing decisions related to health-related information or whether the effect of one aspect can be overshadowed by the other; for instance, whether the nature of the relationships and contexts with 2 different information trustees (ie, vendors and peers) can influence people to share their health information on Twitter. The findings indicate that privacy concerns related to companies play a more significant role in predicting specific health information than in predicting general health information. Privacy concerns related to companies' practices reflect the collection and misuse of health information by vendors, such as concerns about using health information for data mining and research purposes [88].

Our findings demonstrate that Twitter users are more concerned about vendors and companies using or sharing their personal health information than general health information. Thus, the dimensions of CFIP (ie, collection, unauthorized secondary use, improper access, and errors) become salient only when people want to disclose specific health information (such as information about their chronic diseases, signs, and symptoms or personal health status). Sharing public health information about hospitals, medical costs, and medications is not significantly affected by concerns about how companies may use such information. A plausible justification is that general health information cannot reflect any personal information associated with an individual, and even if it is used for data mining or big data analysis, it will not violate the user's privacy needs. Consistent with previous studies [89], individuals may deliberately want to share general health information on social media to increase public awareness and knowledge about a medical situation, such as COVID-19 symptoms and vaccination. Regardless of information accuracy or misinformation, users may engage in sharing their general medical knowledge and public information about treatment options with almost no or minor privacy concerns related to companies and vendors' collection and use practices.

Our results also show that peer-related privacy concerns can significantly shape both general and specific health information sharing on Twitter. Although Twitter is not the same as web-based health communities designed to share health information, many individuals use tweets to share personal and public health information [90]. Web-based interactions with peers may affect their sharing behaviors as they may feel unable to control who can see, comment on, or exchange the health information they share on Twitter. The dimensions of PrPC (ie, psychological, communication, virtual territory, and peer-related information privacy concerns) are important factors in predicting how users may disclose public and personal health information. However, peer-related privacy concerns are more intense for sharing personal than general health information. This finding indicates that when a user wants to share public information about a disease (for instance, cancer, COVID-19, or HIV), they are still concerned about how peers relate such general information to their profile. This concern becomes more salient when the user decides to reveal personal health information about that disease, such as what treatments or medications they are using daily or what medical procedures they will undergo in the future. Previous studies associate sharing personal health information related to physical health problems or mental disorders on web-based P2P networks with stigma [91]. The more sensitive the health information, the more stigma is attached to sharing such information with peers. Thus, being judged by peers (close friends and other web-based users) because of sharing personal health information may prevent them from disclosing that content on Twitter.

Although results show significant impacts of both aspects of privacy concerns on sharing specific health information on Twitter, peer-related privacy concerns are leading factors in shaping personal health information disclosure, more so than privacy concerns associated with companies and third parties. This result confirms the importance of web-based interactions with peers on social media and how to deal with their sharing behaviors, such as commenting or tagging [60]. This finding implies the critical effects of Twitter friends and the circle of people who can see and share tweets about private health information. Individuals may be first concerned about the Twitter circle and how peers would react to the shared personal information about health status and then become worried about how many companies may access such data and how they would use or share them. Thus, secondary dissemination of personal health information by web-based peers through liking, reposting, retweeting, or commenting on posts is more challenging to the maintenance of privacy controls than secondary use of data or unauthorized access to such private information by companies and vendors. Our finding that peer-related privacy concerns have a strong impact on health information sharing compared with privacy concerns associated with companies and third parties offers a counter-narrative to prevalent assumptions in digital privacy research. This could be attributed to Twitter's highly interactive and public nature, which might accentuate peer-related concerns. Previous studies, mainly those conducted in the context of web-based shopping or general social media use, might have overestimated the role of concerns associated with companies and third parties owing to the commercial and private nature of these web-based activities.

Theoretical Implications
This study may offer some theoretical contributions. First, our findings have implications for information privacy research in social media by integrating the existing privacy concern perspectives. This study can open up the discussions through which privacy needs related to companies, third parties, and peer-related aspects can be addressed. Then, this comprehensive mechanism may strongly affect users' sharing behavior patterns. Second, this study distinguishes the differences between sharing public and private health information on Twitter. Although disclosing specific health information may help share personal experiences related to various medical situations that could be useful for peers, it is more challenging than disseminating general health information. The findings demonstrate how
company-related and peer-related privacy concerns could shape the 2 types of information-sharing behaviors. Third, this study investigates the effects of information privacy mechanisms on health information sharing in the context of Twitter. The findings can promote discussions about health information disclosure in other P2P networks such as other social media platforms, virtual worlds, or Metaverse. Fourth, exploring the determinants of information sharing regarding different types of privacy concerns can expand our current understanding of knowledge acquisition. As sharing both general and specific health information on social media can contribute to people’s medical knowledge, addressing the barriers to specific health information sharing and removing the privacy challenges can significantly help the procedures of medical knowledge acquisition from web-based interactions with peers.

The study contributes significantly to the theoretical understanding of privacy concerns in web-based health information sharing. The evidence that peer-related privacy concerns influence more strongly than those related to companies and third parties highlights a potential oversight in theoretical perspectives. Current theories largely view companies as the predominant source of digital privacy concerns, and this may need re-evaluation. The results extend existing theories by emphasizing the role of peer interactions in privacy concerns, particularly in public and highly interactive web-based environments such as Twitter. This recognition of the social dimension of privacy concerns could be integrated into existing theoretical models to provide a more comprehensive framework for web-based privacy behavior. Furthermore, although our study is specific to Twitter and health information, the insights gained may have broad applicability. The potential role of peer-related privacy concerns could be a valuable area of exploration in other social media contexts and in sharing other types of sensitive information. Thus, our findings open up new avenues for theoretical exploration and suggest a need for further studies to fully understand the complexities of privacy behavior in the digital age.

Unlike other research approaches, such as experiments, observational data, or qualitative interviews to assess privacy concerns and information sharing, our study used a quantitative survey approach. This allowed us to capture data from a large and more diverse sample, providing a more robust and generalizable understanding of privacy concerns in web-based health information sharing. The strength of our quantitative approach lies in its ability to establish clear patterns and relationships among various factors influencing privacy concerns. This enabled us to derive a more comprehensive and systematic understanding of the factors that significantly influence privacy concerns and health information sharing on Twitter. In terms of comparison, our findings open up new avenues for theoretical exploration and suggest a need for further studies to fully understand the complexities of privacy behavior in the digital age.

As the 2 aspects (company-related and peer-related aspects) of privacy concerns manifest in several dimensions, different features can be developed to address the need for effective protection mechanisms. Twitter can add a new feature to tweets, enabling users to identify the sensitivity of posts related to health information. The content will be recognized as a private post if the sensitivity score (eg, calculated based on a scale ranging from 1 to 10) is more than average. Then, that post is automatically restricted from exposure to everyone, and users can share their thoughts and experiences only with a small crowd. Users can also define terms and conditions for peers who want to retweet sensitive posts. For instance, a “request for share” button can appear for each sensitive post, and peers can only share the posts when they get approval from the focal users. Given our findings, Twitter could introduce a feature that allows users to select the audience for their health-related posts, thereby addressing peer-related privacy concerns. They could also introduce a “Health Information” mode that automatically

**Practical Contributions**

This study also provides several practical and technical implications for promoting privacy protection on Twitter. To promote the sharing of specific health information, it is essential to address privacy concerns related to both companies and peers. However, addressing peer-related privacy concerns is vital to encourage the disclosure of general health information. This is because concerns related to companies and third parties do not significantly predict general health information sharing. Thus, a robust privacy policy cannot be developed regardless of information type. As the 2 types of health information require different ways of satisfying privacy needs, mechanisms and regulations facilitating general and specific health information sharing cannot be the same. Depending on the type of health information, it is essential to customize the ability of Twitter users to control their self-concept and meet different privacy protection requirements. General procedures and privacy policies to regulate the dissemination and use of personal posts are not sufficient to address the information privacy concerns. Twitter should allow users’ privacy concerns about sharing specific health information using advanced technology and management mechanisms. For example, Twitter can enable individuals to restrict access to their shared personal health information. Punitive regulations can be established for inappropriate behaviors (such as retweeting without consent) that may discourage sharing specific health information. All controlling mechanisms and privacy protection functionalities should be easy to understand and use and should not be an additional burden on the users.

As the 2 aspects (company-related and peer-related aspects) of privacy concerns manifest in several dimensions, different features can be developed to address the need for effective protection mechanisms. Twitter can add a new feature to tweets, enabling users to identify the sensitivity of posts related to health information. The content will be recognized as a private post if the sensitivity score (eg, calculated based on a scale ranging from 1 to 10) is more than average. Then, that post is automatically restricted from exposure to everyone, and users can share their thoughts and experiences only with a small crowd. Users can also define terms and conditions for peers who want to retweet sensitive posts. For instance, a “request for share” button can appear for each sensitive post, and peers can only share the posts when they get approval from the focal users. Given our findings, Twitter could introduce a feature that allows users to select the audience for their health-related posts, thereby addressing peer-related privacy concerns. They could also introduce a “Health Information” mode that automatically
applies high privacy settings for tweets marked as health related. In addition, given the significant role of knowledge in shaping privacy concerns, Twitter should consider educational campaigns or prompts to inform users about these features and the importance of privacy when sharing health information.

In May 2022, Elon Musk called for further investigation into the accuracy of spam and fake account estimates, which Twitter announced to be <5%. Fake and spam accounts could lead to undesirable social interactions with peers, unwanted peer-shared information, and an unpleasant web-based social environment. Twitter needs to use new procedures to detect spam and fake accounts and better control the functionality of Twitter bots to provide a more appropriate web-based social platform. This new mechanism could automatically limit the visibility of private posts containing highly sensitive health information to everyone, even to people who users follow. On the basis of the current Twitter privacy policy, people can mention who can reply to a specific tweet. However, it is hard to confirm who can actually see the posts because of bots and recommendation agents.

Social bots use computer algorithms to artificially create content and interact with people on social media [92]. Twitter bots can be manipulative and purposely change people’s attitudes and opinions about a topic [93]. For instance, bots can share posts with peers who are not following a user but usually read posts with health information content. The existence of bots may be useful for sharing general information but can be harmful because it can increase exposure to private posts with sensitive health information. A plausible recommendation is to add a new category for sensitive content (such as specific health information) besides the photo, graphics interchange format (GIF), and poll categories. Then, users can create a new circle of people who are allowed to see, reply to, or share these sensitive posts. Users can also customize the configuration and limit the possible unwanted interactions by selecting who can see the posts but cannot share them. This small crowd can be saved for future use and can be easily modified later. The advantage of this new category is that people are notified to customize their Twitter circle depending on different content (eg, highly sensitive, semisensitive, and nonsensitive). For instance, a user can select everyone to see and comment on posts containing information about cryptocurrency, high-technology companies, or humanitarian issues. In contrast, the user can select a group of followers to see their thoughts about general health information and choose only a few close friends to see and comment on posts with sensitive health information.

Stringent privacy policies are required to enable Twitter users to limit who (ie, peers) can view, comment on, and share web-based content. People should be able to easily edit with whom they can share health information to exercise control over their personal digital information. Spambots on Twitter should also be controlled, modified, or filtered because they can involve potentially deceptive, harmful, or annoying activities. A more transparent policy is required to detect and deactivate invasive Twitter bots that can automatically like or retweet users’ postings without consent.

Finally, the insights from our study are not only limited to Twitter but also have implications for other social media platforms where users might share health information. Such platforms should recognize the significant role of peer-related privacy concerns and consider introducing similar audience control features. Health professionals and health-related organizations using social media for patient engagement should also be aware of these concerns and take steps to ensure that their communication respects patient privacy. Policy makers should consider our findings when developing regulations for health information sharing on social media to ensure that they address the most significant privacy concerns.

**Limitations and Future Studies**

Our study also has some limitations that can be considered as opportunities for future studies. First, a web-based survey through MTurk was used to collect data, which may be biased toward people familiar with crowdsourcing platforms. Future studies can use other data collection and sampling strategies, such as collecting data directly from Twitter. Second, we collected data from 329 Twitter users, which may not be a good representative of Twitter users. Next, studies can increase the sample size to reduce sampling bias and improve the generalizability of the findings. Third, we did not consider the effects of cultural dimensions (such as individualism, uncertainty avoidance, etc) on sharing health information on Twitter. It can be interesting for future studies to explore the effects of culture on disclosing different types of health information on social media. Fourth, our study tests a model to analyze health information sharing from the perspective of privacy concerns. However, there may be other essential variables. More studies are required to examine other factors inhibiting and promoting sharing behaviors on social media, such as reputation, incentives, trust, stigma, and social support. Fifth, this study did not examine the accuracy of the health information shared on Twitter or the risks of misinformation because it is not within the scope of this study. Future studies could expand upon our results to investigate the role of misinformation risks in information-sharing behaviors. Finally, our study focused on Twitter as a study context. We encourage future studies to extend the proposed model to other social media platforms (eg, Facebook, TikTok, and Instagram) where web-based interactions with peers are essential.

**Conclusions**

This study provides insights into health information sharing on Twitter from a privacy perspective. The findings propose that including CFIP and PrPC constructs can help in better conceptualization of information privacy concerns in the context of social media. The integration of these 2 aspects of information privacy can expand the discussion about internet privacy by addressing the privacy needs associated with the practices of companies, such as collection, unauthorized secondary use, improper access, and errors. It also considers psychological privacy concerns, communication privacy concerns, peers’ sharing behaviors, and territory privacy concerns related to peers in such interpersonal interactions. This interactive approach can provide a more comprehensive analysis of information privacy (related to web-based vendors and web-based peers) and adds...
a more substantial explanation of privacy needs on social media channels (such as Twitter). Privacy concerns may not always prohibit disclosure behaviors on Twitter; it depends on the type of health information. The findings demonstrate that peer-related privacy concerns are more salient to predicting general and specific health information sharing on Twitter than privacy concerns related to companies and third parties. The results could propose practical contributions by shedding more light on the negative impacts of web-based peer behaviors on losing personal control over digital communications and information access. Privacy policies should focus on companies’ practices, such as sharing users’ information with third parties for big data analytics. We suggest mitigating privacy concerns and promoting health information sharing on Twitter by creating policies that tailor privacy needs to the type of health information shared (ie, general or specific).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Web-based survey—measuring items.
[DOCX File; 15 KB - formative_v8i4e5573_app1.docx]

References


Abbreviations

AVE: average variance extracted
Edited by A Mavragani; submitted 07.01.23; peer-reviewed by M Abdalla, Z Zandesh, J Li, YG Kim, Z Wang; comments to author 17.07.23; revised version received 19.07.23; accepted 14.12.23; published 12.01.24.

Please cite as:
Esmaeilzadeh P
Privacy Concerns About Sharing General and Specific Health Information on Twitter: Quantitative Study
JMIR Form Res 2024;8:e45573
URL: https://formative.jmir.org/2024/1/e45573
doi:10.2196/45573
PMID:38214964

©Pouyan Esmaeilzadeh. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Evaluation of the Effectiveness of Suicide.ca, Quebec’s Digital Suicide Prevention Strategy Platform: Cross-Sectional Descriptive Study

Louis-Philippe Côté¹, PhD; Julie Lane², PhD

¹Centre for Research and Intervention on Suicide, Ethical Issues and End-of-life practices, Université du Québec à Montréal, Montreal, QC, Canada
²Centre RBC d’expertise universitaire en santé mentale, Université de Sherbrooke, Sherbrooke, QC, Canada

Corresponding Author:
Louis-Philippe Côté, PhD
Centre for Research and Intervention on Suicide, Ethical Issues and End-of-life practices
Université du Québec à Montréal
100 rue Sherbrooke Ouest
Montreal, QC, H2X 3P2
Canada
Phone: 1 (514) 987 4832
Email: cote.louis-philippe.3@courrier.uqam.ca

Abstract

Background: In 2017, the Quebec government assigned the Association québécoise de prévention du suicide (AQPS) to develop a digital suicide prevention strategy (DSPS). The AQPS responded by creating a centralized website that provides information on suicide and mental health, identifies at-risk individuals on the internet, and offers direct crisis intervention support via chat and text.

Objective: This study aims to evaluate the effectiveness of suicide.ca, Quebec’s DSPS platform.

Methods: This study used a cross-sectional descriptive design. The study population comprised internet users from Quebec, Canada, who visited the suicide.ca platform between October 2020 and October 2021. Various data sources, such as Google Analytics, Firebase Console, and Customer Relation Management data, were analyzed to document the use of the platform. To understand the profile of suicide.ca users, frequency analyses were conducted using data from the self-assessment module questionnaires, the intervention service’s triage questionnaire, and the counselors’ intervention reports. The effectiveness of the platform’s promotional activities on social media was assessed by examining traffic peaks. Google Analytics was used to evaluate the effectiveness of AQPS’ strategy for identifying at-risk internet users. The impact of the intervention service was evaluated through an analysis of counselors’ intervention reports and postintervention survey results.

Results: The platform received traffic from a diverse range of sources, with promotional efforts on social media directly contributing to the increased traffic. The requirement of a user account posed a barrier to the use of the mobile app, and a triage question that involved personal information led to a substantial number of dropouts during the intervention service triage. AdWords campaigns and fact sheets addressing suicide risk factors played a crucial role in driving traffic to the platform. With regard to the profile of suicide.ca users, the findings revealed that the platform engaged individuals with diverse levels of suicidal risk. Notably, users of the chat service displayed a higher suicide risk than those who used the self-assessment module. Crisis chat counselors reported a positive impact on approximately half of the contacts, and overall, intervention service users expressed satisfaction with the support they received.

Conclusions: A centralized digital platform can be used to implement a DSPS, effectively reaching the general population, individuals with risk factors for suicide, and those facing suicidal issues.

(JMIR Form Res 2024;8:e46195) doi:10.2196/46195

KEYWORDS
suicide prevention; public health; information and communication technology; digital mental health; helpline; digital strategy; communication technology; information technology; suicide; psychoeducation; mobile app; suicide risk; risk factor; users; mental health; text; website; prevention strategy; prevention; Google Analytics; Canada; Quebec; questionnaire; mobile phone
**Introduction**

**Overview**

Information and communication technologies (ICTs) are increasingly used for suicide prevention. As demonstrated in a scoping study by Rassy et al. [1], ICTs are used in a multitude of universal, selective, and indicated suicide prevention strategies.

As defined by the World Health Organization (WHO) [2], universal prevention strategies are designed to reach a whole population. They aim to optimize health for all and minimize the risk of suicide by removing barriers to care, expanding access to support services, and strengthening protective factors. Universal suicide prevention strategies using ICTs typically aim to promote mental health literacy, health behaviors, and help seeking [1]. They often take the form of psychoeducational websites and web-based awareness campaigns. Selective prevention strategies, in contrast, are aimed at specific groups or communities [2]. They can be developed based on sociodemographic or geographic characteristics or according to the prevalence of risk factors [3]. They may target individuals who do not exhibit suicidal behaviors but who are at risk for suicide [2]. Selective suicide prevention strategies using ICTs focus on identifying suicidal internet users, offering mental health self-management programs, and providing web-based training [1]. These strategies are, for the most part, included in broader programs involving direct contact between distressed individuals and health professionals. Finally, indicated prevention strategies focus on individuals who have suicidal ideation or who have previously attempted suicide [2]. Indicated suicide prevention strategies using ICTs include web-based suicide risk assessment and triage systems, web-based psychotherapy, chat- or text-based crisis interventions, and digital tools to support face-to-face interventions [1].

Despite the vast body of scientific literature discussing the application of ICT in suicide prevention, few studies have addressed how multiple ICT-based interventions can be combined and articulated together in the development of a digital suicide prevention platform [3] or, more broadly, as part of a digital suicide prevention strategy (DSPS). In a context where an increasing number of countries are developing digital health strategies [4], guidelines must be developed to guide organizations that wish to develop a DSPS. To deploy the 3 levels of prevention defined by the WHO (universal, selective, and indicated), it is crucial that these guidelines contain information on the best strategies for reaching the general population, identifying people with risk factors for suicide on the internet, and providing direct support to people struggling with suicidal issues.

Rassy et al [1] discussed ICT-based interventions aimed at a broad audience, including educational websites, web-based awareness campaigns, and social media platforms. They spotlight the *It Gets Better* campaign launched in 2010, primarily directed at bullied teenagers, especially those identifying as gay, to prevent suicide. The campaign involved videos created by gay adults that conveyed the message that these teens’ lives would improve over time. These videos were hosted on an educational website and YouTube channel and extensively shared on social media, accumulating >50 million views [5]. Despite the limited evaluation of the effects of educational websites, web-based awareness campaigns, and social media initiatives, studies have demonstrated promising outcomes in terms of reducing suicidal ideation and behavior as well as enhancing knowledge, attitudes, and help-seeking behavior [1]. As highlighted by Rassy et al [1], further research is required to determine the specific characteristics of these interventions that are associated with positive outcomes. In addition, there is a need to develop effective methods for identifying beneficial websites and social media posts and to inform users about or guide them toward internet content that may be of help to them.

Using Google searches holds promise in identifying individuals at risk of suicide on the internet and guiding them toward beneficial web-based content, as individuals with suicidal ideations tend to search for suicide-related information on the internet [6-8]. Indeed, multiple studies have established a correlation between suicide-related Google searches and suicide rates across diverse populations [9-11]. The findings from these studies suggest that suicide-related Google searches can be used to identify individuals at risk of suicide on the internet, recognize intervention opportunities, and facilitate real-time monitoring of suicide risk at a population level. However, it is important to acknowledge that only a small number of real-world interventions based on Google searches have been developed to date.

An example of such an intervention is the study conducted by Sueki et al [12], in which they implemented a web-based gatekeeping intervention in Japan. This intervention involved targeted web-based advertisements directed at individuals searching for suicide-related terms on the web, redirecting them to an email-based consultation. The results of the study indicated a significant reduction in users’ suicidal ideation 4 weeks after being identified on the internet and redirected to the email-based intervention. Furthermore, Onie et al [13] conducted 2 studies to assess engagement levels with 2 advertisement campaigns on Google targeting individuals contemplating suicide. The researchers noted significant engagement with both campaigns, as evidenced by high click-through and conversion rates. Similarly, Google displays crisis helpline numbers at the top of the search results when a user performs a suicide-related search. To our knowledge, no empirical data on the actual impact of Google initiative on helplines use worldwide have been published to date. In addition, data on the characteristics of individuals who were identified by Google and directed to help services are also unavailable.

Regarding ICT-based interventions aimed at providing direct support to individuals facing suicidal issues, crisis lines that offer text- and chat-based services are likely to be the primary source of assistance for those seeking help on the internet. Research on the use and effectiveness of crisis text- and chat-based services is limited. Studies so far have shown that they seem effective in reducing users’ distress [14-18]. However, they reported quite different results regarding the suicidality of users. For example, evaluation of the Netherlands’ 113Online chat service [16] and the US Lifeline Crisis Chat [16] suggested that >80% of their users were experiencing suicidal thoughts.
In contrast, evaluation of crisis text lines in the United States and Australia [17,18] indicated that the proportion of users experiencing suicidality was <30%.

In addition, studies suggest that conducting suicide risk assessments during chat- and text-based interactions poses challenges, resulting in fewer assessments completed by counselors [14,19]. These gaps in suicide risk assessments can impact the identification of callers at risk of suicide and potentially bias the reported statistics on suicidal callers by these services. Consequently, providing an accurate description of suicidality among users of these services becomes challenging. Therefore, it is imperative for crisis lines offering text- and chat-based interventions to establish strategies that enable a precise depiction of service users and an assessment of their level of suicidality, rather than solely relying on intervention reports completed by counselors.

In a similar vein, Zabelski et al [20] emphasized the need to investigate the accessibility of chat and text services for high-risk populations in their policy focus review of suicide prevention research on crisis lines. They suggested that if the effectiveness of these services is established, additional studies should be conducted to better understand the demographics and risk profiles of service users as well as to develop effective strategies to promote these services among high-risk populations.

This brief overview of ICT-based interventions highlights the potential of ICTs in the field of suicide prevention. However, it also underscores the existence of significant knowledge gaps that can have important implications for the development of DSPSs. Moreover, essential information required by mental health resources seeking to incorporate ICT-based interventions into their clinical practices is frequently absent from scholarly articles. For example, the technical aspects of developing, maintaining, and operating digital interventions have rarely been discussed [21,22]. Furthermore, there is a dearth of comprehensive research that adequately examines the implementation characteristics of ICT-based interventions in real-world settings, encompassing both the obstacles and factors that aid their implementation [23,24]. Ultimately, to support well-informed decision-making regarding the development of DSPSs, it is vital to establish whether ICT-based suicide prevention strategies are effectively reaching various target audiences or whether they are reaching the same individuals through different means. Therefore, research is necessary to explore digital platforms that incorporate a variety of ICT-based interventions, evaluate the presence of diverse user profiles on these platforms, and analyze the specific interventions they use.

**Background**

In 2017, the Quebec government asked the Association québécoise de prévention du suicide (AQPS) to set up a DSPS. This DSPS was intended to inform the population about suicide and mental health resources, identify people at risk for suicide on the internet, and offer web-based intervention to people who rarely use traditional help services.

The AQPS then launched a broad consultation process with Quebec suicide prevention stakeholders [7]. The objective of this consultation was to paint a picture of the current practices and needs regarding the use of digital technologies in suicide prevention in Quebec. In general, the consultation participants stressed the importance of developing new digital services by building on the programs, services, and intervention practices already in place in Quebec and of using a centralized website that would make it possible to meet several objectives (inform, identify, and intervene).

In this context, the AQPS, in partnership with its technology partner (the Montreal-based technology firm Nventive), developed the suicide.ca digital platform [25]. Suicide.ca is the official digital platform of the Quebec’s DSPS. Through this platform, a multitude of universal, selective, and indicated suicide prevention strategies are deployed.

**Methods**

**Suicide.ca Platform**

This section describes the universal, selective, and indicated suicide prevention strategies implemented through the suicide.ca platform.

**Suicide.ca’s Universal Prevention Strategies**

The components of the suicide.ca platform that enable the deployment of universal prevention strategies are its search engine optimization (SEO), the psychoeducation pages, the blog pages, and the promotion of the platform on social media.

**SEO Strategy**

To fulfill its mission, a digital suicide prevention platform must be known to the public and easily found on the internet. This is why positioning itself favorably in search engine results is at the heart of the suicide.ca platform’s promotional strategy. To do so, an SEO strategy was developed. SEO refers to all the aspects of developing, maintaining, and promoting a website on search engines. For example, all the content on the platform was carefully crafted with the explicit objective of prioritizing search engine ranking.

**Psychoeducation Pages**

When the user arrives at the platform’s home page, they are prompted to choose one of the following sections: *I am thinking about suicide, I’m worried about someone with suicidal thoughts, or I’m grieving a loss by suicide*. The section for suicidal people is presented as a guide to getting better. The content of the psychoeducation pages can be found on the platform [26].
Blog Pages
The suicide.ca platform includes a blog. The blog facilitates the regular generation of fresh content, thereby aiding the enhancement of the platform's natural search engine rankings. It also provides a great deal of flexibility in terms of the topics covered. Content is sometimes developed based on frequent searches by suicide.ca users on search engines.

Promotion of the Platform on Social Media
The AQPS promotes the platform via posts on social media. These posts are generally aimed at promoting the intervention service and communicating information found on the platform. Content from the psychoeducation pages and blog pages is regularly used to create new posts.

Suicide.ca’s Selective Prevention Strategies
The components of the suicide.ca platform that enable the deployment of selective prevention strategies are the fact sheets on suicide risk factors, AdWords campaigns, the self-assessment module, the self-management tools, and the mobile app.

Fact Sheets on Risk Factors Associated With Suicide
The suicide.ca platform includes a section with numerous fact sheets on the risk factors associated with suicide. These fact sheets address mental health problems (eg, depression, anxiety, and schizophrenia) and difficult life situations (eg, relationship problems, money problems, and isolation). These fact sheets have a dual function: to help users cope with their situation and to identify the internet users who present risk factors for suicide on the web.

AdWords Campaigns
AdWords is a paid referencing system developed by Google. This allows organizations to promote their services using paid advertisements. These advertisements appear when internet users perform searches with keywords or expressions that are associated with the services offered. The AQPS frequently uses AdWords campaigns to identify internet users who are conducting research related to suicide, depression, or anxiety.

Self-Assessment Module
A self-assessment module consisting of French-validated questionnaires (refer to the details in the Data Collection Methods and Tools section) is available to users on the Taking Stock of Your Mental Health page.

The self-assessment is performed in 2 stages. During the first stage, users are invited to answer questionnaires allowing them to take stock of their suicidal thoughts, to identify the presence of signs of psychological distress, and to evaluate their positive mental health. If the result of this first evaluation reveals that the user is experiencing suicidal ideation, they are invited to contact the platform’s intervention service via a redirection button. If the user is not referred to the intervention service and the results indicate that they may have a mental health disorder (Kessler Psychological Distress Scale [K10] score [27]), additional questionnaires are offered to identify the presence of symptoms related to various mental health disorders.

Suicide.ca’s Indicated Prevention Strategies
The components of the suicide.ca platform that enable the deployment of indicated prevention strategies are the chat and text intervention service, the triage system, the suicide risk assessment tools, and the follow-up emails.

Intervention Service by Chat and Text Messaging
Suicide.ca is a digital intervention service that operates 24 hours a day, 7 days a week. It is operated by the AQPS and 3 suicide prevention centers. The service can be accessed via the web, the mobile app, or by SMS text messaging.

Triage
The intervention service triage system assigns a priority rating to contact requests. The first questions ask about age, gender, and reason for contact. Individuals at the risk of suicide are then asked, Are you planning to attempt suicide in the coming minutes or hours? If the user answers yes, the triage ends, and they are sent to the queue with a priority rating. If the user answers no or unsure, they are prompted to complete the Suicidal Ideation Attributes Scale—French version (SIDAS-FR [28]). Finally, the user may, if they wish, specify in a few words the reason for contacting the service.

Follow-Up Emails
At the end of the exchange, the counselors can ask users if they would like to receive a follow-up email. These follow-up emails are generally used when the user anticipates a difficult event that could lead to a suicidal crisis (eg, court appearance and meeting with an ex-spouse).

Suicide.ca Prevention Model
Using the classification proposed by Rassy et al [1], Textbox 1 illustrates how the different components of the platform are complementary and how they deploy universal, selective, and indicated prevention strategies.
Textbox 1. Suicide.ca prevention model.

<table>
<thead>
<tr>
<th>Universal prevention strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Search engine optimization</td>
</tr>
<tr>
<td>• Psychoeducation pages</td>
</tr>
<tr>
<td>• Active promotion of the platform on social media</td>
</tr>
<tr>
<td>• Blog pages</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selective prevention strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Fact sheets on suicide risk factors</td>
</tr>
<tr>
<td>• AdWords campaigns</td>
</tr>
<tr>
<td>• Self-assessment</td>
</tr>
<tr>
<td>• Self-management tools and mobile app</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indicated prevention strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Intervention service by chat and text</td>
</tr>
<tr>
<td>• Triage</td>
</tr>
<tr>
<td>• Follow-up email</td>
</tr>
</tbody>
</table>

Study Design
The evaluation of the suicide.ca platform used a cross-sectional design to document the use of the suicide.ca platform; to describe the profile of users; and to estimate the effectiveness of universal, selective, and indicated prevention strategies deployed through the platform.

Population
The population under study in the evaluation of the suicide.ca platform were the internet users who visited the suicide.ca platform.

Recruitment
This study was based on the analysis of secondary data collected by AQPS. Data were collected from internet users from Quebec, Canada, who visited the suicide.ca platform between October 2020 and October 2021.

Data Collection Methods and Tools
Overview
The evaluation of the suicide.ca platform was carefully considered during its development. The questionnaires integrated into the platform’s various components made it possible to measure the platform’s impact. The platform collects various types of data.

Google Analytics
Google Analytics is a statistical tool offered by Google. It allows website administrators to analyze the behavior of their users. The analytics used in this study are the number of times different pages of the platform were visited, the number of visitors (withstanding the number of visits), the sources of traffic acquisition on the platform (eg, search engines, social networks, and hyperlinks on other websites), and the landing pages (the first pages visited by users).

Firebase Console
The Firebase Console platform is a mobile app design tool supported by Google. It was used to count the number of times the My Tools app was downloaded and the number of user accounts created.

Self-Assessment Module Questionnaires
The self-assessment module consists of a series of validated questionnaires in French (Table 1). The results of the self-assessments are recorded in the platform’s Customer Relationship Management (CRM).
Table 1. Self-assessment questionnaires.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Screening questionnaires</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive mental health</td>
<td>MHC-SF&lt;sup&gt;a&lt;/sup&gt; [29]</td>
</tr>
<tr>
<td>Psychological distress</td>
<td>K10&lt;sup&gt;b&lt;/sup&gt; [27]</td>
</tr>
<tr>
<td>Suicidal ideation</td>
<td>SIDAS-FR&lt;sup&gt;c&lt;/sup&gt; [28]</td>
</tr>
<tr>
<td>Depression</td>
<td>PHQ-9&lt;sup&gt;d&lt;/sup&gt; [30]</td>
</tr>
<tr>
<td>Bipolarity</td>
<td>MDQ&lt;sup&gt;e&lt;/sup&gt; [31,32]</td>
</tr>
<tr>
<td>Social phobia</td>
<td>SPIN&lt;sup&gt;f&lt;/sup&gt; [33]</td>
</tr>
<tr>
<td>Generalized anxiety</td>
<td>PSWQ&lt;sup&gt;g&lt;/sup&gt; [34,35]</td>
</tr>
<tr>
<td>Addiction: alcohol</td>
<td>DÉBAA&lt;sup&gt;h&lt;/sup&gt; [36]</td>
</tr>
<tr>
<td>Addiction: drugs</td>
<td>DÉBAD&lt;sup&gt;i&lt;/sup&gt; [36]</td>
</tr>
<tr>
<td>Addiction: gambling</td>
<td>DÉBA-Jeu [36]</td>
</tr>
<tr>
<td>Dependency: screens</td>
<td>DÉBA-internet&lt;sup&gt;k&lt;/sup&gt; [37]</td>
</tr>
<tr>
<td>Posttraumatic stress</td>
<td>PCL-5&lt;sup&gt;l&lt;/sup&gt; [38]</td>
</tr>
<tr>
<td>Social support</td>
<td>SPS&lt;sup&gt;m&lt;/sup&gt; [39,40]</td>
</tr>
<tr>
<td>Reasons to live</td>
<td>RFL&lt;sup&gt;n&lt;/sup&gt; [41,42]</td>
</tr>
</tbody>
</table>

<sup>a</sup>MHC-SF: Mental Health Continuum Short Form.
<sup>b</sup>K10: Kessler Psychological Distress Scale.
<sup>c</sup>SIDAS-FR: Suicidal Ideation Attributes Scale—French version.
<sup>d</sup>PHQ-9: Patient Health Questionnaire-9.
<sup>e</sup>MDQ: Mood Disorder Questionnaire.
<sup>f</sup>SPIN: Spin Phobia Inventory.
<sup>g</sup>PSWQ: Penn State Worry Questionnaire.
<sup>h</sup>DÉBAA: Dépistage et Évaluation du Besoin d’Aide—Alcool.
<sup>i</sup>DÉBAD: Dépistage et Évaluation du Besoin d’Aide—Drogues.
<sup>j</sup>DÉBA-Jeu: Dépistage et Évaluation du Besoin d’Aide—Jeu.
<sup>k</sup>DÉBA-internet: Dépistage et Évaluation du Besoin d’Aide—internet.
<sup>l</sup>PCL-5: PTSD Checklist for DSM-5.
<sup>m</sup>SPS: Social Provisions Scale.
<sup>n</sup>RFL: Reasons for Living Inventory.

**Triage Questionnaire**

The triage questionnaire includes inquiries about age, gender, reason for contact, whether the individual is planning to attempt suicide in the near future, and the SIDAS-FR. The results of the triage questionnaire are recorded in the platform’s CRM.

**Intervention Reports**

In this study, the data used from the intervention reports were the reasons for contact, the needs and issues discussed with the users, and the users’ suicide risk ratings assigned by the counselors (suicide risk ratings: no foreseeable risk for a suicide attempt in the near future, low foreseeable risk for a suicide attempt in the near future, high foreseeable risk for a suicide attempt in the near future, and imminent danger of suicide attempt or attempt in progress). The categories of needs and issues discussed are not mutually exclusive. The intervention report also contains a section where the counselor is asked to indicate whether they think the user feels better and is more able to cope with their difficulties as a result of the intervention (response choice=Yes, unsure, and No). These intervention reports are recorded in the platform’s CRM.

**Postintervention Questionnaire**

This questionnaire includes 9 items in the form of 4-point Likert scales (1=strongly agree to 4=strongly disagree). The items focus on users’ perceptions of the quality of interventions and their impact on distress, feeling of being able to cope with the situation, and suicidal ideation. The items are illustrated in detail in the Results section.

**Analysis Methods**

For each objective, several data sources were consulted, and different types of analyses were performed.
Objective 1: To Document the Use of the Suicide.ca Platform

To document platform use, Google Analytics data from the suicide.ca website [26] were analyzed to describe platform traffic and the number of visits to the psychoeducation pages. The Firebase Console data from the My Tools mobile app were analyzed to document the number of downloads of the mobile app and the number of user accounts created. CRM data were analyzed to describe the number of people who completed the self-assessment, the number of people who accessed the triage of the intervention service, and the number of interventions provided.

Objective 2: To Describe the Profile of Users of the Suicide.ca Platform

To describe the profile of platform users, frequency analyses were conducted using data from the self-assessment questionnaires, the triage questionnaire, and the reasons for contact recorded in the intervention reports.

Objective 3: To Describe the Impact of Universal, Selective, and Indicated Prevention Strategies Deployed Through the Suicide.ca Platform

The impact of the platform’s promotional activities on social media was investigated through an analysis of the platform’s traffic peaks. First, a visual inspection of the traffic curve was performed using Google Analytics. It was then decided to define the traffic peaks as days with >600 visits. This threshold corresponded to twice the average traffic and days when the increase in traffic was evident in the Google Analytics graph. AQPS internal documentation was then inspected to determine if these dates corresponded to dates when AQPS conducted promotional activities on social media or traditional media (eg, web-based publications, radio, or television interviews). The data analyzed to assess the impact of AQPS promotional activities cover the period from October 15, 2020, to October 17, 2021.

With respect to the AdWords campaigns and the suicide risk factor fact sheets, 2 analytical strategies were used to estimate their effectiveness. The first strategy was to use Google Analytics to describe the traffic acquisition sources of suicide.ca. In digital marketing, traffic acquisition sources identify how visitors to a website accessed a site (eg, via search engine queries, social media posts, AdWord advertisements [43]). Thus, traffic acquisition sources were used to describe the number and proportion of users who arrived at the site through an AdWords campaign. The second strategy was to use Google Analytics to identify the top landing pages on the platform. A landing page is the first page a user views when they arrive on the platform (eg, risk factor information page, psychoeducation page, and resource information page). In digital marketing, landing pages are used to better understand what users were looking for on the internet before they arrived at a site or what caught their attention [43]. For example, landing pages were used to describe the number and proportion of users who began their site visit by viewing a suicide risk factor fact sheet. The conversion rate for the AdWords campaigns was calculated based on the number of individuals who clicked on the advertisement and initiated the crisis chat triage. The analytical data on AdWords campaigns and fact sheets cover the period from December 31, 2020, to October 31, 2021.

The impact of the intervention service was assessed via an analysis of counselors’ Perceptions of the impact of their interventions (recorded in intervention reports) and the postintervention survey results. Approximately 3473 interventions were provided between the postintervention survey going live in late January 2021 and October 2021. Therefore, we estimated the response rate to the postintervention survey to be approximately 10%. As the SMS text messaging intervention service was in development at the time of data collection, only the data for the chat-based intervention are presented.

Ethical Considerations

This study is based on the secondary analysis of anonymous data collected by AQPS. Consent for primary data collection was obtained through a waiver form in suicide.ca’s self-assessment module and a consent form in the postintervention questionnaire of the intervention service. These forms specify that data collection is for the purpose of continuous service improvement and that any data shared with third parties will be anonymized and subjected to ethics approval by an institutional research ethics committee. The participants were not offered any compensation for their involvement in the study. Ethics approval for the secondary analysis of all data presented in this study was obtained from the institutional review board of the Centre intégré universitaire de santé et de services sociaux de l’Estrie—Centre hospitalier universitaire de Sherbrooke (project 2021-4016).

Results

In this section, the results concerning platform use; the profile of platform users; and the effectiveness of universal, selective, and indicated prevention strategies deployed through the suicide.ca platform are presented.

Use of the Suicide.ca Platform

Overview

A total of 127,495 visits were made to the suicide.ca platform from October 15, 2020, to October 17, 2021. The average number of visits per day was 346.45 (median 313.5, IQR 257-389). The least busy day had 123 visits, and the busiest day had 1774 visits.

Visits to Psychoeducation Pages

The number of visits to the psychoeducation pages of the 3 main sections of suicide.ca were much lower than the total number of visits to the site. Moreover, the number of page views dropped significantly after the title pages of all 3 sections were visited (Table 2).
### Table 2. Suicide.ca psychoeducation pages.

<table>
<thead>
<tr>
<th>Psychoeducation pages</th>
<th>Visits to the page or website (n=127,495), n (%)</th>
<th>Users who visited the page or website (n=100,114), n (%)</th>
<th>Average visit duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;I’m thinking about suicide&quot; (title page)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taking stock of your mental health</td>
<td>5312 (4.17)</td>
<td>4478 (4.47)</td>
<td>124.66</td>
</tr>
<tr>
<td>Taking care of yourself</td>
<td>3148 (2.47)</td>
<td>2642 (2.64)</td>
<td>106.90</td>
</tr>
<tr>
<td>Finding support services</td>
<td>2051 (1.16)</td>
<td>1778 (1.78)</td>
<td>153.31</td>
</tr>
<tr>
<td>Talking about it with your loved ones</td>
<td>1238 (0.97)</td>
<td>1098 (1.09)</td>
<td>120.84</td>
</tr>
<tr>
<td>&quot;I’m worried about someone with suicidal thoughts&quot; (title page)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognizing signs of distress</td>
<td>5546 (4.35)</td>
<td>4306 (4.30)</td>
<td>105.44</td>
</tr>
<tr>
<td>Supporting someone in difficulty</td>
<td>1788 (1.40)</td>
<td>1573 (1.58)</td>
<td>161.44</td>
</tr>
<tr>
<td>Talking about suicide with someone</td>
<td>1544 (1.21)</td>
<td>1344 (1.34)</td>
<td>161.06</td>
</tr>
<tr>
<td>Respecting your limits and taking care of yourself</td>
<td>1056 (0.83)</td>
<td>955 (0.95)</td>
<td>124.29</td>
</tr>
<tr>
<td>Helping someone recover</td>
<td>1012 (0.79)</td>
<td>872 (0.87)</td>
<td>150.34</td>
</tr>
<tr>
<td>&quot;I’m grieving a loss by suicide&quot; (title page)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning about a death by suicide</td>
<td>857 (0.67)</td>
<td>695 (0.69)</td>
<td>77.92</td>
</tr>
<tr>
<td>Understanding suicide grief</td>
<td>812 (0.63)</td>
<td>653 (0.65)</td>
<td>90.88</td>
</tr>
<tr>
<td>Suicide grief and children</td>
<td>591 (0.46)</td>
<td>441 (0.44)</td>
<td>28.77</td>
</tr>
<tr>
<td>When to seek professional help</td>
<td>538 (0.42)</td>
<td>423 (0.42)</td>
<td>113.32</td>
</tr>
<tr>
<td>Resuming your life after grieving a death by suicide</td>
<td>451 (0.35)</td>
<td>377 (0.37)</td>
<td>129.43</td>
</tr>
<tr>
<td>Talking about it with family and friends</td>
<td>269 (0.21)</td>
<td>250 (0.25)</td>
<td>154.20</td>
</tr>
</tbody>
</table>

**Number of People Who Completed the Self-Assessment Module**

With regard to self-assessment, 2488 users completed the SIDAS-FR. Moreover, 3100 users completed the K10 questionnaire, and 2644 users completed the Mental Health Continuum Short Form questionnaire. Overall, 1471 users completed phase 2 of the self-assessment module.

**Number of Downloads of the Mobile App Versus Number of User Accounts Created**

The analysis of downloads of the mobile app and the creation of user accounts was carried out for the period from November 23, 2020, to October 17, 2021. During this period, there were 5159 downloads of the mobile app, and 2178 user accounts were created.

**Number of Crisis Interventions Provided and Triage Dropout Rate**

From October 15, 2020, to October 17, 2021, suicide.ca's intervention service received 6152 intervention requests. Of these 6152 requests, 1540 (25.02%) did not result in an intervention. The intervention reports provided reasons for not being able to intervene in 1321 of these cases. The most common explanation was the lack of response from the users at the beginning of the contact (840/1321, 63.59%).

Analysis of the intervention service triage data was conducted for the period October 15, 2020, through July 25, 2021 (Figure 1). As the gender question was optional, it was not included in the analysis. Of the 6300 users who responded to the first question, only 2711 (42.86%) completed the triage. The last question, which asked for the user’s first name or alias, resulted in a significant dropout rate. Specifically, only 46.54% (2711/5825) of the users who were presented with the question actually provided a response. As for the questions about suicidality, they resulted in few dropouts. Age, reason for contact, and plan to attempt suicide in the coming minutes or hours were not associated with the probability of abandonment during triage.
Profile of Users of the Suicide.ca Platform

To describe the profile of platform users, frequency analyses were conducted using data from the self-assessment questionnaires, the triage questionnaire, and the reasons for contact recorded in the intervention reports.

Characteristics of Individuals Using the Self-Assessment Module

Table 3 indicates the results of the users who undertook a self-assessment. With regard to suicidality, the average score of the 2488 users who completed the SIDAS-FR questionnaire was 13.47 (SD 14.04; median 10). Among them, 868 (34.89%) reported no suicidal ideation in the last 30 days, whereas 825 (33.16%) reported severe suicidal ideation. In terms of psychological distress, of the 3100 users who completed the K10 questionnaire, 2721 (87.78%) obtained a score indicating the presence of a mental health disorder. With regard to positive mental health, of the 2644 users who completed the Mental Health Continuum Short Form questionnaire, 1024 (38.73%) exhibited languishing mental health.

In the 1471 users who completed phase 2 of the self-assessment, depressive and anxiety disorders were the most prevalent. Specifically, 62.2% (915/1471) of the users obtained a Patient Health Questionnaire-9 score indicating the presence of depression, 61.73% (908/1471) of the users obtained a Penn State Worry Questionnaire score indicating the presence of generalized anxiety disorder, and 53.2% (451/847) of the users obtained a Spin Phobia Inventory score indicating the presence of social phobia disorder.
Table 3. Results of the completed self-assessments.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Self-assessment, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severity of suicidal ideation (SIDAS-FR(^a); n=2488)</strong></td>
<td></td>
</tr>
<tr>
<td>Absence of suicidal ideation (0)</td>
<td>868 (34.89)</td>
</tr>
<tr>
<td>Low severity (1-12)</td>
<td>482 (19.37)</td>
</tr>
<tr>
<td>Medium severity (13-19)</td>
<td>313 (12.58)</td>
</tr>
<tr>
<td>High severity (&gt;20)</td>
<td>825 (33.16)</td>
</tr>
<tr>
<td><strong>Psychological distress (K10(^b); n=3100)</strong></td>
<td></td>
</tr>
<tr>
<td>Probable presence of a mental health problem</td>
<td>2721 (87.77)</td>
</tr>
<tr>
<td><strong>Positive mental health (MHC-SF(^c); n=2644)</strong></td>
<td></td>
</tr>
<tr>
<td>Flourishing mental health</td>
<td>256 (9.68)</td>
</tr>
<tr>
<td>Moderate mental health</td>
<td>1364 (51.59)</td>
</tr>
<tr>
<td>Languishing mental health</td>
<td>1024 (38.73)</td>
</tr>
<tr>
<td><strong>Psychological disorders (n=1471)</strong></td>
<td></td>
</tr>
<tr>
<td>Depression (PHQ-9(^d))</td>
<td>915 (62.2)</td>
</tr>
<tr>
<td>Bipolarity (MDQ(^e))</td>
<td>89 (6.05)</td>
</tr>
<tr>
<td>Social phobia (SPIN(^f); n=847)</td>
<td>451 (53.2)</td>
</tr>
<tr>
<td>Generalized anxiety (PSWQ(^g))</td>
<td>908 (61.73)</td>
</tr>
<tr>
<td>Addiction: alcohol (DÉBBA(^h))</td>
<td>149 (10.13)</td>
</tr>
<tr>
<td>Addiction: drugs (DÉBAD(^i))</td>
<td>246 (16.72)</td>
</tr>
<tr>
<td>Addiction: gambling (DÉBA-Jeu)</td>
<td>30 (2.04)</td>
</tr>
<tr>
<td>Addiction: screens (DÉBA-internet(^k))</td>
<td>320 (21.75)</td>
</tr>
<tr>
<td>Posttraumatic disorder (PCL-5(^l))</td>
<td>414 (28.14)</td>
</tr>
</tbody>
</table>

\(^a\)SIDAS-FR: Suicidal Ideation Attributes Scale—French version.
\(^b\)K10: Kessler Psychological Distress Scale.
\(^c\)MHC-SF: Mental Health Continuum Short Form.
\(^d\)PHQ-9: Patient Health Questionnaire.
\(^e\)MDQ: Mood Disorder Questionnaire.
\(^f\)SPIN: Spin Phobia Inventory.
\(^g\)PSWQ: Penn State Worry Questionnaire.
\(^h\)DÉBBA: Dépistage et Évaluation du Besoin d’Aide—Alcool.
\(^i\)DÉBAD: Dépistage et Évaluation du Besoin d’Aide—Drogues.
\(^j\)DÉBA: Dépistage et Évaluation du Besoin d’Aide—Jeu.
\(^l\)PCL-5: PTSD Checklist for DSM-5.

**Characteristics of Individuals Using the Intervention Service**

Table 4 presents information on the age and gender of intervention service users. Information on user age was available for 4000 interventions. Most users were aged between 18 and 40 years (2829/4000, 70.73%). Information on the gender identity of users was available for 3569 interventions. Most users identified as woman (2555/3569, 71.59%). Users who identified as transgender, nonbinary, or listed other were generally young people aged ≤25 years (235/280, 83.9%). Information on the reason for contact was available for 4000 interventions (Table 5): 84.48% (3379/4000) of the users indicated I am thinking about suicide in the triage questionnaire and 10.1% (404/4000) of the users indicated I am worried about a loved one, and 5.43% (217/4000) indicated I am bereaved by suicide.
Table 4. Sociodemographic profile of the intervention service users.

<table>
<thead>
<tr>
<th>Age group (y)</th>
<th>Gender identity, n (%)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Woman (n=2555)</td>
<td>Man (n=734)</td>
<td>Transgender, nonbinary, or other (n=280)</td>
<td></td>
</tr>
<tr>
<td>≤17</td>
<td>211 (8.26)</td>
<td>106 (14.4)</td>
<td>76 (27.1)</td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>1002 (39.22)</td>
<td>232 (31.6)</td>
<td>159 (56.8)</td>
<td></td>
</tr>
<tr>
<td>26-40</td>
<td>1017 (39.8)</td>
<td>213 (29)</td>
<td>43 (15.4)</td>
<td></td>
</tr>
<tr>
<td>≥41</td>
<td>325 (12.72)</td>
<td>183 (24.9)</td>
<td>2 (0.7)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Reason for contact of the intervention service users.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Reason for contact</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Thinking about suicide</td>
<td>Worried about someone</td>
<td>Grieving a loss by suicide</td>
<td></td>
</tr>
<tr>
<td>Age groups (y), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤17</td>
<td>580 (17.16)</td>
<td>52 (12.9)</td>
<td>21 (9.7)</td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>1251 (37.02)</td>
<td>119 (29.5)</td>
<td>95 (43.8)</td>
<td></td>
</tr>
<tr>
<td>26-40</td>
<td>1196 (35.39)</td>
<td>127 (31.4)</td>
<td>39 (18)</td>
<td></td>
</tr>
<tr>
<td>≥41</td>
<td>352 (10.42)</td>
<td>106 (26.2)</td>
<td>62 (28.6)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3379 (100)</td>
<td>404 (100)</td>
<td>217 (100)</td>
<td></td>
</tr>
</tbody>
</table>

Gender identity, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Woman (n=2555)</th>
<th>Man (n=734)</th>
<th>Transgender, nonbinary, or other (n=280)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2178 (73.06)</td>
<td>250 (64.8)</td>
<td>127 (62.9)</td>
</tr>
<tr>
<td></td>
<td>555 (18.62)</td>
<td>119 (30.8)</td>
<td>60 (29.7)</td>
</tr>
<tr>
<td>Transgender, nonbinary, or other</td>
<td>248 (8.32)</td>
<td>17 (4.4)</td>
<td>15 (7.4)</td>
</tr>
<tr>
<td>Total</td>
<td>2981 (100)</td>
<td>386 (100)</td>
<td>202 (100)</td>
</tr>
</tbody>
</table>

With regard to the suicidal risk of the users of the intervention service, 3490 answered the question (Table 6), *Are you planning to attempt suicide in the coming minutes or hours?* Of the 3490 users, 1800 (51.58%) answered *no*, 972 (27.79%) answered *unsure*, and 718 (20.57%) answered *yes*. A total of 2606 users completed the SIDAS-FR at triage. The mean score was 26.69 (SD 12.79; median 29). Overall, 75.56% (1969/2606) of the users had suicidal ideation of high severity.

Information on the needs expressed by the users was available for 3254 interventions (Table 6). Overall, 75.29% (2450/3254) of the users expressed a need to ventilate. The most frequently discussed issues were suicide (3033/3475, 87.28%), mental health problems (765/3475, 22.01%), emotional problems (657/3475, 18.91%), family problems (455/3475, 13.09%), self-harm (413/3475, 11.88%), loneliness (407/3475, 11.71%), and problems with peers or relatives (361/3475, 10.39%).

Information on the counselors’ suicide risk assessments was available for 1783 interventions (Table 6). The most common assessments were “Low foreseeable risk for a suicide attempt in the near future” (955/1783, 53.56%) and “High foreseeable risk for a suicide attempt in the near future” (602/1783, 33.76%).

https://formative.jmir.org/2024/1/e46195
Table 6. Intervention service users’ suicide risk and presenting problems.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Suicidal users, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plan to attempt suicide in the coming minutes or hours? (n=3490)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1800 (51.57)</td>
</tr>
<tr>
<td>Unsure</td>
<td>972 (27.85)</td>
</tr>
<tr>
<td>Yes</td>
<td>718 (20.57)</td>
</tr>
<tr>
<td><strong>Severity of suicidal ideation (SIDAS-FR⁶; n=2606)</strong></td>
<td></td>
</tr>
<tr>
<td>Absence of suicidal ideation (0)</td>
<td>267 (10.24)</td>
</tr>
<tr>
<td>Low severity (1-12)</td>
<td>120 (4.6)</td>
</tr>
<tr>
<td>Medium severity (13-19)</td>
<td>250 (9.59)</td>
</tr>
<tr>
<td>High severity (&gt;20)</td>
<td>1969 (75.56)</td>
</tr>
<tr>
<td><strong>Expressed needs (n=3254)</strong></td>
<td></td>
</tr>
<tr>
<td>Need to ventilate</td>
<td>2450 (75.29)</td>
</tr>
<tr>
<td>Support for a crisis situation</td>
<td>1093 (33.59)</td>
</tr>
<tr>
<td>Help to ensure safety</td>
<td>870 (26.74)</td>
</tr>
<tr>
<td>Ask for advice</td>
<td>741 (22.77)</td>
</tr>
<tr>
<td>Ask for information</td>
<td>317 (9.74)</td>
</tr>
<tr>
<td><strong>Issues discussed (n=3475)</strong></td>
<td></td>
</tr>
<tr>
<td>Suicide</td>
<td>3033 (87.28)</td>
</tr>
<tr>
<td>Mental health</td>
<td>767 (22.07)</td>
</tr>
<tr>
<td>Emotional problems</td>
<td>657 (18.91)</td>
</tr>
<tr>
<td>Family problems</td>
<td>455 (13.09)</td>
</tr>
<tr>
<td>Self-harm</td>
<td>413 (11.88)</td>
</tr>
<tr>
<td>Loneliness</td>
<td>407 (11.71)</td>
</tr>
<tr>
<td>Problems with peers or relatives</td>
<td>361 (10.39)</td>
</tr>
<tr>
<td>School problems</td>
<td>294 (8.46)</td>
</tr>
<tr>
<td>Death of a loved one</td>
<td>293 (8.43)</td>
</tr>
<tr>
<td>Physical health</td>
<td>179 (5.15)</td>
</tr>
<tr>
<td>Alcohol or substance abuse</td>
<td>165 (4.75)</td>
</tr>
<tr>
<td>Verbal abuse survivor</td>
<td>95 (2.73)</td>
</tr>
<tr>
<td>Perpetrator of physical or verbal abuse</td>
<td>70 (2.01)</td>
</tr>
<tr>
<td>Bullying at school</td>
<td>68 (1.96)</td>
</tr>
<tr>
<td>Problems with the law</td>
<td>44 (1.27)</td>
</tr>
<tr>
<td>Bullying in the workplace</td>
<td>39 (1.12)</td>
</tr>
<tr>
<td>Pregnancy or abortion</td>
<td>34 (0.98)</td>
</tr>
<tr>
<td>Other types of problems</td>
<td>357 (10.27)</td>
</tr>
<tr>
<td><strong>Counselors’ suicide risk assessment (n=1783)</strong></td>
<td></td>
</tr>
<tr>
<td>No foreseeable risk for a suicide attempt in the near future</td>
<td>123 (6.89)</td>
</tr>
<tr>
<td>Low foreseeable risk for a suicide attempt in the near future</td>
<td>955 (53.56)</td>
</tr>
<tr>
<td>High foreseeable risk for a suicide attempt in the near future</td>
<td>602 (33.76)</td>
</tr>
<tr>
<td>Imminent danger of suicide attempt or attempt in progress</td>
<td>103 (5.77)</td>
</tr>
</tbody>
</table>

⁶SIDAS-FR: Suicidal Ideation Attributes Scale—French version.
Effectiveness of Universal, Selective, and Indicated Prevention Strategies Deployed Through the Suicide.ca Platform

Impact of Promotional Activities on the Platform’s Traffic

Analysis of peak traffic (days with >600 visits) for the period from October 15, 2020, to October 17, 2021, indicated that they were primarily associated with promotion of the platform in traditional media and on social media (Figure 2; Table 7). Traffic tended to remain high for 1 or 2 days following these spikes and then returned to its average trend afterward.

Figure 2. Impact of the promotional activities on the platform’s traffic.

Table 7. Promotional activities associated with traffic peaks.

<table>
<thead>
<tr>
<th>Date</th>
<th>Visits to the website, n</th>
<th>Activities or events</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 15, 2020</td>
<td>667</td>
<td>Publication on social media</td>
</tr>
<tr>
<td>October 22, 2020</td>
<td>734</td>
<td>Publication on social media</td>
</tr>
<tr>
<td>November 2, 2020</td>
<td>828</td>
<td>Unknown event (traffic mostly acquired via social media)</td>
</tr>
<tr>
<td>November 4, 2020</td>
<td>1188</td>
<td>Unknown event (traffic mostly acquired via direct acquisition)</td>
</tr>
<tr>
<td>November 26, 2020</td>
<td>1774</td>
<td>Several publications and interviews about suicide.ca on traditional media</td>
</tr>
<tr>
<td>February 1, 2021</td>
<td>722</td>
<td>Suicide Prevention Week</td>
</tr>
<tr>
<td>February 3, 2021</td>
<td>708</td>
<td>Suicide Prevention Week</td>
</tr>
<tr>
<td>May 3, 2021</td>
<td>928</td>
<td>Publication on social media</td>
</tr>
<tr>
<td>June 30, 2021</td>
<td>984</td>
<td>Publication on social media</td>
</tr>
<tr>
<td>July 15, 2021</td>
<td>722</td>
<td>Mass emailing to organizations to promote suicide.ca</td>
</tr>
<tr>
<td>July 26, 2021</td>
<td>1082</td>
<td>Publication on social media</td>
</tr>
<tr>
<td>September 10, 2021</td>
<td>663</td>
<td>World Suicide Prevention Day</td>
</tr>
<tr>
<td>September 23, 2021</td>
<td>1308</td>
<td>Publication on social media</td>
</tr>
</tbody>
</table>

Impact of AdWords Campaigns and Suicide Risk Factors Fact Sheets on the Platform’s Traffic

Table 8 shows the source of traffic and the top landing pages on the platform for the period from December 31, 2020, to October 31, 2021. The source of traffic to the platform was very diverse. AdWords campaigns on suicide risk factors were the second-largest source of traffic acquisition on the platform.
Table 8. Sources of traffic acquisition and landing pages for suicide.ca.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Website visits attributed to traffic acquisition and landing pages, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main sources of traffic acquisition (n=103,402)</strong></td>
<td></td>
</tr>
<tr>
<td>Suicide.ca entered directly in the search bar</td>
<td>33,635 (32.53)</td>
</tr>
<tr>
<td>Targeted advertising on search engines (AdWords)</td>
<td>25,251 (24.42)</td>
</tr>
<tr>
<td>Social media</td>
<td>16,775 (16.22)</td>
</tr>
<tr>
<td>Search engines (excluding targeted advertisements)</td>
<td>11,760 (11.37)</td>
</tr>
<tr>
<td>Hyperlink on another website</td>
<td>9,899 (9.57)</td>
</tr>
<tr>
<td><strong>Landing pages (n=125,677)</strong></td>
<td></td>
</tr>
<tr>
<td>Home page</td>
<td>50,271 (40)</td>
</tr>
<tr>
<td>Suicide risk factor fact sheets</td>
<td>20,497 (16.30)</td>
</tr>
<tr>
<td>My Tools (web and mobile app)</td>
<td>13,717 (11)</td>
</tr>
<tr>
<td>Blog pages</td>
<td>12,124 (10.91)</td>
</tr>
<tr>
<td>Pages for contacting the intervention service</td>
<td>9,892 (7.87)</td>
</tr>
<tr>
<td>Other</td>
<td>9,922 (7.87)</td>
</tr>
<tr>
<td>Section for people who are worried about a loved one</td>
<td>3,860 (3.07)</td>
</tr>
<tr>
<td>Section for suicidal people</td>
<td>3,056 (2.43)</td>
</tr>
<tr>
<td>Section for people bereaved by suicide</td>
<td>670 (0.53)</td>
</tr>
<tr>
<td>Directory of resources</td>
<td>656 (0.52)</td>
</tr>
</tbody>
</table>

*Pages designed to identify internet users at risk for suicide.

The AdWords campaign targeting anxiety and depression was displayed 72,713 times and received 9,576 clicks, resulting in a click rate of 13.7%. This campaign also led to 36 crisis chat interventions, representing a conversion rate of 0.38%. Similarly, the AdWords campaign with suicide-specific keywords was displayed 27,961 times and received 2,647 clicks, resulting in a click rate of 9.47%. This campaign resulted in 55 crisis chat interventions, with a conversion rate of 1.93%.

As for the first pages consulted by users (Table 8), 40% (4,704/11,760) of the visits from search engine queries, excluding AdWords ads, began with a visit to a suicide risk factor fact sheet. Therefore, suicide risk factor fact sheets were an important source of traffic to the platform.

**Effectiveness of the Intervention Service**

Table 9 presents the counselors’ perception of the impact of their interventions. Counselors indicated that the users felt better after the intervention in 56.09% (2,565/4,573) of the interventions. They also reported that the users felt more able to overcome their difficulties in 51.08% (2,336/4,573) of the interventions.

Table 9. Counselors’ perception of the impact of their interventions (n=4,573).

<table>
<thead>
<tr>
<th>Response</th>
<th>The user feels better, n (%)</th>
<th>The user feels more able to overcome their difficulties, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>449 (9.82)</td>
<td>446 (9.75)</td>
</tr>
<tr>
<td>Not sure</td>
<td>1,559 (34.09)</td>
<td>1,771 (38.73)</td>
</tr>
<tr>
<td>Yes</td>
<td>2,565 (56.09)</td>
<td>2,336 (51.08)</td>
</tr>
</tbody>
</table>

Table 10 shows the results of the postintervention survey for 259 suicidal users and 91 users worried for a loved one or bereaved by suicide. Overall, 91.5% (237/259) of the suicidal users indicated that the counselor seemed warm and caring, and 89.5% (232/259) of the users indicated that they felt listened to, understood, and respected. In terms of the effectiveness of the interventions, 82.6% (214/259) of the suicidal users indicated that they felt better, calmer, or more relaxed following the intervention; 76.8% (199/259) of the users indicated that they felt more in control of their suicidal thoughts; and 72.5% (188/259) of the users indicated that they felt more able to overcome their difficulties.
Principal Findings

The objectives of this study were to document the use of the suicide.ca platform; to describe the profile of its users; and to estimate the impact of the universal, selective, and indicated prevention strategies deployed via the platform.

The results for universal prevention strategies indicated that there was a wide range of sources for acquiring traffic to the platform, that actions to promote the platform on social media were directly associated with increased traffic, and that there was little traffic to psychoeducation pages.

The results for selective prevention strategies indicated that AdWords campaigns and fact sheets on suicide risk factors are important sources of traffic acquisition on the platform, that the platform reaches people with varying degrees of suicidal risk, and that the user account was a barrier to the use of mental health self-management tools.

The results for the indicated prevention strategies indicated that triage questions involving personal information result in a significant number of dropouts, that counselors feel that their interventions have a positive impact on the user in half of the contacts, and that intervention service users are generally satisfied with the service they received.

Universal Prevention Strategies

The sources of traffic acquisition on suicide.ca were very diversified, and only 40% (50,271/125,677) of the users started their visit on the home page. Moreover, the platform’s social media promotion was directly associated with traffic peaks. These results highlight the importance of being proactive in promoting the platform and using a wide variety of promotional strategies (eg, SEO strategy, social media posts, AdWords campaigns, and referral links on other websites). They also highlighted the importance of collaborating with digital experts to adapt traditional digital marketing strategies to suicide prevention.

The number of visits to the psychoeducation pages was far less than the total number of visits to the site. For example, the first page of the psychoeducation section for people with suicidal risk was visited only 5312 times in the first year. This number is far less than the 127,495 visits that were recorded on the site for this period. This result suggests that the vast majority of users do little exploration of the platform after viewing the content that brings them to the site. These results highlight the importance of looking beyond the total number of visits to a digital suicide prevention platform to describe its use and of dissecting its analytics data in detail to fully understand the content that is consulted.

Selective Prevention Strategies

AdWords campaigns and suicide risk factor fact sheets were important sources of traffic acquisition on the platform. For example, 40% (4704/11,760) of the visits from search engines, excluding AdWords advertisements, began with a suicide risk factor fact sheet. Although it is challenging to determine from the analytical data whether the individuals who accessed the fact sheets or clicked on the advertisements were at risk for suicide, it is evident that they conducted a Google search related to suicide or suicide risk factors and landed on the suicide.ca platform. Furthermore, the AdWords campaigns targeting anxiety and depression achieved click rates of 13.7% and 9.47%, respectively, which are significantly higher than the average Google Ads benchmarks of 3.17% observed in e-marketing [44] and the click rates of 4% and 5% reported in the study by Onie et al [13]. However, despite the high click rates, the conversion rates for crisis chat interventions were relatively low, with 0.38% for anxiety and depression achieved click rates of 13.7% and 9.47%, respectively, which are significantly higher than the average Google Ads benchmarks of 3.17% observed in e-marketing [44] and the click rates of 4% and 5% reported in the study by Onie et al [13]. However, despite the high click rates, the conversion rates for crisis chat interventions were relatively low, with 0.38% for anxiety and depression campaign and 1.93% for the suicide-specific campaign. It is challenging to interpret why these clicks did not result in more crisis interventions. Although the average conversion rate in e-marketing is 3.75% [44], there are currently no established benchmarks for conversion rates in suicide prevention. Although our findings should be interpreted with caution, they offer initial support for suicide.ca’s strategy of using AdWords campaigns and fact sheets as tools to identify internet users presenting suicide risk factors through search engine platforms.

Previous research has predominantly focused on how Google can enhance support for suicide prevention, paying limited attention to how suicide prevention actors can leverage Google to identify individuals at risk (eg, [12,13,45,46]). More research is required to discover optimal methods for using Google search...
to identify internet users at risk of suicide and to guide them toward support resources.

A major finding of this study is that the platform reaches users with different degrees of suicidal risk. The proportion of users with severe suicidal ideation on the SIDAS-FR was 33.16% (825/2488) for the self-assessment module and 75.56% (1969/2606) for the intervention service. Although the self-assessment module users presented less severe suicidal ideations than the intervention service users, 87.77% (2721/3100) of them obtained a K10 score, indicating the probable presence of a mental health problem. Furthermore, 62.2% (915/1471) of the self-assessment module users obtained a Patient Health Questionnaire-9 score indicating the presence of depression, 61.73% (908/1471) obtained a Penn State Worry Questionnaire score indicating the presence of generalized anxiety disorder, and 53.2% (451/847) obtained a Spin Phobia Inventory score indicating the presence of social phobia disorder.

Our best point of comparison is the 113Online chat service in the Netherlands, which also provides users with access to a self-assessment module and crisis interventions via chat. According to the evaluation of their first 3 years of operation [3], 78.36% (4741/6050) of the users who used the self-assessment module scored positively for depression and anxiety on the Depression and Anxiety Stress Scale. The module also included the Beck Scale for Suicidal Ideation, with results indicating that 7.22% (533/7381) of the users were not suicidal, 1.75% (129/7381) were mildly suicidal, 39.29% (2900/7381) were suicidal, and 51.74% (3819/7381) were severely suicidal. When comparing the results of this study with the observations on the 113Online platform, it appears that the suicide.ca self-assessment module is being used by individuals with lower levels of suicidality. These differences may, in part, be explained by the fact that the 2 platforms used different questionnaires to measure suicidality, depression, and anxiety. In addition, it may be explained in part by the identification strategy of the suicide.ca platform, which aims to reach individuals who have risk factors for suicide but may not be actively suicidal.

With regard to the self-management tools, between 22.99% (1186/5159) and 58% (2992/5159) of the users who downloaded the mobile app did not create a user account. It should be noted that a user can download the app to use the intervention service only and that doing so does not require the creation of an account. Technical problems experienced by users when creating their account could also contribute to this discrepancy [47]. Nevertheless, the discrepancy between the number of downloads and the number of accounts created is so large that it seems to indicate that setting up a user account is a practice that should be avoided as it seems to represent the first barrier to use. This finding highlights an important dilemma for digital suicide prevention interventions. Several studies indicate that users of digital interventions have concerns about the security and privacy of their data [48]. However, although the presence of a user account enhances data security and privacy, it is a barrier to use for some users [49]. These findings highlight the importance of identifying technologies that ensure the security of users’ data while preserving their sense of anonymity.

### Indicated Prevention Strategies

To our knowledge, this study is the first to report on the impacts of triage on the accessibility of a web-based intervention service. Our results indicate that the last question (What is your first name or alias?) was associated with a significant number of dropouts; only 46.54% (2711/5825) of the users who were asked the question answered it. Studies on crisis intervention through chat and text indicate that users prefer to use these services over phone helplines because of the sense of anonymity provided by these mediums [3,50,51]. It is possible that even asking for personal information or requesting an alias could discourage users, emphasizing the importance of maintaining their sense of anonymity.

With regard to the questions about suicidality, it is surprising to observe the low dropout rates they have caused. The requirement to complete the 5 questions of the SIDAS-FR during the triage resulted in a dropout rate of 4%. Web-based crisis service providers are sometimes concerned about implementing a triage. This study provides an estimate of the dropout rate that triage can lead to.

With regard to the profile of the intervention service users, most users (2019/3569, 56.57%) were women aged between 18 and 40 years. The most frequently discussed issues were mental health, emotional problems, relationship problems, self-harm, and isolation. These findings are consistent with those typically observed in digital intervention services [14-17].

It is also interesting to note that 7.85% (280/3569) of the users were from gender minority groups and 83.9% (235/280) of them were young people aged ≤25 years. This percentage is similar to that observed in the Canada Suicide Prevention Service (7.5%) and Lifeline Crisis Chat (7.5%) [14,15]. The age of these users likely reflects the high level of adversity experienced at this time for many gender minority youth [52]. These findings also highlight the relevance of digital intervention services for these youth.

This study provides valuable insights into the suicidality of users of chat intervention services, as multiple sources of information were used to evaluate the risk of suicide. These include data derived from a standardized suicide questionnaire (SIDAS-FR), risk assessments conducted by counselors, and a specific question regarding the risk of suicide attempts in the near future. Approximately half of the users (1690/3490, 48.42%) answered yes or uncertain to the question, Are you planning to attempt suicide in the coming minutes or hours? Moreover, 75.56% (1969/2606) of the users who answered no or uncertain to the question obtained a SIDAS-FR score indicating severe suicidal ideations. Therefore, the intervention service reaches its target population, that is, people at a high risk for suicide. These findings are consistent with those observed in the 113Online crisis chat evaluation [17], where a content analysis of intervention transcripts revealed that 61.1% (320/524) of the users expressed suicidal intentions, 21.2% (111/524) of the users reported both suicidal intentions and a suicide plan, and 3.8% (20/524) of the users were actively attempting suicide at the time of contact.
In terms of the impact of the intervention service, counselors reported that the user felt better after the intervention in 56.09% (2565/4573) of contacts and reported that the user did not feel better in only 9.81% (449/4573) of contacts. The response rate to the postintervention questionnaire was low. Nonetheless, the scores on the various satisfaction indicators were all very high, and 3 (199/259, 76.8%) out of 4 respondents felt that their suicidal ideation had decreased following the intervention. Again, these results are consistent with those typically observed in digital intervention services [15,16,18,53].

Limitations of the Study

This study had several limitations. First, the results on site traffic, traffic acquisition sources, page views, and app downloads were based on an analysis of the platform’s analytics data. However, analytical data do not allow us to understand the needs and motivations behind users’ browsing behavior on the platform. It is therefore important to complement this type of quantitative data in digital platform evaluations with other observational methods, such as qualitative interviews with users, or by directly observing their browsing and help-seeking behavior on the internet [22].

Similarly, analytical data are often inaccurate. For example, the traffic on the platform was analyzed according to the number of sessions. However, a session is the period during which a user was active on the site. If a user is inactive for ≥30 minutes, Google Analytics will attribute any subsequent activity to a new session. Therefore, the number of sessions is not exactly the same as the number of visits, but it is the closest indicator to it.

With respect to the analysis of app downloads, it was difficult to estimate the number of people who downloaded the app but did not create a user account. Therefore, it was impossible to provide an accurate estimate for this important issue. Moreover, it would be important to understand the needs and motivations that explain why so many people downloaded the app but did not open a user account. We explain this phenomenon by the users’ need for anonymity, but it would be important to validate this hypothesis with empirical evidence.

The results on the needs expressed by the intervention service users and the topics discussed during the contacts were based on the analysis of the intervention reports. The completeness of these reports may vary from one counselor to another, which may explain in part why the numbers on issues discussed were low. Furthermore, data on user-identified needs reflect the perceptions of counselors. Therefore, there is most likely a gap between these perceptions and the actual needs of the users. Therefore, it would be important to conduct qualitative interviews with users to understand their needs and motivations for using the intervention service.

An important limitation of this study was the absence of questions on gender and age in the self-assessment module. Therefore, we were unable to compare users of the self-assessment with users of the intervention service based on these characteristics. This information is crucial to understanding who is and is not being reached by suicide.ca and, consequently, by the Quebec DSPS. This highlights the importance of anticipating how the data collected by the various tools offered on a platform will compare with their users.

Conclusions

The results of this study indicate that it is possible to use the same platform to conduct universal, selective, and indicated suicide prevention. They also highlighted the importance of being strategic in the development of a digital platform to anticipate how different target populations will be identified, how different intervention modalities will be offered to them, and how the data collected by the platform will allow for the measurement of its impact.

Finally, the evaluation of suicide.ca produced several important results and was the source of a considerable amount of learning. We hope that sharing these learnings will help guide organizations interested in developing a digital suicide prevention platform and contribute to the development of guidance for the development of DSPSs (eg, the WHO [54]).

Acknowledgments

The authors would like to thank the Association québécoise de prévention du suicide (AQPS) for their valuable and constructive suggestions during the planning and development of this evaluation. A grant was awarded for the publication of the study to LPC and JL by the Centre for Research and Intervention on Suicide, Ethical Issues and End-of-Life Practices and the Fonds de recherche du Québec—Société et culture.

Data Availability

The data sets generated during and analyzed during this study are not publicly available due to confidentiality agreement with the Association québécoise de prévention du suicide (AQPS) but are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors of this study participated in the development of Quebec’s digital suicide prevention strategy.

References


5. It Gets Better home page. It Gets Better. URL: https://getitsbetter.org/about/ [accessed 2024-02-20]


26. suicide.ca home page. suicide.ca. URL: https://suicide.ca/ [accessed 2024-02-21]


44. Irvine M. Google Ads benchmarks for your industry updated! WorldStream. URL: https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks [accessed 2023-01-01]


Abbreviations

AQPS: Association québécoise de prévention du suicide
CRM: Customer Relationship Management
DSPS: digital suicide prevention strategy
ICT: information and communication technology
K10: Kessler Psychological Distress Scale
SEO: search engine optimization
SIDAS-FR: Suicidal Ideation Attributes Scale—French version
WHO: World Health Organization

©Louis-Philippe Côté, Julie Lane. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

Professional Social Media Use Among Orthopedic and Trauma Surgeons in Germany: Cross-Sectional Questionnaire-Based Study

Yasmin Youssef1*; Tobias Gehlen2*, Dr med; Jörg Ansorg3, Dr med; David Alexander Back2,4, Prof Dr Med; Julian Scherer5,6, Dr med

1Department of Orthopaedics, Trauma and Reconstructive Surgery, University Hospital Leipzig, Leipzig, Germany
2Center for Musculoskeletal Surgery, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
3Akademie Deutscher Orthopäden, Berufsverband für Orthopädie und Unfallchirurgie e.V., Berlin, Germany
4Dieter Scheffner Center for Medical Education and Educational Research, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
5Department of Traumatology, University Hospital of Zurich, Zurich, Switzerland
6Orthopaedic Research Unit, University of Cape Town, Cape Town, South Africa

*these authors contributed equally

Corresponding Author:
Julian Scherer, Dr med
Department of Traumatology
University Hospital of Zurich
Rämistrasse 100
Zurich, 8091
Switzerland
Phone: 41 762030775
Email: julian.scherer@usz.ch

Abstract

Background: Social media (SM) has been recognized as a professional communication tool in the field of orthopedic and trauma surgery that can enhance communication with patients and peers, and increase the visibility of research and offered services. The specific purposes of professional SM use and the benefits and concerns among orthopedic and trauma surgeons, however, remain unexplored.

Objective: This study aims to demonstrate the specific uses of different SM platforms among orthopedic and trauma surgeons in Germany as well as the advantages and concerns.

Methods: A web-based questionnaire was developed on the use of SM in a professional context by considering the current literature and the authors’ topics of interest. The final questionnaire consisted of 33 questions and was distributed among German orthopedic and trauma surgeons via the mail distributor of the Berufsverband für Orthopädie und Unfallchirurgie (Professional Association of Orthopaedic Surgeons in Germany). The study was conducted between June and July 2022. A subgroup analysis was performed for sex (male vs female), age (<60 years vs ≥60 years), and type of workplace (practice vs hospital).

Results: A total of 208 participants answered the questionnaire (male: n=166, 79.8%; younger than 60 years: n=146, 70.2%). In total, all of the participants stated that they use SM for professional purposes. In contrast, the stated specific uses of SM were low. Overall, the most used platforms were employment-oriented SM, messenger apps, and Facebook. Instagram emerged as a popular choice among female participants and participants working in hospital settings. The highest specific use of SM was for professional networking, followed by receiving and sharing health-related information. The lowest specific use was for education and the acquisition of patients. Conventional websites occupied a dominating position, exceeding the use of SM across all specific uses. The key benefit of SM was professional networking. Under 50% of the participants stated that SM could be used to enhance communication with their patients, keep up-to-date, or increase their professional visibility. In total, 65.5% (112/171) of participants stated that SM use was time-consuming, 43.9% (76/173) stated that they lacked application knowledge, and 45.1% (78/173) stated that they did not know what content to post. Additionally, 52.9% (91/172) mentioned medicolegal concerns.
Conclusions: Overall, SM did not seem to be used actively in the professional context among orthopedic and trauma surgeons in Germany. The stated advantages were low, while the stated concerns were high. Adequate education and information material are needed to elucidate the possible professional applications of SM and to address legal concerns.

(JMIR Form Res 2024;8:e53336) doi: 10.2196/53336

KEYWORDS
social media; digitalization; digital communication; orthopedics; traumatology

Introduction

In the last decades, social media (SM) use has become increasingly popular among all age groups and affects almost all areas of daily life [1-3]. In 2024, the most popular SM platforms worldwide were Facebook, YouTube, WhatsApp, and Instagram [4]. SM has also gained popularity and importance in the health care sector. Regular SM use among health care providers is as high as 88% [5]. In the health care sector, SM can be used by individual physicians as well as health care institutions (hospitals, journals, societies, etc) to facilitate web-based representation as well as communication with patients, potential patients, and the public [6-9]. One advantage of SM is the low barrier for interactions between users and the fast distribution of information and media from almost anywhere at any time. SM can enhance communication with patients and foster professional development, and has the potential to contribute to the spread of (public) health information [10,11]. In the field of surgery, SM has been shown to serve as a tool for research dissemination and surgical education [12,13]. In particular, it could be demonstrated that there is a correlation between the number of SM posts and academic citations of recent research [14,15]. However, while acknowledging those potential advantages of professional SM use, it must be also noted that there are certain dangers, which include violations of patient privacy and medicolegal, confidentiality, and liability issues [7-9,16]. Furthermore, SM incorporates the risk of spreading health-related misinformation [17,18].

Recent studies have shown that the rate of SM use among orthopedic and trauma surgeons lies between 37% to 100% [19-25]. However, the specific uses of SM in a professional context and the barriers toward comprehensive use within the field of orthopedics and trauma surgery remain unknown. In addition to that, data on the appreciated advantages and concerns of professional SM use in the field of orthopedic and trauma surgery is lacking.

Thus, this study aimed to assess the appreciated advantages and concerns of professional SM use among orthopedic and trauma surgeons in Germany. Further, we aimed to present the current professional use of different SM platforms used by orthopedic and trauma surgeons in Germany.

Methods

Study Design

A web-based questionnaire was created to assess the current use of SM among orthopedic and trauma surgeons in Germany. We used SurveyMonkey (SurveyMonkey Inc), web-based software for creating questionnaires. The link to the questionnaire was shared via the mail distributor of the Berufsverband für Orthopädie und Unfallchirurgie (Professional Association of Orthopaedic Surgeons in Germany). The study was conducted between June and July 2022.

Ethical Considerations

Participation in the web-based questionnaire was voluntary and anonymous, with no identifying data collected, except for age, gender, and occupation. Anonymous questionnaires, by definition, do not collect personal data that can directly or indirectly identify an individual. This means that such surveys do not fall under the scope of the European General Data Protection Regulation (GDPR), as the GDPR only applies to the processing of personal data (Article 5) [26]. Additionally, the presented survey does not require a Data Protection Impact Assessment according to Article 35 of the GDPR [26]. No formal ethical approval by an ethics committee was needed for the study conduction as general waivers apply for surveys with anonymous data in Germany. All participants received written information about the aim and scope of the study as well as how data is collected, processed, and analyzed in the form of a disclaimer before starting the questionnaire (Multimedia Appendix 1). Patients were informed that by answering the questionnaire, they consented to data collection, processing, and use for publication. No incentives were offered for the completion of the questionnaire.

Questionnaire

The development of the questionnaire was described in a preceding publication [25]. The questionnaire was developed by the study team based on a review of the current literature [19,20,22,24,27-29] and was complemented with further areas of interest. The preliminary and digitalized questionnaire was pretested among 5 orthopedic and trauma surgeons. The questionnaire was finalized considering the feedback from the pilot group. The questionnaire (Multimedia Appendix 1) consisted of 33 variables and included two separate sections. A 5-point Likert scale (where 1 was “I strongly disagree” and 5 was “I strongly agree”) was used to assess the advantages and concerns stated regarding professional SM use.

The first publication highlighted the types of SM platforms used for private and professional purposes, the use behavior, and content management of orthopedic and trauma surgeons in Germany [25]. The second section of this study, which includes questions on the specific uses of different SM platforms as well as the appreciated benefits and concerns, will be analyzed.
Data Processing and Statistical Analysis
Statistical analysis was performed using SPSS for Mac (version 26.0; IBM Corp). Both complete and incomplete questionnaires were considered in the final data analysis. Categorical data were presented in frequencies and percentages. Subgroup analysis was performed for sex (male vs female), age (<60 years vs ≥60 years), and type of workplace (practice vs hospital). To assess differences between groups, the chi-square test was used for categorical data. Likert scales were compared using a nonparametric median test. The level of statistical significance was set at a 2-sided P value <.05.

Results
Demographics
The sample is identical to that previously published [25]. In total, 208 participants took part in this survey (male: n=166, 79.8%), of which 70.2% (n=146) were aged <60 years. Most of the participants (n=161, 77.4%) worked in a practice. More female participants were younger than 60 years (<60 years: 39/42, 92.9% vs ≥60 years: 3/42 4.8%; P<.001), and younger participants were more likely to be working in hospital rather than in a practice (<60 years: 39/47, 83% vs ≥60 years: 8/47, 17%; P=.03). Significantly more male participants were in a practice (male: 138/161, 85.7% vs female: 23/161, 14.3%; P<.001). As shown in the preceding study, all participants (<60 years) stated that they used SM for professional purposes [25].

SM for Professional Networking
For professional networking (communication with colleagues), the most used platforms were messenger apps (58/173, 33.5%), employment-oriented SM (47/173, 27.2%), and Facebook (12/173, 6.9%). In addition to that, 39 (22.5%) participants stated that they used conventional websites for communication with colleagues. Female participants were more likely to use Facebook (male: 7/142, 4.9% vs female: 5/31, 16%; P=.03) and Instagram (male: 2/142, 2.8% vs female: 4/31, 13%; P=.02). Participants working in hospital were more likely to use Facebook (hospital: 6/34, 18% vs practice: 6/139, 4.3%; P=.006), Instagram (hospital: 4/34, 12% vs practice: 4/139, 2.9%; P=.03), and employment-oriented SM (hospital: 16/34, 47% vs practice: 31/139, 22.3%; P=.004).

SM for Receiving and Sharing Health-Related Information
The platforms used most frequently for receiving health-related information were YouTube (46/174, 26.4%), employment-oriented SM (28/174, 16.1%), and messenger apps (23/174, 13.2%). More than half of the participants (113/174, 64.9%) used conventional websites. Instagram was used more by female participants (male: 4/143, 2.8% vs female: 5/31, 16%; P=.002). Participants working in a hospital were more likely to use TikTok (hospital: 1/34, 3% vs practice: 0/0, 0%; P=.04). The most often used platforms for sharing health-related information were messenger apps (45/172, 26.2%), employment-oriented SM (23/172, 13.4%), Facebook (10/172, 5.8%), and Instagram (10/172, 5.8%). Conventional websites were only used by a minority of participants (39/172, 22.7%). More female participants used Instagram (male: 5/141, 3.5% vs female: 5/31, 16%; P=.007) and TikTok (male: 0/142, 0% vs female: 1/31, 3%; P=.03). Participants working in the hospital were more likely to use Facebook (hospital: 5/34, 15% vs practice: 5/138, 3.6%; P=.01) and TikTok (hospital: 1/34, 3% vs practice: 0/138, 0%; P=.04).

SM for Sharing Clinical Expertise
Messenger apps were used by 21.8% (37/170) of the participants for sharing clinical expertise and professional skills. Employment-oriented SM was used by 12% (22/170) and Instagram by 4.7% (8/170) of participants. In total, 24.7% (42/170) used conventional websites. Female participants were more likely to use Facebook (male: 3/139, 2.2% vs female: 3/31, 10%; P=.04) and Instagram (male: 4/139, 2.9% vs female: 4/31, 13%; P=.02). Instagram was also used more by younger participants (<60 years: 8/116, 6.9% vs ≥60 years: 0/54, 0%; P=.048). In total, 5.3% (9/169) of participants used Facebook, 4.1% (7/169) used employment-oriented SM, and 4.1% (7/169) used messenger apps for producing content on diseases and treatment methods. Overall, 26.6% (45/169) used conventional websites for sharing content on diseases and treatment methods.

SM for Educational Purposes
The SM platforms used mostly for educational purposes were employment-oriented platforms (eg, LinkedIn: 26/171, 15.2%), YouTube (20/171, 11.7%), and messenger apps (19/171, 11.1%). Conventional websites were used by 60.8% (104/171) of participants. Older participants were significantly more likely to use messenger apps (<60 years: 8/117, 6.8% vs ≥60 years: 11/54, 20%; P=.009).

SM for the Acquisition of and Communication With (Potential) Patients
Only a minority of the participants stated that they use SM to acquire new patients. In total, 8.3% (14/168) of participants used Facebook, 5.4% (9/168) used messenger apps, and 4.2% (7/168) used Instagram. In contrast, 42.9% (72/168) used conventional websites. Overall, messenger apps were significantly more used often by older participants (aged <60 years: 13/118, 11% vs aged ≥60 years: 16/53, 30%; P=.002). A further 5.3% (9/171) of participants used Facebook and 4.1% (7/171) used Instagram. Websites were used by 34.5% (59/171) of participants.

Stated Advantages
In total, 26.6% (47/177) of participants agreed or strongly agreed that professional SM use could help them in the acquisition of new patients. In contrast, 28.8% (51/177) of participants adopted a neutral position, and 44.6% (79/177) disagreed or disagreed strongly. Similarly, only 24.3% (43/177) of participants agreed or strongly agreed that professional SM use could improve their communication with patients, while 53.1% (94/177) disagreed or strongly disagreed with this statement. Additionally, 42.4% (75/177) of participants agreed or strongly agreed that SM could help them stay up-to-date professionally, and 40.1% (71/177) agreed that SM could help them to increase their visibility. There were no significant differences found for age, gender, and type of workplace. The median values are presented in Table 1.
In total, 65.5% (112/171) of participants agreed or strongly agreed that the professional use of SM was time-consuming. Only 8.8% (15/171) of participants disagreed or strongly disagreed with this statement. Overall, 45.1% (78/173) of participants agreed or strongly agreed that they have difficulties assessing what content would interest patients on SM. Furthermore, 52.9% (91/172) of participants agreed or strongly agreed that they were insecure with the medicolegal and data protection regulations concerning SM, while only 25.6% (44/172) disagreed or strongly disagreed.

In addition, 43.9% (76/173) of participants agreed or strongly agreed that they lacked knowledge on the use of SM. In total, 31.2% (54/173) of participants stated that they disagreed or strongly disagreed with this statement. Older participants stated more often that they lack knowledge of SM implementation and use ($P=.006$). The median values are presented in Table 2.

### Table 1. Stated advantages of professional social media use by German orthopedic and trauma surgeons (n=177).

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Score$^a$, median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social media helps me acquire new patients</td>
<td>3.00 (1.63-4.38)</td>
</tr>
<tr>
<td>Social media helps me communicate with my patients</td>
<td>4.00 (3.00-5.00)</td>
</tr>
<tr>
<td>Social media help me stay up-to-date professionally</td>
<td>3.00 (2.00-4.00)</td>
</tr>
<tr>
<td>Social media helps me represent my medical offers</td>
<td>3.00 (1.50-4.50)</td>
</tr>
</tbody>
</table>

$^a$Likert scale where 1 is “I fully disagree” and 5 is “I fully agree.”

### Stated Concerns

In total, 65.5% (112/171) of participants agreed or strongly agreed that the professional use of SM was time-consuming. Only 8.8% (15/171) of participants disagreed or strongly disagreed with this statement. Overall, 45.1% (78/173) of participants agreed or strongly agreed that they have difficulties assessing what content would interest patients on SM. Furthermore, 52.9% (91/172) of participants agreed or strongly agreed that they were insecure with the medicolegal and data protection regulations concerning SM, while only 25.6% (44/172) disagreed or strongly disagreed.

In addition, 43.9% (76/173) of participants agreed or strongly agreed that they lacked knowledge on the use of SM. In total, 31.2% (54/173) of participants stated that they disagreed or strongly disagreed with this statement. Older participants stated more often that they lack knowledge of SM implementation and use ($P=.006$). The median values are presented in Table 2.

<table>
<thead>
<tr>
<th>Concern</th>
<th>Score$^a$, median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The use of social media is too time-consuming (n=171)</td>
<td>2.00 (1.00-3.00)</td>
</tr>
<tr>
<td>I don’t have sufficient knowledge on the efficient use of social media in the workplace (n=173)</td>
<td>3.00 (2.00-4.00)</td>
</tr>
<tr>
<td>I am insecure with legal and data protection regulation (n=172)</td>
<td>2.00 (1.00-3.00)</td>
</tr>
<tr>
<td>I find it hard to estimate which content interests patients (n=173)</td>
<td>3.00 (2.00-4.00)</td>
</tr>
</tbody>
</table>

$^a$Likert scale where 1 is “I fully disagree” and 5 is “I fully agree.”

### Discussion

#### Principal Findings

In the health care system, SM is shaping the ways and dimensions of communication and the dissemination and consumption of (health) information [7,30]. The active use and popularity of SM in the professional context was particularly accelerated by the COVID-19 pandemic [31,32]. Several benefits of professional SM use have been previously identified. SM can enable fast and location-independent sharing of health-related information, which makes it more accessible to health care professionals, patients, and the public [6-8,30,33-35]. Furthermore, SM can improve communication within the health care system and increase the visibility of individual physicians, institutions, or publishing companies [14,15].

SM is also shaping the field of orthopedic and trauma surgery [29,36]. Several studies have investigated the prevalence of professional SM use among orthopedic and trauma surgeons and have reported rates of professional SM use between 37% to 65.7% among the assessed populations [19-24]. In Germany, a recent study showed that 100% of participants used SM for professional purposes. A structured implementation into daily professional work routines, however, seemed to be lacking [25]. Overall, employment-oriented SM like LinkedIn, Facebook, and YouTube are among the most used platforms [19,20,22,24,27,36]. There is, however, limited data concerning the specific use and application of SM and the appreciated benefits and concerns of SM among orthopedic and trauma surgeons.

Overall, the most prevalent use of SM in the professional context of the presented cohort was for professional networking. Almost one-third of the participants used messenger apps and employment-oriented SM for professional networking. Other SM like Facebook, Instagram, and Twitter on the other hand seemed to play a minor role. This low use of SM for professional networking was not expected, as a previous study by Grossman et al [12] highlighted that SM use by surgeons plays an increasingly important role in the dissemination and communication of research results and professional information with colleagues and the public. The sharing of information can, in turn, increase the visibility of a particular surgeon or their research field [14,15]. Furthermore, a study by Justiana et al [37] among 165 Saudi Arabian orthopedic surgeons showed that most of the assessed surgeons used SM for sharing medical knowledge (79.03%), discussing cases with colleagues (72.4%), and sharing work experiences (66.7%). In contrast to this study, Justiana et al [37] did not differentiate between different SM platforms. Analyzing different aspects of professional networking, we found that the most used platforms for receiving health-related information were YouTube, employment-oriented SM, and messenger apps. On the other hand, the most frequently used platforms for sharing health-related information were messenger apps, employment-oriented SM, Facebook, and Instagram. Similarly, messenger apps, employment-oriented SM, and Instagram were the platforms used the most for sharing...
clinical expertise and professional skills. For the above specific uses, Instagram seemed to be used more by female participants, and TikTok was used more often by participants working in a hospital.

The use of SM for both receiving and sharing health-related information and clinical expertise as well as professional skills was low within the presented study population. In contrast, conventional websites occupied a predominant role for those uses.

The presented results, therefore, suggest that the maximum potential of SM in the dissemination of health-related and clinical information and expertise is currently not being reached.

The use of SM for educational purposes also seemed subsidiary in the presented cohort, while conventional websites were used by more than 60% of the participants for this purpose. In this study, employment-oriented SM, YouTube, and messenger apps, which were the SM platforms used the most for educational purposes, were each used by less than 25% of the participants. In contrast, a study by Schneider et al [31] found among a group of 312 orthopedic residents and medical students that the majority showed a high interest in accessing educational information via SM [31]. This difference may be explained by the fact that the presented study included participants from all age groups and experience levels, and not only residents or medical students who might show a higher affinity to SM due to their younger age. No significant difference could be found for the type and prevalence of SM platforms used for educational purposes. The results suggest that there might still be potential for both providers and consumers of educational offers to actively use and develop educational offers via SM. Previous studies have highlighted that the benefits include the connection of learners independently of time and location, and the opportunity to share and actively react to medical information [38-40].

In this study, SM also seemed to play a subordinate role in the communication and acquisition of patients. The most used platforms for these purposes were messenger apps, Facebook, and Instagram. Overall, 17% stated that they use messenger apps for communication with their patients. All other platforms were used by less than 9% of the participants for acquisition of and communication with patients. On the other hand, conventional websites were used by 34.5% and 42.9%, respectively, for those purposes. While comparable data from the field of orthopedic and trauma surgery is missing, a study among 500 US plastic surgeons showed that almost 50% used SM for marketing purposes and 28% acquired patients via SM [41]. However, it must be noted that the study was published in 2013, and plastic surgeons might rely more on SM as they provide private services in most cases. Similarly, in a survey by Brown et al [42] among Australian doctors, only 1 of 187 respondents used SM to communicate with patients and only 21.2% believed it would be appropriate to do so. Of note, the study was published in 2014 and did not clearly distinguish between private and professional SM profiles, so the comparability is limited. A more recent study by Justiana et al [37] showed a higher use of SM for interacting with patients; it was not differentiated between different SM platforms. Duymus et al [43] reported that 61.6% had concerns about communication with patients through SM. Nonetheless, these results suggest that the use of SM for acquisition of and communication with patients is still not yet implemented among orthopedic surgeons. This could be due to the continuing uncertainty concerning medicolegal issues. Further studies are needed to quantify the impact of SM use on patient acquisition.

Furthermore, these results showed that the stated advantages of professional SM use were low. Only the minority agreed that SM could help them in the acquisition of and communication with patients. No clear trend was seen for the stated benefit regarding increased visibility. However, most participants believed that SM could help them stay up-to-date professionally. To the best of our knowledge, this is the first study that evaluated the attitudes of German orthopedic surgeons toward professional SM use. Comparative data in the field of orthopedics and trauma surgery is missing.

Lastly, most participants had concerns about the professional use of SM. This included expenditure of time and an unawareness of what content is of interest to patients. In addition to that, more than half of the respondents had concerns about medicolegal and data protection issues. Similarly, in a study by Brown et al [42], more than 65% of the participants stated that they had concerns about engaging more on SM due to public access and legal concerns. This might be explained by the fact that there is only limited informative and educational material and tools for the professional use of SM. Additionally, there are no clear regulations for the professional use of SM in Germany. For example, whether a patient’s fully anonymized x-ray can be posted online without the patient’s written approval is still a gray area. In the study by Justiana et al [37], over 50% of the respondents had concerns about legal issues related to interaction with patients, and only 42.5% agreed that it was ethically acceptable to anonymously discuss a patient’s case on SM [37]. The appreciated concerns of the presented cohort correspond to the risks and dangers described in the literature—high time requirement, violation of patients’ anonymity, decrease of professionalism, and legal issues [8,44]. Supporting information material and courses provided by political or health institutions or orthopedic and trauma societies could help orthopedic and trauma surgeons understand how they can safely exploit the full advantages of SM in a professional and legal context.

Overall, these results suggest that SM use among German orthopedic and trauma surgeons is predominantly passive. While all participants stated that they use SM for professional purposes, only a minority indicated concrete uses. In a professional context, websites still play an important role. Professional networking was identified as the key benefit of SM. Overall, the most used platforms were employment-oriented SM, messenger apps, and Facebook. This corresponds to previously presented literature that has shown that employment-oriented SM, Facebook, and YouTube are the most used platforms among orthopedic surgeons [19,20,22,24,27,36].

Limitations

This study has certain limitations. First, surveys have minor levels of evidence in general, and their outcome can be affected...
Conclusions

The professional use of SM among the assessed participants of German orthopedic and trauma surgeons is predominantly passive. Only a minority produce their own content. Additionally, the stated advantages of professional SM use were low, while the stated concerns were high. The use of SM for professional purposes seems to play a minor role among orthopedic and trauma surgeons. Adequate education and information material on the professional use of SM is needed to potentially elucidate its applications and benefits, and to address legal concerns. Further studies are needed to validate the described trends.

Data Availability

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

Authors' Contributions

YY is the guarantor of this study. YY, TG, and DAB created the questionnaire. YY and JA recruited the participants. JS and YY performed the statistics and data analysis. YY and TG drafted the manuscript. JS supervised the study. All authors reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire (English translation).

References


Abbreviations

GDPR: General Data Protection Regulation
SM: social media

©Yasmin Youssef, Tobias Gehlen, Jörg Ansorg, David Alexander Back, Julian Scherer. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Abstract

The number of overdose-related fatalities continues to reach historic levels across Canada, despite ongoing efforts by authorities. To reduce mortality, a clinical trajectory ranging from preventative measures to crisis intervention, skill training to treatment, and risk assessment to risk management needs to be supported. The web-based Risk Assessment and Management Platform (RAMP) was developed to realize this concept and to empower people who use drugs through an integrated tool that allows them to better understand and manage their risk of overdose. This paper outlines the architecture and development of RAMP, which is built on the WordPress platform. WordPress components are mapped onto a 3-tier architecture that consists of presentation, application, and database layers. The architecture facilitates the development of a modular software that includes several features that are independent in functionality but interact with each other in an integrated platform. The relatively low coupling and high coherence of the features may reduce the cost of maintenance and increase flexibility of future developments. RAMP’s architecture comprises a user interface, conceptual framework, and backend layers. The RAMP front end effectively uses some of the WordPress’ features such as HTML5, CSS, and JavaScript to create a mobile, friendly, and scalable user interface. The RAMP backend uses several standard and custom WordPress plug-ins to support risk assessment and monitoring, with the goal of mitigating the impacts and eliminating risks together. A rule-based decision support system has been hard-coded to suggest relevant modules and goals to complement each user’s lifestyle and goals based on their risk assessment. Finally, the backend uses the MySQL database management system and communicates with the RAMP framework layer via the data access layer to facilitate a timely and secure handling of information. Overall, RAMP is a modular system developed to identify and manage the risk of opioid overdose in the population of people who use drugs. Its modular design uses the WordPress architecture to efficiently communicate between layers and provide a base for external plug-ins. There is potential for the current system to adopt and address other related fields such as suicide, anxiety, and trauma. Broader implementation will support this concept and lead to the next level of functionality.

(JMIR Form Res 2024;8:e49759) doi:10.2196/49759
KEYWORDS
software designs; risks management; risk assessments; opioid overdose; crisis intervention; substance related disorders

Introduction

The Clinical Problem

In 2016, the British Columbia government declared a state of emergency in response to the increasing number of drug-related overdose fatalities [1]. Despite this measure, more than 14,000 British Columbians have since died due to high potent opioids, and the province has seen a >2-fold increase in the number of overdose deaths, from 20.5 per 100,000 individuals in 2016 to 46 per 100,000 individuals in 2023 [2]. High-risk substance use and overdose risk is often a reflection of complex concurrent mental and physical conditions, which are critical for understanding the risk constellation as well as overdose risk management [3,4]. Integration of informed approaches to address severe and early traumatic experiences, suicidality, and mental illness is necessary for a solution aimed at increasing the chance of survival and recovery among people who use drugs [5]. Web-based interventions have demonstrated effectiveness in improving health outcomes and reducing barriers for people who use drugs who are seeking support [6,7].

A Modular Architecture

Modular architecture is a model of designing web-based psychotherapy platforms and interventions [8,9]. A modular design implies that the content presented can be divided into multiple smaller pieces that can be completed independently or as a set [9]. The strengths of modular systems include the ability for users to focus on 1 area at a time and ease in replicating modules following a common structure. Therefore, this approach allows for the inclusion of new clinical trajectories beyond substance use and overdose for a significantly lower price. It will also reduce the cost of maintenance. This architecture lends itself well to web-based interventions for substance use because sections, topics, and features can work both independently and together to provide care to a unique population with diverse needs. Chronic conditions such as substance use disorder need to be seen as a long-term trajectory—a dynamic and often changing process, which requires a range of tools to adapt to the needs of users.

An ideal web-based platform addressing substance use would create an environment that provides tools to assess risks, gain deeper knowledge of risks, prevent deterioration and crises, and manage specific behavioral factors and symptoms. The tasks for the development team would be to add content and functionalities to make the system engaging and interactive. Our vision is a platform that can become a long-term tool providing an engaging learning environment where user choices and outcomes influence the trajectory and functionality of the platform. Tools such as machine learning and embedded trials would allow for a systematic internal evaluation of user choices as well as better referrals and decision support tools. Users can adapt the content and functionalities to their needs, add content and their own experiences, and provide feedback to become active drivers of the system.

Risk Assessment and Management Platform

The Risk Assessment and Management Platform (RAMP) was developed by the Addictions and Concurrent Disorders Research Group at the University of British Columbia, supported by funding from Health Canada. The goal of this web-based platform is to provide a low-barrier tool that educates and empowers users to make informed decisions for their own recovery journey. The platform consists of self-assessments, educational and therapeutic modules, and tools to help users plan and monitor their risk of overdose, for example, “Goal setting” and “Lifechart.” Interactions among RAMP features to identify and manage the risk of overdose are shown in Figure 1.
Figure 1. Interaction among different features of the Risk Assessment and Management Platform to identify and manage the risk of overdose. HRA: High-Risk Assessment.

Users must create a free account in order to use the full functionalities of RAMP, including having their scores and data tracked over time. New users have the option to explore some selected features, such as modules and “Art with a Meaning” before they register. This allows users to better understand the system and feel comfortable with it enough to create an account. It also enables public access to modules, which, on its own, will have significant psychotherapeutic benefits for users.

Upon registration, the High-Risk Assessment (HRA) is proposed to the user. The HRA is a screening tool developed by the Addictions and Concurrent Disorders Research Group to identify the risk of overdose among people who use drugs through a rule-based algorithm. Development of the HRA will be explained in a separate publication.

Based on the results of each user’s HRA score, a clinical decision support system suggests the most relevant modules and goals to the user. The user may comply with the recommendations or choose to work on additional or substitute modules of their choice. The modules will help the user to gain knowledge about their risky behaviors and how to prevent or mitigate those risks. The Goals setting function will allow the user to make their risk mitigation plan and monitor their progress through the Lifechart feature. An email notification is sent to the user to inform them about their progress and to encourage them to continue interacting with the software by reminding them about their recent usage and how they can resume.

The core concepts and theoretical frameworks of RAMP will be presented in a series of publications. Topics will include the clinical concept underlying RAMP, development of the HRA, development of modules, and user interface evaluation. This paper is focused on the design and architecture of RAMP. The modular architecture of RAMP uses WordPress’ multilayer architecture, which lends itself well to the dynamic and customizable application design.

Design of RAMP

Overview

RAMP is a web application developed using the WordPress content management system (CMS). Therefore, description of the architecture of WordPress is essential to describe RAMP’s design and architecture.

WordPress is built on a PHP and MySQL architecture [10]. It uses a multilayer architecture to provide a flexible platform for building and customizing websites [11]. The purpose of a layered architecture is to organize the components of an application into horizontal logical layers and physical tiers. A layer is a logical unit that separates a specific role and responsibility within an application. Each layer manages its own software dependencies. A tier is a physical unit where the code runs, for example, a web server or a database. Scalability, maintainability, and resiliency are increased when tiers are physically separated; however, latency increases due to additional network communication.

Though WordPress does not strictly follow the traditional 3-tier architecture to separate the presentation, application, and database tiers, the WordPress components that are used to develop web applications can be mapped on a 3-tier design model [12,13].
Presentation Tier

The presentation tier is responsible for showing the content and functionality to the user, using HTML, CSS, and JavaScript. HTML provides a way to create and structure the elements of a web page, such as headings, paragraphs, images, links, etc. CSS is a powerful tool for customizing the appearance and layout of WordPress websites. JavaScript is a scripting language that is used to add interactivity, dynamic behaviors, and effects such as animations, interactive forms, and user interface elements such as sliders and pop-up menus to web pages.

WordPress Themes

Themes control the appearance of a WordPress website. They include templates, CSS files, and other components that determine how the site is displayed.

Application Tier

Overview

The application tier in WordPress refers to the core PHP code that provides the underlying functionality and controls the flow of data between the presentation and database tiers. The application tier includes the WordPress core, plug-ins, and custom code that interacts with the database to store and retrieve data.

Plug-ins are add-ons that extend the functionality of WordPress. They can provide additional features such as contact forms, e-commerce functionality, etc.

The application tier is responsible for implementing the business logic as well as accessing the data stored in the database through the data access layer.

Business Logic Layer

The business layer in WordPress architecture refers to the layer that handles the business logic and processes in a WordPress website. This layer is responsible for performing tasks such as data validation, calculations, and decision-making.

In the context of WordPress, the business layer can be implemented in a number of ways, including custom plug-ins, custom code within the theme, or through the use of third-party plug-ins that provide specific functionality.

The business layer acts as the intermediary between the presentation and database tiers. It takes input from the presentation layer, performs the necessary business logic, and then communicates with the database tier to retrieve or store data. This architecture allows for greater flexibility and maintainability of WordPress products, as well as the ability to easily modify and extend the functionality of the website without affecting other tiers.

Data Access Layer

The data access layer in WordPress architecture refers to the layer responsible for accessing and manipulating data stored in the database. This layer is responsible for retrieving data from the database, performing operations on the data, and storing the results back in the database.

In WordPress, the data access layer is typically implemented using the WordPress database application programing interface (API), which provides a set of functions and methods for interacting with the database. These functions and methods allow developers to perform common database operations, such as inserting, updating, and retrieving data, in a consistent and efficient manner.

Database Tier

The database tier in WordPress is where all the data are stored, including posts, pages, users, and metadata. WordPress uses a Relational Database Management System, MySQL, to store this information, and it is accessed by the application tier to retrieve and manipulate data.

The database layer in WordPress works by creating tables and storing data in those tables. Each table contains rows and columns that represent the different data types stored in the database. For example, WordPress has separate tables for posts, users, comments, etc.

When a user interacts with a WordPress site, the database layer retrieves data from the appropriate table and returns it to the application tier. The data can then be displayed on the user interface of the site or being used to perform various operations in the backend.

In conclusion, the database layer in WordPress applications is a critical component that provides a structured way to store and retrieve data and plays an important role in the overall architecture of a WordPress site.

Overall, the architecture of WordPress promotes modularity. This modular design approach allows for easier customization, scalability, and maintainability of software products built on the WordPress platform. Developers can create independent modules that can be combined or modified without significant impact on other components, facilitating the creation of flexible and extensible software systems.

RAMP software design architecture follows the WordPress multilayer architecture.

Ethical Considerations

The RAMP HRA pilot project has undergone research ethics review and received ethics approval from the University of British Columbia Behavioural Research Ethics Board (H19-02231). Collection of data on RAMP is limited to only those necessary for the research objectives. Participant identities are kept confidential, and the research protocol is aligned with the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2). Informed consent forms include information about data storage, data security, data sharing, privacy protections, legal limits to confidentiality, and contact information of the University of British Columbia Behavioural Research Ethics Board. It clearly delineates how the data will be used and shared, and the process to report participant concerns.

Upon registration, the system stores their registered email IDs, names, last names, usernames, passwords (encrypted), login times, and IP addresses. Data are stored on secure cloud servers,
and security features such as reCAPTCHA and temporary login blocks are used to minimize the risk of brute force attacks. Upon login, their interactions with the system are stored and tracked. This includes assessments results, symptoms check, completed modules, goals, journal diaries, and achievements. Data are not shared with third parties unless for maintenance, software improvements, or legal obligations. The overall data governance plan of the project is modeled after the highest local, federal, and international standards of privacy, data management, and data sharing.

**System Description**

**Overview**

RAMP is designed as an engaging, learning, and open access web application that empowers users to manage their risk of overdose and concurrent mental health disorders. RAMP is a modular software that includes several features and functionalities.

The architecture and design of RAMP features is mapped on a 3-tier architecture model that uses WordPress components. **Figure 2** demonstrates the mapping.

---

**Figure 2.** Mapping RAMP features on a 3-tier software architecture model that uses WordPress components. RAMP: Risk Assessment and Management Platform.
**Data Tier**

The Data Tier uses relational databases on the MYSQL 5.7.30 database management system. RAMP stores and retrieves various types of data including text, video, audio, and images, which can become bulky over time. The data layer is managed by the WordPress database tier, which provides secure and efficient methods to store, retrieve, and organize data. It uses indexing, caching, lazy loading, query optimization, and other techniques to improve its efficiency.

---

**The Backend**

The backend uses the WordPress 5.4 development framework. The system is a web application software compatible with most of the standard browsers such as Chrome, Safari, and Edge, and is responsive to mobile apps. The system’s administrative panel allows the administrator to add new reports, reorganize some user interface elements, and define new assessments (Figure 3). This panel also allows the administrator to monitor users’ activities and progresses and add new modules and other content. Certain key performance indicators can be found on the administrative panel.

![RAMP’s administrative panel on WordPress](https://formative.jmir.org/2024/1/e49759)

**Figure 3.** RAMP’s administrative panel on WordPress (accessible to administrators only).

The RAMP backend also uses some standard and custom plug-ins. Some examples of the plug-ins that RAMP is using are listed below:

- **Disable XML-RPC-API**: a lightweight plug-in to disable XML-RPC API, Pingbacks, and Trackbacks for faster and more secure browsing of a website.
- **Duplicate Page**: a plug-in to duplicate posts, pages, and custom posts with a single click.
- **LoginWP**: a plug-in to redirect users to different URLs based on their role, capability, and more.
- **Quform**: a plug-in to create powerful and engaging quizzes, tests, and examinations within minutes, and to build an unlimited number of quizzes and questions.
- **WP File Manager**: a plug-in that manages WP files.
- **WP SMTP**: a plug-in that helps developers to send emails via SMTP instead of the PHP mail function. This plug-in
is particularly useful to send email reminders to the user, which would less likely end up in junk mail; support superior authentication and encryption methods; and is more compatible with other WordPress plug-ins.

There are custom-built plug-ins created for different RAMP features such as Achievements, Assessments, Symptom Diary Forms, Journal, and Lifechart to enhance user experience.

The inference engine of a rule-based decision support system is custom-coded in the business layer. User responses to the HRA are used to determine relevant modules and goals, and to suggest them via the user interface.

**Conceptual Framework**

RAMP’s conceptual framework is part of the backend and is currently tailored to substance use but can be modified to manage the risk of other mental health conditions such as suicide and trauma. The principles are in line with the core concept of the risk management process, which includes identifying and assessing risks, making a plan, and monitoring the plan to be able to mitigate or eliminate the risk as needed (Figure 4).

![Figure 4. Risk management process.](https://formative.jmir.org/2024/1/e49759)

**The Front End**

RAMP front end uses WordPress presentation layer components such as themes, HTML5, CSS, and JavaScript to communicate with the user and provide a pleasant experience when showing content.

RAMP uses BuddyBoss as its core theme. However, a subtheme was derived and customized from BuddyBoss to enhance user experience and use consistent visual themes across RAMP’s branding.

Using HTML5 in RAMP has created more modern and effective web pages for the system with better accessibility, improved search engine optimization, higher performance, and mobile friendliness. Additionally, pages have been further fine-tuned by the developers for a better mobile experience. HTML5 has also allowed the developers to embed audio and video content directly into the web pages, eliminating the need to rely on third-party plug-ins such as Adobe Flash. RAMP is also using native lazy loading of images and script execution in parallel, which can improve the performance of web pages and reduce their load times.

RAMP uses CSS effectively to create websites that are visually appealing, flexible, responsive to mobile devices, and easy to use.

**Interaction With External Systems**

RAMP uses different WordPress plug-ins and features to interact with different external systems. YouTube videos in the modules are embedded into Elementor, a popular visual page builder plug-in for WordPress that allows users to create custom designs for their website and ensures the correct publication of the videos.

The mental health resource map is another example of an external system that is integrated into RAMP. The development team used Google My Maps to pin the location of mental health services, including clinics, youth services, and urgent care centers, onto a Google map. This software enables users to view services and facilities in their area, find their locations, hours of operation, contact information, and more on a Google map in a layered format. The system embeds this software using a simple script method by adding its link to RAMP.

The development team has also been working on a parallel project to create a machine learning model that would predict the risk of fatal overdose based on a comprehensive data set called the British Columbia Provincial Overdose Cohort curated by the British Columbia Centre for Disease Control. The model will subsequently be embedded into an API, which allows the predictive model to be used as a stand-alone tool. The aim is to use the developed predictive model in RAMP using a service-oriented architecture, where RAMP calls upon the predictive model’s API using WordPress REST API plug-in to send user-specific risk factors to the model. The model will then process the overall risk of fatal overdose based on user inputs, determine the likelihood of overdose, and send the results back to RAMP. The RAMP interface will display the results and provide tailored recommendations based on each individual’s risk factors.
Discussion

Background

The aim of RAMP is to empower people who use drugs to become more aware of their overdose risk and to provide a tool that promotes agency in helping them manage their risk factors. Although risk management is broad and may include several activities, the core concepts include key steps to identify, plan, and monitor risks, as well as tools to execute the plan to eliminate or mitigate the impact of these risks [14].

RAMP is a web application. In contrast to native apps, web applications have certain advantages that RAMP benefits from:

- Cross-platform compatibility and being installation-free: RAMP can run on any device with a web browser, which means that it is platform-independent and can be accessed from a wide range of devices, including desktops, laptops, tablets, and smartphones without the need to install the software on the device. In contrast, native apps are developed for a specific platform, such as iOS or Android, and need to be downloaded and installed on the user's device. This is helpful when considering the population of people who use drugs, who might not have consistent access to a smartphone but can access public or shared computers [15].
- Easier maintenance: RAMP maintenance and updating is easier and less costly than native apps, as changes can be made on the server side, and users are not required to download updates. This can save time and resources for developers and ensure that all users are using the most up-to-date version of RAMP.
- Lower development costs: developing a web application typically requires fewer resources than developing a native app, as there is no need to develop separate versions for different platforms such as iOS and Android. This can result in lower development costs and less time to deploy the application.

While web applications offer certain benefits, native apps offer advantages such as better performance, native features, push notifications, and access to device-specific capabilities. Web applications always need an active internet connection to the server, but native apps may not need constant access to the internet if there is no need to communicate with a centralized server on a continuous basis.

RAMP features effectively use the WordPress components and their multilayer architecture to support the user to make informed decisions. One example is the design of HRA, a 29-question screening tool that helps users identify several overdose risk factors across different domains.

The integration of a dynamic color-coded heat map on the user interface of the presentation tier, featuring a sliding score bar that slides as the questions are answered, assists the user in comprehending the relative significance of individual factors and whether they represent risks or protective factors. Upon responding to each question, the inference engine of a rule-based decision support system will use the responses that are stored in the RAMP database layer to determine and suggest the most relevant modules and goals that are critical to the management of the most pressing risks. The use of these features across layers empowers the user to accurately identify their risks, gain knowledge of such risks, establish relevant goals, and monitor progress—all through the use of RAMP.

RAMP’s modular design allows novice users to follow the decision support system recommendations and create relevant goals and educational plans, while providing more freedom to experienced users to focus on specific areas that are most related to their personal risk of overdose and lifestyle. This design is important for the risk planning phase. When the user executes the plan and performs the given tasks, the system provides rewards to further engage the user.

Strategies to Enhance User Engagement

Reward systems are one of the most common gamification strategies designed to encourage users. The system awards badges as the user makes progress. This provides reinforcement as an individual continues their recovery journey. There is broad interest in the use of digital badges to enhance learners’ motivation [16]. These web-enabled tokens of learning and accomplishment have the potential to induce excitement and elicit powerful forms of engagement and learning. Badges are one of the most powerful gaming elements that are widely used in digital health products to motivate users to keep interacting with the system [17]. This strategy is particularly important for the population of substance users where retention and long-term interaction might be difficult [18,19].

RAMP does not support on-screen or push notifications. Instead, the system uses reliable SMTP email notifications regularly to inform users about their progress and motivate them to continue their interaction with the system. Notifications are an important gamification tool to promote engagement with the platform [20]. The frequency and content of RAMP notifications is different for active and inactive users. Active users receive more frequent notifications with suggestions regarding next steps they might benefit from based on their current patterns of use, while inactive users may receive positive reinforcement notifications and encouragements to start interacting with the system or continue where they have stopped. Previous studies have shown that tailoring notifications to each user is an effective way of boosting the use of web-based health tools [21].

Application of RAMP to Other Disorders and Diseases

RAMP content is currently developed to support those struggling with substance use and who are at risk of overdose. However, the tool can also be used to empower patients in other related domains such as navigating trauma, managing anxiety, and preventing self-harm and suicide by adding relevant assessments, modules, and goals. Through the development of proper content, the same logic can be applied to other mental health domains without a need to customize or redevelop the features, which maximizes the reusability of the developed system. Furthermore, RAMP features are fairly independent but interact with each other based on the business logic and through interfaces. This is in line with the principles of software
engineering in which subsystems of a modular software should have high internal cohesion and low coupling [22].

Limitations
The knowledge of the authors about the application of WordPress plug-ins and add-ons is limited to RAMP. RAMP is a web application; therefore, general limitations of web applications such as requiring continuous internet connection, etc, are also applicable to the present software.

RAMP is developed on the WordPress CMS. Security and performance of the software are limited to the WordPress development framework. Although many successful and secure applications are developed on the same CMS platform, there is news about possible security exploitations of the websites developed on WordPress such as being infected by malwares, backdoors, or being hacked [23,24]. The RAMP development team has attempted to use security best practices such as using a secure host and server, security protocols, reCAPTCHA, reducing log-in attempts, disabling XML RPC, and many more features to create a safe and secure environment for the users.

RAMP is developed with a limited number of users willing to share their ideas. Wider use of RAMP may reveal other limitations.

RAMP neither substitutes or replaces medical advice by professionals nor provides emergency services during times of crisis. RAMP is designed for adults and is not yet adapted for adolescents and young adults.

Conclusions
Web-based platforms such as RAMP can be powerful assets to address significant gaps in the mental health and substance use care systems and can address a major public health crisis. It can integrate different settings of care from prevention to crisis management, all on one accessible platform. In terms of its architecture, RAMP uses a modular design that allows for maximal personalization and integration of user engagement strategies. WordPress components that are used to develop RAMP are mapped onto a 3-tier architecture, where the presentation, application, and database layers communicate to ensure proper functionality for users.

Although web applications seem to be an appropriate development technology for the current and future developments of RAMP, users may also benefit from having certain features such as “Assessments” and “Goals” on a mobile app. This would allow users to access these features even without a reliable internet connection and to receive on-screen notifications as appropriate.

RAMP is currently focused on substance use disorders and specifically reducing overdose, but the tool is flexible enough to include many more domains such as suicide, anxiety, self-harm, and trauma. Broader implementation will support this concept and lead to the next level of functionality. RAMP’s modular design makes it a promising tool for addressing a breadth of mental health conditions among vulnerable populations, increasing the capacity of clinical settings and reducing barriers to care. Further research will need to investigate whether this architecture style is appropriate for eHealth tools addressing different populations and appropriate implementation strategies for eHealth tools addressing people who use drugs.

Acknowledgments
The authors appreciate and recognize Health Canada for funding the RAMP project and its publications (1819-HQ-000122). We also appreciate TelusHealth for their cooperation on providing detailed information on the design and development technologies of RAMP.

Data Availability
Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

Conflicts of Interest
None declared.

References


Abbreviations

API: application programming interface
CMS: content management system
HRA: High-Risk Assessment
RAMP: Risk Assessment and Management Platform
Decentralizing Health Care: History and Opportunities of Web3

Aditya Narayan¹, BSc; Kydo Weng², BSc, MS; Nirav Shah¹, MPH, MD

¹Clinical Excellence Research Center, Palo Alto, CA, United States
²Computer Science Department, Stanford University, Stanford, CA, United States

Abstract

This paper explores the relationship between the development of the internet and health care, highlighting their parallel growth and mutual influence. It delves into the transition from the early, static days of Web 1.0, akin to siloed physician expertise in health care, to the more interactive and patient-centric era of Web 2.0, which was accompanied by advancements in medical technologies and patient engagement. This paper then focuses on the emerging era of Web3—the decentralized web—which promises a transformative shift in health care, particularly in how patient data are managed, accessed, and used. This shift toward Web3 involves using blockchain technology for decentralized data storage to enhance patient data access, control, privacy, and value. This paper also examines current applications and pilot projects demonstrating Web3’s practical use in health care and discusses key questions and considerations for its successful implementation.

JMIR Form Res 2024 | vol. 8 | e52740 | p.3221
doi:10.2196/52740

KEYWORDS
Web3; health care; patient-centric; data ownership; decentralization; interoperability; electronic health record (EHR); privacy; blockchain; digital transformation; digital health care; digital health; patient; patients; technological framework; security and privacy; security; privacy

Introduction

The relationship between the development of the internet and health care is a complex one, marked by parallel growth and mutual influence. As we have witnessed with the recent surge in interest surrounding generative artificial intelligence (AI), merely responding to technological innovations within the health care sector is insufficient; instead, we must encourage proactive readiness for the future. Accordingly, in this piece, we aim to explore the current trajectory of internet technologies alongside health care and emphasize the need for a shared vision of the path forward.

In many ways, the evolution of the internet parallels the iterative changes that have defined health care (Table 1). The internet’s earliest form was dubbed Web 1.0, or the “read-only web,” and was filled with static pages, unidirectional flows of information, and minimal opportunities for engagement. This is not unlike the early days of medicine, in which siloed physician expertise left little space for patients to have a voice in their care. Web 1.0 is exemplified by the early adoption of electronic health records (EHRs), which digitized patient records for easier access, albeit in a static form. Over the past several decades, we have transitioned from Web 1.0 to Web 2.0, which emphasizes bidirectional information flows with user-generated content, social networks, and the democratization of information. In the same way, medicine has evolved in light of advances in medical technologies and a growing recognition of the critical role of patients making decisions about their own care. Web 2.0 is characterized by enhanced interactivity and user-generated content, with platforms such as patient portals adding another layer of patient engagement. Regarding public consumption, social media platforms, for instance, facilitated health communities wherein patients could share experiences and build support networks. Further, this era saw the beginning of digital information dissemination in health care. Web-based health information repositories like WebMD provide patients and health care providers with access to medical information. Thus, as a field, health care has made significant progress. However, in the context of Web 2.0, we remain bound by limited interoperability while various entities benefit from the storage,
use, and sale of personal information. We are now on the cusp of a new era of the internet—and potentially a new era of health care—with the advent of Web3.

Table 1. Parallels between the evolution of the web and medical practice.

<table>
<thead>
<tr>
<th>Evolution of the web</th>
<th>Evolution of health care</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Web 1.0</strong></td>
<td><strong>Early medicine</strong></td>
</tr>
<tr>
<td>- Static webpages have minimal opportunities for users to engage with content on the page or with other users.</td>
<td>- Physician-centric—the early stages of medicine were largely dominated by physicians and their expertise. The flow of information and decision-making were unidirectional, from physician to patient. Additionally, there were fewer roles within health care teams, thereby limiting the division of labor and wraparound service delivery.</td>
</tr>
<tr>
<td>- There is a unidirectional flow of information from the web page to the user.</td>
<td>- Expertise is siloed by specialty or organ system, with limited crosstalk.</td>
</tr>
<tr>
<td>- It is text-based with limited media.</td>
<td>- Limited patient engagement—patients had minimal input or agency in their own health care. They were often passive recipients of care.</td>
</tr>
<tr>
<td><strong>Web 2.0</strong></td>
<td>- Limited technology—the use of technology in health care was limited or rudimentary, with only foundational electronic health record platforms. Technology primarily focused on direct patient care with little emphasis on data management or communication.</td>
</tr>
<tr>
<td>- Web 2.0 introduced a read-write web where users could both consume and contribute information.</td>
<td>- Medicine in the era of technology</td>
</tr>
<tr>
<td>- Social networks—the emergence of social networking platforms allowed users to connect, collaborate, and share information.</td>
<td>- Increased patient engagement—this era saw a shift toward patient-centered care, where patients have become more active participants in their health care journey with the ability to access (but not own) their health care data.</td>
</tr>
<tr>
<td>- User-generated content—websites began to allow users to upload, modify, and share content.</td>
<td>- Integration of technology—novel technologies have become deeply integrated into health care, leading to the development of telemedicine, patient portals, and wearable technologies.</td>
</tr>
<tr>
<td>- User-driven applications—the development of more advanced web technologies allowed for user-driven applications and services, like blogs, wikis, and media-sharing platforms.</td>
<td>- Interdisciplinary approach—the practice of medicine has become increasingly interdisciplinary, involving teams of diverse health care professionals working together to provide care.</td>
</tr>
<tr>
<td><strong>Web3</strong></td>
<td><strong>The future of medicine</strong></td>
</tr>
<tr>
<td>- Web3 uses machine learning, artificial intelligence, and natural language processing to understand and interpret information.</td>
<td>- Data ownership—patients may have full control and ownership of their health data, which could be securely stored and accessed on decentralized networks, improving privacy and interoperability.</td>
</tr>
<tr>
<td>- Decentralization—data are stored on decentralized networks of computers rather than controlled by individual entities.</td>
<td>- Advanced health care technologies—the integration of artificial intelligence, blockchain, large language models, and other advanced tools may lead to novel health care solutions like smart contracts for health insurance, predictive health analytics, precision treatments, and so on.</td>
</tr>
<tr>
<td>- User control—it is designed to give end users more control over their own data and web-based interactions. It eliminates the need for intermediaries, enabling secure, peer-to-peer interactions.</td>
<td>- Patient autonomy—patients may have increased agency over their involvement in research and clinical care, reinforcing the shift toward truly patient-centered approaches.</td>
</tr>
</tbody>
</table>

Web3 is also referred to as the decentralized web, or the blockchain web, and represents the next stage of the internet wherein data are stored on decentralized networks of computers rather than by individual, centralized entities. Web3 aims to create a more secure, transparent, and user-owned paradigm built on blockchain technology and peer-to-peer networks which enable users to securely interact with one another without the need for intermediaries. Web3 broadly may be thought of as a “new” internet—in which data and web-based interactions are owned and controlled by the end users.

Sharing of patient data initially collected for clinical care or research between health care institutions and third-party companies where it might be used for commercial gain is common [1]. While being cared for, patients may unknowingly agree to the deidentified sharing of their data with external entities. Explicit permission for such sharing is sometimes not required if the data are deidentified but nonetheless limits patient agency. As health care institutions increasingly recognize the commercial value of such data, they engage in activities to refine and monetize these data through specialized third-party organizations. For example, Truveta, founded by a consortium of health systems, exists to structure and commercialize their data assets on behalf of the health systems [2]. The trend extends to hospitals creating and divesting their own spin-off companies for specific data like genetics and annotated pathology. Moreover, companies like Komodo Health aggregate diverse data sets to develop products and services such as clinical trial planning [3]. These commercial use cases highlight the value of the data itself—and the potential for greater patient agency in deciding how their health care data are used. Web3’s architecture could potentially offer more transparent and secure data management solutions for such uses.

With a move to Web3, we may further shift power to patients from insurers, the government, and health systems. For health
Several health care standards, platforms, and techniques are actively paving the way for the implementation of Web 3.0 in health care. For example, the Solid (Social Linked Data) framework, pioneered by Sir Tim Berners-Lee (creator of the World Wide Web), seeks to facilitate a decentralized web by enabling individuals to store their personal health data in “pods” or personal data stores. This grants them the autonomy to manage and share their data securely and efficiently. This democratization of data management resonates with a vision of an open, collaborative internet [15].

Notably, many early Web3 use cases center largely around generating financial value—for example, buying and selling Bitcoin. In fact, countries such as Estonia, have accelerated the Web 2.0 to Web3 transition with significant public investment and widespread adoption of Web3 within the financial sector [16]. However, we wish to focus our discussion on the technology undergirding these transactions and the unique prospect of individual ownership of data. Here, we seek to raise the promise and potential pitfalls of Web3, alongside the associated implications for health care.

**Potential of Web3**

Central to the discussion of Web3 is the value of personal ownership of health care data. In the current health care landscape, there is growing concern regarding the ownership and use of patient data. Typically, these data are collected and managed by health systems, service providers, and health insurers. While primarily intended for clinical purposes, there have been discussions and concerns in the public and academic domains about the potential for these data to be leveraged in ways that extend beyond these original intentions [17]. These concerns underscore the importance of clear data governance policies and highlight the ethical implications of data management in health care. The evolution toward systems like Web3 seeks to address these concerns by offering enhanced data ownership and control to patients, thus potentially mitigating the risks associated with data misuse. The global market for data monetization is projected to reach more than US $15 billion by 2030 [18]. This growth highlights the increasing value and potential revenue from data monetization, though it must be noted that business models and strategies for data monetization vary significantly across industries and companies.

While there are instances where health data may be consolidated and disseminated, potentially for profit, there exists a legal and ethical landscape governing these practices. Ethical and policy guidelines, such as those outlined by the US Centers for Disease Control, mandate the responsible use of such data, ensuring the protection of individual privacy and adherence to confidentiality norms. For example, we might consider the UK’s National Health Service. There, a national data opt-out system is in place for the secondary use of confidential health data for research and planning. However, a study revealed that there is limited public awareness of this opt-out system, with many participants being unaware that their anonymized health data could be used for secondary purposes such as research and health planning by...
entities outside of the National Health Service, including academic and commercial organizations [19].

In addition to facilitating personalized ownership, Web3 offers the potential for digital transformation and an opportunity to overcome the limitations of Web 2.0. Perhaps the most appealing target lies in the reform of EHRs—originally a Web 1.0–based health care technology. Imagine a future in which any patient can view their health records on their cell phone that is hosted by a decentralized network that only they can access. They open the app and it not only shows all their records but also all the scientific papers they contributed to and the insights those papers have generated. Lastly, they can easily change the sharing access of different data points on the device itself. This level of engagement will not just change the relationship between the patient and the health care system, but also the general public with science and research.

Current trends in EMRs already hint at the transition toward a decentralized web as they are evolving to become more patient-focused, with enhancements in data interoperability and security. These developments suggest a shift toward a model where patients play a more active role in managing their health care data. In the foreseeable future, EMRs are likely to further align with decentralized web concepts. This includes the adoption of blockchain technology for secure and transparent data management to streamline health care procedures and empower patient consent. Additionally, the incorporation of decentralized identity solutions in EMRs will enable secure and independent verification of patient identities, reducing dependency on centralized systems. These advancements, which are already being piloted, represent a significant step toward the next generation of web technologies in health care [20]. In contrast to traditionally siloed patient data, Web3 offers the potential of reimagining EHRs as mutable, patient-centric, and patient-owned, where linked data will allow tailored treatments that account for a patient’s unique health history, genetics, environment, and lifestyle. This collective potential may be too great to ignore.

**Five Key Questions to Answer**

Despite its theoretical benefits, at present the decentralized web is nascent, and the benefits are largely unrealized. Before Web3 reaches widespread adoption, we believe health care leaders should consider several key questions.

**How Can We Secure Privacy in a Web3 Context?**

While Web3 technologies enable ownership, without privacy protections for information sharing, ownership can only go so far. Even with the enhanced security features of Web3 technologies, they are not immune to vulnerabilities. The decentralized nature of blockchain, while reducing certain types of security risks, can also introduce new vulnerabilities. These vulnerabilities may be significant, particularly in light of blockchain’s immutable nature. Researchers have attempted to tackle the challenge of privacy through numerous technical approaches—though many current solutions sacrifice computational speed for privacy [21]. However, emergent solutions have begun to reach the stage of adoption [22-24]. For example, companies such as Onai achieve privacy protection through cryptographic techniques such as secure multiparty computation [25]. Broadly, compliance with legal standards like HIPAA (Health Insurance Portability and Accountability Act) and General Data Protection Regulation is also a must, necessitating regular updates to the governance framework in response to evolving legal requirements. An example is the need to consider data fidelity in the context of General Data Protection Regulation’s “right to be forgotten” in which individuals have the right for their personal data to be erased [26]. Additionally, policies should be patient-centric, prioritizing patient needs and privacy, and include guidelines for ethical data use, especially in research.

**How Will We Facilitate Change Management?**

The health care field is conservative by design, and Web3-based technologies often mandate a high level of technical knowledge to implement and use. Further, the centralized model of Web 2.0 is largely incapable of operating alongside Web3’s decentralized approach, and so both will exist in parallel for a time. Thus, there will exist challenges to seamless interoperability between blockchain systems and traditional EHRs leading to data siloes. Beyond this, existing legacy systems within health care institutions may not align seamlessly with the advanced capabilities of Web3 technologies, necessitating significant technical expertise and likely infrastructure upgrades. Practical considerations regarding infrastructure include establishing a robust network infrastructure with adequate bandwidth, setting up blockchain nodes with the requisite server capacity and storage solutions, decentralized identity verification systems to securely manage patient identities, and experience design for user-friendly interfaces. Accordingly, change management toward Web3 systems will require a staged approach supported by iterative stages of implementation.

**How Can Web3 Scale While Maintaining an Equity Lens?**

As new technologies arise and each offers novel opportunities, we must remain centered on the ethical mandates of medicine. Access to Web3 mandates the presence of new tools and technical expertise which few health systems possess [27]. Thus, as technologies scale, an opportunity lies in creating an incentive schema that reimburses for equitable adoption without limiting the innovation potential. That is, payers, including insurance companies and government health programs, could establish incentive structures that financially reward health care providers for adopting and effectively using Web3 technologies. These incentives might be linked to specific metrics including, but not limited to, the level of data interoperability achieved, the extent of patient data control facilitated, or the efficiency gains in patient care and administrative processes. Similar to the meaningful use criteria established for EMRs, payers could define a set of criteria or benchmarks for Web3 implementation, offering reimbursements or bonuses to providers who meet these standards. This approach would not only accelerate the adoption of Web3 technologies across different health care settings but also ensure that their integration is aligned with improved patient outcomes and system efficiency. Additionally,
there is likely to be user resistance due to the complexities of blockchain technology and the digital divide. Patients in underresourced or rural areas, or those who are less tech-savvy, might find it challenging to engage with Web3-based health care systems, potentially exacerbating existing health care disparities.

**How Can We Avoid Wasteful Spending?**

The adoption of Web3 in health care has the potential to help address the persistent “grand challenge” of escalating expenditure in US health care. Examples include enabling more personalized treatment plans to reduce wasteful procedures, improving interoperability to enable efficient care delivery, and enhancing fraud prevention with transparent transactions. However, to realize these benefits, the initial adoption of Web3 must prioritize empirical, high-value principles—focusing on efficiency, appropriateness, and patient-centeredness—to avoid generating new expenses in the rollout process. The integration of Web3 into health care systems, akin to any substantial systemic transition, presents a unique opportunity to reevaluate and realign the system with its foundational principles. Web3’s decentralized architecture, characterized by technologies like blockchain, offers distinct advantages that can be harnessed to enhance these principles more effectively than current systems. Most notably, Web3 technologies provide a framework that can shift the focus back to patient-centered care by giving patients greater autonomy and control over their health data. Such a systemic reorientation during the transition to Web3 not only aligns with the ongoing evolution of health care but also ensures that these fundamental principles are more deeply ingrained in the fabric of health care systems. Through this transition, as described previously, it will be important to curtail costs. The scalability and performance of blockchain in processing large data volumes, given limitations in managing high-throughput data, must be assessed [27]. This scalability challenge could lead to slower transaction processing times and increased costs, potentially hindering the widespread adoption of Web3 in large health care systems.

**How Can the Health Care System Support Policy Makers in Preparing for the Arrival of Web3?**

As a field, we must acknowledge that policy regulation is unlikely to keep pace with technological innovation. Accordingly, legal clarity and guardrails will be necessary to ensure industry players, health systems, and academic experts collaborate to create shared endpoints that center on consumer protection, data privacy, and improved care. With fast-growing interest in generative AI and the creation of large language models specific to health care, questions of data ownership, use, and incorporation into new Web3-based tools bring new challenges that have yet to be addressed at scale.

**Considerations for Implementation**

**Overview**

In the process of transitioning to Web3 systems in health care, a phased approach is paramount for effective change management, implementation, and adoption. Integrating Web3 technologies into health care systems begins with a thorough assessment of existing IT infrastructure to identify areas needing upgrades or changes for Web3 compatibility, which is already a prerequisite for any health system addressing cybersecurity and regulatory compliance. Implementation will also require data migration and ensuring seamless integration with existing health care databases and applications, a step crucial for maintaining data continuity and integrity. Further, it will be essential to define clear rules regarding data ownership, explicitly outlining patient rights and access conditions. Incorporating blockchain technology can facilitate granular control, allowing patients to specify access permissions, thereby creating an auditable trail of data access. This initial phase may be followed by defining specific use cases where Web3 can add significant value, such as in patient data management, supply chain transparency, or facilitating research collaborations. These use cases help to focus the direction of Web3 implementation.

For a health system, initially, strategic planning forms the cornerstone of this process. After prioritizing Web3 opportunities, a detailed implementation roadmap with a sufficient budget must be developed. Stakeholder engagement should include technology experts and vendors, health care providers, IT staff, and patients.

Throughout this process, it will be necessary to scaffold rollout with clinical staff training. As Web3 introduces advanced technologies like blockchain and smart contracts, health care providers must undergo specialized training to become adept in these new systems, focusing on a basic technical understanding and practical clinical application. This shift will also lead to changes in clinical workflows; for instance, the processes for accessing and sharing patient data will evolve, necessitating adaptations to new data retrieval and sharing protocols. Health care professionals will need to adapt to more dynamic decision-making processes due to the real-time nature of data updates on blockchain platforms. Additionally, the accuracy and accessibility improvements provided by Web3 could boost clinical efficiency but will require new competencies in data management. Crucially, the patient-centric model of Web3, which grants patients greater control over their data, will transform patient-provider dynamics, placing a greater emphasis on shared decision-making and patient engagement.

The challenges of enforcing privacy protection laws in the Web 3.0 era, where patient control over health care data is paramount, may be addressed through a combination of approaches. These include implementing robust encryption, smart contracts, and security protocols to safeguard patient data against unauthorized access, a measure that gains importance given the existence of extant and ongoing challenges with health care data breaches. To monitor such breaches, comprehensive incident response plans by technology platforms may be developed alongside IT leaders. At a policy level, the development of legal frameworks specifically designed for decentralized data management in health care will be central for providing clear guidelines on liability and actions in the event of a data breach. Finally, the launch of the Web3 system should be coupled with ongoing postimplementation support to address any technical issues or concerns, ensuring the long-term effectiveness and efficiency of Web3 technologies in the health care environment.
Conclusions

Ultimately, the advent of Web3, with its revolutionary capabilities of decentralization, user data ownership, and advanced privacy, presents transformative possibilities for health care. It portends a shift of power from traditional entities like insurers and health systems to patients—granting individuals the agency to make informed decisions regarding their involvement in research and clinical care. Successful integration of Web3 in health care will hinge on a multidisciplinary approach that combines technological innovation with practical, ethical, and regulatory considerations. By focusing on these areas, we can pave the way for a health care ecosystem that is not only more efficient and patient-centric but also adaptable to the evolving digital landscape.

Conflicts of Interest

None declared.

References


Abbreviations

AI: artificial intelligence
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
Solid: Social Linked Data

©Aditya Narayan, Kydo Weng, Nirav Shah. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
AI Analysis of General Medicine in Japan: Present and Future Considerations

Nozomi Aoki¹, MD; Taiju Miyagami¹, MD, PhD; Mizue Saita¹, MD, PhD; Toshio Naito¹, MD, PhD
Department of General Medicine, Juntendo University Faculty of Medicine, Tokyo, Japan

Corresponding Author:
Taiju Miyagami, MD, PhD
Department of General Medicine
Juntendo University Faculty of Medicine
2-1-1 Hongo
Tokyo, 113-8421
Japan
Phone: 81 3 3813 3111
Email: miya0829gami@gmail.com

Abstract

This paper presents an interpretation of artificial intelligence (AI)–generated depictions of the present and future of general medicine in Japan. Using text inputs, the AI tool generated fictitious images based on neural network analyses. We believe that our study makes a significant contribution to the literature because the direction of general medicine in Japan has long been unclear, despite constant discussion. Our AI analysis shows that Japanese medicine is currently plagued by issues with polypharmacy, likely because of the aging patient population. Additionally, the analysis indicated a distressed female physician and evoked a sense of anxiety about the future of female physicians. It discusses whether the ability to encourage the success of female physicians is a turning point for the future of medicine in Japan.

(Int.) doi:10.2196/52566

KEYWORDS
artificial intelligence; physicians; hospitalists; polypharmacy; sexism; Japan; AI; artificial intelligence; medicine; Japan; gender-biased; physicians; physician; medical care; gender; polypharmacy; women; Pharmacology; older adults; geriatric; elderly; Japanese

Introduction

The concept of “general medicine in Japan,” born around the 1970s, led to 2 main physician categories in Japan: family doctors and hospitalists. Unlike general medicine departments overseas, there are no barriers to practicing within either family medicine or hospital medicine; both are unique specialties that require the adaptation of work to the size and conditions of the institutions they belong to, such as hospitals and clinics [1]. Additionally, some general medicine department managers are physicians who have transferred from specialized departments, and the mindset and competencies differ across institutions. Thus, the state of general medicine in Japan is unclear, and we have been unable to provide a clear answer despite more than 50 years of unremitting discussion. Since then, the field of general medicine has gradually expanded, and as of 2017, 86% of university hospitals have a department of general medicine [2]. Although the Japanese government and patients themselves have recently begun to recognize the importance of general medicine doctors, the public still lacks awareness of general medicine practitioners who do not require patients or their families to visit the hospital [3,4]. Therefore, we used artificial intelligence (AI) to assess the society’s perception of general medicine doctors in Japan. What answer will AI provide?

What Is the Answer Shown by AI?

We used a tool called DALL-E (Open AI), which analyzes text from a neural network and produces a graphical representation. We inputted the text “General medicine in Japan” and “Future of general medicine in Japan” to generate fictitious images [5].

General Medicine in Japan

The entry “General medicine in Japan” generated a picture of several drugs (Figure 1), which we interpreted as being indicative of the polypharmacy observed in Japanese medicine. Compared to the rest of the world, Japan is a superaging society. As of 2013, 25% of the population is older than 65 years [6]. Polypharmacy, the combination of various drug therapies for comorbid diseases, is one of the problems seen within an aging
society [7]. In Japan, polypharmacy has other causes. For example, all citizens are covered by medical insurance, providing ease of access to hospitals and medical care at reasonable fees at any time. One of the disadvantages of this system is that patients can see specialists immediately, which is why family doctors have fewer patients. This further encourages polypharmacy as patients visit multiple departments [8], leading to a growing need for doctors who can provide broad-spectrum diagnostic and treatment opportunities to patients. Both general doctors with diverse work backgrounds and polypharmacy are complex issues in Japan, which require total reform. Based on this concept, AI may have created polypharmacy as a graphic representation of the most common medical issue seen in Japan.

Figure 1. General medicine in Japan.

Future of General Medicine in Japan

Interestingly, “Future of general medicine in Japan” presented a completely different graphic. Despite approximately 70% of doctors in Japan being male, only 4 female doctors were depicted to be distressed (Figure 2). Although the percentage of female doctors in Japan is on the rise, in 2018, it was reported that women are less likely to be accepted into some medical schools than men [9], therefore raising awareness of sex discrimination in the Japanese medical community. Several reports describe the difficulties of working as a female doctor; during promotions and evaluations, women in Japan often feel discriminated against [10-13]. Moreover, the gender gap at work in Japan is by far the largest among high-income countries [14]. Currently, all the 19 medical specialty societies in Japan are primarily represented by men; a survey study of the actual state of general medicine found that the ratio of men to women was 9:1, indicating that women are grossly underrepresented in general medicine [15]. Many within general medicine have expressed concern about the impact of this issue on future health care. In response, the Japanese Society of Hospitalists and General Medicine established a division for female physicians in February 2021. In 2022, as part of the Women Physicians Subcommittee’s activities, national and international comparisons of female hospitalists were proposed [16]. However, activity still remains limited: the inclusion of women has increased only slightly since inception, and the number of female leaders remains low [16]. It must also be stated that an absence of women in leadership positions can cause female doctors to lose their role models, create an even wider gender gap, and encourage further bias in the medical workforce [17].
The future that AI envisioned may have focused on female physicians, who are currently reported to be oppressed in Japan. As a solution, it is necessary to support the activities of female physicians in general medicine and develop a workplace where women can work comfortably as general physicians. It has been reported that older hospitalized patients treated by female internists have lower mortality and readmission rates than those cared for by male internists [18]. The activities of female physicians may have important clinical significance for future medical care in the superaged Japanese society. Social change in the current gender-biased image of physicians and acceptance by the population is a great hope for Japanese medicine.

Limitations
Using AI, we assessed the current and future status of general medicine practice in Japan. However, the criteria used to select the AI image output are unclear, making the academic evidence less credible. Scientific publications on the subject of AI have been submitted since 2023. While it has been reported that AI may be an effective tool in medical education [19], careful evaluation of the preparation of the article is recommended [20]. However, the picture painted by the AI in this report is consistent with previously reported social facts. It is a perspective that we humans are unaware of or avoid looking at; we believe that this view cannot be easily ignored.

Conclusions
AI depicts “General medicine in Japan” as polypharmacy, the “Future of general medicine in Japan” as distressed female physicians. While this AI-generated picture cannot be verified or avoided, it shows the future challenges associated with the practice of general medicine in Japan.

Conflicts of Interest
None declared.

References


Abbreviations

AI: artificial intelligence
Assessment of Qatar’s Health Care Community Call Center Efficacy in Addressing COVID-19 Pandemic Health Care Challenges: Cross-Sectional Study

Muhammad Atif Waheed¹, MBBS, MRCS, MRCGP, DPD, MBA; Lolwa Al Mannai¹, MD, ABCM; Hanan Khudadad¹, MPH; Jamil Alenbawi¹, PhD; Mariama Aminata Mansaray¹, LLB, MA; Samya Al Abdulla¹, MBBCh, ABFM
Primary Healthcare Corporation, Qatar, Doha, Qatar

Corresponding Author:
Muhammad Atif Waheed, MBBS, MRCS, MRCGP, DPD, MBA
Primary Healthcare Corporation, Qatar
Al Mina Street
Doha, 40000
Qatar
Phone: 974 33015895
Email: dratifwaheed@gmail.com

Abstract

Background: The global COVID-19 pandemic caused by SARS-CoV-2 created many unprecedented challenges for health care organizations worldwide, placing a great deal of strain on the health care systems, especially access to health care services. To address these challenges, Qatar established a centralized digital platform as a community call center, initially offering digital consultations via its hotline (number: 16000) and later expanding to include a COVID-19 vaccination hotline (number: 7077) for mass immunization.

Objective: This study aims to comprehensively examine the community call center’s operations and their significant role during the COVID-19 pandemic.

Methods: Retrospective data were collected from the Health Information and Technology Department of the Primary Health Care Corporation, Qatar, from March 29, 2020, to January 27, 2022. Data analysis for the hotline (number: 16000) focused on telephone and video call volumes, call response rates, abandonment rates, and call classification. In addition, data from the COVID-19 vaccination hotline (number: 7077) were analyzed for call volumes, call response rates, abandonment rates, appointment booking rates, confirmations, rescheduling, and cancellations.

Results: The hotline (number: 16000) received a substantial total of 429,212 calls, with 284,849 (66.37%) calls effectively answered. The average number of calls received per day during the study period was 640.61 (SD 470.53), and the average number of calls answered per day was 425.14 (SD 206.64). Notably, of the total 128,468 consultations, video consultations were conducted for 3810 (2.96%). Among the diverse call categories, diabetes mellitus (6284/84,299, 7.45%), prescriptions and medications (4709/84,299, 5.59%), hypertension (3874/84,299, 4.6%), vitamin D-related issues (3770/84,299, 4.47%), upper respiratory tract infections (2690/84,299, 3.19%), and COVID-19–related inquiries (2590/84,299, 3.07%) were most frequently addressed. For the COVID-19 vaccination hotline (number: 7077), an impressive total of 1,512,354 calls were received, with a 58.27% (n=881,305) call response rate. The average number of calls per day during the study period was 3828.74 (SD 2931.94), and the average number of calls answered per day was 2231.15 (SD 1496.02). Appointment booking accounted for 26.37% (265,721/1,007,596), appointment confirmation accounted for 10.24% (103,136/1,007,596), rescheduling accounted for 7.95% (80,124/1,007,596), and cancellations accounted for 1.6% (16,128/1,007,596) of the calls.

Conclusions: The findings of this research highlight the crucial significance of the community call center hotline (number: 16000) and the COVID-19 vaccination hotline (number: 7077) in effectively addressing the multifaceted challenges posed by the global COVID-19 pandemic. In Qatar, the community call center emerged as an indispensable and accessible centralized resource, facilitating streamlined digital consultations and vaccination appointments. The impressive call response rate highlights its operational efficiency, adeptly managing a diverse range of health-related issues. This study emphasizes the critical role of community call centers in health care emergency response, signaling their potential as invaluable assets for future preparedness and effective mitigation strategies during similar public health crises.
Introduction

Background

SARS-CoV-2, an enveloped single-stranded RNA virus with an epicenter in the Hubei province of China, has caused the current COVID-19 pandemic [1]. It initially appeared as an outbreak of an unknown cause of pneumonia and marked the seventh coronavirus outbreak [2]. Following its emergence, it rapidly spread globally, resulting in significantly high morbidity and mortality rates. At the time of writing this paper, on August 18, 2022, there were 589,680,368 confirmed cases of COVID-19 globally, with 6,436,519 deaths and a fully vaccinated population of 4,857,273,828 [3]. In Qatar, 7,355,038 doses of COVID-19 vaccination were administered until the date of writing this paper.

The COVID-19 pandemic poses many challenges to health care systems worldwide and places considerable strain on them. One of the main challenges is to respond to the crisis and simultaneously maintain the provision of essential health care services, which is crucial for a high-quality and resilient health care system. During the COVID-19 pandemic, many essential health care services were disrupted, including cancer screening, tuberculosis screening, HIV testing, outpatient services, and maternal and child health services [4]. The lack of health care workers, diversion of health care staff to COVID-19 management, cancellation of planned treatments, and risk of viral transmission during on-site patient visits have disrupted health care services [5]. It is essential to take necessary steps to curtail the spread of COVID-19 using available resources and the best use of digital technologies [6]. Different countries have responded differently to the COVID-19 crisis. In the United Kingdom, the National Health Services established a COVID Response Service, accessible via 111 phone lines, with the recruitment of 5000 call handlers and 1500 retired clinicians [7]. The Gulf Cooperation Council (GCC) countries implemented various measures in response to the pandemic, including lockdowns of major cities, airline suspensions, school and university closures, restrictions on social gatherings and sporting events, free health care provision, and active screening for COVID-19 [8]. In Saudi Arabia, the Umrah pilgrimage was suspended, and travel restrictions were placed on GCC citizens who had visited COVID-19–affected countries [9]. To protect public health as a national health strategy in Qatar, the Ministry of Public Health has provided digital solutions by providing remote access channels to health care services at the Primary Health Care Corporation (PHCC) and Hamad Medical Corporation in collaboration with the TASMU Smart Program Qatar, Ministry of Transport and Communication, Hukoomi, and Qatar Post, along with notable digital solution providers [10]. A hotline (number: 16000) was set up to provide 24/7 service to patients’ inquiries regarding COVID-19. In addition, the PHCC established an inbound community call center on March 29, 2020, accessible via the hotline (number: 16000), to provide digital virtual primary care consultations as an alternative to face-to-face health center visits. Physicians and nurses working in health centers who were deemed to be at a high risk for COVID-19, such as those with chronic conditions (hypertension, diabetes, ischemic heart disease, chronic kidney disease, pregnancy, immunocompromised state, etc), were given the option of working in the community call center. On January 5, 2021, the community call center added a COVID-19 vaccination hotline (number: 7077) for booking, cancelling, and rescheduling COVID-19 vaccination appointments [11]. The community call center hotline (number: 16000) provides telephone and video consultation services to all registered patients from 28 health centers in Qatar. Individuals can access the call center through hotline (number: 16000) and book appointments for telephone or video consultation with physicians, dentists, and ophthalmologists from 7 AM to 11 PM. Initially, the calls were triaged by the nurses. For patients who are not registered with the health centers, such as visitors and single workers, the triage nurses direct them to the appropriate service. In cases of emergencies requiring immediate medical attention, patients are guided to dial 999 for ambulance services as appropriate. For nonemergency situations, visitors are directed to local health centers, whereas single male workers are referred to worker’s health centers (HC-21), which are operated by the Qatar Red Crescent Society (QRCS). A nurse-led telephone triage service is available from 11 PM to 7 AM, which signsposts patients to appropriate services [12]. Although primarily established for the COVID-19 pandemic, it provided all types of consultations, whether urgent or nonurgent. There were 7 physicians and 9 nurse stations. There was 1 ophthalmology station, whereas for the COVID-19 vaccination hotline (number: 7077), there were 22 stations. The staffing level of the community call center varied according to its operational needs during the study period.

Digital consultations offer several advantages for patients, providers, and health care systems [13]. Patients benefit from avoiding waiting in queues, reducing travel burdens [14], convenience, cost efficiency [15], accessibility [16], and high levels of satisfaction in primary care settings [17]. Satisfaction levels are particularly high in digital consultations, which include communicating with physicians, addressing patients’ concerns and queries, developing treatment plans, improving illness comprehension, and offering usefulness and reliability [5]. During pandemics, digital consultations provide an excellent alternative to traditional face-to-face consultations for patients [18]. Providers also experience advantages, including flexible working hours, the ability to work from anywhere; a reduced risk of infection; increased job satisfaction [13,19,20]; and less psychological distress, burnout [13], and sickness, which can be a burden on the organization. From an organizational perspective, digital consultations provide centralized operations with weekly statistics and demand forecasting, eliminating unnecessary patient visits to health centers and resulting in...
smooth operations and reduced clinic congestion [13]. Moreover, they enable health care services in remote areas and offer an opportunity for service expansion whenever possible. In addition, they contribute to lower CO₂ emissions and cost savings [20].

Call centers can serve as central hubs to respond to public health emergencies by providing rapid information transfer to health care providers [21]. During the COVID-19 pandemic, telehealth call centers played a significant role in supporting rural community health workers in Uganda, facilitating prompt identification and referral of COVID-19 cases for appropriate care [22]. In South Korea, the telehealth system provided up-to-date information to callers, helping them protect themselves and others from COVID-19 effectively [23]. Similarly, China established psychological support hotlines to offer mental health assistance during the pandemic [24], and in Paris, France, health care workers benefited from psychological support services [25].

**Objective**

Despite the importance of call centers during the pandemic, there is limited information available on their operations and outcomes, especially in relation to the COVID-19 pandemic and the vaccination hotline. Existing literature does not include reports from GCC countries detailing the role and impact of community call centers during the pandemic. This study aims to fill this gap by examining call center operations; documenting their contributions; and analyzing call volumes, patterns, response rates, categories of calls, priorities, and problems or diagnoses. By investigating the effectiveness of community call centers in Qatar during the COVID-19 pandemic, this research seeks to provide valuable insights into their performance during public health emergencies.

**Methods**

**Study Design**

A cross-sectional study design was used to assess the community call center hotline operations and performance during the study period.

**Disease and Study Population**

This research mainly evaluated community call center services to handle the COVID-19 pandemic health care challenges, specifically focusing on their utilization and effectiveness. The study population consisted of patients seeking routine and urgent consultations and advice regarding COVID-19 and its vaccination.

**Location**

Data were retrospectively collected from the Health Information and Technology Department of the PHCC, Qatar headquarters, where the community call center is located.

**Time Frame**

Data for the community call center hotline (number: 16000) were collected from March 29, 2020, to January 27, 2022. Similarly, data for the community call center vaccination hotline (number: 7077) were collected from December 29, 2021, to January 27, 2022. This time frame encompassed critical phases of the COVID-19 pandemic and the subsequent mass vaccination campaign.

**Data Collection and Analysis**

This study involved meticulous analysis of 2 separate data sets: one pertaining to the hotline service (number: 16000) and the other concerning the COVID-19 vaccination hotline (number: 7077). The hotline service (number: 16000) data set was analyzed for various essential metrics, including call volume, percentage of calls answered or abandoned, call patterns, categories of calls, and specific health concerns addressed. Similarly, the COVID-19 vaccination hotline (number: 7077) data set was thoroughly scrutinized, considering call volume; call response rates; call patterns; and appointment-related information such as bookings, rescheduling, confirmations, and cancellations.

**Data Cleaning**

To ensure utmost data quality and reliability, a robust data cleaning process was meticulously executed. The 3-step approach encompassed vigilant screening of the databases to detect and address any suspicious features, precise diagnosis of faulty data to ensure accurate results, and appropriate treatment of identified discrepancies. Moreover, a separate copy of the original data was meticulously created in a new file to maintain data integrity during the cleaning process.

**Data Reporting**

Conforming to the highest standards of scientific reporting, this study adhered to the RECORD statement, an extension of the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement checklist. This ensured accurate and transparent reporting of the secondary data analysis, further enhancing the credibility of the study findings [26].

**Ethical Considerations**

This study obtained ethical approval from the PHCC Research Subcommittee, adhering to research ethics and ensuring participant protection (PHCC/DCR/2021/11/068). As the study used existing deidentified data, informed consent was waived in accordance with the approved protocol. Stringent privacy and confidentiality measures were implemented to safeguard the data. The study strictly complied with data protection regulations to maintain the confidentiality of sensitive information. No compensation was provided to human participants, as the study involved secondary analysis of deidentified data and did not involve direct interaction with participants. The research team took utmost care to handle the data responsibly and ethically throughout the study.

**Results**

**Community Call Center Hotline (Number: 16000)**

Table 1 shows the quarterly performance of the community call center hotline (number: 16000). During the study period, 66.37% (284,849/429,212) of the calls were handled, with the highest being handled in the third quarter (33,395/34,376, 97.15%), followed by the fourth quarter (34,543/35,885, 96.26%) of 2021.
The average number of calls per day during the study period was 640.61 (SD 470.53), and the average number of calls answered per day was 425.14 (SD 206.64).

**Figure 1** shows the hotline (number: 16000) call volume patterns during the study period. The peak of calls occurred on April 19, 2020; April 7, 2021; and January 5, 2022.

**Figure 2** shows the volume and pattern of the video consultations conducted during the study period. Video consultations accounted for 2.96% (3810/128,468) of the total consultations conducted. The number of video consultations conducted was highest in the beginning of the COVID-19 pandemic reaching to its peak in the first week of July 2020 and then gradually declined.

**Figure 3** illustrates problem-based summary of consultations over hotline (number: 16000). The highest number of calls was related to diabetes mellitus (6284/84,299, 7.45%), prescriptions or medications (4709/84,299, 5.59%), hypertension (3874/84,299, 4.6%), vitamin D-related issues (3770/84,299, 4.47%), upper respiratory tract infection (2690/84,299, 3.19%), COVID-19 (2590/84,299, 3.07%), thyroid-related problems (2237/84,299, 2.65%), dermatological problems (1807/84,299, 2.14%), dyslipidemia (1358/84,299, 1.61%), pregnancy (1320/84,299, 1.57%), back pain (1216/84,299, 1.44%), fever (1089/84,299, 1.29%), cough (1060/84,299, 1.26%), dry eyes (1004/84,299, 1.19%), gastritis (984/84,299, 1.17%), and many other problems or diagnoses.

### Table 1. Quarterly (Q) performance of community call center hotline (number: 16000).

<table>
<thead>
<tr>
<th>Time period</th>
<th>Total calls, n (%)</th>
<th>Total calls answered, n (%)</th>
<th>Total calls abandoned, n (%)</th>
<th>Calls/d, mean (SD)</th>
<th>Calls answered/d, mean (SD)</th>
<th>Calls abandoned/d, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2020</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1(^a) (n=3146)</td>
<td>3146 (100)</td>
<td>1906 (60.58)</td>
<td>1240 (39.42)</td>
<td>1048.67 (125.43)</td>
<td>635.33 (75.74)</td>
<td>413.33 (200.97)</td>
</tr>
<tr>
<td>Q2 (n=57,628)</td>
<td>57,628 (100)</td>
<td>37,732 (65.48)</td>
<td>19,896 (34.52)</td>
<td>633.27 (365.44)</td>
<td>414.64 (159.40)</td>
<td>218.64 (223.79)</td>
</tr>
<tr>
<td>Q3 (n=64,172)</td>
<td>64,172 (100)</td>
<td>37,224 (58.01)</td>
<td>26,948 (41.99)</td>
<td>697.52 (318.96)</td>
<td>404.61 (127.12)</td>
<td>292.91 (218.14)</td>
</tr>
<tr>
<td>Q4 (n=43,642)</td>
<td>43,642 (100)</td>
<td>27,834 (63.78)</td>
<td>15,808 (36.22)</td>
<td>474.37 (214.21)</td>
<td>302.54 (98.53)</td>
<td>171.83 (137.15)</td>
</tr>
<tr>
<td><strong>2021</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 (n=68,364)</td>
<td>68,364 (100)</td>
<td>35,070 (51.3)</td>
<td>33,294 (48.7)</td>
<td>759.60 (386.49)</td>
<td>389.67 (138.61)</td>
<td>369.93 (283.37)</td>
</tr>
<tr>
<td>Q2 (n=70,255)</td>
<td>70,255 (100)</td>
<td>49,831 (70.93)</td>
<td>20,424 (29.07)</td>
<td>772.03 (421.53)</td>
<td>547.59 (164.05)</td>
<td>224.44 (304.45)</td>
</tr>
<tr>
<td>Q3 (n=34,376)</td>
<td>34,376 (100)</td>
<td>33,395 (97.15)</td>
<td>981 (2.85)</td>
<td>373.65 (133.23)</td>
<td>362.99 (128.06)</td>
<td>10.66 (10.31)</td>
</tr>
<tr>
<td>Q4 (n=35,885)</td>
<td>35,885 (100)</td>
<td>34,543 (96.26)</td>
<td>1342 (3.74)</td>
<td>390.05 (258.57)</td>
<td>375.47 (212.67)</td>
<td>14.59 (59.67)</td>
</tr>
<tr>
<td>Q1 2022(^b) (n=51,744)</td>
<td>51,744 (100)</td>
<td>27,314 (52.79)</td>
<td>24,430 (47.21)</td>
<td>1916.44 (944.09)</td>
<td>1011.63 (220.61)</td>
<td>904.81 (764.96)</td>
</tr>
<tr>
<td><strong>Total</strong> (n=429,212)</td>
<td>429,212 (100)</td>
<td>284,849 (66.37)</td>
<td>144,363 (33.63)</td>
<td>640.61 (470.53)</td>
<td>425.14 (206.64)</td>
<td>215.46 (311.91)</td>
</tr>
</tbody>
</table>

\(^a\)March 29 until March 31, 2020.

\(^b\)January 1 until January 27, 2022.
Figure 1. Call volume patterns of hotline (number: 16000).

Figure 2. Volume and pattern of video consultations over the hotline (number: 16000).

Figure 3. Problem-based summary of consultations over hotline (number: 16000). GERD: gastroesophageal reflux disease; URTI: upper respiratory tract infection; UTI: urinary tract infection.
Community Call Center Vaccination Hotline (Number: 7077)

Table 2 shows the quarterly performance of the COVID-19 vaccination hotline (number: 7077). The COVID-19 hotline (number: 7077) handled approximately 58.27% (881,305/1,512,354) of the total calls, whereas 41.73% (631,049/1,512,354) were abandoned during the study period. The average number of calls per day during the study period was 3828.74 (SD 2931.94), whereas the average number of calls answered per day was 2231.15 (SD 1496.02). The highest percentage of calls answered was during the third quarter of 2021 (112,445/118,372, 94.99%), while the lowest number of calls answered was during the first quarter of 2021 (180,205/311,364, 42.12%).

Figure 4 shows the volumes of calls and the pattern received by the COVID-19 vaccination hotline (number: 7077) during the study period. The highest number of calls was received around mid-May and November and the last week of December 2021.

Table 3 displays the quarterly call categories for the COVID-19 vaccination hotline (number: 7077). The highest number of appointments booked occurred in the fourth quarter of 2021 (86,924/1,007,596, 8.63%), followed by that in the second quarter (81,331/1,007,596, 8.07%). Similarly, the highest number of confirmed appointments occurred during the second (51,122/1,007,596, 5.07%) and fourth quarter (16,735/1,007,596, 1.66%) of 2021. The highest number of appointment cancellations occurred in the second (7028/1,007,596, 0.7%) and third quarter (2072/1,007,596, 0.21%) of 2021. Notably, the highest call volume identified for the workers’ health centers (HC-21) operated by QRCS occurred during the second quarter of 2021 (62,208/1,007,596, 6.17%), followed by the fourth quarter of 2021 (43,735/1,007,596, 4.34%). Calls from patients who did not meet the age vaccination criteria, as announced by the Ministry of Public Health of Qatar, were the highest in the second quarter of 2021 (57,081/1,007,596, 5.67%), followed by the third quarter of 2021 (8337/1,007,596, 0.83%).

Table 4 illustrates the distribution of call categories for the COVID-19 vaccination hotline (number: 7077) during the study period. Among the incoming calls, approximately 26.37% (265,721/1,007,596) were associated with appointment bookings, 10.24% (103,136/1,007,596) were for confirming existing appointments, 7.95% (80,124/1,007,596) involved rescheduling appointments, and 1.6% (16,128/1,007,596) were calls to cancel previously scheduled appointments. In addition, 14.8% (149,117/1,007,596) of the calls were identified as related to the workers’ health centers (HC-21) operated by QRCS, while 7.11% (71,600/1,007,596) of the calls received did not meet the age criteria for vaccination.

Table 2. Quarterly (Q) performance of COVID-19 vaccination hotline (number: 7077).

<table>
<thead>
<tr>
<th>Time period</th>
<th>Total calls, n (%)</th>
<th>Calls answered, n (%)</th>
<th>Calls abandoned, n (%)</th>
<th>Calls/d, mean (SD)</th>
<th>Calls answered/d, mean (SD)</th>
<th>Calls abandoned/d, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 2020a (n=852)</td>
<td>852 (100)</td>
<td>297 (34.86)</td>
<td>555 (65.14)</td>
<td>284 (36.01)</td>
<td>99 (21.52)</td>
<td>185 (57.42)</td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 (n=310,842)</td>
<td>310,842 (100)</td>
<td>131,159 (42.19)</td>
<td>179,683 (57.88)</td>
<td>3453.80 (1735.19)</td>
<td>1457.32 (849.34)</td>
<td>1996.48 (1214.87)</td>
</tr>
<tr>
<td>Q2 (n=555,336)</td>
<td>555,336 (100)</td>
<td>356,259 (64.15)</td>
<td>199,077 (35.85)</td>
<td>6102.59 (2315.49)</td>
<td>3914.93 (885.59)</td>
<td>2187.66 (1788.70)</td>
</tr>
<tr>
<td>Q3 (n=118,372)</td>
<td>118,372 (100)</td>
<td>112,445 (94.99)</td>
<td>5927 (5.01)</td>
<td>1286.65 (887.29)</td>
<td>1222.23 (836.57)</td>
<td>64.42 (52.90)</td>
</tr>
<tr>
<td>Q4 (n=361,461)</td>
<td>361,461 (100)</td>
<td>190,959 (52.83)</td>
<td>170,502 (47.17)</td>
<td>3928.92 (3627.34)</td>
<td>2075.64 (1546.77)</td>
<td>1853.28 (2217.60)</td>
</tr>
<tr>
<td>Q1 2022b (n=165,491)</td>
<td>165,491 (100)</td>
<td>90,186 (54.5)</td>
<td>75,305 (45.5)</td>
<td>6129.30 (2040.44)</td>
<td>3340.22 (650.37)</td>
<td>2789.07 (1792.58)</td>
</tr>
<tr>
<td>Total (n=1,512,354)</td>
<td>1,512,354 (100)</td>
<td>881,305 (58.27)</td>
<td>631,049 (41.73)</td>
<td>3828.74 (2931.94)</td>
<td>2231.15 (1496.02)</td>
<td>1597.59 (1790.57)</td>
</tr>
</tbody>
</table>

aDecember 29 until December 31, 2020.

bJanuary 1 until January 27, 2022.
Table 3. Quarterly (Q) call categories of COVID-19 vaccination hotline (number: 7077; n=1,007,596).

<table>
<thead>
<tr>
<th>Time period</th>
<th>Rescheduled, n (%)</th>
<th>Canceled, n (%)</th>
<th>Booked, n (%)</th>
<th>Confirmed, n (%)</th>
<th>&gt;1 appointment, n (%)</th>
<th>Worker’s health centers (HC-21), n (%)</th>
<th>Unmet vaccination age criteria, n (%)</th>
<th>Other n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 2020a</td>
<td>81 (0.01)</td>
<td>14 (0)</td>
<td>0 (0)</td>
<td>78 (0.01)</td>
<td>13 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>68 (0.01)</td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>17,164 (1.7)</td>
<td>3514 (0.35)</td>
<td>42,559 (4.22)</td>
<td>13,195 (1.31)</td>
<td>823 (0.08)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>62,762 (6.23)</td>
</tr>
<tr>
<td>Q2</td>
<td>24,306 (2.41)</td>
<td>7028 (0.70)</td>
<td>81,331 (8.07)</td>
<td>51,122 (5.07)</td>
<td>1349 (0.13)</td>
<td>62,208 (6.17)</td>
<td>57,081 (5.67)</td>
<td>121,467 (12.06)</td>
</tr>
<tr>
<td>Q3</td>
<td>11,856 (1.18)</td>
<td>2072 (0.21)</td>
<td>20,442 (2.03)</td>
<td>14,737 (1.46)</td>
<td>656 (0.07)</td>
<td>16,006 (1.59)</td>
<td>8337 (0.83)</td>
<td>50,256 (4.99)</td>
</tr>
<tr>
<td>Q4</td>
<td>17,815 (1.77)</td>
<td>1943 (0.19)</td>
<td>86,924 (8.63)</td>
<td>16,735 (1.66)</td>
<td>240 (0.02)</td>
<td>43,735 (4.34)</td>
<td>4132 (0.41)</td>
<td>60,223 (5.98)</td>
</tr>
<tr>
<td>Q1 2022b</td>
<td>8902 (0.88)</td>
<td>1557 (0.15)</td>
<td>34,465 (3.42)</td>
<td>7269 (0.72)</td>
<td>221 (0.02)</td>
<td>27,168 (2.7)</td>
<td>2050 (0.2)</td>
<td>23,692 (2.35)</td>
</tr>
<tr>
<td>Total</td>
<td>80,124 (7.95)</td>
<td>16,128 (1.6)</td>
<td>265,721 (26.37)</td>
<td>103,136 (10.24)</td>
<td>3302 (0.33)</td>
<td>149,117 (14.80)</td>
<td>71,600 (7.11)</td>
<td>318,468 (31.61)</td>
</tr>
</tbody>
</table>

aDecember 29 until December 31, 2020.
bJanuary 1 until January 27, 2022.

Table 4. Distribution of call categories for COVID-19 vaccination hotline (number: 7077).

<table>
<thead>
<tr>
<th>Call category</th>
<th>Values (n=1,007,596), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booked</td>
<td>265,721 (26.37)</td>
</tr>
<tr>
<td>Confirmed</td>
<td>103,136 (10.24)</td>
</tr>
<tr>
<td>Rescheduled</td>
<td>80,124 (7.95)</td>
</tr>
<tr>
<td>Cancelled</td>
<td>16,128 (1.6)</td>
</tr>
<tr>
<td>Worker’s health centers (HC-21)</td>
<td>149,117 (14.8)</td>
</tr>
<tr>
<td>Unmet vaccination age criteria</td>
<td>71,600 (7.11)</td>
</tr>
<tr>
<td>Other</td>
<td>318,468 (31.61)</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

According to the findings of this study, the community call center hotline (number: 16000) and the COVID-19 vaccination hotline (number: 7077) have played a pivotal role during the ongoing COVID-19 pandemic. These hotlines effectively answered 66.37% (284,849/429,212) of the calls for urgent consultation requests and 58.27% (881,305/1,512,354) of the calls for COVID-19 vaccination appointment inquiries. Notably, the community call center serves as a centralized and versatile resource for citizens and residents of Qatar, offering consultations for a wide range of cases, whether they are emergencies, urgent, or nonurgent, and enabling centralized operations for the mass vaccination program against COVID-19. It has provided a hub during challenging times when health care organizations have witnessed a rapid shift from face-to-face to virtual consultations.

The results of this study showed that after 4 weeks of the first COVID-19 case in Qatar, which was announced on February 29, 2020, the community call center was fully operating. The trajectory of the call volumes received is in line with the peaks of COVID-19 cases, with the first one occurring on May 30, 2020, the second occurring in the middle of April 2021, and the third occurring in the middle of January 2022. As a result, during the COVID-19 pandemic including the lockdown phases, the people of Qatar were able to receive the best medical care. This has also prevented the spread of COVID-19 while providing alternative solutions for patients to access primary health care services [27]. With the community call center, patients have been empowered in terms of flexibility and convenience to use virtual consultations as an alternative to traditional health care services.

The community call center’s diagnostic categories revealed its capacity to handle a diverse array of medical concerns, encompassing inquiries related to COVID-19; vaccination inquiries; acute problems; chronic diseases; medication and repeat prescription requests; requests for laboratory investigations; and various medical conditions related to ear, nose and throat, ophthalmology, dermatology, psychiatry, gynecology, gastrointestinal, neurology, cardiovascular, orthopedics, hematology, pediatrics, and endocrine and metabolic conditions. Interestingly, patients with diabetes mellitus accounted for the highest number of consultations, and virtual consultations have proved effective for managing type 2 diabetes with outcomes comparable with face-to-face consultations, as demonstrated in primary care settings in Australia [28]. Similarly, the management of hypertension via virtual consultations during the COVID-19 pandemic has offered a holistic approach to achieve hypertension control in the United States [29].

It is intriguing to note that video consultations experienced a significant surge in demand during the early phase of the COVID-19 pandemic, especially in the first week of July 2020, reaching a peak of 30 video consultations per day. However, following this peak, the use of video consultation gradually declined over time. Although there is some evidence indicating that video consultations were more sought-after during periods of higher COVID-19 cases, such as in early July 2020 and mid-April 2021, it is essential to highlight that this correlation is not consistently observed throughout the data set. Notably, there were instances of relatively low video consultation counts during times of higher COVID-19 cases, such as in mid-January 2022. The lower use of video consultations can be attributed to various factors, including patients’ preferences influenced by cultural and language barriers, their comfort or inclination toward telephone or face-to-face consultations, privacy and security concerns, and technical limitations. Cultural reasons play a role in shaping patient preferences; some individuals may feel more at ease and familiar with consultation methods such as telephone calls. The content, duration, and quality of video consultations might resemble those of telephone consultations, which could affect patient choices [30], aligning with the findings reported in existing literature. Interestingly, in the United Kingdom, video consultations constitute <1% of consultations in general practice [31]. This preference for other consultation methods, including telephone and face-to-face consultations, might be influenced by physicians’ perceptions that video consultations offer minimal advantages compared with the alternatives.

The COVID-19 vaccination hotline (number: 7077) played a crucial role in Qatar’s mass immunization program, facilitating services for booking, cancellations, and rescheduling initially based on professional group and age criteria. The program was launched on December 21, 2020, introducing the BNT162b2 (Pfizer-BioNTech) messenger RNA vaccine initially, followed by the mRNA-1273 (Moderna) vaccine 3 months later. Initially, priority was given to vaccinating frontline health care workers, individuals with severe or multiple chronic diseases, and those aged ≥70 years. Subsequently, the program was extended gradually by 1 age group at a time, along with selected professional groups, using age as the primary eligibility criterion throughout the rollout [32]. The data from the vaccination hotline revealed noteworthy call volume spikes during mid-May, mid-November, and the last week of December 2021, coinciding with new announcements related to the vaccination program’s expansion. However, the percentage of vaccination hotline’s answers was observed to be lower than that of digital consultation phones. This disparity in the response rates can be attributed to multiple factors. Periods of high demand, such as vaccination eligibility expansions or public announcements about vaccination campaigns, led to a surge in hotline calls, potentially affecting response times due to the increased call volume. Specific eligibility criteria for vaccination appointments also played a role, as some callers may not have met these criteria, resulting in redirection or absence of appointments. Technical issues, waiting times, and call congestion further influenced the hotline’s response rate. Notably, the third quarter of 2021 demonstrated a higher response rate for the vaccination hotline. This could be attributed to a relatively lower volume of calls received during this period, possibly influenced by factors such as expats traveling during summer vacations and the vaccination hotline’s capacity to meet the demand effectively. Understanding these dynamics is crucial for optimizing hotline performance and enhancing vaccination service accessibility. Both vaccines in Qatar have been found
to elicit strong protection against COVID-19, prevent hospital admission, and reduce mortality [33]. Due to mass vaccination supported by the community call center and other measures taken by the Ministry of Public Health, Qatar’s mortality remained very low at 0.0016%, whereas globally, it was 1% at the time of writing this paper.

One of the most significant outcomes of the COVID-19 pandemic is that many outpatient appointments can now be managed efficiently via telemedicine without affecting patient care [34]. The COVID-19 pandemic has shown that health care workers can swiftly adjust to the new technologies required to use telemedicine [13]. In a study involving 23 primary care providers and 1692 patients, both providers and patients reported a desire to continue telemedicine visits after the pandemic in primary care settings in the United States [17]. Even before the COVID-19 pandemic, the telehealth business flourished, with a market value of >US $50 billion in 2019 and an anticipated growth rate of more than 9-fold over the next decade [35]. Virtual consultant jobs have also been advertised [36].

The significance of the community call center and vaccination hotline in providing consultations, vaccinations, and health care services during a challenging period demonstrates the potential and value of telehealth solutions in health care systems worldwide. By sharing the successful experiences and best practices of the community call center hotlines, this study can contribute to the enhancement of telehealth services and call centers globally. The lessons learned from operating the hotline can be invaluable for improving crisis response strategies and optimizing health care delivery during public health emergencies. By leveraging digital technologies, health care systems can enhance access to services, improve patient satisfaction, and manage public health emergencies effectively. Integrating call centers and telehealth into routine health care services and emergency response strategies can provide long-term benefits beyond the pandemic.

Limitations

This study has several limitations. The data presented in this study involved 3 peaks of COVID-19 that affected the overall percentage of calls answered by the hotline (number: 16000). The hotline answered 100% of all calls on several days of the week during the study period. Perhaps, the percentage of calls answered improved to 71.71% at the time of writing this paper. Similarly, for the COVID-19 vaccination hotline (number: 7077), a surge in the volume of calls due to a new announcement of rolling out of the vaccination program affected the overall percentage of the calls answered. COVID-19 vaccination hotline (number: 7077) also answered >97% of the calls on many days of the week during the study period. The percentage of calls handled by the COVID-19 vaccination hotline (number: 7077) improved to 62.17% by the date of writing this paper.

It is important to acknowledge that although this study provides valuable insights into the role of community call center during the COVID-19 pandemic in Qatar, the applicability of these findings to low- and middle-income countries (LMIC) may be subject to limitations. LMIC often face unique challenges, including resource constraints, technological disparities, and varied health care infrastructures. The digital divide prevalent in many LMIC could significantly exacerbate the access-to-technology gap, hindering the establishment and effectiveness of call centers. Moreover, the lack of access to robust health systems or timely responses in LMIC might impact the feasibility and sustainability of implementing similar call center solutions. The success of such initiatives in Qatar, being one of the richest countries globally, may not necessarily translate seamlessly to LMIC because of the nuanced economic, infrastructural, and health care disparities prevalent in those regions. Therefore, although call centers can be a valuable tool, particularly in health care emergencies, their feasibility and effectiveness must be carefully evaluated in the context of LMIC’s unique challenges and constraints.

Conclusions

This study highlights the significant role played by the community call center hotline (number: 16000) and the COVID-19 vaccination hotline (number: 7077) in effectively managing the challenges caused by the COVID-19 pandemic in Qatar. The community call center responded well to these challenges in Qatar by providing patients with an accessible, centralized resource as an alternative platform for telephone and video consultations and for managing a mass vaccination program. The findings demonstrate the call center’s operational efficiency to handle high call volumes and offer diverse consultations, effectively addressing a wide range of health concerns throughout different waves of the COVID-19 pandemic. Virtual consultations have emerged as a practical solution to empower patients with flexibility and convenience. Although video consultations experienced a temporary surge, their overall use gradually declined during the specified study period, reflecting patients’ preference for other consultation methods. Nevertheless, the community call center remained instrumental in managing various health care aspects, including chronic disease management and COVID-19–related queries. Moreover, the COVID-19 vaccination hotline played a major role in executing a successful mass immunization program. Prioritizing frontline health care workers and susceptible groups initially, the vaccination drive significantly contributed to reducing mortality rates in Qatar. This accomplishment can be attributed to the coordinated efforts of the community call center and other strategic measures implemented by the Ministry of Public Health of Qatar. As the COVID-19 pandemic recedes, it will be essential to adapt the community call center’s capacity and its services to address future health care challenges. Expanding community call center services to encompass specialized health services such as dermatology, counseling, health education, and psychological support could further enhance Qatar’s health care infrastructure. Our study emphasizes the critical significance of telemedicine and digital solutions in crisis response and health care delivery. The insights gained from the community call center in Qatar during the pandemic will provide valuable lessons for improving health care accessibility and emergency response strategies in the future. As health care systems transition to postpandemic norms, leveraging the knowledge and experience gained will foster resilience and efficacy in addressing upcoming challenges, ensuring the well-being of the population.
Acknowledgments
Publication fees was provided by Primary Health Care Corporation, Qatar. The authors received no specific funding for this work.

Data Availability
All data generated or analyzed during this study are included in this published paper.

Conflicts of Interest
None declared.

References


**Abbreviations**

- GCC: Gulf Cooperation Council
- LMIC: low- and middle-income countries
- PHCC: Primary Health Care Corporation
- QRCS: Qatar Red Crescent Society
- STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
Waheed et al

©Muhammad Atif Waheed, Lolwa Al Mannai, Hanan Khudadad, Jamil Alenbawi, Mariama Aminata Mansaray, Samya Al Abdulla. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

Jim Swift¹, BA, MSc; Noel O’Kelly¹, MB BCh; Chris Barker¹, BSc; Alex Woodward², BSc, MSc; Sudip Ghosh²,³, MD, PhD

¹Spirit Health, Leicester, United Kingdom
²Leicestershire Partnership NHS Trust, Leicester, United Kingdom
³Department of Allied Health Sciences, De Montford University, Leicester, United Kingdom

Corresponding Author:
Jim Swift, BA, MSc
Spirit Health
Spirit House, Saffron Way
Leicester, LE2 6UP
United Kingdom
Phone: 44 1162865000
Email: jim.swift@spirit-health.com

Abstract

Background: The COVID-19 pandemic stressed global health care systems’ acute capacity and caused a diversion of resources from elective care to the treatment of acute respiratory disease. In preparing for a second wave of COVID-19 infections, England’s National Health Service (NHS) in Leicester, Leicestershire, and Rutland sought to protect acute capacity in the winter of 2020-2021. Their plans included the introduction of a digital ward where patients were discharged home early and supported remotely by community-based respiratory specialists, who were informed about patient health status by a digital patient monitoring system.

Objective: The objective of the digital ward was to maintain acute capacity through safe, early discharge of patients with COVID-19 respiratory disease. The study objective was to establish what impact this digital ward had on overall NHS resource use.

Methods: There were no expected differences in patient outcomes. A cost minimization was performed to demonstrate the impact on the NHS resource use from discharging patients into a digital COVID-19 respiratory ward, compared to acute care length of stay (LOS). This evaluation included all 310 patients enrolled in the service from November 2020 (service commencement) to November 2021. Two primary methods, along with sensitivity analyses, were used to help overcome the uncertainty associated with the estimated comparators for the observational data on COVID-19 respiratory acute LOS, compared with the actual LOS of the 279 (90%) patients who were not discharged on oxygen nor were in critical care. Historic comparative LOS and an ordinary least squares model based on local monthly COVID-19 respiratory median LOS were used as comparators. Actual comparator data were sourced for the 31 (10%) patients who were discharged home and into the digital ward for oxygen weaning. Resource use associated with delivering care in the digital ward was sourced from the digital system and respiratory specialists.

Results: In the base case, the digital ward delivered estimated health care system savings of 846.5 bed-days and US $504,197 in net financial savings across the 2 key groups of patients—those on oxygen and those not on oxygen at acute discharge (both P<.001). The mean gross and net savings per patient were US $1850 and US $1626 in the base case, respectively, without including any savings associated with a potential reduction in readmissions. The 30-day readmission rate was 2.9%, which was below comparative data. The mean cost of the intervention was US $223.53 per patient, 12.1% of the estimated gross savings. It was not until the costs were increased and the effect reduced simultaneously by 78.4% in the sensitivity analysis that the intervention was no longer cost saving.

Conclusions: The digital ward delivered increased capacity and substantial financial savings and did so with a high degree of confidence, at a very low absolute and relative cost.
Introduction

Overview

The COVID-19 pandemic stressed global health care systems, diverting resources from elective care and prioritizing the care of people with acute respiratory disease [1].

The introduction of a digital ward that commenced enrolling patients in November 2020 in Leicester, Leicestershire, and Rutland (LLR) was a key part of National Health Service (NHS) preparations for an expected surge in COVID-19 infections in the winter of 2020-2021.

The background to this decision was the impact of the first wave on LLR acute bed availability and the use of digital technology to remotely monitor patients to support their chronic conditions, both at the start of the COVID-19 pandemic and previously [2,3]. There was evidence to support the care of people after discharge with a high risk of readmission [2-4], and digitally supported patients with infectious respiratory disease had a reduced length of stay (LOS) [5-7].

The primary objectives for the COVID-19 digital ward were to maintain acute bed capacity through safe, early acute discharge and step-down into a specialist respiratory–managed service in the patients’ homes. Patients with COVID-19 respiratory infections were discharged into the care of the Leicester Partnership Trust (LPT) specialist respiratory team who were supported by digital technology (Clinitouch, Spirit Health), which electronically conveyed clinical observations and health status information to clinicians in an algorithm-based, traffic light (red, amber, and green)–prioritized basis. The standard operating procedures (service description), inclusion and exclusion criteria, objectives, and outcome measures for the digital ward are displayed on the internet [8], and more detail related to the intervention can be found in a study that regarded the first 65 patients that accessed the digital ward [9].

Paper-based, phone-based, digitally-based, and wearable device–based digital wards were used during COVID-19 as vehicles for admission avoidance and for stepping down patients who had been acutely admitted. There were mixed results [10].

The findings could have implications for future pandemics and how health systems allocate resources to better recover from the pandemic.

Objective

The objective of this study was to demonstrate the impact of a digital COVID-19 ward on NHS resource use. Specifically, the study aimed to establish if the digital ward achieved its primary goal of freeing up beds and the extent to which it reduced or increased overall NHS resource use.

Methods

Participants

There were 310 patients admitted to the University Hospitals Leicester (UHL) NHS Trust with COVID-19 respiratory disease; they were either discharged home between November 2020 and November 2021 into a digital ward to support their oxygen weaning (31/310, 10%) or had not required oxygen at discharge and needed additional support to recover (279/310, 90%). The patients’ mean age was 55.0 (SD 13.7; median 56; range 22-86) years, 3.2% (n=10) of patients were 80 years or older, and 40.6% (n=143) of patients were female. No ethnicity, comorbidity, or socioeconomic status information was collected. Patients were given the option to go home early and be supported digitally when they met the inclusion criteria [8].

Resource Use, Perspective, and Time Horizon

This study was based on observational data and represents a cost minimization of a service evaluation. All patients discharged into the digital ward had confirmed COVID-19–related respiratory disease. The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 guidelines for economic evaluations were followed [11] (see Multimedia Appendix 1). This cost minimization analysis only evaluated NHS resource use and the outcomes of interest were acute bed-days saved, their costs offset, and the costs of the digital ward. It was considered that the digital ward would not deliver any change in patients’ long-term, health-related quality of life or health outcomes.

The perspective taken was that of the English NHS. Any savings would have been in the acute sector and additional costs in the intermediate care sector. The resource use only considered the digital ward costs and the costs of its comparator, acute hospital care.

The time horizon was slightly more than over 12 months, and no discounting of costs was conducted. All costs were in 2020-2021 pounds sterling and converted to US dollars.

Resource Use Data Sources

Comparator data for patients not on oxygen with acute LOS were sourced from patients discharged immediately prior to the introduction of the digital ward in November 2020 and a published NHS data set [12]. The difference between the duration of the LOS for patients discharged into the digital ward for oxygen weaning and those discharged routinely were acquired from UHL by LPT and NHS X (now part of the NHS Transformation Directorate). Mean clinical consultation durations and staff seniority were sourced from LPT (both correspondence: JS). Individual acute LOS, readmissions data, and the cost of a day in a respiratory ward were sourced from...
UHL (correspondence: SG). Staff unit costs were sourced from the Personal Social Services Research Unit (PSSRU) data set for 2020-2021 [13].

There were 2 principal costs associated with participants’ stay in the digital ward: the duration of patients’ digital ward LOS, which influenced the costs of the digital technology and was calculated on a per diem basis; and the number and duration of respiratory specialist–patient digital contacts. The duration of the digital ward stays and the number of contacts were sourced from the digital technology database. The method for estimating the mean duration of a clinical contact was described in the paragraph above.

Resource use data in both units and costs can be found in Multimedia Appendix 1.

**Comparison of Acute Ward LOS Versus Actual and Imputed Comparators**

There were 2 different populations discharged into the digital ward; those on oxygen at discharge and those who were not.

Those on oxygen were subjected to an analysis conducted by NHS X and LPT. They were discharged from the hospital, on average, 9.9 days earlier than similar patients who had not accessed the digital ward.

The first 65 patients not on oxygen discharged into the digital ward left acute care in 3.3 days, 2.2 days (40% relative reduction in LOS) earlier than controls who did not have the potential to access the digital ward [9]. The information on controls’ LOS (5.5 days) was sourced immediately prior to the digital ward’s introduction. This is an analysis of all patients admitted into the LLR COVID-19 digital ward prior to the end of November 2021 and prior to the availability of disease-modifying medicines that reduced acute COVID-19 illness severity. This study includes the first 65 patients.

**Data Used to Estimate LOS, Patients Not on Oxygen—Comparators**

The LOS in UHL was reported alongside the median and mean monthly LOS for COVID-19 discharges between March 20 and December 21 [12]. The linear relationship between UHL discharges and median acute LOS is illustrated in Figure 1. The red line displays the result of pulling the data on the median LOS (gray hashed line) 1 month back in time. In doing so, the $r^2$ improved from 0.31 to 0.75 in an ordinary least squares (OLS) model.

Mean LOS included a minority of patients with intensive resource use and very long LOS, skewing the LOS upward (see Figure 2 for differences between mean and median LOS for England and UHL). The estimated comparator LOS was calculated for patients who were discharged without requiring oxygen. The authors considered that the median LOS better reflected patients’ estimated LOS.

Figure 2 shows the mean and median LOS for England and for UHL [12]. The solid blue and orange lines show the mean and median English LOS. The hashed gray and yellow lines show the mean and median UHL LOS.
Figure 2. UHL and England monthly mean and median length of stay for patients with COVID-19: September 2020 to November 2021. UHL: University Hospitals Leicester.

Patients Discharged Acutely Into the Digital Ward Not on Oxygen—Comparators

The 4 comparators below were then used to populate 2 simple simulations, creating 1000 random iterations for each of the 279 patients in both uniform and normal distributions, and the mean of both was taken as the base case.

1. An OLS regression was used to establish the quality of the relationship between monthly median LOS with the number of monthly acute discharges as the independent variable ($r^2=0.31$; see Figure 1).

2. A regression similar to method 1 was used, but it used the better-fitted median data (1 month in arrears, $r^2=0.75$). This improved relationship reflected that the data observed acute discharges instead of admissions in calendar months (see Figure 1).

3. The LOS in the comparator group in November 2020 was 5.5 days. This comparator assumed that LOS in November 2020 remained constant.

4. The comparator LOS data (5.5 days, prior to the introduction of the service) was 10% (5.0 days) longer than the median LOS for November 2020. This relative margin was continued in this median monthly LOS-based OLS comparator.

The data model used the 4 comparators and was conducted in Microsoft Excel. The `RANDBETWEEN` function was used to create a uniform model with 1000 iterations between the lowest and highest values of the 4 comparators for each patient’s LOS and the mean of the 1000 iterations was taken for all 279 patients. The same process was conducted to create a normal distribution of those data using the `NORMINV` and `RAND` functions, using the mean of the 4 comparators and their SD to develop 1000 iterations, from which the mean was taken. The final mean was taken from the 2 mean parameters and provided the base case.

Patients Treated With Oxygen—Comparator

The NHS X and LPT analysis was used as the basis of the comparator for all patients treated with oxygen discharged into the digital ward.

Sensitivity Analysis

A deterministic 1-way analysis was conducted to establish the extent to which varying the parameters influenced the savings. A 2-way analysis was developed to establish the extent to which the acute LOS and savings would have to be varied to reach the net savings threshold value of 0.

Statistical Methods

The data sets were not normally distributed for both study subgroups (patients on and not on oxygen) on acute discharge. To evaluate the data, 2-tailed Wilcoxon signed-ranks tests for paired samples were used. Individual patients’ LOS was compared with their comparators. Normal approximation and ties correction were used, and $\alpha$ was set to .05. These methods were used consistently to compare paired data.

Study Size

A total of 55 pairs were required for a 90% power to establish a difference between the 5.5-day LOS (the lowest comparator) for patients not on oxygen. The scale of the difference between the patients who were oxygen weaning meant that the number of pairs required was only 4. There were 310 patients in the study: 279 (90%) patients who were not discharged on oxygen and 31 (10%) patients who were discharged home and into the digital ward for oxygen weaning.

Resource Use in the Digital Ward

Patients in the digital ward were monitored against a range of clinical criteria and classified by an algorithm into red, amber, and green ratings, which were highlighted on the LPT specialist respiratory clinicians’ dashboards. Red alerts provoked clinical contact within 24 hours throughout the patients’ stay in the digital ward; amber alerted the same, but in the first week only; and green rated patients who had not been contacted previously but were contacted in their second week in the digital ward. Patients were monitored using digital technology, which was billed on a per diem basis.

Ethical Considerations

The study was evaluated by the Institutional Ethics Review body of DeMontfort University, with a decision reference of HLS FREC Ref: 2091/22. Approval was waived for the protocol.
as it was an economic analysis of a service that eligible patients routinely accessed as part of standard care. All analyzed data were deidentified. All patients accessed usual care. No payments were made to authors or patients. Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research.

Software Used for Analyses
All analyses were conducted in Microsoft Office Excel, Excel Analysis Toolpak, and The R Foundation’s R Studio.

Results
Overview
The 2 different patient populations had differing acute LOS. For patients not on oxygen when discharged, the mean acute LOS was 4.2 (SD 2.1; median 4.2) days. For patients treated with oxygen, the mean acute LOS was 13.3 (SD 4.5; median 12.4) days. Overall, the mean acute LOS was 4.3 (SD 2.1; median 4.4) days.

The phasing of patients discharged into the digital ward reflected demand in UHL [14]. The number of discharges into the COVID-19 ward (bars in Figure 3) corresponded with key system stressors (lines in Figure 3) rising and falling. The 3 key stressors reported were the ratios of COVID-19–related absences divided by all absences, total COVID-19–related beds occupied divided by all beds occupied, and COVID-19–related mechanically ventilated beds divided by all mechanically ventilated beds occupied [14].

Figure 3. Digital ward admissions and supply stressors in University Hospitals Leicester between November 2020 and November 2021: acute beds, mechanically ventilated (MV) beds, and staff absences.

Patients Discharged Not on Oxygen—Acute LOS
Table 1 demonstrates the difference in acute LOS versus the imputed non–oxygen weaning LOS comparators.

Table 1. Comparison of estimated comparators with acute University Hospitals Leicester length of stay prior to discharge into the COVID-19 digital ward in patients not weaning on oxygen between November 2020 and November 2021.

<table>
<thead>
<tr>
<th>Wilcoxon signed rank tests with normal approximations and ties correction</th>
<th>Acute LOS(^a) prior to discharge to digital ward</th>
<th>Estimated LOS of comparator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median LOS</td>
<td>5.5 days LOS</td>
<td>Modified median LOS</td>
</tr>
<tr>
<td>(W_{\text{min}})(^b)</td>
<td>N/A(^c)</td>
<td>7192</td>
</tr>
<tr>
<td>W (95% region of acceptance)</td>
<td>N/A</td>
<td>16,886.34-22,173.66</td>
</tr>
<tr>
<td>Z-statistic ((zc=1.96))</td>
<td>N/A</td>
<td>(-9.146816)</td>
</tr>
<tr>
<td>Test (P) value</td>
<td>N/A</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mean</td>
<td>4.2</td>
<td>6.26</td>
</tr>
</tbody>
</table>

\(a\): LOS: length of stay.
\(b\): \(W_{\text{min}}\): Wilcoxon signed rank test statistic.
\(c\): N/A: not applicable.
A UK study of patients with COVID-19 not admitted to an intensive care unit estimated the acute LOS as 8.0 to 9.1 days [15]. All acute LOS comparators in the non–oxygen weaning group fell below this range, with the base case being almost 2 days beneath the lower value [15], supporting the conservatism of the model. The mean digital ward LOS was 14.2 days (SD 4.9; median 15) for patients not on oxygen.

Patients Discharged on Oxygen—Acute LOS

The mean acute LOS for patients discharged into the digital ward on oxygen was 13.3 days, and the difference was 9.9 days (42.7%) between the actual LOS and the comparator LOS ($P<.001$). The mean digital ward LOS for patients on oxygen was 24.5 (SD 12.6; median 23) days.

Digital Ward Resource Use

Digital ward resource use was driven by 2 elements: the number of clinical contacts and the LOS; the details are shown in Table 2. Multimedia Appendix 2 contains a breakdown of unit costs. How they were calculated is outlined in the Methods section.

### Table 2. Total and per-patient costs of specialist respiratory consultations and digital technology in the COVID-19 digital ward between November 2020 and November 2021.

<table>
<thead>
<tr>
<th>Specialist respiratory nurse or physiotherapist calls (n=310)</th>
<th>Contacts or days</th>
<th>Cost (US $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reds (total duration)</td>
<td>852</td>
<td>32,326</td>
</tr>
<tr>
<td>Ambers (week 1 only)</td>
<td>442</td>
<td>16,770</td>
</tr>
<tr>
<td>Greens (not previously contacted in week 1)</td>
<td>58</td>
<td>2201</td>
</tr>
<tr>
<td>Total consultations (at US $37.94 per contact)</td>
<td>1352</td>
<td>51,294</td>
</tr>
<tr>
<td>Number of digital ward days&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4711</td>
<td>17,999</td>
</tr>
<tr>
<td>Cost of clinical consultations per patient</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>165.47</td>
</tr>
<tr>
<td>Monitoring costs per patient</td>
<td>N/A</td>
<td>58.06</td>
</tr>
<tr>
<td>Total patient contact and monitoring costs per patient</td>
<td>N/A</td>
<td>223.53</td>
</tr>
</tbody>
</table>

<sup>a</sup>Costs of Clinitouch per diem=US $3.82.

<sup>b</sup>N/A: not applicable.

Estimated Resource Use and Savings

The range of estimated bed-days saved in the 279 patients not on oxygen varied from 363.8 (constant 5.5 days) to 749.9 (median OLS+10%); the comparison of digital ward patients versus all imputed controls was significant ($P<.001$). The base case was estimated to have saved 490.3 bed-days in the patients not on oxygen ($P<.001$). The estimated bed-days saved in the 31 patients on oxygen was 306.9, with which there was greater certainty ($P<.001$). The total estimated bed-days saved were between 670.7 and 1056.8 (2.2-3.4 days per patient). The base case estimate was 846 bed-days saved.

The cost of a bed day in a UHL acute respiratory ward in November 2020 was US $678.

Estimated implications on costs are summarized below:

- The total estimated gross savings related to bed-days were between US $454,404 and US $715,991 (US $1466-US $2310 per patient). The base case was US $573,490 (US $1850 per patient).
- Health care professional interventions were made to address patient symptoms. There were 1355 digital specialist respiratory consultations costing US $51,294 (US $165.47 per patient).
- The digital technology costs of the digital ward were US $17,999 (US $58.06 per patient).
- The total costs of the digital ward were US $69,293 (US $223.53 per patient).
- The total net savings were between US $385,411 and US $646,697 (US $1242-US $2086 per patient). The net savings in the base case were US $504,197 (US $1626 per patient).
- The intervention was cost saving in all scenarios. The costs of US $54,420 were between 9.7% (US $715,991) and 15.2% (US $454,404) of the estimated gross savings.

Readmissions

There were 9 hospital readmissions (9/310, 2.9% of digital ward admissions) within 30 days. All readmissions were in the non–oxygen weaning cohort. In a systematic review, 10.3% (n=265,590) of those admitted with COVID-19 respiratory infections had a 30-day readmission [16]. Of the included English studies, 3 reported on 30-day readmission rates but included patients who had accessed critical care or had a high percentage of older adult patients [17-19] and reported readmission rates of 10.2% to 17.1%.

A total of 7.1% (n=154) of patients with mild or moderate COVID-19 disease, similar to the digital ward patients not on oxygen, discharged from a Turkish tertiary center were readmitted within 30 days [20]. Another systematic review [21] found that readmissions ranged from 4.2% [22] to 19.9% [23]. Evidence suggested that there was no difference between the rates of 30-day readmissions between the first and second COVID-19 waves [24]. If it was assumed the readmission rate for a comparator was 7.1% [20], the costs of readmissions would have been US $69,050 or 99.6% (US $69,293) of the total costs of running the digital ward. Using the lowest plausible comparator of 4.2% [22] would have offset 30.8% (US $21,342)
of the costs of the digital ward. Any potential savings associated with readmissions were excluded from the results of the analysis because of the uncertainty associated with any potential comparator and to aid in the overall conservatism of this study.

**Estimated Carbon Dioxide Emissions**

The carbon footprint associated with acutely hospitalized patients has been described as the most carbon-intensive care pathway and contributed 125 kg of carbon dioxide equivalent per day [25]. The gross reduction in 2019 carbon dioxide equivalent was 341 kg per patient in the base case and totaled 105.75 metric tonnes.

**Deterministic Sensitivity Analysis**

There were 8 variables in this analysis; the differences in acute LOS for those who were on the digital ward versus comparators, the cost of an acute bed-day, the number of clinical contacts, the clinical contact duration, the cost per hour of clinicians, the digital ward LOS, and the digital technology costs.

A 1-way sensitivity analysis demonstrated that the 2 variables that changed more than the relative input parameters were the LOS of patients who were and were not oxygen weaning. There was a linear 3.18% and 2.34% increase or decrease in the savings or costs for every 1% change in the value for the patients who did or did not access oxygen, respectively. All other resource-use variables increased or decreased at the same rate as the input variable. The parameter with the greatest uncertainty was the one that had the greatest impact on potential savings.

The cost of clinical consultations (respiratory specialists, LPT) and the cost of digital technology (CliniTouch, Spirit Health) were the 2 components of the digital ward costs. When the costs of both were simultaneously increased in a 2-way sensitivity analysis in the 279 patients not on oxygen by the same percentage as the LOS was reduced, it required a 75.4% change to reach the 0-threshold value. Performing the same for patients who accessed remote oxygen weaning, the threshold value was reached by simultaneously increasing costs and reducing the LOS cost offset by 87.5%. Overall, it took a reduction in effect and increase in costs of 78.4% (savings reduced from their estimated base case value of US $573,490 to 21.6% of that US $123,874 and costs to rise from US $69,293 to US $123,618) to reach a 0-threshold value.

**Discussion**

Resource use in the digital ward was lower than all of the potential comparators. The net savings were estimated to have been between US $835,111 and US $846,697, with an estimated saving of US $1850 per patient in the base case. The digital ward costs were relatively low compared with the estimated gross savings (9.7% to 15.2%). The risk to health care systems of the digital ward not being cost saving was low. The UHL mean and median acute LOS were tracked at or below the England overall mean and median LOS between November 2020 and November 2021 (see Figure 2) [12]. The gap was largest after times of peak acute COVID-19 bed demand, which is also when the digital ward was used most to reduce pressure on UHL beds (see Figure 3). Both lend face validity to the findings. The COVID-19 digital ward intervention achieved its primary goal of increasing acute capacity. It also reduced overall NHS resource use, had a very low rate of readmissions, and reduced carbon dioxide emissions. Patients were released from acute wards earlier but clinically monitored for longer in the digital ward, which may have accounted for a plausible reduction in 30-day readmissions.

The main limitations associated with this analysis were the observational nature of the data and the associated use of imputed indirect comparators for 90% (279/310) of patients. The findings were most influenced by the parameter around which there was the greatest uncertainty. This renders a degree of uncertainty around the savings. Leicester City has a greater population of South Asian ethnicity than those of White ethnicity [26]. It is documented that people of South Asian ethnicity were at a higher risk of hospitalization from COVID-19 than their White counterparts [27]. The lack of socioeconomic status, comorbidities, and ethnicity data made the representativeness of the results difficult to interpret for wider settings.

The United Kingdom has a crisis in demand for elective care driven by an aging population [28]. It also had a shortfall in elective health care supply during the COVID-19 pandemic [1]. The supply-side constraint has been exacerbated by the lowest number of hospital beds per capita of all G7 countries [29] and social care’s inability to accommodate patients medically fit for discharge [30].

Specialist community respiratory teams exist widely across the United Kingdom. In addition, 82% and 99% of people older and younger than 55 years, respectively, owned a smartphone in the United Kingdom in 2022 [31]. The low financial and safety risk to health systems, digital scalability, widespread digital literacy, and high access to smartphones has meant there is an opportunity to be better prepared for future pandemics and enable wider digital diffusion into health care to support patients with other diseases, especially when there is continued capacity constraint. The findings were broadly consistent with other evaluations of digital wards where reductions in LOS were observed [32,33], and cost savings were found [32].

**Acknowledgments**

The authors would like to thank Irene Valero-Sanchez, respiratory consultant from University Hospitals Leicester National Health Service (NHS) Trust, for help in developing and steering the project. The authors would also like to thank Zoe Harris from NHS England’s Transformation Directorate, who helped develop the project and shared resource use analysis while at the Leicester Partnership Trust and NHS X. NHS Leicester City Clinical Commissioning Group (CCG), NHS East Leicestershire and Rutland CCG, NHS West Leicestershire CCG and Ageing Well funded the intervention.
Data Availability

All data analyzed during this study are included in this published article and its supplementary information files. An anonymized data set is included in Multimedia Appendix 3.

Conflicts of Interest

JS, NOK, and CB work for Spirit Health, the manufacturers of Clinitouch, the digital tool used in the intervention. AW and SG have no competing interests.

Multimedia Appendix 1
Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 22 guidelines table.

Multimedia Appendix 2
Resource use unit tables.

Multimedia Appendix 3
Anonymized raw data file with original and comparator data.

References


Abbreviations

CHEERS: consolidated health economic evaluation reporting standards
Original Paper

User-Friendly Chatbot to Mitigate the Psychological Stress of Older Adults During the COVID-19 Pandemic: Development and Usability Study

Ya-Hsin Chou¹; Chemin Lin²,³,⁴, MD; Shwu-Hua Lee³,⁵, MD; Yen-Fen Lee⁶, PhD; Li-Chen Cheng⁶, Prof Dr¹

¹Department of Psychiatry, Taoyuan Chang Gung Memorial Hospital, Taoyuan County, Taiwan
²Department of Psychiatry, Keelung Chang Gung Memorial Hospital, Keelung City, Taiwan
³College of Medicine, Chang Gung University, Taoyuan County, Taiwan
⁴Community Medicine Research Center, Chang Gung Memorial Hospital, Keelung, Taiwan
⁵Department of Psychiatry, Linkou Chang Gung Memorial Hospital, Taoyuan County, Taiwan
⁶Department of Information and Finance Management, National Taipei University of Technology, Taipei, Taiwan

Corresponding Author:
Li-Chen Cheng, Prof Dr
Department of Information and Finance Management
National Taipei University of Technology
1, Sec 3, Zhongxiao E Rd
Taipei, 10608
Taiwan
Phone: 886 2771 2171
Email: lijen.cheng@gmail.com

Abstract

Background: To safeguard the most vulnerable individuals during the COVID-19 pandemic, numerous governments enforced measures such as stay-at-home orders, social distancing, and self-isolation. These social restrictions had a particularly negative effect on older adults, as they are more vulnerable and experience increased loneliness, which has various adverse effects, including increasing the risk of mental health problems and mortality. Chatbots can potentially reduce loneliness and provide companionship during a pandemic. However, existing chatbots do not cater to the specific needs of older adult populations.

Objective: We aimed to develop a user-friendly chatbot tailored to the specific needs of older adults with anxiety or depressive disorders during the COVID-19 pandemic and to examine their perspectives on mental health chatbot use. The primary research objective was to investigate whether chatbots can mitigate the psychological stress of older adults during COVID-19.

Methods: Participants were older adults belonging to two age groups (≥65 years and <65 years) from a psychiatric outpatient department who had been diagnosed with depressive or anxiety disorders by certified psychiatrists according to the Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) (DSM-5) criteria. The participants were required to use mobile phones, have internet access, and possess literacy skills. The chatbot’s content includes monitoring and tracking health data and providing health information. Participants had access to the chatbot for at least 4 weeks. Self-report questionnaires for loneliness, depression, and anxiety were administered before and after chatbot use. The participants also rated their attitudes toward the chatbot.

Results: A total of 35 participants (mean age 65.21, SD 7.51 years) were enrolled in the trial, comprising 74% (n=26) female and 26% (n=9) male participants. The participants demonstrated a high utilization rate during the intervention, with over 82% engaging with the chatbot daily. Loneliness significantly improved in the older group ≥65 years. This group also responded positively to the chatbot, as evidenced by changes in University of California Los Angeles Loneliness Scale scores, suggesting that this demographic can derive benefits from chatbot interaction. Conversely, the younger group, <65 years, exhibited no significant changes in loneliness after the intervention. Both the older and younger age groups provided good scores in relation to chatbot design with respect to usability (mean scores of 6.33 and 6.05, respectively) and satisfaction (mean scores of 5.33 and 5.15, respectively), rated on a 7-point Likert scale.

Conclusions: The chatbot interface was found to be user-friendly and demonstrated promising results among participants 65 years and older who were receiving care at psychiatric outpatient clinics and experiencing relatively stable symptoms of depression and anxiety. The chatbot not only provided caring companionship but also showed the potential to alleviate loneliness during the challenging circumstances of a pandemic.
Introduction

As COVID-19 spread worldwide, many governments implemented stay-at-home, social distancing, and self-isolation measures to protect the most vulnerable people, including older adults, for whom deteriorating physical health could result in both physical and mental challenges [1,2]. As one of the most vulnerable groups to experience negative outcomes of COVID-19, older adults were instructed to self-quarantine and isolate themselves from others to mitigate the risk of infection. Consequently, they missed out on their standard interactions with others and spent substantial amounts of time in isolation. Limited contact with other people can lead to loss of social support, which is especially important for older adults [1,2]. Indeed, older adults who experienced social isolation during the pandemic have been negatively impacted by feelings of increased loneliness and reduced quality of life. Furthermore, the risks faced by individuals living alone, including older adults, are particularly pronounced with more challenging economic conditions [3,4]. Loneliness is defined as a subjective feeling manifesting as uneasiness and unhappiness linked to a lack of connection or inclusiveness with others, whereas social isolation describes the objective state of an individual’s social environment and interactional patterns [5]. Growing evidence shows that “loneliness” is closely related to physical and mental health in older adults; stronger feelings of loneliness can lead to the development of unhealthy lifestyle habits (eg, smoking and drinking) and a reduction in engagement with health-positive behaviors (eg, exercise and a nutritious diet) [6,7]. Loneliness is also associated with decreased sleep and sleep efficiency and poor subjective sleep quality in middle-aged and older adults, along with correlations to cognitive dysfunction, disability, cardiovascular disease, increased mortality, and even suicide [8-10]. Loneliness is also associated with negative emotion processing in the brain [11] and is a significant risk factor for anxiety and depression in older adults [12,13].

The prevalence of anxiety disorders and depression among older adults is high. Approximately 2% of adults older than 55 years are diagnosed with major depression, and 10%-15% of older adults have depressive symptoms [14]. Anxiety disorders are the most common psychiatric diseases, with a lifetime prevalence of up to 30%. Treatment of depressive and anxiety disorders in older adults is difficult and common challenges include poor drug efficacy, adverse drug reactions, and common drug interactions caused by polypharmacy. Coping with loneliness among older adults has potential benefits for depression and anxiety, as loneliness can affect older people’s physical and mental health, triggering psychological vulnerability, causing brain alterations beyond those associated with normal aging, and reducing intellectual or psychological flexibility caused by discrete and chronic stress [15]. Given the serious impact of social isolation and loneliness on older adults’ mental and physical health, researchers must investigate ways to mitigate these adverse health outcomes. As an assistive technology, chatbots and conversational agents may provide companionship and reduce the loneliness experienced because of the social restrictions imposed during a pandemic [7].

In recent years, there have been breakthroughs in the application of deep-learning techniques for natural language processing, allowing a variety of applications to be used in medical communication, disease detection, and treatment. For instance, chatbots have been developed to assist patients with autism and to provide social-skill training for specific fields such as mental health [16]. There is also a publicly listed application that provides cognitive behavioral therapy to improve anxiety and depression symptoms. Chatbots are also used for veterans as a screening tool for posttraumatic stress syndrome, and users of such applications have compared its use to interacting with real people, highlighting the chatbot’s ability to more freely and accurately reflect their own symptoms [17]. Additionally, chatbots used for mental illness discharge preparation can offer time-saving benefits for nurses and allow users to repeat questions without time pressure [18]. Upon reviewing previous research, it is evident that most chatbot users are relatively young, with limited availability of chatbots specifically developed for older adults in psychiatric clinics. While an overwhelming number of mental health apps are available to the public, these apps seldom focus on the needs of the older adult population. Furthermore, older adults can be apprehensive about using apps containing too many complex functions. This situation highlights the need to design and develop an older adult–user-friendly chatbot geared toward the specific interests of this population, while promoting their physical and mental health.

To address this gap, this study investigated the perspectives of geriatric psychiatric outpatient users of a mental health chatbot during the COVID-19 pandemic and sought to develop a chatbot for older adult users with anxiety and depressive disorders. This chatbot aims to offer companionship and promote self-awareness regarding emotions, sleep patterns, and physical activity to improve overall health. Therefore, our primary research objective was to examine whether chatbots can mitigate the psychological stress that older adults experienced during COVID-19.

Methods

Ethical Considerations

Ethical approval was granted by the Institutional Review Board of Chang Gung Memorial Hospital (2020019150B). The participants were recruited from the psychiatric outpatient...
department of the hospital between September 2021 and March 2022. Informed consent was obtained from each participant before study initiation. Data were deidentified by assigning a unique identifier code to each participant; only the study moderator has access to the file linking the code to the personal information, which is stored in a locked cabinet. Participants were provided with a travel compensation of ~US $15 for completing the postintervention questionnaires; no other compensation was provided for participation in the study.

Participants

Participants comprised two groups of individuals aged either 65 years and older or younger than 65 years who were diagnosed with depressive or anxiety disorders, including persistent depressive disorder, major depressive disorder, generalized anxiety disorder, panic disorder, or adjustment disorder with anxiety or depression. Certified psychiatrists made diagnoses according to Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) (DSM-5) criteria.

During the time of the study, none of the participants had experienced suicidal ideation or had an increase in medication 3 months prior to the beginning of the study. Participants included in the study were literate, active mobile phone users with internet access, who had used the chatbot regularly for at least 4 weeks. We excluded patients who were not literate, those with severe physical illness, and those diagnosed with a major neurocognitive illness (eg, schizophrenia, bipolar disorder, brain injury, substance abuse, stroke history, and dementia). During the case referral period, only one patient was excluded due to a diagnosis of bipolar disorder. All participants completed a questionnaire regarding their demographic characteristics.

Health Promotion Chatbot

Chatbot Design

The LINE Official Account Manager was used to build the “Health Promotion” chatbot. LINE was used as it is currently the most popular chatbot for older adults in Taiwan. Previous research found that some older adults have difficulty keeping pencil-and-paper sleep diaries [19]. This difficulty can be ameliorated using mobile technologies; thus, we designed a user-friendly interface for older adults using buttons in the chatbot app. The proposed chatbot is multifunctional and includes health diary data collection and hygiene education functions. Figures 1-4 provide screenshots of the Health Promotion chatbot functions translated into English.
Figure 2. Sleep tracking function of the Health Promotion chatbot.

Figure 3. Daily activity tracking function of the Health Promotion chatbot.
Monitoring and Recording Health Data

The Health Promotion chatbot sent daily messages to all participants and collected their health diary responses. The diaries tracked three key response areas, including mood, sleep, and daily activity. Figure 1 depicts the participants’ rating of their mood using five status terms: great, not bad, indifferent, sad, and angry. Sleep quality was rated as good, not bad, or sleepless (Figure 2). In addition to keeping an activity diary, the chatbot sent messages such as “Did you go out today?” or “Did you do some exercise? What kind of exercise?” (Figure 3). These messages helped guide older adult users to complete their activity information in the formatted “Message buttons.”

Providing Health Information and Advice

The main health promotion education program consisted of three topics: sleep hygiene education, food education, and exercise videos. The educational videos were provided to help the older adults practice good sleep hygiene and exercise habits. The sleep hygiene education video helped users practice “abdominal breathing” for relaxation. The exercise videos were recorded by a professional rehabilitator to guide older adults in performing different types of body movements. Overall, the app proved to be an effective tool to help the users perform training courses at home. The health promotion education programs are shown in Figure 4.

Measures

Self-Assessments

Self-assessment scales for loneliness, depression, and anxiety were administered before and after chatbot use. Using a 7-point Likert scale, we also assessed individuals’ attitudes toward the chatbot post intervention.

Loneliness was measured using the University of California Los Angeles Loneliness Scale (UCLA-LS), which has demonstrated good reliability and validity for all ethnic groups as well as for older adults [20]. Participants rated themselves on a 1-4 scale for the 20 items, with a possible total score ranging from 20 to 80. As a one-dimensional measure, higher scores indicate greater loneliness.

Depression was measured using the Geriatric Depression Scale (GDS) 15-item short form; a cutoff score of 15 indicates clinical depression [21].

Anxiety was assessed using the Hospital Anxiety and Depression Scale anxiety subscale (HADS-A), which consists of 7 items. The depression subscale, also comprising 7 items, was not used in this study owing to the duplication of targets measured by the GDS [22].

Attitudes Toward the Chatbot

Based on the measure proposed by Bickmore et al [23], the questionnaire was modified to evaluate users’ attitudes toward the chatbot. The questionnaire covered four dimensions regarding satisfaction, usability, willingness to continue use, and adherence. All participants were asked to rate a series of
attitude statements using a 7-point Likert rating scale, ranging from 1 (strongly disagree) to 7 (strongly agree), to assess their attitudes toward the chatbot’s features, such as “I was satisfied with the chatbot,” “It was easy to use the chatbot,” “I would like to continue using the chatbot,” “I will follow the chatbot’s recommendations,” and others [23-26].

Utilization Rate
The participant utilization rate was determined according to the frequency of interactions with the chatbot, specifically measured as the ratio of the number of interaction days to the total test days for each participant. This score encompasses the percentage of responses to inquiries regarding emotions, sleep quality, sport participation, and daily outings. For example, if a participant responded to emotional inquiries 15 times (with only one interaction per day counted) during a 30-day testing period, their emotional inquiry response rate would be 50%.

Statistical Analysis
We compared the effectiveness of the chatbot between the age groups of ≥65 and <65 years. In Taiwan, individuals 65 years and older can enjoy old-age benefits (such as old-age pensions and bus discounts). Therefore, the division of participants into these two age groups allowed for a comparison of the effects of the chatbot on anxiety and depression in individuals who have already reached the age of eligibility for old-age benefits and those who have not. This distinction is important as it enables researchers to explore differences in psychological well-being between the two groups considering the impact of access to benefits and associated social support networks.

All statistical analyses were performed using SPSS for Windows 20.0 (IBM Corp). The Fisher exact test, Wilcoxon signed rank test, analysis of covariance, and Spearman rank correlations were applied as appropriate.

Results
Sample Characteristics
The age of the participants ranged from 55 to 82 years, with a mean age of 65.21 (SD 7.51) years; the study sample comprised 74% (n=26) female participants and 26% (n=9) male participants. There were no significant differences in the demographic variables between the groups, except that a greater percentage of participants in the older age group (≥65 years) were living alone (Table 1).
Table 1. Demographic characteristics of the participants in each group.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Entire cohort (N=35)</th>
<th>Aged ≥65 years (n=15)</th>
<th>Aged &lt;65 years (n=20)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>65.2 (7.4)</td>
<td>72.3 (4.9)</td>
<td>59.9 (3.2)</td>
<td>N/A</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>Female</td>
<td>26 (74)</td>
<td>13 (87)</td>
<td>13 (65)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (26)</td>
<td>2 (13)</td>
<td>7 (35)</td>
<td></td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.38</td>
</tr>
<tr>
<td>Married</td>
<td>20 (57)</td>
<td>6 (40)</td>
<td>14 (70)</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>5 (14)</td>
<td>3 (20)</td>
<td>2 (10)</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>5 (14)</td>
<td>3 (20)</td>
<td>2 (10)</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>5 (14)</td>
<td>3 (20)</td>
<td>2 (10)</td>
<td></td>
</tr>
<tr>
<td>Living situation, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>Living alone</td>
<td>8 (23)</td>
<td>7 (47)</td>
<td>1 (5)</td>
<td></td>
</tr>
<tr>
<td>Accompanied by family occasionally</td>
<td>4 (11)</td>
<td>1 (7)</td>
<td>3 (15)</td>
<td></td>
</tr>
<tr>
<td>Accompanied by family mostly</td>
<td>23 (66)</td>
<td>7 (47)</td>
<td>16 (80)</td>
<td></td>
</tr>
<tr>
<td>Current employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Employed</td>
<td>5 (14)</td>
<td>0 (0)</td>
<td>5 (25)</td>
<td></td>
</tr>
<tr>
<td>Unemployed/retired</td>
<td>25 (71)</td>
<td>12 (80)</td>
<td>13 (65)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5 (14.3)</td>
<td>3 (20)</td>
<td>2 (10)</td>
<td></td>
</tr>
<tr>
<td>Systemic disease, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.73</td>
</tr>
<tr>
<td>Any systemic disease</td>
<td>19 (54)</td>
<td>9 (60)</td>
<td>10 (50)</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>5 (14)</td>
<td>1 (7)</td>
<td>4 (20)</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>13 (37)</td>
<td>8 (53)</td>
<td>5 (25)</td>
<td></td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>10 (29)</td>
<td>5 (33)</td>
<td>5 (25)</td>
<td></td>
</tr>
<tr>
<td>Psychiatric diagnosis, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>N/A^c</td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>9 (26)</td>
<td>4 (27)</td>
<td>5 (25)^b</td>
<td></td>
</tr>
<tr>
<td>General anxiety disorder</td>
<td>6 (17)</td>
<td>2 (13)</td>
<td>4 (20)</td>
<td></td>
</tr>
<tr>
<td>Other anxiety disorder</td>
<td>20 (57)</td>
<td>9 (60)</td>
<td>11 (55)</td>
<td></td>
</tr>
<tr>
<td>Persistent depressive disorder</td>
<td>3 (9)</td>
<td>0 (0)</td>
<td>3 (15)</td>
<td></td>
</tr>
</tbody>
</table>

^aN/A: not applicable.
^bOne participant had comorbid major depressive and anxiety disorders.
^cStatistical analysis was not applicable in this case since a single participant may have multiple psychiatric diagnoses.

Attitude Toward the Chatbot

In the postintervention assessment, we collected the participants’ attitudes toward the chatbot using a 7-point Likert scale. The mean scores for satisfaction, usability, willingness to continue using, and adherence were 5.33, 6.33, 4.87, and 5.07, respectively, in the ≥65 years age group; the corresponding mean scores in the <65 years age group were 5.15, 6.05, 4.7, and 4.95, respectively. There were no group differences in satisfaction, usability, willingness to continue using, and adherence scores.

Use and Response Rates

The use rate was determined by the frequency of a patient’s interaction with the chatbot, specifically measured as the ratio of interacting days to the total testing days for each participant. This included the rate of response to the inquiries for mood, sleep quality, and participation, as well as the “I went out today” inquiry.

There was no significant difference in the response rate between the two age groups. However, the response rate to the sport participation inquiry was less than 50%. Furthermore, during the 1-month intervention, the participants had a high use rate, with over 82% of the participants reporting being active daily (Table 2).
Table 2. Use and response rates in each group.

<table>
<thead>
<tr>
<th>Use and response rates</th>
<th>Aged ≥65 years (n=15)</th>
<th>Aged &lt;65 years (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chatbot use, %</td>
<td>82.55</td>
<td>83.38</td>
</tr>
<tr>
<td>Emotional inquiry response, %</td>
<td>84.19</td>
<td>80.88</td>
</tr>
<tr>
<td>Sleep quality response, %</td>
<td>82.55</td>
<td>83.37</td>
</tr>
<tr>
<td>Sport participation response, %</td>
<td>46.95</td>
<td>48.16</td>
</tr>
<tr>
<td>“I went out today” response, %</td>
<td>78.25</td>
<td>66.33</td>
</tr>
</tbody>
</table>

aCalculated as the ratio of interacting days to the total testing days for each participant.

Self-Assessment Scales

The improvement in the UCLA-LS score post intervention was not significant in the <65 age group ($P=.98$) but was significant in the ≥65 age group ($P=.006$). HADS-A and GDS scores showed no significant improvements in either group (Table 3). Figures 5-7 show the changed trends in the scales before and after the intervention. The slopes flattened after the intervention.

Table 3. Comparison of outcome measures at baseline and post intervention.

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>Entire cohort (N=35), mean (SD)</th>
<th>Aged ≥65 years (n=15), mean (SD)</th>
<th>Aged &lt;65 years (n=20), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UCLA-LS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>40.4 (10.1)</td>
<td>38.9 (7.7)</td>
<td>41.5 (11.7)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>38.1 (12)</td>
<td>34.9 (9.2)</td>
<td>40.4 (13.5)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.12</td>
<td>.006</td>
<td>.98</td>
</tr>
<tr>
<td><strong>GDS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>4.9 (3.4)</td>
<td>3.6 (2.8)</td>
<td>5.8 (3.6)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>4.1 (4.1)</td>
<td>2.5 (1.8)</td>
<td>5.4 (4.9)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.07</td>
<td>.13</td>
<td>.30</td>
</tr>
<tr>
<td><strong>HADS-A</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>6.3 (4.2)</td>
<td>4.3 (2.7)</td>
<td>7.8 (4.6)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>6.0 (4.5)</td>
<td>3.7 (2.1)</td>
<td>7.8 (5.0)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.61</td>
<td>.32</td>
<td>.92</td>
</tr>
</tbody>
</table>

aUCLA-LS: University of California Los Angeles Loneliness Scale.
bGDS: Geriatric Depression Scale.
cHADS-A: Hospital Anxiety and Depression Scale anxiety subscale.
**Figure 5.** UCLA-LS scores post intervention versus baseline with LOESS fit and 95% confidence bands. UCLA-LS: University of California Los Angeles Loneliness Scale.

**Figure 6.** GDS-15 scores post intervention versus baseline with LOESS fit and 95% confidence bands. GDS-15: Geriatric Depression Scale 15-item short form.
Discussion

Principal Findings

Usability and User Satisfaction of the Health Promotion Chatbot Among Older Adults During the COVID-19 Pandemic

This study took place during the COVID-19 pandemic, which greatly impacted people's daily lives and led to notable reductions in interpersonal interactions and outdoor activities [27]. The pandemic also substantially affected the economy, impacting employment and raising anxiety for many who experienced both the disease itself and the protocols set in place to counter the spread of the virus [28,29].

The results of this study indicate that both age groups, ≥65 and <65 years, were equally satisfied with the chatbot. Its design was rated highly for usability and user satisfaction by both groups. The ≥65 age group provided a mean satisfaction rating of 5.33 and a mean usability rating of 6.33, whereas the corresponding scores for the <65 age group were slightly lower at 5.15 and 6.05, respectively. These high satisfaction ratings reflect the chatbot’s effectiveness in catering to a wide range of users and their positive reception of the user-friendly interface.

The use rate among the older adults was remarkably positive. Previous studies have suggested that systems that are not user-friendly for older adults are often met with resistance and ineffectiveness [30]. However, this system, which was specifically designed for older adults, features a simple and user-friendly interface to target the barriers that older adult users might face in terms of digital literacy and technical assistance [31,32]. Our Health Promotion chatbot was designed to have concise functions, aiming to provide companionship and support even to those who are less familiar with technological products. Additionally, the use and response rates showed that age did not affect the willingness to use the chatbot, further suggesting that the content and interface of the chatbot are well-received and accessible for older adult users.

Therefore, the findings of this investigation on chatbot applicability in geriatric psychiatric outpatient clinics indicate that the chatbot interface is user-friendly and meets the desired criteria.

Chatbots Mitigate Loneliness in Older Adults

Some studies have explored the potential of assistive technology in addressing social isolation and loneliness among older adults. For example, one prospective cohort study reported that humanoid robots improve patient outcomes by providing companionship and entertainment in acute hospital settings and may be an acceptable outlet for socialization among older adults [33,34]. In a systematic review, Chen and Schulz [7] found that information and communication technology interventions can effectively reduce social isolation among older adults and potentially improve their quality of life by enhancing social support, promoting social engagement, and reducing feelings of loneliness and depression. However, a cross-sectional study also showed that problematic social media use was positively
associated with perceived social isolation among older adults [35].

Four possible mechanisms by which information technology alleviates social isolation and loneliness have been proposed in the literature [7,36]: connection to the outside world, gaining social support, engaging in interesting activities, and boosting self-confidence. Boosting self-confidence was interpreted based on reports of older adults feeling younger and more equipped with new skills when taught how to use communication technologies. In our study, the chatbot greeted older adults daily and encouraged them to stay active. Some participants reported benefitting from the chatbot’s exercise videos; engaging with the chatbot and performing daily exercises made them feel rejuvenated and more purposeful.

Overall, our results suggest that the chatbot helped to reduce feelings of loneliness; the average UCLA-LS score of the total cohort at baseline was 40.4, which decreased to 38.1 post intervention. For the ≥65 age group, the average baseline score was 38.9, which decreased to 34.9 post intervention, representing a statistically significant difference (P=.006). However, the <65 age group had an average baseline score of 41.5, which remained stable post intervention with a nonsignificant difference (P=.98). Therefore, although the participants had decreased UCLA-LS scores post intervention on the whole, only the older participants (≥65 years) showed a significant improvement in loneliness.

Generally, the ≥65 age group responded well to the chatbot, as reflected in the UCLA-LS score changes, indicating that this age group can benefit from chatbots. Some potential reasons to explain the lack of an effect on loneliness for the <65 age group may be related to this group experiencing elevated stress in their daily lives [4,37] due to financial burdens and concerns about the future. Evidently, the effectiveness of chatbots in alleviating such stress is limited. Moreover, the <65 age group likely also had more opportunities for interpersonal interaction, even during the pandemic. Members of this group tend to not live alone and to have more avenues (eg, different types of social media) to access information. Consequently, the influence of chatbot interventions on their lives was comparatively less pronounced.

Our study demonstrates the positive impact of chatbot interventions on older adults enrolled in psychiatric outpatient departments and with relatively stable symptoms during the pandemic. Previous studies have mentioned the importance of using technology to reduce the impact of loneliness on older adults, particularly in the context of pandemic-related restrictions [2,38,39]. Although this study was conducted during a period when the pandemic was more severe, it offers valuable insights into older adult care. To enhance the generalizability and applicability of future trials, we recommend expanding the participant pool to include older adults outside of clinical settings.

**Limitations**

This pilot study has some limitations. When assessing loneliness and depression, it is important to include in-depth interviews to understand users’ thoughts. However, due to the pandemic, interviews could not be conducted during the study period. We therefore aim to further enhance this aspect by including interviews in future research. In addition, given that the study was conducted during the COVID-19 pandemic, it was challenging to enroll participants, as many patients were unwilling to visit the hospital, resulting in a limited sample size. The improvement in loneliness may be due in part to the placebo effect. Additionally, the study sample was derived solely from the psychiatric outpatient department of a single medical center, potentially introducing bias in the referral process and limiting the generalizability of the findings to a broader population. Finally, the follow-up period was relatively short (within half a year); thus, determining the long-term effect of the intervention requires further research.

**Conclusions**

The chatbot interface developed in this study was simple and easy to use. The results demonstrated that older adults in psychiatric outpatient clinics with relatively stable depressive and anxiety symptoms could benefit from the chatbot’s care, companionship, and potential reduction in loneliness, particularly during a pandemic. Simultaneously, the accessibility and user satisfaction of chatbots, which are crucial for promoting the health of older adults through using such apps, enabled the participants to easily access health information and advice, thereby enhancing their health awareness and quality of life.

Future research can explore how more personalized mental health support can be provided by expanding on the functionality of chatbots to effectively manage depression and anxiety among older adults. Simultaneously, there should be in-depth exploration of how chatbots can be integrated with traditional health care services to establish a more comprehensive medical support system. Qualitative investigations through individual or group interviews can add further insights into their use and adaptation.

**Acknowledgments**

The authors would like to thank Tzu-Tao Huang, Chen-Ju Lu, and Tzu-Yun Tseng for their assistance in developing the chatbot. The authors are also grateful to Wei-Yang Chang and the Clinical Trial Center of Chang Gung Memorial Hospital for assistance with the statistical analysis. This study was supported by the Ministry of Science and Technology of Taiwan under grants MOST 109-2511-H-216-001-MY3 (to LCC), 109-2314-B-182A-107-MY3 (to CL), 109-2314-B-182A-084-MY2, and 110-2314-B-182A-042 (to SHA). The study was supported in part by the National Taipei University of Technology and Chang Gung Memorial Hospital Joint Research Program (grant NTUT-CGMH-110-07), Chang Gung Memorial Hospital Research Program (grant CORPG3L0131), and Chang Gung Memorial Hospital Medical Research Grant (CRRPG2K0021, 22, 23).
Data Availability
All data generated or analyzed during this study are included in the published article.

Conflicts of Interest
None declared.

References
18. Chou et al. JMIR FORMATIVE RESEARCH 2024 | vol. 8 | e49462 | p.3266https://formative.jmir.org/2024/1/e49462


Abbreviations

- **DSM-5**: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
- **GDS**: Geriatric Depression Scale
- **HADS-A**: Hospital Anxiety and Depression Scale anxiety subscale
- **UCLA-LS**: University of California Los Angeles Loneliness Scale
Original Paper

Novel Web-Based Drop-In Mindfulness Sessions (Pause-4-Providers) to Enhance Well-Being Among Health Care Workers During the COVID-19 Pandemic: Descriptive and Qualitative Study

Mary Elliott\textsuperscript{1,2}, MD; Camille Khallouf\textsuperscript{3}, BA; Jennifer Hirsch\textsuperscript{2,3}, MD; Diane de Camps Meschino\textsuperscript{2,4}, MD; Orit Zamir\textsuperscript{2,3,5,6}, MD; Paula Ravitz\textsuperscript{2,3,7}, MD

\textsuperscript{1}Department of Supportive Care, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
\textsuperscript{2}Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
\textsuperscript{3}Department of Psychiatry, Sinai Health System, Toronto, ON, Canada
\textsuperscript{4}Women's College Hospital, Toronto, ON, Canada
\textsuperscript{5}Toronto Academic Pain Medicine Institute, Women's College Hospital, Toronto, ON, Canada
\textsuperscript{6}Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada
\textsuperscript{7}Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada

Corresponding Author:
Mary Elliott, MD
Department of Supportive Care
Princess Margaret Cancer Centre
University Health Network
610 University Ave
Toronto, ON, M5G 2C1
Canada
Phone: 1 416 946 2000
Email: mary.elliott@uhn.ca

Abstract

Background: The COVID-19 pandemic exerted extraordinary pressure on health care workers (HCWs), imperiling their well-being and mental health. In response to the urgent demand to provide barrier-free support for the health care workforce, Pause-4-Providers implemented 30-minute live web-based drop-in mindfulness sessions for HCWs.

Objective: This study aims to evaluate the use, feasibility, satisfaction, and acceptability of a novel mindfulness program aimed at enhancing the well-being of HCWs during the COVID-19 pandemic.

Methods: Accrual for the study continued throughout the first 3 pandemic waves, and attendees of \( \geq 1 \) session were invited to participate. The evaluation framework included descriptive characteristics, including participant demographics, resilience at work, and single-item burnout scores; feedback questionnaires on reasons attended, benefits, and satisfaction; qualitative interviews to further understand participant experience, satisfaction, benefits, enablers, and barriers; and the number of participants in each session summarized according to the pandemic wave.

Results: We collected descriptive statistics from 50 consenting HCWs. Approximately half of the participants (24/50, 48\%) attended 1 session. The study participants were predominantly female individuals (40/50, 80\%) and comprised physicians (17/50, 34\%), nurses (9/50, 18\%), and other HCWs (24/50, 48\%), who were largely from Ontario (41/50, 82\%). Of 50 attendees, 26 (52\%) endorsed feeling burned out. The highest attendance was in May 2020 and January 2021, corresponding to the first and second pandemic waves. The participants endorsed high levels of satisfaction (43/47, 92\%). The most cited reasons for attending the program were to relax (38/48, 79\%), manage stress or anxiety (36/48, 75\%), wish for loving kindness or self-compassion (30/48, 64\%), learn mindfulness (30/48, 64\%), and seek help with emotional reactivity (25/48, 53\%). Qualitative interviews with 15 out of 50 (30\%) participants identified positive personal and professional impacts. Personal impacts revealed that participation helped HCWs to relax, manage stress, care for themselves, sleep better, reduce isolation, and feel recognized. Professional impacts included having a toolbox of mindfulness techniques, using mindfulness moments, and being calmer at work. Some participants
noted that they shared techniques with their colleagues. The reported barriers included participants’ needing time to prioritize themselves, fatigue, forgetting to apply skills on the job, and finding a private place to participate.

**Conclusions:** The Pause-4-Providers participants reported that the web-based groups were accessible; appreciated the format, content, and faculty; and had high levels of satisfaction with the program. Both novel format (eg, drop-in, live, web-based, anonymous, brief, and shared activity with other HCWs) and content (eg, themed mindfulness practices including micropractices, with workplace applications) were enablers to participation. This study of HCW support sessions was limited by the low number of consenting participants and the rolling enrollment project design; however, the findings suggest that a drop-in web-based mindfulness program has the potential to support the well-being of HCWs.

(JMIR Form Res 2024;8:e43875) doi:10.2196/43875

**KEYWORDS**
COVID-19; pandemic; health care worker; resilience; mental health; burnout; well-being; mindfulness meditation; web-based group; drop-in; mindfulness; health care staff; meditation; worker; job; occupational health

**Introduction**

**Background**

The COVID-19 pandemic has jeopardized the mental health and well-being of health care workers (HCWs). Rapidly shifting public health guidelines have been implemented in response to the emergent scientific findings. The uncertain landscape of this pandemic has produced a sense of uncertainty, fear, and psychological distress in HCWs [1]. This was compounded by the realities HCWs faced, including inequitable access to rationed resources such as personal protective equipment and needed acute care beds [2]. Changes in procedural guidelines and service priorities, ramping down of nonessential services, and shifting standards of care have contributed to moral distress [3-5]. Risk factors for HCWs’ burnout included heavy workload, shifting roles, need for additional training, redeployment, and hybrid models of web-based care with feeling disconnected from their usual support and community of practice [6,7]. Greater COVID-19 exposure risks forced many HCWs into isolation owing to fears of infecting family members [1,8,9]. Early in the pandemic, significant levels of fatigue, insomnia, anxiety, depression, and psychological distress among HCWs were reported [8,10-15]. Informed by the literature on previous pandemics, it was anticipated that as the pandemic endured, moral distress, posttraumatic stress, and burnout would become problematic [16]. All combined, the COVID-19 pandemic amplified the challenges faced by HCWs within an already-strained workforce and was thus critically important to address [1,3,9,17-22].

Before the pandemic, evidence-based mindfulness interventions for HCWs were found to enhance stress management, awareness, emotion regulation, connection, self-compassion, empathy, and resilience [23-27]. Research demonstrates that mindfulness training for HCWs can improve well-being by promoting self-care and cultivating self-compassion [28]. The term “mindfulness” is defined as the “awareness that arises through paying attention, on purpose, in the present moment, and nonjudgmentally” and implies the cultivating of compassion [29]. Most interventions for HCWs evolved from manualized formats of mindfulness-based stress reduction, mindfulness-based cognitive therapy, and mindful self-compassion, which require a large time commitment [29-31].

“Pause-4-Providers” was created during the first wave of the COVID-19 pandemic to support HCWs and be accessible using a web-based drop-in platform. This extended the program’s reach to underserved areas and those in quarantine, with adherence to physical distancing requirements [17,21,32-34]. To establish a safe space for weary HCWs, no registration was required, and anonymity was allowed as we did not require participants to turn on their cameras or introduce or disclose information about themselves [35]. Unique to this program was the flexible curriculum, making it adaptable to the varied and changing needs of HCWs with time [36]. We implemented Pause-4-Providers drop-in sessions at the beginning of the COVID-19 pandemic to provide momentary refuge and respite from the intensity associated with working in health care during the pandemic.

**Objective**

The objective of this study was to evaluate the implementation of Pause-4-Providers, a 30-minute evening web-based drop-in mindfulness program for HCWs during the COVID-19 pandemic. We designed a descriptive and qualitative study to examine the feasibility, satisfaction, and experiences of the participants. The research questions for the descriptive study were as follows:

1. What is the feasibility and acceptability of the web-based drop-in mindfulness program?
2. What is the level of burnout and resilience among the consenting participants?

Research questions for the qualitative study included the following: What were the experiences, enablers, and barriers to participating in the Pause-4-Providers program?

**Methods**

**Program Development and Format**

Within the first weeks of the COVID-19 pandemic in Canada, our team of university-affiliated psychiatrists (ME, JH, DdCM, and OZ) with expertise in mindfulness met and collaboratively decided to offer free, 30-minute live web-based drop-in mindfulness sessions for HCWs with the objective of supporting their well-being. We defined HCWs as anyone working in a health care or social care setting, including those providing direct patient-facing services and those performing indirect
roles. Along with health care providers and trainees, the sessions were open to those providing administrative facilities, those working in support services, and research staff. Our sessions used a secure, privacy-compliant Zoom (Zoom Video Communication, Inc) platform multiple times per week, starting on March 24, 2020. We e-blasted Pause-4-Providers program announcements to >400 health care institutions, hospitals, long-term care homes, shelters, university departments, and associations (eg, Ontario Medical Association, Royal College of Physicians and Surgeons of Canada, Canadian Mental Health Association, and Registered Practical Nurses Association of Ontario) for distribution of session links to their HCWs.

Our group of faculty facilitators had extensive mindfulness training and clinical experience in leading manualized mindfulness group interventions in health care settings for HCWs and patients in an oncology health setting (ME) [21,35,37], patients with pain disorders (OZ), women with mental health concerns (DdCM and JH), and hospital-based health leaders and policy makers (DdCM). As mental health professionals working in hospital settings, we were also attuned to the psychological impacts the unfolding public health crisis had on HCWs. Notably, because we used the term “facilitators” when referring to the faculty who led the web-based drop-in sessions, we chose to use the term “enablers” when referring to factors in the qualitative analysis that facilitated or promoted participation. The themed agenda for each drop-in session was based on mindfulness, compassion, or resiliency skills to fit the emerging needs of HCWs, distressing local and global events and the time of the year or holidays. An evolving curriculum allowed for flexibility and responsiveness to the changing needs of HCWs [36]. We canvassed attendees using a situational needs assessment with routine quick check-ins at the start of each session to tailor the agenda accordingly. Given the time pressures and competing demands placed on HCWs, the sessions were intentionally designed to be of short duration. Each 30-minute session followed a semistructured agenda as presented in Figure 1 and Multimedia Appendix 1 [27,28,38,39]. The session themes evolved to reflect the immediate needs of HCWs, with a curricular structure of the sessions that was maintained. This structure included 2 short practices, including a micropractice and a closing. Maintaining a consistent format, the sessions introduced different practices and themes that were aligned with the objectives and principles of mindfulness and self-compassion. In small doses, with time, during the prolonged pandemic, faculty facilitators introduced a wide variety of mindfulness practices based on HCW attendees’ needs and levels of emotions (eg, hyperarousal vs emotional numbing). Drawing from positive psychology and resilience factors [40-45], the faculty facilitators used mindfulness techniques to promote positive emotions, compassion, and coping in the midst of challenges. Sessions included at least 1 guided micropractice (ie, a practice that was distilled into ≤1 minute meant for integration by participants throughout their workday) to emphasize how participants could apply aspects of the practices for self-care moments in the workplace [35,44,46,47]. Many sessions concluded with poems, followed by invited reflections and feedback. To support the program, we also created a website for Pause-4-Providers, where HCWs could find the schedule and session links, and access 4 prerecorded mindfulness practices for use at any time and participate in the research study.

**Figure 1.** Sample agenda for Pause-4-Providers sessions.

<table>
<thead>
<tr>
<th>Suggested agenda</th>
<th>Sample content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session theme</strong></td>
<td><strong>Grounding</strong></td>
</tr>
<tr>
<td>Welcome, brief check-in and assess needs</td>
<td>Description of practice</td>
</tr>
<tr>
<td>Introduce the theme</td>
<td>Use of breath to ground and restore</td>
</tr>
<tr>
<td>Short practice 1</td>
<td>3-Minute Breathing Space</td>
</tr>
<tr>
<td>Short practice 2</td>
<td>Mindfulness of Breath</td>
</tr>
<tr>
<td>Micropractice (≤1 minute)</td>
<td>3-5 breaths</td>
</tr>
<tr>
<td>Reflection prompts (examples)</td>
<td>What did you notice?</td>
</tr>
<tr>
<td>Encouraging use of practices on-the-go at work</td>
<td>Walk Slowly by Dana Fuchs</td>
</tr>
</tbody>
</table>

**Ethical Considerations**

This study was approved by the Sinai Health Research Ethics Board (20-0158-E). On the project website, participants were given a complete description of the study, and informed consent was obtained. The research assistant emailed the respondents who consented to be contacted for participation in a semistructured interview and obtained consent before the start of the interview. Participation in the study was voluntary and anonymous. The interviews were transcribed and deidentified to protect the identity of the study participants. The participants did not receive any compensation.

**Research Methods**

This study aimed to evaluate the feasibility and acceptability of a web-based drop-in mindfulness program. We assessed feasibility quantitatively with attendance and qualitatively by exploring participants’ experiences of barriers and facilitators or enablers of attending such a program. We assessed acceptability using a brief satisfaction questionnaire and...
semistructured interviews with a subset of participants to explore their experiences of the program sessions. The burnout and resilience measures characterized the sample to learn about the potential suitability or appropriateness of the Pause-4-Providers program [48].

The study was open to all HCWs who attended ≥1 Pause-4-Providers session. Faculty facilitators informed HCW attendees in the sessions that they could visit the project website if they wished to participate or learn more about the research study. On the website, a banner informed HCWs of the research study with a link to a consent form and the confidential questionnaire package. Study participants could choose to participate in the descriptive portion of the study or both the descriptive and qualitative studies. The descriptive measures were collected once at study enrollment and included a questionnaire package consisting of demographics (eg, profession, age, sex, province or country of origin, and primary care setting), burnout and resilience measures, and a self-report feedback questionnaire focused on the participants’ reasons for attending, satisfaction, and utility of the sessions. At the end of the questionnaire, respondents could consent to be contacted for participation in a semistructured interview about their experiences of the program. There were no exclusion criteria for this study.

Descriptive Measures

The Single-Item Burnout Measure [49] instructed respondents to define burnout for themselves with the following question: “Overall, based on your definition of burnout, how would you rate your level of burnout?” Participants scored their responses on a 5-category ordinal scale. The self-defined burnout measure has been useful as an abbreviated burnout assessment tool in medical professionals [50,51]. The self-defined burnout item correlates with the emotional exhaustion subscale of the Maslach Burnout Inventory (r=0.64; P<0.001), defining burnout as a score of ≥3, with a correlation of 0.79, sensitivity of 83.2%, specificity of 87.4%, and an area under the curve of 0.93 (SE 0.004) [49].

The Resilience at Work Scale consists of 25 items measuring 7 aspects of workplace resilience, focusing on everyday behaviors. Participants rated their responses on a Likert scale, ranging from 0 (strongly disagree) to 6 (strongly agree) [52]. The scale is a validated tool for measuring and developing resilience among HCWs, with 67% of variance and an acceptable confirmatory factor analysis model fit (goodness of fit 0.97, Tucker-Lewis Index 0.98).

The Participant Self-Report Feedback Questionnaire was developed by us and asked about the number of sessions attended, reasons and motivation to attend (eg, to relax, improve sleep, be in a community, and learn mindfulness), benefits of attending (eg, increased awareness of emotional states, enhanced coping skills, and improved capacity to work with your team on a day-to-day basis), and overall satisfaction with the web-based mindfulness sessions. There were several open-text questions where respondents could describe how Pause-4-Providers influenced their coping (eg, positive or negative), describe the impact the pandemic had on their role and job, and provide feedback on the Pause-4-Providers program as presented in Multimedia Appendix 2.

Qualitative Methodology

Overview

To explore the participants’ experiences and views of the Pause-4-Providers program, we used a qualitative research design, which does not test or generate theory but instead gathers explicit insight into clinical or behavioral interventions [53,54]. Qualitative methods are able to gather detailed insights into what does and does not work well to reveal the impacts, enablers, and barriers of program participation [55]. This is widely used in health service research to evaluate the determinants and impacts of interventions. The study complied with the standards for reporting qualitative research in accordance with COREQ (Consolidated Criteria for Reporting Qualitative Research) [54]. All participants consented before the interviews. There was no prior relationship between the qualitative researcher (Anna R Gagliardi) and the participants.

Sampling and Recruitment

At the end of the web-based questionnaire, respondents were asked if they were willing to be contacted for subsequent participation in an interview. Using a convenience sampling technique, all interested participants were emailed and invited to be interviewed by the research assistant (CK). The study and recruitment continued until June 2021, coinciding with the first 3 waves of the COVID-19 pandemic as defined by Public Health Ontario and presented in Multimedia Appendix 3. We sampled across 3 time frames corresponding to waves 1, 2, and 3 of the COVID-19 pandemic in Canada, with equal numbers of participants in each wave. We sampled concurrently with data collection and qualitative analysis and ceased when thematic saturation was achieved.

Data Collection

The audio-recorded (17-38 min), telephone-based, transcribed interviews conducted by the research assistant (CK) included questions about expectations, impacts, enablers, barriers, and recommendations to improve the program (Multimedia Appendix 4). Participants were asked about their motivation to attend the program, their experience of the program sessions, associated benefits or challenges, and feedback or suggestions for the future. Interview transcripts were independently analyzed by Anna R Gagliardi (a PhD-trained consultant with expertise in qualitative research), following which the findings were reviewed by all members of the research team.

Data Analysis

Descriptive Analysis

The total weekly attendees at drop-in sessions were tallied from April 20, 2020, to June 21, 2021, as shown in Multimedia Appendix 3; a single participant would count multiple times in these tallies if multiple sessions were attended. These totals were plotted against time, with the ends of the first 3 pandemic waves on August 14, 2020; April 1, 2021; and July 6, 2021, respectively. The remaining analyses pertained only to survey respondents. Counts and percentages were used to summarize their demographic characteristics. Responses to questions on burnout, motivation, and experience were summarized, and the motivation and experience scales were collapsed into 3 levels
(ie, very much [significantly or definitely], somewhat, and not at all or a bit). The 7 Resilience at Work subscale scores and the total score were each calculated according to the scoring guide, a 2-tailed $t$ test was used to compare the mean overall score to that from a study on the psychological impact of COVID-19 on >500 hospital workers by Maunder et al [56] (email, August 13, 2021), and a 95% CI for the mean difference in overall scores was calculated. The relationship between the number of sessions attended (1 vs $\geq 2$) and respondents’ feedback on the benefits of attending was assessed by cross-tabulating the 2 variables and calculating the $P$ value using a Fisher exact test. Participants with a missing value for a variable were omitted from the analyses of that variable.

Qualitative Analysis

We used content analysis to identify themes for inductive coding in the qualitative interviews. This was done through constant comparison to develop a preliminary codebook of themes and exemplar quotes (first-level coding) [53]. We shared themes for discussion with all members of the research team to verify them. For subsequent interview transcripts, the qualitative expert researcher (Anna R Gagliardi) performed coding to expand or merge themes (second-level coding), tabulated participant characteristics and data (eg, themes and quotes), and examined similarities or differences by participant sampling characteristics. The research team reviewed and discussed the findings and further grouped them into categories related to motivation and facilitators, which we refer to as enablers, barriers, and impacts associated with participation. The qualitative findings were triangulated using descriptive findings.

Results

Descriptive Results

Between April 20, 2020, and June 27, 2021, there were 1333 attendance entries at the drop-in sessions. Consent was obtained from 88 respondents, most of whom (66/88, 75%) were enrolled during the second wave of the pandemic. Of 88 respondents, we excluded 38 (43%) from the analysis because they did not submit the questionnaire, were not HCWs, or completed the questionnaire a second time to arrive at a total sample of 50 (57%) participants, as presented in Figure 2. At the time of survey completion, 50% (25/50) had attended 1 session, 24% (12/50) 2 to 5 sessions, 8% (4/50) 6 to 10 sessions, and 16% (8/50) $\geq 11$ sessions. If the participants indicated at the start of the questionnaire that they were completing it a second time, they were automatically taken to the end of the survey and only their first set of responses was included in the analysis.

Of 88 participants, the demographics were similar between the 50 (57%) included and 38 (43%) excluded participants. There were 5 sections in the survey on motivation, experience, satisfaction, burnout, and resilience. Overall, 90% (45/50) of the participants filled out all 5 sections, 10% (5/50) did not report on their prior experiences with mindfulness, and 6% (3/50) did not report on satisfaction. Of the 50 participants who submitted the survey, not all items or sections of the survey were completed. We included participants in the analysis if they provided data on $\geq 1$ section of the questions we examined. Of the 50 HCWs, 24 (48%) had participated in $\geq 1$ session; 17 (34%) were physicians, 9 (18%) were nurses, and the remaining 24 (48%) were other HCWs (eg, administrative or education assistant, occupational therapist, pharmacist, physiotherapist, psychotherapist, social worker, and others). Most respondents (40/50, 80%) were female individuals, 43% (21/50) were aged 20 to 39 years, 45% (22/50) 40 to 59 years, and 12% (6/50) $\geq 60$ years, with participation primarily from the province of Ontario (41/50, 82%) and other Canadian provinces (eg, Alberta, British Columbia, New Brunswick, Quebec, and Saskatchewan).

Half of the HCWs (26/50, 52%) endorsed feeling burned out (score $\geq 3$). Regarding Resilience at Work scores, participants scored highly (mean score >70) on the subscales of “finding your calling” (mean score 78, SD 17), “building networks” (mean score 78, SD 17), “living authentically” (mean score 76, SD 12), and “interacting cooperatively” (mean score 75, SD 15). Participants also scored moderately (mean score 50-69) on the workplace resilience subscale items of “staying healthy” (mean score 73, SD 22) and “managing stress” (mean score 65,
SD 19) and reported difficulty with the subscale item “maintaining perspective” (mean score 48, SD 17). The overall mean resilience score was 69.4 (SD 12.6) [56].

In the self-report feedback questionnaire, participants’ most frequently reported motivations to attend the program included relaxing, managing stress or anxiety, wishing for loving kindness or self-compassion, and learning mindfulness, as presented in Table 1.

Table 1. Self-reported motivation for attending the Pause-4-Providers sessions (N=48).

<table>
<thead>
<tr>
<th>Motivation for attending the program</th>
<th>Very much or definitely motivated, n (%)</th>
<th>Moderately motivated, n (%)</th>
<th>Not at all or a bit motivated, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>To relax</td>
<td>38 (79)</td>
<td>6 (13)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>To manage stress or anxiety</td>
<td>36 (75)</td>
<td>9 (19)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Wish for loving kindness or self-compassion</td>
<td>30 (64)</td>
<td>9 (19)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>To learn mindfulness</td>
<td>30 (64)</td>
<td>11 (23)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>To seek help with emotional reactivity</td>
<td>25 (53)</td>
<td>9 (19)</td>
<td>13 (28)</td>
</tr>
<tr>
<td>To improve sleep</td>
<td>22 (47)</td>
<td>10 (21)</td>
<td>15 (32)</td>
</tr>
<tr>
<td>To gain energy</td>
<td>21 (45)</td>
<td>7 (15)</td>
<td>19 (40)</td>
</tr>
<tr>
<td>To learn tips to cope with work</td>
<td>20 (42)</td>
<td>12 (25)</td>
<td>16 (33)</td>
</tr>
<tr>
<td>To cope at home</td>
<td>19 (40)</td>
<td>12 (26)</td>
<td>16 (34)</td>
</tr>
<tr>
<td>To be in a community</td>
<td>14 (30)</td>
<td>12 (26)</td>
<td>21 (45)</td>
</tr>
<tr>
<td>To feel less isolated</td>
<td>10 (21)</td>
<td>9 (19)</td>
<td>28 (60)</td>
</tr>
</tbody>
</table>

Table 2. Self-reported benefits of attending the Pause-4-Providers sessions (N=45).

<table>
<thead>
<tr>
<th>Benefits of attending the program</th>
<th>Very much or significantly benefited, n (%)</th>
<th>Somewhat benefited, n (%)</th>
<th>Not at all or a bit benefited, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased self-awareness of emotional states</td>
<td>25 (56)</td>
<td>12 (27)</td>
<td>8 (18)</td>
</tr>
<tr>
<td>Reduced stress level</td>
<td>24 (53)</td>
<td>12 (27)</td>
<td>9 (20)</td>
</tr>
<tr>
<td>Enhanced ability to be responsive rather than reactive to challenges as they arise</td>
<td>15 (33)</td>
<td>19 (42)</td>
<td>11 (24)</td>
</tr>
<tr>
<td>Enhanced coping skills</td>
<td>14 (31)</td>
<td>17 (38)</td>
<td>14 (31)</td>
</tr>
<tr>
<td>Improved capacity to work with teams on a day-to-day basis</td>
<td>11 (24)</td>
<td>15 (33)</td>
<td>19 (42)</td>
</tr>
</tbody>
</table>

Qualitative Results

Participants

A total of 21 participants agreed to be contacted for participation in an interview and received a standardized invitation letter via email. Qualitative interviews were conducted with 15 consenting participants who answered the email and agreed to be interviewed, including 5 physicians, 2 administrative assistants, 2 physiotherapists, 2 psychotherapists, and 1 each of nurse practitioner, social worker, executive manager, and a speech pathologist. By age, this included 9 (60%) persons aged between 40 and 59 years and 6 (40%) 20 and 39 years. A considerable number of qualitative study respondents were women (13/15, 87%) than men, reflecting the fact that, out of 15 participants, 40 (80%) women and 7 (14%) men completed the descriptive questionnaire. Participants were evenly divided in terms of when they were interviewed during the COVID-19 pandemic: 5 (33%) in the first wave, 5 (33%) in the second wave, and 5 (33%) in the third wave. Interview participants completed a mean of 7 sessions (median 10, range 1 to 11) at the time of completing the questionnaire. Most participants had some prior exposure to mindfulness (11/15, 73%).

Themes

Overview

Themes, experiences, enablers, and barriers to participation were related to respondents’ motivation, satisfaction, impact, and program design. Qualitative findings are described with select quotes and summarized below. Where participants’ quotes are used, we have specified their study number, role, age, number of sessions attended, whether they had applied the mindfulness skills taught, and the pandemic wave during which they were interviewed. A summary of findings is presented in Figure 3 and all data are included in Multimedia Appendix 5.
Motivation to Attend and Expectations

A variety of reasons motivated the participants to take part in the program. Most were looking for ways to relax and manage their stress:

*Trying to help deal with the COVID and all the stress behind it. Dealing with the patients and their anxiety with it. I was kind of trying to be more relaxed.* [#02 administrative assistant, aged 40-59 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 1]

Some participants wanted to decrease their sense of isolation created by the pandemic and have a way to connect with other HCWs:

*I had been looking all over the place just for some kind of event or experience or connection of some kind for healthcare staff, healthcare teams because I felt really isolated.* [#46 psychotherapist, aged 20-39 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

Others said that they were curious about mindfulness, had been exposed to it in the past, and desired to gather with others and be a role model to encourage other staff to participate:

*I am in a leadership position at the hospital and it was passed on to me as something that we could share with our staff. I wanted to encourage them to take it so I thought I should take it myself.* [#14 manager, aged 40-59 years, ≤5 sessions, not applied learning, techniques, or tools in daily life, wave 1]

When asked about expectations, participants largely said that they had none and came with an open mind. Several participants anticipated that it was likely to be beneficial but did not specify how:

*I didn’t honestly think about it that much other than I know, at least on some level, that mindfulness practice would be good for me, that just by showing up it would be a positive thing.* [#10, physiotherapist, aged 20-39 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 1]

*Feeling quite unmotivated, just finding things at work difficult for the winter and the early part of spring, so it wasn’t necessarily fatigue, it was more feeling a bit overwhelmed with work.* [#75, physician, aged 40-59 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 3]

Satisfaction

All participants said that the sessions were helpful, and some said that they surpassed their expectations and felt it was essential for getting through the pandemic:

*It was more than what I expected. I thought I would try, to help out with the burnout and with the pandemic. But it was a lot more relaxing than I thought it would be. I was able to get into a really deeply relaxed state.* [#22, psychotherapist, aged 20-39 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]
In order to keep working and go on in my life, it’s been one of the very important things for me. [#35, physician, aged 40-59 years, applied learning, techniques, or tools in daily life, wave 3, did not indicate the number of sessions they attended]

When asked to rate the importance of the program to them on a scale from 1 to 10, in which 10 was very important, most participants articulated a moderately high score (mean 7.7, SD 1.48; median 8.0, range 5.0–10.0). Some participants commented on how this program was one of the several resources or supports, such as family, and that they were new to the program and could not yet fully appreciate how important the sessions may be:

I needed multiple things to help me through this time and so it was very important to me to have that regular time twice a week, where I would check-in and do these techniques to make sense of what’s going on, being able to deal with the difficult situation. In order to keep working and go on in my life, it’s been one of the very important things for me. [#35, physician, aged 40-59 years, applied learning, techniques, or tools in daily life, wave 3, did not indicate the number of sessions they attended; level of importance in life: score=10]

I’ve barely scratched the surface... it’s the beginning of a learning. I’m hoping that if and when it ends, I will be able to continue this. A 10 out of 10 would be a sort of life-changing, I don’t think it’s life-changing. I would call it life-enhancing. [#78, physician, aged 40-59 years, >11 sessions, applied learning, techniques, or tools in daily life, wave 3; level of importance in life: score=8]

I can’t say yet. I think I’m still pretty new. I’ve only done two sessions. Yes and no. In the sense that I think it’s nice to have that put in to your schedule a little bit but I can’t say that it’s translated into anything else in my life yet. [#44, physician, aged 40-59 years, ≤5 sessions, not applied learning, techniques, or tools in daily life, wave 2; level of importance in life: score=5]

Impact

The participants articulated the numerous positive benefits of the mindfulness program. Most participants said that participation in the program helped them to be more present and calm and to better manage stress or anxiety. Many participants applied mindfulness techniques in their daily lives, including at work:

Being present, being consistent and calm when you are working in very high stress environments...I did try that out when I was at work, and especially when you’re feeling overwhelmed or just tired. And so, I think just being able to use those skills can help you think more clearly, tackle things in a calmer manner. [#01, physician, aged 20-39 years, >11 sessions, applied learning, techniques, or tools in daily life, wave 1]

I’m much more productive with my time...and when you start to get more things done you don’t stress as much about it. It’s still pretty intense and it’s been a year and half...it just made me more level and able to work. [#78, physician, aged 40-59 years, >11 sessions, applied learning, techniques, or tools in daily life, wave 3]

I found it a very useful tool to use, in my personal toolbox, to help me be more grounded, to be more productive and actually slow down and listen to others. [#78, physician, aged 40-59 years, >11 sessions, applied learning, techniques, or tools in daily life, wave 3]

Engaging in a shared activity with other HCWs reduced feelings of isolation and loneliness, with gratitude for acknowledging HCWs as those who needed help to combat the undue stress of the pandemic:

I didn’t know any of the other people but you still kind of felt, even if you don’t know them everyone is facing similar challenges and probably having similar worries. That was nice to sit together. [#01, physician, aged 20-39 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

It’s really comforting to see the same people coming, even if I don’t see them, sometimes I see their comments...I always like the idea of a community so feeling included or part of the group, where I’m not lost in a big system, so that helps. [#35, physician, aged 40-59 years, applied learning, techniques, or tools in daily life, wave 3]

Most participants said that participation in the program helped them to be more present and calm and to better manage stress or anxiety. Many participants applied mindfulness techniques in their daily lives, including at work:

Being present, being consistent and calm when you are working in very high stress environments...I did try that out when I was at work, and especially when you’re feeling overwhelmed or just tired. And so, I think just being able to use those skills can help you think more clearly, tackle things in a calmer manner. [#01, physician, aged 20-39 years, >11 sessions, applied learning, techniques, or tools in daily life, wave 1]

I’m much more productive with my time...and when you start to get more things done you don’t stress as much about it. It’s still pretty intense and it’s been a year and half...it just made me more level and able to work. [#78, physician, aged 40-59 years, >11 sessions, applied learning, techniques, or tools in daily life, wave 3]

I found it a very useful tool to use, in my personal toolbox, to help me be more grounded, to be more productive and actually slow down and listen to others. [#78, physician, aged 40-59 years, >11 sessions, applied learning, techniques, or tools in daily life, wave 3]

Engaging in a shared activity with other HCWs reduced feelings of isolation and loneliness, with gratitude for acknowledging HCWs as those who needed help to combat the undue stress of the pandemic:

I didn’t know any of the other people but you still kind of felt, even if you don’t know them everyone is facing similar challenges and probably having similar worries. That was nice to sit together. [#01, physician, aged 20-39 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

It’s really comforting to see the same people coming, even if I don’t see them, sometimes I see their comments...I always like the idea of a community so feeling included or part of the group, where I’m not lost in a big system, so that helps. [#35, physician, aged 40-59 years, applied learning, techniques, or tools in daily life, wave 3]

A few participants mentioned additional positive impacts, such as sharing mindfulness techniques with patients whom they thought might benefit.

Design

When asked about the program design and format, participants appreciated the web-based, drop-in nature requiring no advance registration, the flow and variety of activities during each session, and the interactive components. Participants also appreciated the soothing voices, reassurances and reminders.
offered by faculty facilitators, the timing of the evening sessions, and the short 30-minute session length:

I love that it’s drop-in, you don’t have to register or sign up or make a plan. [#04, nurse practitioner, aged 20-39 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 1]

I joined a couple sessions where it seemed to flow really nicely. I thought that there was kind of an initial activity and then a little interlude moment and then a second activity. [#04, nurse practitioner, aged 20-39 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 1]

I really also enjoyed that there was a little bit of dialogue at the end, and I thought that was so nice for other people to share their comments, to be able to share mine. [#14, manager, aged 40-59 years, ≤5 sessions, not applied learning, techniques, or tools in daily life, wave 1]

Furthermore, the participants noted specific aspects of the content they found helpful. In most cases, this referred to mindfulness exercises or techniques, for example, breathing exercises or body scans. In other cases, this referred to a concept, such as the need for self-care:

Self-compassion practices are probably the ones that stick out most. [#10, physiotherapist, aged 20-39 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 1]

The whole practice was around being kind to yourself and kind to others even though we may differ in our opinions or views etcetera. [...] And that was really impactful given the social-political state the world is in right now, so much divide and social injustice and political divide. It was really refreshing to talk about community and generosity and focus on that, when we’re bombarded everyday with the news and tv about what’s different about us right now. [#22, psychotherapist, aged 20-39 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

My personality is that I always helped other people, family. That’s me. That’s my job and that’s also my personality. I always put everybody in front of me instead of taking care of myself. Which I think I do have to change a little bit as I am getting older [...] if I am forced to get into a place where I am the center and not everybody else but just me only, I think that would help me tremendously. [#46, psychotherapist, aged 20-39 years, ≤5 sessions, not applied learning, techniques, or tools in daily life, wave 2]

Enablers

Participants noted multiple factors that enabled or challenged participation in Pause-4-Providers and the application of mindfulness outside the scheduled sessions.

These included a well-designed website and facilitated, web-based drop-in sessions. Although most participants appreciated the group nature of the sessions, they also valued their anonymity:

It’s convenient because it’s online, you don’t have to go somewhere, you can just Zoom in. It’s hard to get myself going somewhere to an actual place, but it is good to be with people. You kind of feel like you are with them, even if you don’t see them, so for me, the virtual format was perfect. [#35, physician, aged 40-59 years, applied learning, techniques, or tools in daily life, wave 3, did not indicate the number of sessions they attended]

I think that’s a really powerful aspect of the program that invites more people to attend because you can change your screen name and be anonymous, so I thought that was cool. I think that anonymity aspect is really important so that everybody can just feel safe to access support and not have to reveal their identity from fear of judgement from coworkers. [#76, administrative assistant, aged 20-39 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 3]

Enablers to adopt or use mindfulness included faculty facilitators giving advice and examples on how to integrate practice into daily life, demonstrating techniques that were easy to adopt, and offering self-directed resources on the website to enable practice outside the sessions:

The way the facilitators are able to link the practices to real, realistic, day-to-day moments where the skills might come in handy [...] the suggestions, gentle reminders of where these skills can be useful in daily life. [#10, physiotherapist, aged 20-39 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 1]

Even throughout the day I think of some of the exercises. I’m seeing that start to change in a positive way. I’m taking a break from work at my desk, I stand up and I’ll take a moment with the practices in mind. [#46, psychotherapist, aged 20-39 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

Barriers

Barriers to participation included competing demands that made it difficult to be a priority, being tired and feeling easily distracted, difficulty finding a quiet and private space at home, internet connection issues, and being seen by others in a susceptible state:

I guess the time of day. Although it is very convenient for me to attend, I do have to make a choice to do. Either I fit in time with my husband, or I prep for tomorrow or clean up. It’s a choice. You have to make a choice to go cause there’s always something else that you could be doing. [#04, nurse practitioner, aged ≥40 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 1]

Finding a quiet space in the house. I have to be really intentional about that part of things. I really need to
set that up. They compare to in-person! But now, even if I get in the room, even if I close the door and set it up, there’s still this possibility of other sounds in the house or somebody coming in. I think that’s a barrier to full participation, a distraction. [#46, psychotherapist, aged <40 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

Some of the participants expressed a concern about appearing susceptible:

The one thing did cross my mind, this is a time I was hoping to be in my pajamas, wash my face, maybe lying down, and I wondered am I going to have to be on camera? I don’t want other people to see me in this kind of vulnerable state but that was quickly cleared up at the beginning of the Zoom call. [#22, psychotherapist, aged <40 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

Barriers to adopting, applying, or using mindfulness outside of the sessions included forgetting or feeling unable to invoke learned skills and finding a quiet place at work to practice mindfulness:

It hasn’t been something that I have been trained to do so it takes a lot more effort and add-on to implement these new learned behaviors. [#22, psychotherapist, aged <40 years, ≤5 sessions, applied learning, techniques, or tools in daily life, wave 2]

Participants variably preferred some practices over others, with none of the different practices being universally favored or disliked by the respondents. Participants offered 2 suggestions for promoting and supporting the adoption of mindfulness: offering quiet spaces at work and providing or referring participants to other resources that support mindfulness practice. The sessions were open to all HCWs, and some saw this as a barrier to feeling comfortable, with several participants offering the opinion that to create a safe space, consideration should be given to restricting the sessions to licensed clinical health care providers:

If one of the goals of the intervention was more engagement or discussion, or people sharing their stories, keeping it to the regulated professionals might have been more of a safe space, I guess. Not knowing who was on the call. I think at times kind of kept me from speaking up or asking questions or participating in discussions or feedback. Whereas, if it was just my peers, then it may have felt more of a safer space to talk. [#04, nurse practitioner, aged 20-39 years, 6-10 sessions, applied learning, techniques, or tools in daily life, wave 1]

**Discussion**

**Principal Findings**

The strain placed on HCWs was evident from the start of the COVID-19 pandemic and persisted as the pandemic endured. This newly conceived well-being support program, Pause-4-Providers, consisted of synchronous, web-based, brief, drop-in mindfulness sessions for HCWs. It was feasible to implement as evidenced by its rapid deployment early in the first phase of the pandemic, with good use and continued session attendance during the first 3 waves of the pandemic (total attendance=1333). This program reached a wide range of HCWs, including physicians, a group typically more reluctant to attend well-being sessions, and more so during the pandemic [57]. It was acceptable to participants, with a large majority endorsing overall satisfaction of 5 out of 5 (32/47, 68%) and 4 out of 5 (11/47, 23%).

The Pause-4-Providers program was intentionally planned for implementation for HCWs with competing demands, time pressures, and exhaustion during the COVID-19 pandemic. Participants corroborated that they valued the program content and format that did not require registration, took place on the internet in the evening, was a drop-in session, and was of short duration (30 min) with the option of anonymity (eg, cameras could be switched off). The nonprescriptive nature of drop-in sessions allowed participants the choice to attend any number of sessions, unlike other structured mindfulness programs. Consistent with mindfulness literature, we observed a potential dose effect [31]. Participants who attended ≥2 sessions reported more positive benefits in coping skills (P<.001) and reduced perceived stress (P=.008).

Half of the participants attending Pause-4-Providers had moderate to high levels of burnout, as measured by the Single-Item Burnout Measure (26/50, 52%), similar to what others had during the COVID-19 pandemic [56]. This highlights the fact that some participants were in need of well-being interventions. In other words, some may have been motivated to alleviate burnout symptoms, while others sought to cultivate their well-being. We observed that resilience and burnout coexisted in this sample of HCWs during the COVID-19 pandemic.

The Pause-4-Providers study participants scored moderately high overall on the Resilience at Work Scale, with high subscale scores of “building networks, interacting cooperatively, finding your calling, and living authentically.” This may reflect HCWs’ need to pull together teams, with a shared purpose and a “calling”—resiliency factors that are essential during a pandemic. There is no ceiling on resilience, which can continue to grow. However, participants endorsed difficulty with the “maintaining perspective” subscale, scoring 48.1 (difficulty ≤50). This subscale represents an ability for flexible thinking and optimism. The pandemic has challenged HCWs to be adaptable and maintain positivity. The use of mindfulness training can help stressed individuals maintain perspective and shift from primitive survival stress responses (eg, fight, flight, freeze, or fold) to adaptive coping with mindful and reflective problem solving skills. The participants scored moderately on the subscale of “managing stress” aligned with their endorsed reasons for taking the course to relax and manage stress or anxiety. This is consistent with areas that mindfulness sessions can address by helping individuals to keep calm and reduce perceived stress [23,24,29-31]. These findings on the Resilience at Work measure were similar to those in a study by Mauder et al [56] with Canadian HCWs that evaluated the psychological impact of COVID-19.
The reported benefits of the program in the qualitative study included being better able to disconnect and decompress after work, seeing a need to prioritize self-care, and an improved ability to manage stress at work. This aligned with the survey findings of increased self-awareness of emotional states (25/45, 56%) and reduced stress (24/45, 53%). By providing a toolbox of mindfulness strategies and skills, participants were empowered to be calmer and more present. Brief mindfulness micropractices provided a potential mechanism for HCWs to navigate stressful and demanding workplace environments, as demonstrated in previous work done by one of the authors (ME) [13,35,36]. The themed agendas of each session focused on workplace applications of mindfulness practices. Respondents described how the particular use of micropractices allowed them to be more present, calmer, and more balanced at work. Some participants shared these mindfulness tools with their coworkers and patients. Participants also valued the faculty facilitators’ encouragement, tips, and examples of ways to use practices in the workplace, on the go.

Many participants did not turn on their cameras, although many did offer their thoughts and feedback (often gratitude) in the chat at the end of each session. Without their cameras turned on and in their homes in the evenings, the participants reported feeling less susceptible and more relaxed. Despite the anonymity of participation, the attendees endorsed a decreased sense of isolation in the context of the pandemic and felt connected to the group. We speculate that the combination of the live format, the commonality of being with HCWs during the pandemic, and the shared experience of practicing self-care in a group setting was sufficient to promote a sense of connection. This is similar to other mindfulness skills–based groups in which the commonality of being with HCWs during the pandemic, and the shared experience of practicing self-care in a group setting was sufficient to promote a sense of connection. This is similar to other mindfulness skills–based groups in which participants do not talk about their life experiences and yet endorse an increased sense of connectedness [37].

The use of an evolving curriculum allowed the team of faculty facilitators to perform practical ongoing needs assessment by checking in with attendees at the start of each session and flexibly adapting the session in response to the relevant needs of the attendees [36]. This was important as HCWs’ needs changed with the course of the pandemic. In the first wave, HCWs were frightened and needed to find ways to cope with uncertainty and change, whereas, with time, participants shared experiences of moral distress, exhaustion, and elements of posttraumatic stress [6,22]. The faculty facilitators could garner an iterative understanding of challenges and burnout factors with time as mental health experts, mindfulness specialists, and HCWs.

Limitations
We largely relied on a top-down approach to invite HCWs to Pause-4-Providers sessions, contacting leadership in a wide array of institutions and health care settings. This dissemination strategy included outreach to interprofessional stakeholder organizations at the provincial, university, and hospital health care system levels. We did not track how or if they disseminated the information, and we did not have a consistent strategy to follow up with these organizations. Although this session recruitment strategy supported an expedient deployment strategy, it may have affected attendee recruitment.

Prioritizing anonymity and open access to drop-in web-based sessions with no required registration meant that we were not able to provide attendance reminders and relied on participant engagement and institutions to notify and remind potential attendees. This likely limited session attendance, retention, and study recruitment. Nonetheless, the study design was pragmatic for the drop-in nature of the program, with rolling recruitment during the first 3 waves of the pandemic. Approximately half (24/50, 48%) of the participants attended >2 sessions at the time of survey completion. Participants’ reports on resilience and burnout could have been affected by how many sessions they had attended before study enrollment and how far into the pandemic they started to attend sessions. We do not know if the consenting study participants were representative of the entire Pause-4-Providers attendee population or the HCWs’ population at large, limiting generalizability.

When the program was developed, we did not expect the pandemic to endure and have multiple waves. Attendance in the program was well distributed across the first 3 waves of the pandemic, and survey responses were concentrated between the end of wave 1 and the end of wave 2. It is possible that HCWs became more accustomed to the COVID-19 pandemic, with reduced stress and improved coping with time. Alternatively, the cumulative effect of the pandemic may have adversely affected the well-being of HCWs (eg, with increased levels of moral distress). The sample size was too small to determine how the motivations, benefits, satisfaction, and experiences of the participants may have differed across time points. Consistent with qualitative research standards, saturation of the findings was identified, and our sample size was consistent with the qualitative methodology [55,58].

It is unknown how representative the 50 study HCW participants were in relation to the larger sample of all attendees. The demographics did not include data on race, ethnicity, or intersectional identity, nor did they distinguish between work practice settings (eg, intensive care unit vs outpatient clinics) or whether HCWs were in COVID-19 patient-facing practice settings. Any of these factors could have affected or moderated the findings on burnout, workplace resilience, and pandemic-related stresses. The faculty facilitators in this project were White women psychiatrists. Future iterations of this program would benefit from a more culturally diverse and larger group of interprofessional faculty facilitators and participants. With respect to generalizability, qualitative findings were obtained from participants in the province of Ontario, Canada, and thus may not represent other health settings in other locations. Nurses represent approximately half of the health care workforce in Canada, and yet only 18% of the study participants were nurses. This low level of nurse participation may have been because of long, 12-hour shifts and competing priorities, such as family or childcare during the pandemic. This may also be related to free alternative support provided through their professional organizations (eg, individual psychotherapy) [6].

The format of live web-based mindfulness sessions is resource intensive as it relies on faculty facilitators; therefore, it may not be sustainable without funding or a dedicated team. Early in the COVID-19 pandemic, with heightened uncertainty and major
changes in health care, faculty facilitated 8 evening sessions per week. As the pandemic continued, to reduce the burden on the faculty facilitators, the frequency of sessions was decreased to once a week.

**Conclusions**

This descriptive, qualitative study and program evaluation demonstrated that the Pause-4-Providers program was feasible, acceptable, and appropriate, with high levels of satisfaction. Half of the participants endorsed moderate to high burnout, which may have motivated their attendance. Most participants endorsed positive benefits enabled by the curriculum, the faculty facilitators and the drop-in, low-dose, open-access format of the program. Importantly, participants reported feeling more relaxed after the sessions and using mindfulness tools at work and in their daily lives.

As Dana Faulds, an American poet, writes, “It only takes a reminder to breathe, a moment to be still, and just like that, something in me settles, softens, makes space for imperfection...I can make the choice to stop, to breathe, and be, and walk slowly into the mystery” [38]. The Pause-4-Providers program reminded HCWs to breathe, create space to be still, reflect on mindful self-compassion, and use microskills in the workplace and at home. To address the unprecedented negative psychological effects and burnout during the COVID-19 pandemic and support the well-being of HCWs beyond the pandemic, this program of brief web-based drop-in mindfulness merits further study.

**Acknowledgments**

The Pause-4-Providers program was supported by the Innovation Fund of the Alternative Funding Plan for the Academic Health Sciences Centres of Ontario. The authors gratefully acknowledge the Sinai Health System, University Health Network, Women’s College Hospital and University of Toronto, Department of Psychiatry, the Lunenfeld-Tanenbaum Research Institute, and Drs Jon Hunter and Robert Maunder for their mentorship; the Ontario Medical Association; and the contributions of Anna Gagliardi for the qualitative analysis, George Tomlinson and Fuzo Yosh for descriptive analyses in this study, and Natalie Heeney for the design of study figures.

**Data Availability**

The data sets generated and analyzed during this study are not publicly available because participants’ approval for public data sharing was not explicitly requested in the consent form approved by the research ethics board.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Definitions of mindfulness concepts.
[DOCX File, 20 KB - formative_v8i1e43875_app1.docx]

Multimedia Appendix 2
Pause-4-Providers questionnaire package.
[DOCX File, 34 KB - formative_v8i1e43875_app2.docx]

Multimedia Appendix 3
Weekly number of attendees during the first 3 waves of the COVID-19 pandemic.
[PNG File, 66 KB - formative_v8i1e43875_app3.png]

Multimedia Appendix 4
Pause-4-Providers qualitative interview guide.
[DOCX File, 23 KB - formative_v8i1e43875_app4.docx]

Multimedia Appendix 5
Themes and exemplar quotes on motivation, expectations, design, and impact.
[DOCX File, 81 KB - formative_v8i1e43875_app5.docx]

**References**


**Abbreviations**

COREQ: Consolidated Criteria for Reporting Qualitative Research

HCW: health care worker
Latino Parents’ Reactions to and Engagement With a Facebook Group–Based COVID-19 Vaccine Promotion Intervention: Mixed Methods Pilot Study

Anna I González-Salinas¹, BSc, MBA, MM; Elizabeth L Andrade¹, BA, MPH, DrPH; Lorien C Abroms¹, BSc, MA, DSC; Kaitlyn Gómez², BA; Carla Favetto¹; Valeria M Gómez¹, BS; Karen K Collins¹, PhD

¹George Washington University, Washington, DC, United States
²California State University, Fullerton, CA, United States

Corresponding Author:
Anna I González-Salinas, BSc, MBA, MM
George Washington University
950 New Hampshire Avenue
Washington, DC, 20052
United States
Phone: 1 (202) 994 7400
Email: agonzalez985@gwmail.gwu.edu

Abstract

Background: Misinformation in Spanish on social media platforms has contributed to COVID-19 vaccine hesitancy among Latino parents. Brigada Digital de Salud was established to disseminate credible, science-based information about COVID-19 in Spanish on social media.

Objective: This study aims to assess participants’ reactions to and engagement with Brigada Digital content that sought to increase COVID-19 vaccine uptake among US Latino parents and their children.

Methods: We conducted a 5-week intervention in a private, moderator-led Facebook (Meta Platforms, Inc) group with Spanish-speaking Latino parents of children aged ≤18 years (N=55). The intervention participants received 3 to 4 daily Brigada Digital posts and were encouraged to discuss the covered topics through comments and polls. To assess participants’ exposure, reactions, and engagement, we used participants’ responses to a web-based survey administered at 2 time points (baseline and after 5 weeks) and Facebook analytics to calculate the average number of participant views, reactions, and comments. Descriptive statistics were assessed for quantitative survey items, qualitative responses were thematically analyzed, and quotes were selected to illustrate the themes.

Results: Overall, 101 posts were published. Most participants reported visiting the group 1 to 3 times (22/55, 40%) or 4 to 6 (18/55, 33%) times per week and viewing 1 to 2 (23/55, 42%) or 3 to 4 (16/55, 29%) posts per day. Facebook analytics validated this exposure, with 36 views per participant on average. The participants reacted positively to the intervention. Most participants found the content informative and trustworthy (49/55, 89%), easy to understand, and presented in an interesting manner. The participants thought that the moderators were well informed (51/55, 93%) and helpful (50/55, 91%) and praised them for being empathic and responsive. The participants viewed the group environment as welcoming and group members as friendly (45/55, 82%) and supportive (19/55, 35%). The 3 most useful topics for participants were the safety and efficacy of adult COVID-19 vaccines (29/55, 53%), understanding child risk levels (29/55, 53%), and the science behind COVID-19 (24/55, 44%). The preferred formats were educational posts that could be read (38/55, 69%) and videos, including expert (28/55, 51%) and instructional (26/55, 47%) interviews. Regarding engagement, most participants self-reported reacting to posts 1 to 2 (16/55, 29%) or 3 to 4 (15/55, 27%) times per week and commenting on posts 1 to 2 (16/55, 29%) or <1 (20/55, 36%) time per week. This engagement level was validated by analytics, with 10.6 reactions and 3 comments per participant, on average, during the 5 weeks. Participants recommended more opportunities for engagement, such as interacting with the moderators in real time.

Conclusions: With adequate intervention exposure and engagement and overall positive participant reactions, the findings highlight the promise of this digital approach for COVID-19 vaccine–related health promotion.

(JMIR Form Res 2024;8:e51331) doi:10.2196/51331
Introduction

Background

US Latino individuals have experienced distressing COVID-19 disparities related to morbidity and mortality. Recent studies have demonstrated this heightened risk in Latino adults, who have a 1.5 times greater risk of infection, 2.3 times greater risk of hospitalization, and 1.8 times greater risk of death [1-5]. These increased risks are likely related, in part, to slower vaccine uptake during the pandemic [6,7]. Although adult COVID-19 vaccination rates have increased, with 57.1% of US Latino adults having completed a primary vaccine series as of March 2023, only 8.5% of Latino adults have received a bivalent booster dose, the lowest coverage across racial and ethnic subgroups [8].

What is also alarming is the underrepresentation of Latino children in some age groups among the population vaccinated against COVID-19. The vaccination rates of Latino children continue to lag, with only 5.1% of Latino children aged 6 months to 4 years having completed a 2-dose vaccine series as of December 2022 and 28.8% of Latino children aged 5 to 11 years, 57.8% of Latino adolescents aged 12 to 15 years, and 70.4% of Latino adolescents aged 16 to 17 years having completed the primary series as of August 2022 [9,10]. Booster dose coverage among Latino children is also worrisome, with only 4.6% of Latino children aged 5 to 11 years, 20.7% of Latino adolescents aged 12 to 15 years, and 32.3% of Latino adolescents aged 16 to 17 years having received a booster dose as of August 2022 [9,10]. These low rates are particularly concerning, given that research has demonstrated an increased risk among Latino children and adolescents aged <18 years for COVID-19 infection, hospitalization, multisystem inflammatory syndrome, and related health complications [11-16]. These findings emphasize the need for targeted interventions to mitigate these disproportionate risks and increase immunization rates.

Research has revealed that COVID-19 vaccine hesitancy has been a significant obstacle to reducing these risks for US Latino communities [4,17-19], especially among Latino parents, who have exhibited considerable reluctance toward vaccinating their children against COVID-19 [20-23]. This hesitancy has been partly fueled by exposure to misinformation regarding COVID-19 vaccines, primarily through social media networks [20,24-27]. This exposure has created more mistrust and concerns among Latino parents regarding the safety and efficacy of COVID-19 vaccines for children [28-32], and studies have shown an association between exposure to misinformation on social media and increased hesitancy and decreased intention to vaccinate [33-39].

Although there have been efforts to address vaccine hesitancy among Latino individuals, including communication campaigns and community-based outreach initiatives [25,40-42], few social media–based intervention studies have focused on Latino parents. Ramirez and colleagues [43] implemented a Facebook (Meta Platforms, Inc)–based pilot intervention to promote COVID-19 vaccine uptake among rural community residents in South Texas, whereby intentions to vaccinate were assessed following exposure to video testimonials that acknowledged misinformation and promoted vaccination using peer modeling. Although not explicitly focused on parents, this study demonstrated significantly higher vaccine intentions following exposure to such video testimonials compared with exposure to standard Centers for Disease Control and Prevention advertisements. In addition, a study by Panameno and colleagues [44] assessed Latino parents’ experiences with MiVacunaLA, a mobile phone–delivered digital intervention for improving Latino parents’ vaccination intentions and Latino children’s vaccination rates. The study found that digital technology was beneficial for delivering language-tailored and culturally tailored pediatric COVID-19 vaccine information to Latino parents, which addressed their specific informational needs and increased their confidence in COVID-19 vaccines. This study highlighted the importance of concise, accessible information delivered by trusted sources, such as videos portraying physicians or community health workers, to overcome barriers to health literacy or other barriers to COVID-19 vaccination. Furthermore, this study revealed a persistent need for culturally tailored information about COVID-19 vaccines and boosters, specifically information delivered in Spanish and through trusted sources.

Objectives

In response to information gaps among US-based Spanish-speaking audiences, we established Brigada Digital de Salud in 2021 to disseminate accurate, credible, and science-based COVID-19 information and prevention messages in Spanish based on the latest research and public health guidelines. Brigada Digital aims to serve as a social media resource for Spanish-speaking individuals who want to make sound decisions about their health and the health of their families in the face of web-based misinformation [45]. This paper details the findings from a 5-week study among Latino parents in a private Facebook group to assess their reactions to and engagement with Brigada Digital social media content and to elicit their recommendations for future iterations of the intervention.

Methods

Participant Recruitment

Eligible study participants included Latino adults aged ≥18 years who were parents of at least 1 child aged <18 years, spoke fluent Spanish, and reported using Facebook at least once daily. We recruited participants through 8 targeted Facebook advertisements between August 12 and August 22, 2022. All interested individuals completed an 8-item screener to determine eligibility, and eligible participants were automatically directed to an informed consent form and a baseline survey administered using Qualtrics (Qualtrics International Inc). The participants who completed the baseline survey were assigned to either the intervention Facebook group (N=55) that received Brigada
Digital content or the control Facebook group (n=65) that received only a link to the Centers for Disease Control and Prevention web-based COVID-19 information in Spanish (standard of care) but did not receive the intervention content. This paper discusses the reactions and engagement of only the participants in the intervention group.

**Intervention Overview**

A 5-week intervention delivered through Facebook groups was created to educate participants on the risks of COVID-19, COVID-19 prevention methods, and the importance of adult and child vaccination. The intervention aimed to address common misinformation often spread through social media. Intervention content development was informed by the Theory of Planned Behavior, according to which beliefs about vaccination, social norms, perceived control over vaccination, and attitudes about vaccination influenced COVID-19 vaccination intentions [46]. The intervention content was developed before the study and was carefully crafted to be accessible to a broad Latino audience, regardless of education or health literacy levels. For example, scientific concepts were explained in simple terms, prevention recommendations were accompanied by visual aids, and the text was audio narrated in Spanish (Figure 1).

Once participants were accepted into the Facebook group, they received prerecorded welcome videos from moderators, 3 posts, and an overview of group expectations and guidelines and were invited to introduce themselves to the group. This point onward, posts were scheduled to be shared with the group on a daily basis. In total, 101 posts were shared during the 5-week period, with at least 2 posts per day being educational and the remainder alternating between posts intended to counter COVID-19 misinformation and posts intended to engage group members.

The content was delivered using various formats, such as a narrated slide carousel, animated images with text and music, and videos featuring health professionals and subject matter experts (Figure 2).

Furthermore, the post content was ordered to progress through various themes related to COVID-19, beginning with fundamental scientific concepts behind disease transmission and then covering topics with increasing complexity. For example, we incorporated content explaining the connection between carbon dioxide levels in indoor environments and COVID-19 transmission risk, which educated participants regarding the risks associated with gathering indoors in crowded spaces and encouraged them to improve ventilation indoors to reduce their risk (Figure 3).

When promoting the use of web-based resources, posts included tools for assisting participants with navigating these resources, such as tutorials and step-by-step instructions (Figure 4).

**Figure 1.** Brigada Digital social media post using Spanish audio narration to emphasize the importance of wearing a mask to reduce respiratory viral transmission.
**Figure 2.** Brigada Digital TikTok video featuring a Latina health expert explaining the cumulative monthly COVID-19-related hospitalizations among children ages 5-17 years old from June 2021 to March 2022.

**Figure 3.** Brigada Digital educational post using animated images with text and Spanish audio narration to educate about the link between carbon dioxide levels and COVID-19 risk and provide guidelines for indoor mask use. NIH: National Institutes of Health.

**WHAT DOES (CO₂) TELL US?**

- The air that people exhale (CO₂) has small particles. If someone has a COVID infection, those particles can also contain the virus.
- With poor ventilation (airflow), CO₂ and infectious particles accumulate in the air.
- We can measure CO₂ levels in closed spaces. A high CO₂ level indicates that it is a riskier place for COVID-19 transmission.
The post content covered a wide range of topics related to COVID-19, including the science of COVID-19 and its transmission, risks, and prevention (36/101, 35.6%); COVID-19 testing (5/101, 4.9%); COVID-19 vaccine information, such as the safety and efficacy of adult and pediatric vaccines and boosters (28/101, 27.7%); risks associated with pregnancy or breastfeeding and COVID-19 (4/101, 4%); COVID-19 treatment and post–COVID-19 condition (6/101, 5.9%); and understanding COVID-19 misinformation (15/101, 14.8%). Furthermore, (7/101, 6.9%) posts were included on related topics for the purpose of participant engagement (refer to Multimedia Appendix 1 for the complete 5-week schedule of intervention post topics). The posts were scheduled to be delivered every morning, afternoon, and evening. Furthermore, the members of the group had unrestricted access to the Facebook group content, allowing them to view the posts and engage with all the content at any time.

The Facebook intervention was facilitated by 2 Spanish-speaking moderators. The moderators engaged group members in discussions about the posts by posing questions, eliciting opinions, and encouraging the sharing of relevant experiences. Group activity was monitored daily to promptly respond to comments and acknowledge group member participation in a timely manner to encourage continued engagement. Furthermore, the moderators aimed to be helpful in responding to questions by providing detailed responses and links to additional information and resources. In addition, 5 weekly polls were administered to inquire about the group members’ risk perceptions, attitudes, and vaccine intentions. Given the timing of the intervention leading up to the start of the school year for children, the poll questions inquired about parents’
concerns about sending their children back to school with COVID-19 still circulating as well as barriers and intentions to vaccinating their children. During their interactions with participants, the moderators prioritized adherence to culturally appropriate expectations of respect, attentiveness, and kindness [47-49].

Data Collection

Survey
A brief, web-based survey was self-administered in Spanish using the Qualtrics software. Following approval by the George Washington University Institutional Review Board committee and participants’ informed consent, the survey was administered at 2 time points, namely at baseline and after 5 weeks. The participants were incentivized for survey completion. This paper reports intervention group participants’ survey responses at the 5-week follow-up and participants’ sociodemographics reported at baseline, including age, sex, education, income, employment status, household composition, and whether they were born in the United States.

The survey instrument included an item for self-reported adult COVID-19 vaccination, with the response options of “I received 1 dose of a 2-dose series,” “I received both doses of a 2-dose series,” “I received a one-dose vaccine (for adults only),” and “I have not been vaccinated against COVID-19.” Furthermore, a survey item for child COVID-19 vaccination was included for parents with children in the age groups of <5 years, 5 to 11 years, and 12 to 17 years, with response options similar to those provided for the item for adult COVID-19 vaccination. In addition, participants were asked whether they had been required to vaccinate.

To assess participants’ intervention exposure, 2 categorical survey items focusing on self-reported exposure asked how frequently participants visited the Brigada Digital Facebook group and saw the group posts in an average week during the intervention. The survey included standardized and open-ended items related to participants’ reactions to the intervention content, moderators, and group environment. A total of 4 categorical survey items assessed participants’ opinions regarding the information received in the group (ie, how informative or trustworthy the information was, the most useful topics, and their preferences regarding post formats); 2 categorical survey items assessed participants’ opinions about the group moderators; and 2 categorical survey items assessed participants’ opinions of the Facebook group environment and members. These items asked the participants to select their level of agreement with a series of statements, using a 5-point Likert scale ranging from disagree to strongly agree. For participant engagement, we used 3 categorical survey items focusing on self-reported engagement, asking whether they introduced themselves to the group and how frequently they reacted to and commented on posts. Furthermore, 3 open-ended items were included to inquire about participants’ reasons for not engaging in each of these actions. Finally, to elicit participants’ feedback about the intervention, we included 3 open-ended survey items that asked what they liked or did not like about the Facebook group and what suggestions they had for improving the intervention.

Facebook Analytics
Facebook analytics facilitated the tracking of views, reactions, and comments for each post and each participant. We assessed participants’ exposure to the intervention using the metrics of total post views and total views for each participant. We assessed participants’ engagement using the metrics of total post reactions and comments and total reactions and comments for each participant.

Ethical Considerations
All study protocols were reviewed and approved by the George Washington University Institutional Review Board (study #NCR213586). All study participants provided informed consent before enrollment, including their acknowledgment that their survey responses were confidential and that data would be deidentified for analysis and reported in aggregate. The participants were compensated with US $65 in retail gift cards for enrolling in the group and completing the surveys.

Data Analysis
We performed a mixed methods analysis. The survey items were assessed cross-sectionally among intervention participants (N=55) only at follow-up. The survey results for participants’ intervention exposure and reactions as well as participants’ self-reported engagement and opinions about the intervention content, moderator, and group environment were assessed using response frequencies for 13 categorical survey items. The participants’ qualitative responses to 6 open-ended survey items were thematically analyzed to identify the predominant reasons for not engaging with the intervention and areas of concordance for what was liked or disliked about the intervention and to summarize suggestions for improving the intervention. The participants’ quotes were selected and translated into English to illustrate these themes and topics.

Facebook metrics were cumulatively assessed for the duration of the intervention. The Facebook metrics of total post views, reactions, and comments were summed to further assess participants’ intervention exposure and engagement. Views, reactions, and comments were also summed for each participant, and participant averages were tabulated for each of the 3 metrics. The Stata (version 17; StataCorp) software was used for quantitative analysis, and the NVivo (version 14; QSR International) software was used for qualitative analysis.

Results

Overview
The 55 intervention group participants had an average age of 38 (SD 7) years, 91% (n=50) were female, 67% (n=37) had less than a college education, and 91% (n=50) were foreign born. Among the intervention group participants, 64% (n=35) reported having children aged 0 to 4 years living in their household, 64% (n=35) had children aged 5 to 11 years, and 49% (n=27) had adolescents aged 12 to 17 years. Regarding employment status and income, 44% (n=24) of the participants were employed, 49% (n=27) reported being unemployed or staying at home, and 58% (n=32) reported earning an annual household income of ≤US $35,000. At baseline, 85% (n=47) of the parent...
participants had received an initial 1-dose COVID-19 vaccine or completed a 2-dose vaccine series, with 29% (n=16) of the participants saying that they had been required to vaccinate. Among the intervention participants, 3% (1/35) of the parents with children aged 0 to 4 years said that their children had completed the 2-dose vaccine series, with 31% (11/35) of the parents with children aged 5 to 11 years and 70% (19/27) of the parents with adolescents aged 12 to 17 years also saying that their children had completed the 2-dose vaccine series (Table 1).

### Table 1. Intervention group participants’ sociodemographics and COVID-19 vaccination status (N=55).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographics</strong></td>
<td></td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>38 (7)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>50 (91)</td>
</tr>
<tr>
<td>Education lesser than a bachelor’s degree, n (%)</td>
<td>37 (67)</td>
</tr>
<tr>
<td>Foreign born, n (%)</td>
<td>50 (91)</td>
</tr>
<tr>
<td>Employed, n (%)</td>
<td>24 (44)</td>
</tr>
<tr>
<td>Unemployed or staying at home, n (%)</td>
<td>27 (49)</td>
</tr>
<tr>
<td>Annual household income ≤US $35,000, n (%)</td>
<td>32 (58)</td>
</tr>
<tr>
<td><strong>COVID-19 primary vaccination series completion, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Parents</td>
<td>47 (85)</td>
</tr>
<tr>
<td>Children aged 0-4 y (n=35)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Children aged 5-11 y (n=35)</td>
<td>11 (31)</td>
</tr>
<tr>
<td>Adolescents aged 12-17 y (n=27)</td>
<td>19 (70)</td>
</tr>
</tbody>
</table>

### Participants’ Intervention Exposure

When asked how many times they visited the private Facebook group per week during the study, 40% (22/55) of the participants reported visiting the group 1 to 3 times per week, a substantial proportion (18/55, 33%) of the participants reported visiting the group 4 to 6 times per week, a smaller proportion (8/55, 15%) of the participants reported visiting <1 time per week, and an even smaller proportion (7/55, 13%) of the participants reported visiting the group ≥7 times per week (Table 2).

### Table 2. Participants’ average weekly exposure to the intervention content (N=55).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of group visits/wk</strong></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>22 (40)</td>
</tr>
<tr>
<td>4-6</td>
<td>18 (33)</td>
</tr>
<tr>
<td>&lt;1</td>
<td>8 (15)</td>
</tr>
<tr>
<td>&gt;7</td>
<td>7 (13)</td>
</tr>
<tr>
<td><strong>Number of posts viewed/d</strong></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>23 (42)</td>
</tr>
<tr>
<td>3-4</td>
<td>16 (29)</td>
</tr>
<tr>
<td>&lt;1</td>
<td>12 (22)</td>
</tr>
<tr>
<td>≥5</td>
<td>4 (7)</td>
</tr>
</tbody>
</table>

In addition, when asked how many posts, on average, they viewed per day, 42% (23/55) of the participants reported viewing 1 to 2 posts per day, approximately one-third (16/55, 29%) of the participants reported viewing 3 to 4 posts per day, approximately one-fifth (12/55, 22%) of the participants reported viewing <1 post per day, and a small proportion (4/55, 7%) of the participants reported viewing ≥5 posts per day. Facebook group analytics from the 5-week period validated this level of participant exposure, with a total of 2004 post views or an average of 36.4 post views per participant.

### Reactions to the Intervention Content

The participants generally had positive reactions to the content of the intervention. Most participants (49/55, 89%) indicated that they agreed or strongly agreed that the intervention content was informative, whereas only 7% (4/55) disagreed or strongly disagreed. When asked what they liked about the intervention, 78% (43/55) of the participants responded that they liked the information provided and found it to be important and beneficial to them:
They give a lot of very important information, and more for parents, it’s super important to keep us informed.

Very good information for our benefit.

Everything. I like it because they inform us well about the vaccine.

Specifically, 13% (7/55) of the participants responded that they liked how the group content kept them informed of the latest, up-to-date information about COVID-19:

It is a very helpful group informing and trying to keep people up to date on COVID.

Furthermore, 13% (7/55) of the participants expressed that they liked that the topics were covered in detail yet in a way that was easy to comprehend, concise, and interesting. One of the participants commented on how the information provided in the group helped her make COVID-19 vaccination decisions:

All the information was explained well and the posts were understood. I had no problems understanding the topics.

I like the way the topics are fully explained.

All the information that was given to me in the group seemed very interesting to me.

This helped me be confident in decisions about the COVID vaccine.

In addition, most participants (49/55, 89%) agreed or strongly agreed that they trusted the information presented in the posts, whereas very few (2/55, 4%) disagreed or strongly disagreed. When asked what they liked about the intervention, 3 (5%) of the 55 participants echoed this viewpoint, indicating that they liked receiving “official” information about COVID-19 vaccines that was from a trusted source or supported by data:

That it addresses issues about vaccines and gives official information that many times one does not know.

That I can get information and opinions from a safe place.

I liked that they talked about the importance of vaccines with data since there are a lot of incredulous people that give out different kinds of information.

When asked for feedback on the quantity of posts received, 89% (49/55) of the participants thought that they had received the correct quantity of posts, and 11% (6/55) thought that they had not received enough posts. Furthermore, the intervention participants were asked which post topics they deemed to be the most useful. According to the participants, the 2 most useful topics were the safety and efficacy of adult COVID-19 vaccines and understanding the levels of COVID-19 risks for children, with over half (29/55, 53%) of the participants selecting each of these 2 topics (Table 3).

Furthermore, 44% (24/55) of the participants found the information about the science behind COVID-19 and its variants to be the most useful, and 38% (21/55) thought that the adult and child COVID-19 booster information was the most useful. Approximately one-fifth to one-quarter of the participants thought that the topics of COVID-19 testing (14/55, 25%), pregnancy and COVID-19 (12/55, 22%), and understanding COVID-19 misinformation (13/55, 24%) were the most useful.

When asked for feedback on how to improve the Facebook group experience, 5 (9%) of the 55 participants suggested having more content in general and continuing to provide information on COVID-19, COVID-19 vaccine side effects, COVID-19 vaccines for children, and other health topics.

Regarding preferred post types or formats, the participants indicated that they liked video content, with 51% (28/55) of the participants liking video interviews with experts, 47% (26/55) of the participants liking instructional videos, and 29% (16/55) of the participants liking audio-narrated educational posts (Table 4).

Most participants (38/55, 69%) indicated favoring educational content that could be read, and a substantial proportion also liked formats that were meant to be engaging, including group polls (25/55, 45%) and funny posts (16/55, 29%).
Table 3. The most useful topics covered in the posts (N=55).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COVID-19 test</td>
<td>14 (25)</td>
</tr>
<tr>
<td>COVID-19 and pregnancy and breastfeeding</td>
<td>12 (22)</td>
</tr>
<tr>
<td>COVID-19 transmission and prevention</td>
<td>17 (31)</td>
</tr>
<tr>
<td>Information on the selection and use of masks</td>
<td>19 (35)</td>
</tr>
<tr>
<td>Purpose of adult and child COVID-19 boosters</td>
<td>21 (38)</td>
</tr>
<tr>
<td>Safety and efficacy of the COVID-19 vaccine for adults</td>
<td>29 (53)</td>
</tr>
<tr>
<td>Safety and efficacy of the COVID-19 vaccine for children</td>
<td>19 (35)</td>
</tr>
<tr>
<td>Science behind COVID-19 and its variants</td>
<td>24 (44)</td>
</tr>
<tr>
<td>Understanding COVID-19 risk levels for adults</td>
<td>20 (36)</td>
</tr>
<tr>
<td>Understanding COVID-19 risk levels for children</td>
<td>29 (53)</td>
</tr>
<tr>
<td>Understanding misinformation about COVID-19</td>
<td>13 (24)</td>
</tr>
<tr>
<td>Other topics</td>
<td>3 (5)</td>
</tr>
</tbody>
</table>

*a*The response format was *select all that apply.*

Table 4. Preferred post types and formats (N=55).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational posts that could be read</td>
<td>38 (69)</td>
</tr>
<tr>
<td>Educational posts that were audio narrated</td>
<td>16 (29)</td>
</tr>
<tr>
<td>Instructional videos (eg, tutorials and demonstrations)</td>
<td>26 (47)</td>
</tr>
<tr>
<td>Videos of interviews or presentations with experts</td>
<td>28 (51)</td>
</tr>
<tr>
<td>Funny posts</td>
<td>16 (29)</td>
</tr>
<tr>
<td>Facebook group polls</td>
<td>25 (45)</td>
</tr>
<tr>
<td>“I did not like any of the posts”</td>
<td>2 (4)</td>
</tr>
</tbody>
</table>

*a*The response format was *select all that apply.*

Reactions to the Moderators and Group Environment

When asked what they liked about the group, 7 (13%) of the 55 participants said that they liked everything about the group. Furthermore, the results revealed that most participants *agreed* or *strongly agreed* that the moderators were well informed (51/55, 93%) and helpful (50/55, 91%), whereas 2% (1/55) and 5% (3/55) *disagreed* or *strongly disagreed* with these statements, respectively. When asked what they liked about the intervention, 6 (11%) of the 55 participants specifically praised the moderators for being empathetic and responsive:

*I really liked the moderator of the group.*

The information that they provided and the empathy that was shown.

The information and the attentiveness to answer comments.

Regarding the group environment, a large proportion (51/55, 93%) of the participants *agreed* or *strongly agreed* that joining the group was helpful and had positive opinions about the group members. Most participants said that their perceptions of other group members were that they were friendly (45/55, 82%) and supportive (19/55, 35%) during their interactions with the group (Table 5).
A little more than one-third (18/55, 33%) of the participants thought that other group members were similar to them; one-fifth (11/55, 20%) of the participants thought that other group members were tolerant; and few participants thought that other group members were critical (3/55, 5%), bothersome (1/55, 2%), or different than me (3/55, 5%).

When asked what they liked about the intervention, 2 (4%) of the 55 participants said that they liked how the group supported the Latino community with an opportunity to connect with people from many locations and with diverse ways of thinking:

I like the interest that exists to inform and support the Latino community.

It is very entertaining and friendly. You meet people from all over with different ways of thinking.

When asked whether they had recommendations for the intervention, 3 (5%) of the 55 participants indicated that they thought that the posts were not always reaching participants and suggested disseminating more posts or repeating posts:

Be more insistent in repeating the posts so that they go out to all the participants.

More posts, so if I miss one, I can look at others.

Participants’ Engagement

The participants were asked about their average weekly engagement with the post content. The levels of engagement were relatively evenly distributed in terms of self-reported reactions to posts, with 13% (7/55) of the participants reacting to ≥5 posts in an average week, 29% (16/55) reacting to 1 to 2 posts per week, 27% (15/55) reacting to 3 to 4 posts per week, and 27% (15/55) reacting to ≤1 post per week. A small proportion (2/55, 4%) of the participants did not react to any posts. This level of participant engagement through post reactions was validated by Facebook analytics; the group had a cumulative total of 584 reactions or an average of 10.6 post reactions per participant. When asked about the reasons for having never reacted to posts, the most common reasons were that the participants preferred only receiving information from the posts, disliked commenting, or were shy.

We also inquired about self-reported group engagement, including whether participants had interacted with group members and posts. When asked whether they had introduced themselves to the group upon joining, 89% (49/55) of the participants revealed that they had. Among those who had not introduced themselves, the reasons varied. Some were interested only in staying up to date with the group’s posts and did not want to engage with others. Of the 55 participants, 2 (4%) mentioned that they were timid, whereas another 2 (4%) said they were unsure of how to introduce themselves.

When asked about recommendations for improving the intervention, 3 (5%) of the 55 participants suggested adding more opportunities for engagement, 1 (2%) requested adding more polls, and 2 (4%) suggested having opportunities to interact in real time with the moderators and group members:

Maybe they could do something to get to know each other a little more and resolve certain doubts.

Discussion

Principal Findings

This intervention entailed the delivery of health promotion messaging and engagement with Spanish-speaking Latino parents in a Facebook group to increase the uptake of adult and child COVID-19 vaccination. The study results suggest that this is a promising approach that should be further tested. The digital setting was particularly appealing because of the flexible opportunities it offered parents to interact both with the educational content at their own pace and with the group members and moderators to resolve concerns, ask for advice, and seek social support. This group-based approach also

Table 5. Participants’ opinions of other group members (N=55).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values, n (%)a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friendly</td>
<td>45 (82)</td>
</tr>
<tr>
<td>Bothersome</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Supportive</td>
<td>19 (35)</td>
</tr>
<tr>
<td>Similar to me</td>
<td>18 (33)</td>
</tr>
<tr>
<td>Critical</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Tolerant</td>
<td>11 (20)</td>
</tr>
<tr>
<td>Different than me</td>
<td>3 (5)</td>
</tr>
</tbody>
</table>

aThe response format was select all that apply.
facilitated the use of multimedia educational formats, which included accessibility features, such as audio narration or video-based delivery, that are preferred by parents and address barriers related to literacy or health literacy.

A substantial quantity of daily posts was shared in the group over the 5-week period, with content tailored to a Latino parent audience and covering a wide range of topics related to COVID-19 risk, prevention, and vaccination. Intervention content was intended to provide updated, accurate information to influence parents’ attitudes and beliefs about adult and child COVID-19 vaccination and, ultimately, increase intentions to vaccinate. It is important to note that all content was developed before the intervention, with great attention paid to the selection and sequencing of topics, accessibility to diverse educational levels, and cultural aspects of messaging and delivery. This enabled the scheduling and delivery of many posts within a short time frame. However, given that the intervention was implemented during an ongoing pandemic, which was a rapidly evolving context, it was necessary to review and update all content immediately before scheduled delivery to ensure that the messaging was still consistent with current public health guidelines, overall risk levels, and vaccine recommendations.

With the abundance of misinformation and mixed messaging on social media, this is highly important for establishing and maintaining trust and credibility in participants’ eyes [50-53]. Future interventions seeking to replicate this approach, especially in rapidly changing contexts, should also prioritize the review of scheduled content before delivery to ensure the accuracy and consistency of messaging and alignment with the latest guidelines. In addition, given that the group members resided in different areas of the United States, messaging was developed to be relevant to national COVID-19 community transmission trends and risk levels; geographically focused interventions should tailor content to reflect local circumstances more closely.

**Participants’ Intervention Exposure**

The participants reported viewing posts and visiting the group a fair number of times during the intervention, with 78% (43/55) of the participants saying that they viewed at least 1 post on an average day and 85% (47/55) of the participants saying that they visited the group at least once during an average week. However, 22% (12/55) of the participants reported viewing <1 post on an average day and 15% (8/55) reported visiting the group <1 time in an average week. Although there was considerable variability in participants’ self-reported exposure to post content, posts received an average number of 36.4 views per participant during the 5 weeks.

In addition to creating content that was varied in format and tailored to the Latino parent audience, posts were scheduled to be delivered at different times of the day (10 AM, 3 PM, 5 PM, and 7 PM) to increase the likelihood that participants would receive the intervention posts in their Facebook feed, regardless of when they were using the platform. However, some participants had still not been exposed to all intervention content. This was also reflected in the participant responses, and 3 (5%) of the 55 participants stated that they thought that the posts were not always reaching group members. This could be because of a number of barriers experienced by the participants, such as competing demands for their time, or platform-specific factors, such as the Facebook algorithm for user feed content curation. Given that the Facebook algorithm is personalized and operates based on the user’s history and patterns of content interaction, more exposure will be attained based on post contents, post formats (eg, videos and photos), and posts from sources with which an individual frequently interacts [54]. For example, if intervention participants do not engage with or view posts delivered earlier in the intervention, they are less likely to be exposed to similar content in the later weeks of the intervention. Similarly, if users predominantly engage with video content, they are less likely to be exposed to content containing only images. This algorithm that dictates content in users’ feeds has implications for group-based public health interventions that rely on content exposure to improve knowledge, shift attitudes and beliefs, and increase access to resources and tools for health behavior changes. Because exposing participants to content is imperative for public health impact, future interventions could take actions to increase the likelihood that this algorithm works in their favor, such as placing content or formats that have demonstrated (or are theorized to have) higher levels of engagement at the beginning of the intervention to boost early views and engagement; sending content that varies in terms of format each day or week; and prioritizing early, active engagement strategies as well as continued engagement opportunities each week. Furthermore, as suggested by the study participants, the moderators could post content multiple times to maximize reach or use targeted engagement strategies to prompt the viewing of posts, including tagging or direct messaging specific users. In addition, the Facebook platform, in particular, has added features that allow users to customize their feed to a certain extent, including the option to select up to 30 people and pages to add to “Favorites.” Future interventions should include protocols as part of participant recruitment that request users to select the intervention group as one of their Favorites during the intervention period to increase content exposure. Furthermore, interventions can include information and prompts that explain different ways in which the group members can access content, including through their feed or direct access to the group.

**Participants’ Reactions**

Overall, the Latino parent participants had positive reactions to the intervention. The participants liked various aspects of the group environment, including the opportunity to come together with other parents like them, support the Latino community, stay informed about COVID-19, and learn the perspectives and experiences of others. The participants’ perception of other group members was very positive in that they viewed others as friendly, supportive, tolerant, and “similar to them.” For future interventions that aim to recruit participants with distinctly opposing views about vaccination or any health issue with polarized viewpoints, group moderators should be trained and prepared to navigate interactions that could become less friendly and tolerant, thus potentially stifling group members’ willingness to interact. This intervention did establish group guidelines at the outset, which did not need to be reinforced at any point during the 5 weeks; however, future interventions...
should consider explicitly stating in the guidelines the grounds for removal from the group to prevent guideline violations.

Another aspect that appeared to contribute to participants’ positive reactions was their perceptions of the group moderators. In addition to being deemed well informed and helpful, the participants provided feedback that the moderators were attentive to the group members’ needs and questions and showed empathy. We prioritized the selection of group moderators to include individuals who were familiar with and adhered to social norms and expectations of positive interpersonal dynamics common to Latin culture, including respect for others, kindness, and attentiveness [47-49]. The moderators were also parents of children aged <18 years, enabling them to relate to the participants’ concerns about their children’s health and safety. Future interventions should seek to replicate a similar group environment with Latino parent audiences, including by having culturally competent moderators who adhere to these expectations for interactions with participants and who are highly responsive to group members’ inquiries.

Overall, participants’ feedback indicated that the intervention content was perceived to be informative and useful. There were 3 topics in particular that the participants thought were especially useful, namely the safety and efficacy of adult COVID-19 vaccines, understanding the levels of COVID-19 risks for children, and the science behind COVID-19 and its variants. Future interventions should explore these topics in depth and, as recommended by others, such as de Vere Hunt and Linos [55], should directly address COVID-19 misinformation in these areas through audience-tailored, culturally representative messaging. In addition, Latino parent participants appreciated the importance of staying up to date with COVID-19 information by regularly receiving concise content, which has also been reported elsewhere, including by Panameno and colleagues [44]. Interventions seeking to replicate this approach, especially in the context of a rapidly evolving public health emergency, should capitalize on how the interventions can address this need for parents by delivering accurate and convenient messaging at regular intervals. Furthermore, when asked what they liked about the intervention, many participants highlighted that they liked that the information was presented in detail yet in an interesting manner that was easy to comprehend. The main goal of Brigada Digital content was to address the gaps in Spanish-language COVID-19 information that was appealing of Brigada Digital content was to address the gaps in Spanish-language COVID-19 information that was appealing.

Participants’ Engagement
Intervention participants demonstrated an early willingness to engage with the group following recruitment, with 89% (49/55) of the participants introducing themselves to the Facebook group when prompted by the moderators. In addition, with averages of 10.6 post reactions and 3 post comments per participant, engagement was similar to or higher than that with other parent-focused Facebook interventions [56,60], although comparable interventions reaching Latino parents specifically are very limited. Combined with positive feedback regarding the group dynamics, these are promising signs of Latino parents’ willingness to engage with Brigada Digital content and learn about and discuss health-related topics in this digital group format. Similar willingness has been observed in other Latino parent–focused digital health promotion interventions [61]. The moderators of future interventions should aim to similarly establish rapport and expectations for group interactions that can promote participants’ willingness to engage with the group. This rapport might be further fostered by offering additional opportunities for engagement throughout the intervention, for example, through live discussions and events, as suggested by the participants, or through polls, which almost half of the participants said that they preferred. The kind of engagement observed among the intervention participants through platform analytics as well as from participant self-report is consistent with typical patterns of engagement among social media platform users, including Latino individuals, who are most likely to simply view content and less likely to comment on posts than react [62,63]. Future efforts might explore ways to boost more active engagement, including using polls as catalysts for discussion or addressing individuals directly with targeted engagement tactics. Moreover, based on feedback from the participants, future interventions should explore ways to offer more opportunities to interact with and get to know group moderators and members, which would further develop rapport, establish trust in the information provided, address parental concerns, and increase engagement by making experts available to answer questions that are of the most interest to participants.

In addition, delivering content in a wide variety of formats and then inquiring about preferences helped us hone in on formats that participants gravitated toward the most. Similar to the study by Panameno and colleagues [44], which demonstrated Latino parents’ preferences for COVID-19 educational information in video formats from trusted sources, approximately half (28/55, 51%) of the study participants indicated that video interviews with experts were preferred, as were instructional videos, which often portrayed a Spanish-speaking health professional giving a demonstration. Formats intended to bridge potential literacy barriers, such as videos with audio-narration features, were also well received. One surprising finding, given social media audiences’ increasing affinity for multimedia content, was that most parents liked educational posts they could read; it is possible that the inclusion of an audio-narration feature helped increase parents’ preferences for these types of posts. This may also indicate that parent audiences interested in a particular health topic, especially for their children’s well-being, may be willing to invest time and effort in reading social media content if they perceive the content to be beneficial, trustworthy, and
valuable to them. Implementers of future social media group–based interventions should also aim to offer a diversity of content delivery formats, with an emphasis on video and other multimedia formats” (in consonance with “content delivery formats”) accessible to audiences with diverse levels of literacy and health literacy.

Limitations
This study has some limitations that should be considered when interpreting the results. This study included a relatively small sample (N=55) of Spanish-speaking Latino parents, with an overrepresentation of female participants. Therefore, the results may not be generalizable to a broader population of Latino parents, particularly fathers. In addition, the study was only 5 weeks in duration, potentially limiting insights into the sustainability of participant engagement and the long-term impact on vaccine uptake. Furthermore, given that the participants were recruited through their response to Facebook advertisements, there is a potential for selection bias that could have resulted in a sample that was more interested in health-related topics and COVID-19 in particular, was more supportive of vaccination in general, and had the time and ability to participate in the group. In addition, it was not possible for this pilot study to fully track all participants’ engagement, and we did not track Facebook log-ins, the length of time using the platform, or the number of times accessing the group. Besides platform metrics of active participant exposure and engagement, such as views, reactions, and comments, data were limited to participant self-reports of group visits and post views.

Conclusions
The findings of this study highlight the promise of a Facebook group–based intervention approach to engage Spanish-speaking Latino parents in health promotions related to COVID-19 and vaccination. The intervention participants experienced adequate exposure to the intervention content and engagement and reported overall positive reactions to the intervention. Future efforts should consider using this digital health promotion approach while integrating strategies to augment participants’ exposure to health promotion content; cultivating regular participant engagement, including using concise, audience-tailored messaging; and moderating the group with culturally aligned facilitators to foster trust, address misinformation, and promote child vaccine acceptance among parents.

Acknowledgments
This research was funded by the National Institutes of Health Community Engagement Alliance Against COVID-19 Disparities (NIH CEAL; grant 1OT2HL156812-01; multiple Principal Investigators: Mark Edberg and ELA), and the George Washington University Institute for Data, Democracy, and Politics (internal grant; multiple Principal Investigators: LCA and ELA).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Schedule of posts.

[XLSX File (Microsoft Excel File), 63 KB - formative_v8i1e51331_app1.xlsx]


on social media. Vaccines (Basel) 2023 Aug 09;11(8):1346 [FREE Full text] [doi: 10.3390/vaccines11081346] [Medline: 37631914]


54. Why are posts I've already seen still appearing in my Facebook feed? Facebook. URL: https://www.facebook.com/help/727563780642344/?helpref=sf_share [accessed 2023-07-17]


57. Barua Z. COVID-19 misinformation on social media and public’s health behavior: understanding the moderating role of situational motivation and credibility evaluations. Hu Arenas (Forthcoming) 2022 May 14 [FREE Full text]


Barriers to COVID-19 Vaccination in a Troop of Fleet Antiterrorism Security Team Marines: Observational Study

E Susanne Blazek¹, PhD; Amy Bucher¹, PhD
Behavioral Reinforcement Learning Lab (BReLL), Lirio, Knoxville, TN, United States

Corresponding Author:
E Susanne Blazek, PhD
Behavioral Reinforcement Learning Lab (BReLL)
Lirio
320 Corporate Drive
Knoxville, TN, 37923
United States
Phone: 1 865 839 5539
Email: sblazek@lirio.com

Abstract

Background: In 2019, the World Health Organization declared the reluctance to vaccinate despite the availability of vaccination services as one of the top 10 threats to global health. In early 2021, self-reported reluctance to vaccinate among military personnel might have been considered a significant threat to national security. Having a choice architecture that made COVID-19 vaccination optional rather than required for military personnel could have inadvertently undermined military readiness if vaccination uptake did not reach an acceptable threshold.

Objective: The purpose of this observational study was to examine Marines’ self-reported reasons for planning to decline the COVID-19 vaccine to understand their barriers to vaccination.

Methods: As the vaccination became available to 1 company of Fleet Antiterrorism Security Team (FAST) Marines in early 2021, company command required those planning to decline vaccination to write an essay with up to 5 reasons for their choice. These essays provided the data for this study. Qualitative descriptive analysis with elements from grounded theory was used to thematically categorize FAST Marines’ written reasons for planning to decline the COVID-19 vaccine into a codebook describing 8 key behavioral determinants. Interrater agreement among 2 qualitatively trained researchers was very good (κ=0.81).

Results: A troop of 47 Marines provided 235 reasons why they planned to decline the COVID-19 vaccine. The most frequent reasons were difficulty understanding health information (105/235, 45%), low estimates of risk (33/235, 14%), and fear of physical discomfort (29/235, 12%). Resulting interventions directly targeted Marines’ self-reported reasons by reducing barriers (eg, normalized getting the vaccine), increasing vaccine benefits (eg, improved access to base gyms and recreational facilities), and increasing nonvaccine friction (eg, required in writing 5 reasons for declining the vaccine).

Conclusions: Understanding the barriers military personnel experience toward COVID-19 vaccination remains critical as vaccine acquisition and availability continue to protect military personnel. Insights from subpopulations like FAST Marines can enhance our ability to identify barriers and appropriate intervention techniques to influence COVID-19 vaccination behaviors.

(JMIR Form Res 2024;8:e50181) doi:10.2196/50181

KEYWORDS
behavioral barriers; benefits; COVID-19; Marine Corps; military; vaccine reluctance

Introduction

In 2019, the World Health Organization declared the reluctance to vaccinate despite the availability of vaccination services as one of the top 10 threats to global health [1]. Since COVID-19 emerged in late 2019, COVID-19 vaccination specifically has been deemed a critical public health priority [2]. Vaccination is associated with reduced hospitalization and death rates following COVID-19 infection [3-5], as well as reductions in transmission and severity of infection [5,6]. Yet, it has proven challenging to achieve the desired rates of vaccination in the United States generally [7] and within specific subpopulations such as rural dwellers [8], political conservatives [9], and people with previous diagnoses of COVID-19 [10].
Combating uneven and inadequate uptake of COVID-19 vaccines requires an understanding of the behavioral determinants (ie, barriers and facilitators of the behavior) leading people to vaccinate or not [11]. Behavioral design frameworks, such as the COM-B (capability, opportunity, motivation, behavior) model of behavior change which suggests that capability, opportunity, and motivation are essential for behavior change, provide a categorization scheme for behavioral determinants that then allows interventionists to select techniques that target the specific barriers preventing people from completing a desired behavior [12-16]. While there are some behavioral determinants that may be addressable through directives from credible sources, as traditional public health campaigns may be described, these campaigns are unlikely to be broadly successful or sufficient in isolation to promote widespread behavior change. This is because they do not adequately account for the breadth of people’s barriers to performing the behavior or adequately enhance the benefits of the behavior [17,18]. Counteracting a crisis like the COVID-19 pandemic necessitates understanding how people make decisions, what gets in the way, and what gets people to take action [19]. Qualitative research approaches that yield richer insights might contribute to a better understanding of these nuances [20-23].

Barriers to COVID-19 vaccination in the general population include, among others, beliefs of low susceptibility to infection, perceived low severity of COVID-19, beliefs of low vaccine effectiveness, concerns about side effects, desires for more information, and not having explicit endorsement from a health care provider [24,25]. Not having received a vaccine in the past year is also associated with lower vaccination rates [26]. Barriers differ for population subgroups based on characteristics such as geographical location, work responsibilities, income level, educational attainment, and insurance status [25], several of which are social determinants of health. To effectively influence vaccination behavior within a population subgroup, researchers should seek to understand the specific barriers prevalent in that group.

One such group is military personnel and veterans. A specific study of the behavioral determinants of COVID-19 vaccination in military personnel is critical, as vaccination against infectious diseases supports operational readiness [27]. Moreover, given military veterans’ high rates of mental health issues and suicide, there has been some effort to consider veteran status as a social determinant of health [28]. A better understanding of the health decisions of military personnel therefore has the added potential of supporting improved health equity.

In early 2021, medical experts and government leaders considered the lack of COVID-19 vaccination among military personnel a potential threat to national security [29,30]. Unlike influenza, tetanus, diphtheria, pertussis, smallpox, and other vaccines, which are required aspects of force readiness, the COVID-19 vaccine was initially optional for military personnel before becoming mandatory [31] and finally becoming optional again on January 10, 2023 [32]. Having a choice architecture that makes COVID-19 vaccination optional for military personnel [33] could inadvertently undermine military readiness if vaccination uptake does not reach an acceptable threshold.

Therefore, it is important to understand the reasons why military personnel might not vaccinate to appropriately intervene on those barriers.

In the United States, military personnel are socialized to a specific set of values, including honor, bravery, and personal sacrifice [34]. Each branch of the US military then has characteristics that may influence the specific behavioral determinants of its troops toward vaccination. For example, within the United States Marine Corps, values such as the need for control and independence are reinforced, sometimes to the detriment of health-promoting behaviors [35,36]. Marines report a lower likelihood of vaccine receipt than other branches of service [37,38]. These differences may reflect true differences between the branches or, more likely, may reflect underlying differences in the characteristics of individuals choosing to serve in different branches.

Junior Marines (E1-E4) experiencing a high degree of external control and loss of agency due to service [39] might feel an enhanced need to establish autonomy when given the chance, such as by declining an optional vaccine. Research has found that military personnel ranked E1-E4 were the most likely to reject an avian influenza vaccine [37]. While the Marine Corps has embraced an approach of informed health care choice for its members, there is a rocky history of medical paternalism (eg, anthrax vaccination in the 1991 Gulf War, but see [40]) that may negatively influence modern Marines’ perceptions of vaccination campaigns.

Given their overall health and fitness and the risks that come with their jobs, Marines may also have a low perceived personal risk from the COVID-19 infection and a tolerance for risk that others may find uncomfortably high. Although data were limited at the time, it was projected in early 2021 that those who contracted COVID-19 had a 0.5%-5% chance of developing a critical illness requiring intensive care or resulting in death [41]. Those odds might have seemed acceptable to Fleet Antiterrorism Security Team (FAST) Marines, who are trained to provide security forces to guard high-value naval installations, most notably those containing nuclear vessels and weapons. FAST Marines find themselves in high-risk situations as part of their regular training, for example with live fire, extreme exposure to the elements, and sleep and nutrition deprivation. FAST Marines’ daily work likely undermines their perceived severity of contracting COVID-19. In contrast, the perceived severity of contracting anthrax among 22 Department of Defense (DOD) personnel in 2015 resulted in near 100% adherence to a complex emergency postexposure prophylaxis following accidental exposure to live Bacillus anthracis spores [42].

Americans’ willingness to get the COVID-19 vaccine declined from 72% in May 2020 to 51% in September 2020 [43] and rebounded to 60% in November 2020 [44]. Surveys from early 2021 found that fewer than half of military troops and military spouses planned to receive the vaccine [45]. At the time of the study, FAST Marines in Yorktown, Virginia, were in a privileged position: they were 1 of 12 bases nationwide to be offered the COVID-19 vaccine before widespread availability [46]. Yet an informal poll of Marines in 1 FAST company showed that only 10% planned to get the COVID-19 vaccine.
much lower than other estimates at the time [43,44]. The FAST Marines’ company command introduced key friction to those FAST Marines who planned to decline the COVID-19 vaccine: they had to write formal papers outlining up to 5 reasons why. This observational study makes use of those written data to understand FAST Marines’ barriers to COVID-19 vaccination.

Methods

Ethical Considerations
We followed the Standards for Reporting Qualitative Research [47]. This research was approved by an Exception Determination Official with the Naval Medical Center Portsmouth Research Subjects Protection Division. The study does not meet the Naval Medical Center Portsmouth Research Subjects Protection Division’s definition of research in accordance with 32 CFR 219.102 and DoDI 3216.02. Therefore, no written consent was obtained for the qualitative analyses. Study data were deidentified.

Sample
The FAST platoon is the core operational unit of the Marine Corps Security Force Battalion. There are 18 FAST platoons divided into 3 FAST companies that are based in Norfolk and Yorktown, Virginia. The battalion commander of 1 of the 3 FAST companies asked his Marines to either obtain the COVID-19 vaccine or to provide, in writing, at their own leisure but with a deadline, up to 5 reasons why they did not want the COVID-19 vaccine. One platoon of 47 male junior FAST Marines (ie, E1-E4) exclusively chose to provide written reasons for declining the COVID-19 vaccine. Their responses were collected by the battalion commander, entered into Microsoft Excel (Microsoft Corp) without demographic information, and shared with the researchers. Demographic data were not collected, and Marines did not receive compensation because the writing task was originally intended as a troop exercise and an intervention to promote COVID-19 vaccination, not a research project. Data were collected in February 2021. There were no efforts to retroactively identify which FAST Marine submitted which reasons or determine their demographics or ranks.

Qualitative Descriptive Analysis
The purpose of our study, to examine FAST Marines’ self-reported reasons for planning to decline the COVID-19 vaccine to understand their barriers to vaccination, guided our choice of qualitative descriptive analysis. Qualitative description is an appropriate approach to gathering insights on novel or poorly understood phenomena [20]. We conducted a rapid review to develop a codebook of empirically supported barriers to obtaining a COVID-19 vaccine that we coded the FAST Marines’ responses against. We used thematic analysis [48,49] with elements from grounded theory [50,51]. The 2 qualitatively trained researchers independently coded the responses in Excel. We calculated interrater agreement with Cohen’s $\kappa$ [52], which accounts for change agreements. The interrater agreement was very good ($\kappa=0.81$). See Table 1 for the final codebook consisting of 8 codes.
Table 1. The codebook used for qualitative descriptive analyses of 47 Yorktown, Virginia, Fleet Antiterrorism Security Team Marines’ 235 written reasons for why, in February 2021, they planned to decline the COVID-19 vaccine.

<table>
<thead>
<tr>
<th>Label</th>
<th>Definition</th>
<th>Description of how to know when code occurs</th>
<th>Keywords or phrases</th>
<th>Positive example (is this code)</th>
<th>Negative example (is not this code)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty understanding health information</td>
<td>Marines do not understand or misunderstand</td>
<td>Responses that inadequately represent (most likely due to honest lack of understanding) what is currently known about the vaccine’s ingredients, effectiveness, side effects, outcomes, etc.</td>
<td>FDA(^a), CDC(^b), mRNA(^c); effective; effectiveness; study; studies; too soon; early to know; immunity; side effects; takes years; not approved; not proven</td>
<td>“I’ve heard that [the vaccine] makes you very susceptible to sickness and weakens your immune system.”</td>
<td>“It is my right to refuse it”</td>
</tr>
<tr>
<td>Perceived low susceptibility of COVID-19 and its severity</td>
<td>Marines underestimate their risk of contracting COVID-19 and of a serious negative outcome.</td>
<td>Responses that suggest a low expected likelihood of contracting COVID-19 and of falling seriously ill.</td>
<td>No symptoms; young; healthy; likeliness; slim; survival rate; no danger</td>
<td>“I don’t have COVID and have never had a single symptom of it, so therefore I don’t need it.”</td>
<td>“Flu hasn’t affected me any different”</td>
</tr>
<tr>
<td>Concern over vaccine safety and potential side effects</td>
<td>Marines expect the vaccine or side effects will cause physical pain or discomfort.</td>
<td>Responses mentioning known and unpleasant side effects.</td>
<td>Needles; side effects; symptoms; allergic reaction; getting sick</td>
<td>“I don’t like needles”</td>
<td>“Can cause you to be sterile”</td>
</tr>
<tr>
<td>Status quo bias may impact preventive pandemic behaviors</td>
<td>Marines postpone getting the vaccine if doing so does not have an immediate benefit for them.</td>
<td>Responses mentioning drawbacks to getting vaccinated and no benefits.</td>
<td>No benefit; not a requirement; voluntary; wear mask; socially distance</td>
<td>“Not a requirement for the Marine Corps or deployment.”</td>
<td>“I’ve already received a flu shot”</td>
</tr>
<tr>
<td>Distrust in health care</td>
<td>Marines distrust the health care system in general or have had bad experiences in health care settings.</td>
<td>Responses mentioning trust in the health care system, in the CDC, and in the vaccine.</td>
<td>Trust; Pfizer; questionable history; rushed; wrong about the virus</td>
<td>“I don’t trust it at all”</td>
<td>“I’d rather wait and see how people’s body respond to it”</td>
</tr>
<tr>
<td>Social norms may influence attitudes and beliefs about vaccines</td>
<td>Marines have no role models—family members, community members, etc—who are seen endorsing the behavior.</td>
<td>Responses indicating that trusted friends, family, and health care providers are not getting vaccinated.</td>
<td>Family; peers; Marines; doctor; friends</td>
<td>“Other marines I trust aren’t getting it”</td>
<td>“Side effects are negative”</td>
</tr>
<tr>
<td>Believing in COVID-19 conspiracies</td>
<td>Marines have strong, irrational, and disproven beliefs against the vaccine.</td>
<td>Responses that indicate believing in nonvalidated, mis- or disinformation about the COVID-19 vaccine.</td>
<td>Test dummy; guinea pig; political gain; election time; media; democrats</td>
<td>“Don’t want to be a test dummy for the military.”</td>
<td>“The vaccine is not guaranteed to work”</td>
</tr>
</tbody>
</table>

\(^{a}\) FDA, Food and Drug Administration, \(^{b}\) CDC, Centers for Disease Control and Prevention, \(^{c}\) mRNA, messenger RNA
Results

Frequencies of coded responses can be found in Table 2. FAST Marines’ reasons for planning to decline the COVID-19 vaccine were categorized into the following 7 barriers: difficulty understanding health information (105/235, 44.7%), perceived low susceptibility to COVID-19 and its severity (33/235, 14%), concern over vaccine safety and potential side effects (29/235, 12.3%), status quo bias may impact preventive pandemic behaviors (28/235, 11.9%), distrust in health care (12/235, 5.1%), social norms may influence attitudes and beliefs about vaccines (12/235, 5.1%), and believing in COVID-19 conspiracy theories (5/235, 2.1%). The last 4.7% (11/235) of responses were categorized as other or unclear.

Table 2. Summary of the behavioral determinants of COVID-19 vaccination uncovered by qualitative descriptive analyses of 47 Yorktown, Virginia, Fleet Antiterrorism Security Team Marines’ 235 written reasons for why, in February 2021, they planned to decline the COVID-19 vaccine.

<table>
<thead>
<tr>
<th>Determinant</th>
<th>Values (n=235), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty understanding health information</td>
<td>105 (44.7)</td>
</tr>
<tr>
<td>Perceived low susceptibility to COVID-19 and its severity</td>
<td>33 (14)</td>
</tr>
<tr>
<td>Concern over vaccine safety and potential side effects</td>
<td>29 (12.3)</td>
</tr>
<tr>
<td>Status quo bias may impact preventive pandemic behaviors</td>
<td>28 (11.9)</td>
</tr>
<tr>
<td>Distrust in health care</td>
<td>12 (5.1)</td>
</tr>
<tr>
<td>Social norms may influence attitudes and beliefs about vaccines</td>
<td>12 (5.1)</td>
</tr>
<tr>
<td>Believing in COVID-19 conspiracy theories</td>
<td>5 (2.1)</td>
</tr>
<tr>
<td>Other or unclear</td>
<td>11 (4.7)</td>
</tr>
</tbody>
</table>

A total of 45% (105/235) of all responses (each FAST Marine gave up to 5), and 3 times as many as the second largest barrier, had to do with difficulty understanding medical information. In fact, 96% (45/47) of FAST Marines—all except 2 FAST Marines—indicated difficulty understanding medical information, ranging from the inability to understand basic medical information to the inability to grasp nuances in the medical literature.

Discussion

Overview

A qualitative descriptive analysis of 235 reasons for planning to decline the COVID-19 vaccine provided a contemporaneous understanding of the barriers to vaccination experienced by an infrequently studied population subgroup: junior FAST Marines. FAST Marines’ barriers to COVID-19 vaccination included low health literacy, perceived low susceptibility to COVID-19 and its severity, concern over vaccine safety and potential side effects, and other less frequently reported barriers. These findings are consistent with work on Air Force personnel who self-identified as vaccine-hesitant and indicated concerns about efficacy, side effects, and vaccine-induced illness as top barriers to vaccination [61], while adding the insight that low health literacy may be associated with such concerns.

Our findings extend previous work on barriers to COVID-19 vaccination in part because we captured in-the-moment attitudes, beliefs, and cognitions about the COVID-19 vaccine in a less frequently studied population subgroup: junior FAST Marines. The barriers described in the written responses were not subject to the same demand characteristics as barriers elicited through formal research. That is, FAST Marines were not restricted to a preselected list of barriers to COVID-19 vaccination as might have been the case in survey research, nor were they aware of a research framing that might have prompted more socially acceptable responses.

Impact on Future Interventions Suggested by Findings

Fear is a common, if not primary, motivator used by Marine Corps leadership. Instead of fear-based messaging, which might induce reactance [62,63], and based on our results from the qualitative descriptive analysis, we recommended to the company command placing emphasis on the fact that because FAST Marines are young and healthy, they must get the COVID-19 vaccine. This kind of messaging emphasizes resilience and self-efficacy [55,63] and builds on FAST Marines’
esprit de corps. Future research should evaluate building upon the positive and strong military identity [34] to promote health behaviors in military personnel.

What makes Marines great at their jobs might get in the way of success elsewhere. That is, feeling invincible could positively support junior Marines in performing their job but could become a barrier when they decide whether or not to get the COVID-19 vaccine. Our results do not suggest or endorse changing traits or skills that make Marines great at their jobs. Instead, the results suggest using that information to better speak the Marines’ language and motivate them to get vaccinated. In this case, intervention design should not treat some of these qualities as barriers to overcome so much as consider them to be facilitators to be leveraged.

The study results provided insights for behavioral intervention at the troop and company levels. Company command developed interventions that directly targeted FAST Marines’ self-reported reasons by increasing nonvaccine friction (eg, required in writing 5 reasons for declining the vaccine), reducing barriers (eg, normalized getting the vaccine by encouraging senior personnel to get vaccinated), and increasing vaccine benefits (eg, provided fringe benefits to vaccinated FAST Marines such as the higher likelihood of approval for out-of-area leave).

Limitations and Directions for Future Research
The primary limitation stems from the nature of the data. The 235 written reasons to decline the COVID-19 vaccine were intended as a troop exercise and not a scientific study. We were unable to explore demographics known to be determinants of COVID-19 vaccination because of the deidentified data.

The second limitation involves our inability to determine our results’ direct impact on COVID-19 vaccination in the troop of FAST Marines. Company command introduced a number of behavioral interventions following this study but implemented them inconsistently across the troops within the company. The goal of managing COVID-19 transmission supersedes scientific rigor. It is therefore difficult to determine why only 10% (11/114) of FAST Marines in the company expressed intentions to accept the COVID-19 vaccine in January of 2021 but, 8 weeks later, 84% (105/114) of the company had chosen to receive the vaccine. Some FAST Marines likely chose to vaccinate due to the interventions implemented by their leadership to address their barriers, but it is also likely that the intention-behavior gap accounts for some of the discrepancy [23]. Future research on vaccine campaigns in the military should explore the impact of behavioral interventions like the ones implemented here and should also capture how vaccines are administered: with informed choice, under time pressure, etc (see Murphy et al [40]).

Conclusion
Understanding the barriers military personnel experience toward COVID-19 vaccination remains critical despite the success of COVID-19 vaccination to date, with 96% of US military personnel fully vaccinated as of January 2023 [32], compared to roughly 70% for all Americans [64]. It is hard to overstate the impact of even small wins when it comes to the health of military personnel. Vaccine acquisition and availability continue protecting military personnel, and the DOD currently offers 17 different vaccines for the prevention of infectious diseases among military personnel, where appropriate [65]. The historical and cultural context of the Marine Corps, and specifically the Marine Corps’ FAST, likely gave rise to a particular constellation of barriers to the COVID-19 vaccine. Identifying these barriers early and as the COVID-19 vaccine was made available helped shape how a FAST company command successfully ensured vaccine participation. With the return to a choice architecture that makes ongoing COVID-19 vaccination voluntary, coupled with the knowledge that ongoing boosters will be necessary to maintain resistance to infection, military leadership has ample opportunity to continue to offer behavioral interventions tailored to their troops’ needs.

Acknowledgments
This paper was not commissioned or requested by the United States Marine Corps or the Department of Defense and should not be construed as an endorsement of Lirio’s services by either organization. The views expressed in this study are those of the authors and do not necessarily reflect the official policy or position of the Marine Corps, the Department of Defense, or the US government. We would like to acknowledge the invaluable support of our second qualitative coder, Signe M Spencer.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to a lack of ethics approval to share.

Conflicts of Interest
ESB and AB are paid employees of Lirio, which developed the intervention studied.

References


Abbreviations

**COM-B:** capability, opportunity, motivation, behavior

**DOD:** Department of Defense

**FAST:** Fleet Antiterrorism Security Team

---

Edited by A Mavragani; submitted 21.06.23; peer-reviewed by G Goldman, B Eshrati; comments to author 01.11.23; revised version received 21.11.23; accepted 19.02.24; published 19.03.24.

Please cite as:
Blazek ES, Bucher A
Barriers to COVID-19 Vaccination in a Troop of Fleet Antiterrorism Security Team Marines: Observational Study
JMIR Form Res 2024;8:e50181
URL: https://formative.jmir.org/2024/1/e50181
doi:10.2196/50181
PMID:38502179
©E Susanne Blazek, Amy Bucher. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
A Snapshot of COVID-19 Vaccine Discourse Related to Ethnic Minority Communities in the United Kingdom Between January and April 2022: Mixed Methods Analysis

Nazifa Ullah¹, MBBS, BSc; Sam Martin², LLB, MSc, PhD; Shoba Poduval³, MBBS, PhD

¹Research Department of Primary Care & Population Health, University College London, London, United Kingdom
²Vaccines and Society Unit, Oxford Vaccine Group, University of Oxford, Oxford, United Kingdom
³Institute of Health Informatics, University College London, London, United Kingdom

Corresponding Author:
Shoba Poduval, MBBS, PhD
Institute of Health Informatics
University College London
222 Euston Road
London, NW1 2DA
United Kingdom
Phone: 44 (0)20 3549 5969
Email: s.poduval@ucl.ac.uk

Abstract

Background: Existing literature highlights the role of social media as a key source of information for the public during the COVID-19 pandemic and its influence on vaccination attempts. Yet there is little research exploring its role in the public discourse specifically among ethnic minority communities, who have the highest rates of vaccine hesitancy (delay or refusal of vaccination despite availability of services).

Objective: This study aims to understand the discourse related to minority communities on social media platforms Twitter and YouTube.

Methods: Social media data from the United Kingdom was extracted from Twitter and YouTube using the software Netlytics and YouTube Data Tools to provide a “snapshot” of the discourse between January and April 2022. A mixed method approach was used where qualitative data were contextualized into codes. Network analysis was applied to provide insight into the most frequent and weighted keywords and topics of conversations.

Results: A total of 260 tweets and 156 comments from 4 YouTube videos were included in our analysis. Our data suggests that the most popular topics of conversation during the period sampled were related to communication strategies adopted during the booster vaccine rollout. These were noted to be divisive in nature and linked to wider conversations around racism and historical mistrust toward institutions.

Conclusions: Our study suggests a shift in narrative from concerns about the COVID-19 vaccine itself, toward the strategies used in vaccination implementation, in particular the targeting of ethnic minority groups through vaccination campaigns. The implications for public health communication during crisis management in a pandemic context include acknowledging wider experiences of discrimination when addressing ethnic minority communities.

(JMIR Form Res 2024;8:e51152) doi: 10.2196/51152

KEYWORDS
COVID-19; ethnic minorities; vaccine; hesitancy; social media; discourse; minority groups

Introduction

In December 2020, the United Kingdom began its COVID-19 vaccination program [1]. Population-level studies have demonstrated persistently lower levels of uptake of the first dose of the COVID-19 vaccine among all minority groups compared with the White British ethnic group, and these differences widened over time [2]. Vaccine hesitancy, that is, the delay or refusal of vaccinations despite the availability of services [3], has been recognized as a significant contributing factor to the unequal uptake of COVID-19 vaccines across
different ethnic groups [4]. This contrasts with anti-vaxxers, defined as a group of individuals who may refuse most if not all vaccinations [5].

Several studies have investigated the causes of hesitancy within the context of the COVID-19 vaccine [6-8]. A summary provided by the World Health Organization 3 C’s model identifies the following 3 core areas that influence hesitancy: confidence (toward the vaccine or its provider), complacency (perceived risk toward vaccine-preventable diseases), and convenience (access to vaccines) [8].

Historical mistrust toward governments, racism within the health care system, and lack of diversity in medical research are further points in the literature noted as contributing factors to hesitancy specifically in relation to ethnic minority groups [5,9-11].

Among the many factors noted, the rising use of social media as an interactive health ecosystem fuelling vaccine hesitancy remains a recurrent concern contributing to vaccine confidence, as identified in the literature [12]. During the COVID-19 pandemic, there was noted to be the issue of an “infodemic,” that is, the spread of large volumes of information including false or misleading information [13,14]. Social media played a significant role in this infodemic via the rapid dissemination of information. A total of 49% (n=980) of respondents to an online survey in the United Kingdom used social media as a source of information about COVID-19 [15] and Twitter reported COVID-19–related posts every 45 milliseconds in 2020 [16-18].

Social media was a prominent source of misinformation during the pandemic, particularly in relation to vaccinations. Misinformation refers to false or inaccurate information shared unknowingly and without the intention to cause harm [19].

Studies have highlighted the correlation between exposure to misinformation on social media and both vaccine hesitancy and reduced compliance to public health measures and disease prevention activities such as vaccinations [20-22]. Loomba et al [21] conducted a randomized control trial in the United Kingdom and the United States quantifying exposure to misinformation related to COVID-19 vaccines on social media platforms and vaccination intent and found that exposure to misinformation resulted in a 6.2% decline in the number of respondents who would “definitely” take the vaccine in comparison to control groups. Betsch et al [20] found that even brief exposure to vaccine-critical websites would increase an individual’s overall perception of vaccine risk in comparison to exposure to control websites. Furthermore, negative information related to vaccinations tends to encourage greater user engagement in comparison to positive content [23,24]. This may be due to the persuasive narratives and powerful imagery often used in antivaccination content [22]. Another important note is the creation of “echo chambers” on social media, that is, the network of users in which each user encounters beliefs supporting pre-existing opinions without being exposed to opposing viewpoints [12]. Such effects further reinforce antivaccine perspectives by connecting like-minded individuals thereby amplifying antivaccine narratives and potentially dampening vaccination attempts [7].

Despite the clear role of social media as a source of information during a pandemic and its impact on hesitancy and consequently vaccination attempts, there is little published data on its role in the public discourse on COVID-19 vaccines among ethnic minority groups in the United Kingdom. For ethnic minority communities, there may be a greater reliance on social media as a source of information because of barriers in accessing health care and health information due to, for example, language barriers or poor health literacy [25]. Dickson et al [26] study looking at culturally and linguistically diverse communities found that individuals would often receive information by “word of mouth,” that is, from peers from the same cultural groups as these would be seen as a trusted source of information. Those peers themselves would receive information from social media platforms such as Facebook and WhatsApp [26]. It is for these reasons that understanding hesitancy in minority communities through the lens of social media is of paramount importance.

Our article aims to address the following research question by obtaining data from social platforms Twitter and YouTube between January and April 2022:

1. What was the discourse related to and within ethnic minority communities on social media platforms?
2. What were the general sentiments and stances of social media posts analyzed?
3. What were the most frequent topics of conversation seen on social media?

Methods

Study Design

The study analyzed Twitter and YouTube comments using a mixed method approach. This approach is often used when analyzing social media posts due to the large quantity of data that allows for both quantitative exploration as well as qualitative analysis of the contents, further enhancing the understanding of a research topic [27].

Data Source

We obtained publicly available data from Twitter and YouTube comments in the United Kingdom between January 17, 2022, and April 7, 2022, using the software Netlytics (for Twitter posts) and YouTube Data Tools (for YouTube data) [28,29]. The platforms Twitter and YouTube were selected because they provided the most openly available application programming interfaces (APIs) [16]. APIs are mechanisms that enable 2 software components to communicate with one another and allow third-party developers to access data such as tweets and YouTube comments [30].

Drawing on prior research studies on social media and the COVID-19 vaccine [6-8], a pilot Boolean search strategy was created (Multimedia Appendix 1) and applied to Netlytics and YouTube Data Tools in order to understand recurrent topics of discussion around the vaccine. As of April 7, 2022, which was the end date for our data retrieval, the UK government was in the midst of its booster vaccination rollout and proposed that all eligible adults older than 18 years would be offered the booster vaccine [31]. The government had also proposed mandatory vaccinations for National Health Service workers during this period, a proposal that was later reversed by January 31 [31,32].
Consequently, our preliminary search of the data informed the development of a second more focused Boolean search (Multimedia Appendix 1) containing keywords and relevant hashtags such as those related to booster vaccination campaigns. All posts were screened by NU according to our inclusion and exclusion criteria (Textbox 1) and duplicates were removed. To note, the platform Netlytics provided geo-coded social media data, enabling us to ensure that tweets analyzed are from those based in the United Kingdom [28]. Geo-coded data may be provided through the Netlytics filtering system that allows users to include or exclude tweets based on a given radius and latitude

**Textbox 1. Inclusion and exclusion criteria.**

**Inclusion criteria**
- Posts where ethnic minority groups are discussed or mentioned (all ethnic groups except White British groups). For example use of the terms: (“BAME” OR “BME” OR “Black” OR “Asian” OR “minority ethnic” OR “ethnic minority” OR “minority” OR “non-white” OR “raci*” OR “MinorityHealth” OR “Minority” OR “race”; see Multimedia Appendix 1 for full list).
- Posts referring specifically to COVID-19 vaccination. For example use of the terms: (“vaccin*” OR “immunis*” OR “vax*” OR “jab*” OR “covidvaccin*” OR “COVIDVaccine” OR “COVID19” OR “COVID19Vaccine” OR “CovidVaccine” OR “covid-19” OR “covid vaccine,” OR “covid vax”).
- Example of a hypothetical post to include: “Racism and historical injustices fuelled low COVID vaccine uptake by minorities.” We would include this as it refers specifically to COVID-19 vaccinations in relation to ethnic minority groups.

**Exclusion criteria**
- Posts that discussed the COVID-19 vaccine without mention of ethnic minority groups.
- Posts that do not mention the COVID-19 vaccine.
- Posts by individuals not based in the United Kingdom.
- Example of hypothetical post to exclude: “COVID has caused many deaths.” This would be excluded as there is no mention of the COVID-19 vaccines and minority groups.

**Measures of Variables**

**Themes**
In order to identify key themes of topics in our social media posts, we developed an initial analytic coding framework (Multimedia Appendix 1) that was informed by previous literature investigating causes of hesitancy as discussed previously [6-11]. We added the coding framework to a Microsoft Excel spreadsheet, with codes in the columns and social media posts entered as individual comments in the rows. The framework was refined during team discussions and all authors applied the same framework to the data.

**Sentiment and Stance Coding**
In order to appreciate the nuanced opinions on social media and how they may be contributing to positive or negative attitudes toward COVID-19 vaccines [37], individual posts were manually coded into sentiment and stance categories (Multimedia Appendix 1). Sentiment refers to the overall tone of a post and the categorization of opinions toward a subject expressed by the author (being positive, negative, neutral, or ambiguous) [37,38]. Whereas stance refers to the process of determining the author’s attitude or stance toward a target [37]. For example, a post may have a positive sentiment toward anti-vaxxers but a negative stance toward vaccines.

Our stance framework (Multimedia Appendix 1) was informed by the Leask et al [39] categorization of vaccine intention. Hypothetical examples derived from our social media data and the sentiment and stance assigned to each text can be found in Multimedia Appendix 1. We coded the sentiment and stance of the data and added both to our Microsoft Excel spreadsheet. For example, a positive sentiment post would be coded as P and a positive stance post would be coded PS (codes clarified in Multimedia Appendix 1). Any disagreements regarding the classification of the social media posts were discussed during team meetings where a unanimous decision was made regarding their categorization.

**Data Analysis Procedure**
The percentage of Twitter and YouTube posts assigned to each sentiment and stance category was calculated and posts coded according to themes derived were quantified for further analysis. Where a post was assigned more than 1 theme, they were calculated as separate entities, for example, if a post was assigned to themes 1 and 2, they would be included separately when calculating overall percentages.

Taking an inductive approach, we collated the results from our analysis and a random sample of posts (30 tweets and 20 YouTube comments) were selected for a more detailed analysis and qualitative interrogation of the data to address our overall aims [40,41]. Our thematic analysis framework was updated...
accordingly to ensure themes identified were grounded in the data obtained.

**Network Analysis and Visualization**

Network analysis was also carried out to highlight the most prevalent topics of discussion on social media. Text network analysis software Infranodus (Nodus Labs) was used [42]. Data were imported into Infranodus and semantic networks were generated with data organized by specific topics and subtopics. Word clusters were derived using modularity measurement tools to highlight the most common topics of conversations within a group of tweets or YouTube comments. Keywords obtained using betweenness centrality, an analysis of connections between subtopics or words that link different clusters of conversations together, helped deepen our understanding of influential words or topics that may be linking clusters of conversations together. Word frequency analysis was then performed to highlight the most frequent and weighted subtopics of conversations.

**Ethical Considerations**

Ethics and data protection approval were obtained from University College London’s (UCL’s) Research Ethics Committee (ID 21773/001) and UCL Data Protection Officer (registration number: Z6364106/2021/10/60). All data obtained were publicly available via YouTube’s and Twitter’s APIs.

**Results**

**Overview**

Using our search strategy, a total of 44,144 Twitter posts and 9 YouTube videos were identified. After the deduplication process and screening according to inclusion and exclusion criteria, a total of 416 deduplicated social media posts were identified. This included 260 tweets (that were retweeted 15,331 times) and 156 comments from 4 YouTube videos that were included in our analysis. Video 1 was a documentary exploring whether ethnic minorities and in particular minority health care workers, should be prioritized in having the COVID-19 vaccine. Video 2 was a debate on a news channel looking at the causes of vaccine hesitancy specifically in minority communities. Videos 3 and 4 were documentaries that focused on similar topics and explored the causes of vaccine hesitancy within ethnic minority communities. Out of the 4 videos analyzed, video 2 had the highest number of total comments of 1273.

**Themes**

The following 3 themes were identified in the discourse on both Twitter and YouTube: concerns related to vaccination implementation, questions over vaccine development and safety, and wider systemic issues such as institutional racism and historical abuse of power.

Concerns related to vaccine implementation were mostly related to mandatory vaccinations for health care workers and the wider public and public health messaging, in particular the use of the phrase “BAME.” Most of the topics connected to vaccine development were related to the speed of development of the vaccine and its side effects. Comments alluding to conspiracy theories suggesting the vaccine to be gene-altering, experimental in nature, and causing infertility as a side effect were particularly seen on YouTube. Moreover, systemic racism, that is, a form of racism that is embedded within the laws and regulations of a society [43], was a prevalent topic noted under the theme of wider systems. The topic of racism was particularly prevalent when discussing the strategies of vaccine rollout such as mandatory vaccinations for health care workers and the focus on vaccinating ethnic minorities. Table 1 summarizes the themes noted in our discourse analysis with hypothetical examples of tweets and YouTube comments derived from our data. Hypothetical examples are given in order to abide by ethics guidelines preventing any identifiable information from being published.

<table>
<thead>
<tr>
<th>Themes or topics of discussion</th>
<th>Examples of tweets</th>
<th>Example of YouTube comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vaccination implementation</strong></td>
<td>“There is a witch hunt against black and Asian NHS workers who don’t want the vaccine”</td>
<td>“Why are we obsessing over BAME groups. It’s just another form of segregation and racism.”</td>
</tr>
<tr>
<td></td>
<td>“We need to stop this race baiting by focusing all the time on BAME”</td>
<td></td>
</tr>
<tr>
<td><strong>Vaccine development and safety</strong></td>
<td>“The vaccine is an experiment aimed at wiping out humanity”</td>
<td>“The vaccine is just part of a scheme to reduce the population of minorities. Look at the testing they have done on African women causing infertility”</td>
</tr>
<tr>
<td></td>
<td>“Stay away from the jab, it’s likely to cause autism in Black people”</td>
<td>“The vaccine isn’t safe, it was made too quickly”</td>
</tr>
<tr>
<td></td>
<td>“You’re less likely to die from having the vaccine than from Covid- vaccines undergo a lot of testing to rule out long term side effects”</td>
<td></td>
</tr>
<tr>
<td><strong>Wider systemic issues</strong></td>
<td>“It’s not just about mistrust towards institutions but health racism is a real issue too.”</td>
<td>“Remember the Tuskegee trials? They are just trying to use Black and Asian people as guinea pigs now”</td>
</tr>
</tbody>
</table>

**Sentiment and Stance**

Table 2 demonstrates the sentiment and stance of tweets and YouTube comments as a percentage. In general, YouTube depicted a higher percentage of negative sentiment and stance posts over Twitter.

---

https://formative.jmir.org/2024/1/e51152
Network analysis identified the following 4 key most frequent topics of conversations on Twitter, portrayed as clusters of tweets: booster vaccinations (cluster 1), mistrust (cluster 2), vaccination risk (cluster 3), and racism (cluster 4). Cluster 1 was the most popular group of conversations, with 23% (n=60) of tweets containing the keywords “vaccine,” “covid,” and “uptake.” Most tweets were related to encouraging booster vaccinations in minority communities. Some comments were noted addressing issues around misinformation and hesitancy in an endeavor to encourage booster vaccinations. Cluster 2 of tweets contained keywords, “trust,” “pandemic,” and “work.” There were discussions related to vaccinations for health workers as well as topics related to health inequalities in minority communities and historical abuses of power. Cluster 3 contained the keywords “jab,” “risk,” and “explain,” with conversations focused on the side effects of vaccines. Cluster 4 contained keywords “black,” “people,” and “race,” paying particular attention to the uptake of vaccines within the Black African and Caribbean community.

The word “vaccine” had the greatest betweenness centrality of 0.71, meaning that it was this phrase that predominantly linked the different clusters of conversations in the network, that is, the distribution of tweets, together. This is followed by “minority,” “covid,” and then “black.” The network structure, that is, the network of comments that people have made around a specific topic and their relation to one another, had a focused influence distribution of 80% with a modularity measure of 0.28. This illustrates that although there were several conversations related to the vaccine, the conversations were focused on a small number of key themes discussed above. Further statistics discussing the average degree and weighted betweenness between nodes can be found in the Multimedia Appendix 2. Further images of individual clusters can be found in Multimedia Appendix 3.

Regarding YouTube, video 1 keywords found in the most popular cluster of conversations included “race,” “worker,” and “baiting.” “Worker,” in this context was related to debates on whether ethnic minority health care workers should be prioritized in having the vaccine. Most conversations were hence related to racial tension and worries that vaccine campaigns prioritizing specific ethnic minority communities would seed further division. Video 2 had a similar focus in that the main topic of conversation was related to critiquing the vaccination program and the focus on ethnic minorities. Both videos 1 and 2 had a focused network structure reflecting that opinions expressed were similar in nature, hence a focused structure.

Video 3 had the most popular cluster of conversations containing the keywords “time,” “treatment,” and “American.” This video had multiple mentions of the Tuskegee trial, an experiment from 1932 to 1972 that involved the unethical testing of over 400 Black men who were falsely led to believe that they were being treated for syphilis [44]. Video 4 contained similar clusters of conversations related to mistrust. Both videos had a diverse network structure due to the various causes of vaccine hesitancy discussed in the comments sections. Multimedia Appendix 4 provides examples of conversation clusters in Videos 1 and 3 related to themes of racism and mistrust.

**Discussion**

**Principal Findings**

Our study aimed to provide an overview of the discourse related to and within ethnic minority communities on the social media platforms Twitter and YouTube. This is one of the few studies to our knowledge that has explored the use of social media in this manner specifically related to ethnic minority communities. The summary of our results can be found in Table 3. The themes identified were intrinsically linked with one another. For example, concerns raised around vaccination strategy such as the use of the term “BAME” were associated with themes related to wider systemic issues such as racism. The higher percentage of negative sentiment and stance seen on YouTube may be attributed to the fact that on YouTube, there is less of a focus on the individual profile page giving rise to the perception of greater anonymity when posting negative statements [45]. However, the nature of the YouTube videos may have also contributed to the higher levels of negative posts. For example, video 1 explored the potential public health policy of prioritizing minority groups for vaccination, a controversial debate that was not mentioned in the Twitter data we had collected. Naturally,

<table>
<thead>
<tr>
<th>Stance</th>
<th>Twitter, n (%)</th>
<th>YouTube, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>61 (23.5)</td>
<td>12 (7.7)</td>
</tr>
<tr>
<td>Neutral</td>
<td>130 (50)</td>
<td>79 (50.6)</td>
</tr>
<tr>
<td>Stance</td>
<td>YouTube, n (%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>69 (26.5)</td>
<td>18 (11.5)</td>
</tr>
<tr>
<td>Neutral</td>
<td>109 (41.9)</td>
<td>16 (10.3)</td>
</tr>
</tbody>
</table>
with such contentious issues, greater negative comments could be seen as users shared their opinions via the comment section.

### Table 3. Summary of results*

<table>
<thead>
<tr>
<th>Codes grouped to form themes and topics</th>
<th>Percentage of Twitter posts assigned, n (%)</th>
<th>Percentage of YouTube posts assigned, n (%)</th>
<th>Keywords from Twitter and YouTube obtained from network analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vaccine implementation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Booster vaccinations</td>
<td>301 (60.2)</td>
<td>110 (44.9)</td>
<td>Vaccine, COVID, and uptake</td>
</tr>
<tr>
<td>• Communication strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Control measures, for example, manda-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tory vaccinations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vaccine development and side effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Speed of development</td>
<td>15 (3)</td>
<td>18 (5.3)</td>
<td>Jab, risk, and explain</td>
</tr>
<tr>
<td>• Side effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wider systems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Systemic racism</td>
<td>184 (36.7)</td>
<td>122 (49.8)</td>
<td>Black, people, race, trust, pandemic, work, worker, and bai-</td>
</tr>
<tr>
<td>• Mistrust toward institutions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Historical abuses of power</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Where a post was assigned more than 1 theme, they were calculated as separate entities, for example, if a post was assigned to themes 1 and 2, they would be included separately when calculating overall percentages.

Our paper contributes to the existing literature by highlighting that one of the predominant and most popular topics of conversations within ethnic minority communities during the 3 months of data analyzed, was related to the implementation of the vaccination program, particularly communication strategies adopted during part of the booster vaccine rollout. The focus on ethnic minority communities and alienation of these groups through terms such as “BAME,” in vaccination campaigns, has been highlighted from our data as being divisive and causing further racial division. Policies such as mandatory vaccinations for health workers were deemed as particularly discriminatory toward minority workers. In fact, it seemed to have further propelled discussions related to trust or mistrust toward institutions on social media. Few papers explicitly highlight this connection between vaccination implementation strategies propelling concerns related to racism and fuelling further mistrust toward institutions specifically within minority communities.

There is an acknowledgment in the literature about the importance of adopting appropriate communication strategies in public health messaging. The Commission on Race and Ethnic Disparities has acknowledged the generalization that terms such as “BAME,” can cause [46]. Consequently, recent government guidelines have recommended using the terms “ethnic minorities” or “people from ethnic minority backgrounds” in communication [46]. Coccia’s [47] study looking at the maximum level of COVID-19 vaccinations that could be achieved across 150 countries without social impositions, for example, restrictions on public gatherings and government lockdowns, found the maximum level to be 70% based on normal hesitancy in society. In order to increase the number of vaccinated above this threshold, communicating effectively using “humble inquiry, compassionate listening and storytelling,” is a more effective approach than implementing strict health policies that impact individual freedoms [47].

Our study depicted how opinions toward vaccination strategies were linked to concepts related to trust and mistrust toward institutions. Mistrust toward institutions contributing to hesitancy in minority communities has already been noted by previous studies [5,7]. A systematic review looking at factors influencing vaccination uptake in minority communities found 6 studies that attributed lower uptake to “mistrust including pre-existing lower scientific or medical trust, conspiracy suspicions and attitudes” [48].

Moreover, our paper suggested that mistrust was mostly directed toward governments and pharmaceutical companies. These findings are echoed in a study looking at survey data from 100 countries and the roles of different forms of trust in predicting vaccine hesitancy [49]. Trust in political institutions was a consistent predictor of vaccine hesitancy [49]. The impact of a lack of trust in governments in vaccination attempts is exemplified in Romania’s COVID-19 vaccination strategies. Romania has one of the lowest confidence rates in their national government and this played a significant role in the low COVID-19 vaccination rates seen [50]. Analysis of comments from Romania’s #storiesfromvaccination campaign, found politically independent sources, for example, from health experts and laypersons, were deemed as being legitimate sources of information and consequently more trustworthy [50].

The continued concerns over vaccine development and side effects identified in our data are consistent with other studies investigating the causes of hesitancy in ethnic minority communities [11]. Fertility was a side effect most often noted to be of concern on YouTube and this is further supported by existing literature. A qualitative study of 12 participants done in 2021 exploring vaccine hesitancy in ethnic minority communities found infertility as a prevalent theme, as well as concerns related to period irregularities and breastfeeding [4]. Conspiracy theories suggesting the vaccine to be gene-altering...
or experimental in nature were also reported in our results. The spread of such theories could have a detrimental effect on vaccination efforts as studies have suggested that vaccination conspiracy belief was the most relevant predictor in willingness to get vaccinated (i.e., a negative association) [51,52].

**Strengths and Limitations**

The findings of our paper must be considered in light of the limitations of our study. First, our data extraction was limited by Twitter and YouTube API policies that restrict the amount of data that can be extracted over a certain period of time [28,29]. Moreover, our data set was small with a total of 416 social media posts analyzed. However, a rigorous qualitative analysis was conducted with multiple steps taken to validate our analysis as described in the methodology. Furthermore, the demographic of individuals using Twitter and YouTube tends to represent a younger and more politically engaged population [53]. Consequently, the data obtained may be biased toward this population. However, there are few studies looking at social media as data and by capturing the opinions of a younger population, a greater understanding may be gained of the view points behind a traditionally low uptake group [7]. Moreover, it is difficult to be conclusive about our sentiment and stance analysis since we were limited by looking at each individual tweet and YouTube comment as isolated posts rather than in the context in which the comment was made (i.e., as part of a thread of posts). This was to ensure our methodology of analysis remained consistent and prevented us from making assumptions about the users’ beliefs or opinions.

Much of the literature looking at social media data tends to focus on Twitter, although few on YouTube, due to the ease of accessing Twitter data when compared with other social media platforms [37]. Hence, we acknowledge that this study is not representative of all social media activity on this topic. Data from other social media sites such as Facebook, Instagram, and TikTok, remain underexplored, and more research is needed looking at such platforms and their contribution to vaccine hesitancy.

**Implications and Future Directions**

Our findings suggest that when planning communication strategies for future public health interventions such as vaccination, policymakers and health practitioners must acknowledge the wider experiences of individuals, particularly in minority groups. Public health messaging around the COVID-19 vaccine has placed the emphasis on ethnic minorities to become less hesitant and more trusting rather than acknowledging the systemic racism and experiences of discrimination raised by individuals [54]. Future vaccination strategies and public health messages targeted toward minority groups must make greater concerted efforts to acknowledge the historical abuses of power and contextualize hesitancy accordingly [43,54]. The sources in which information is disseminated should also be considered.

Minority communities may have a greater trust in information obtained from peers from similar cultural groups [26] and trusted community sources [55]. Hence, it is critical that communication strategies are not only culturally sensitive and tailored toward individual groups but also use trusted sources of information.

Using existing social networks would also be useful in framing vaccinations as a social norm. Descriptive norms, aka what other people do, say, and believe, have an impact on an individual’s intentions to accept a vaccine [56]. People often underestimate vaccine acceptance by others making hesitancy more noticeable and influencing their decision-making [56]. However, presenting vaccinations as a descriptive norm and correcting people’s overestimation of the prevalence of vaccine hesitancy, may have a positive impact on improving uptake [56].

When deciding what messages to communicate, having an understanding of the discourse on social media in “real time” may help guide communication strategies. By identifying trends in opinions, changes in sentiments, and stances toward a particular intervention such as vaccinations, behaviors can be anticipated [57]. Policymakers may then be able to intervene in a timely manner to encourage and sustain support [58].

What is clear to see is the importance of effective communication that explains the reasons for targeted approaches toward vaccinations in the context of a pandemic in a manner that is sensitive and addresses misconceptions from minority communities [59]. Underlying this is the need for strong governance led by effective leadership that engages communities and adjusts to a population’s needs [60,61]. Good governance not only allows for timely and effective vaccination campaigns [60-62] but also greater investment in research and development and higher public spending [60]. Greater public spending may address some of the wider socioeconomic factors that contribute to reduced vaccine uptake in minority communities [7]. More investment in research may allow governments to better address and implement nonpharmaceutical methods of control during a pandemic such as effective contract systems and stronger early warning systems [63]. Such steps will help improve governments’ prevention and preparedness for future pandemic threats.

**Conclusions**

Our study highlighted the concerns minority groups had in relation to the vaccine implementation, specifically the targeting of minority groups through vaccination campaigns. The shortfalls in the communication strategies adopted to relay public health information to minority groups during the pandemic must be acknowledged if any meaningful improvement is to be made moving forward to strengthen future interventions targeted toward these groups.

https://formative.jmir.org/2024/1/e51152

Jmir Form Res 2024 | vol. 8 | e51152 | p.3317

(page number not for citation purposes)
Acknowledgments
The authors are appreciative of ongoing collaborations at University College London (UCL) and the University of Oxford that have enabled this social media research on COVID-19. NU was funded by the National Institute for Health and Care Research School for Primary Care Research Undergraduate Student Internship Programme. The views expressed are those of the authors and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Data Availability
The data sets generated or analyzed during this study are not publicly available due to restrictions from the UCL Research Ethics Council that prohibit the publication of any identifiable information from social media platforms.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Initial Boolean search term, final Boolean search term, thematic analysis framework, sentiment analysis coding, stance analysis coding, hypothetical examples of social media posts, and their sentiment and stance.

Multimedia Appendix 2
Statistics from network analysis.

Multimedia Appendix 3
Most frequent clusters of tweets and key themes noted in relation to the COVID-19 vaccines between Jan 2022 to April 2022 in the UK obtained from network analysis.

Multimedia Appendix 4
Conversation clusters in YouTube videos 1 and 3 related to themes of racism and mistrust in the United Kingdom.

References


36. VACMA GitHub repository. GitHub. URL: https://github.com/peregilk/VACMA_PUBLIC [accessed 2022-08-26]


Abbreviations

API: application programming interface
UCL: University College London
Factors Impacting Chinese Older Adults’ Intention to Prevent COVID-19 in the Post–COVID-19 Pandemic Era: Survey Study

Huixin Guan¹, MA; Wei Wang¹, PhD
USC-SJTU Institute of Cultural and Creative Industry, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Huixin Guan, MA
USC-SJTU Institute of Cultural and Creative Industry
No155. Tanjiatang Road
Shanghai, 200241
China
Phone: 86 13840128403
Email: guan_hx@sjtu.edu.cn

Abstract

Background: Understanding the factors influencing individuals’ health decisions is a dynamic research question. Particularly, after China announced the deregulation of the COVID-19 epidemic, health risks escalated rapidly. The convergence of “no longer controlled” viruses and the infodemic has created a distinctive social period during which multiple factors may have influenced people’s decision-making. Among these factors, the precautionary intentions of older individuals, as a susceptible health group, deserve special attention.

Objective: This study aims to examine the intention of older adults to engage in preventive behaviors and the influencing factors, including social, media, and individual factors, within the context of the postepidemic era. Drawing upon the structural influence model of communication, this study tests the potential mediating roles of 3 different types of media exposure between cognitive and structural social capital and protective behavior intention, as well as the moderating role of negative emotions between social capital and media exposure.

Methods: In this study, a web survey was used to collect self-reported quantitative data on social capital, media exposure, negative emotions, and the intention to prevent COVID-19 among older adults aged ≥60 years (N=399) in China.

Results: The results indicate that cognitive social capital significantly influenced protective behavior intention (P<.001), with cell phone exposure playing an additional impactful role (P<.001). By contrast, newspaper and radio exposure and television exposure mediated the influence of structural social capital on protective behavior intention (P<.001). Furthermore, negative emotions played a moderating role in the relationship between cognitive social capital and cell phone exposure (P<.001).

Conclusions: This study suggests that using tailored communication strategies across various media channels can effectively raise health awareness among older adults dealing with major pandemics in China, considering their diverse social capital characteristics and emotional states.

(JMIR Form Res 2024;8:e53608) doi:10.2196/53608

KEYWORDS
COVID-19; SARS-CoV-2; health protection; social capital; media exposure; negative emotions; structural influence model of communication; SIM; protect; protection; protective; intent; intention; prevention; preventative; restriction; restrictions; public health measures; safety; news; newspaper; media; radio; health communication; influence; influencing; infectious; infection control; pandemic; gerontology; geriatric; geriatrics; older adult; older adults; older person; older people; aging

Introduction

Background

In December 2022, the Chinese State Council issued the Notice on Further Optimizing the Implementation of COVID-19 Prevention and Control Measures, announcing the liberalization of COVID-19 control in China [1]. Following the policy change, there was a rapid and widespread surge in virus transmission. During this period, the health status of older adults became a matter of concern. Older adults, characterized by degenerating body functions, weakened immune systems, and underlying
diseases [2], are generally at a disadvantage in fighting viral infections. Survey data showed that the mortality rate of COVID-19 is significantly higher among older adults aged ≥65 years than among younger adults globally. In China, the COVID-19 mortality rate is 2.3%, with >50% of deaths occurring in patients aged ≥50 years [3]. Thus, older adults are considerably more susceptible to the detrimental effects of COVID-19 than the general population. In the postepidemic era, it is crucial to examine the behavioral intentions of older adults to cope with disease risk.

This study aimed to address this research problem by introducing the structural influence model of communication (SIM), a research framework proven to be reliable in the context of large-scale health risks. In addition, this study aimed to enhance the SIM by delving into specific variables in depth, seeking to establish a more comprehensive research perspective on investigating influencing factors. This approach further examined the applicability of the model across a broader range of research contexts.

Theory Framework
The SIM suggests that health cognition and behavior should be considered within the elements of the social system based on the social-ecological theory [4]. It builds on the basic idea that communication is controlled by power dynamics, similar to the knowledge gap theory, which suggests that individuals with lower socioeconomic status have limited access to information [5]. Accordingly, the model states that social determinants and the conditions of mobilizable resources greatly shape the information environment, leading to communication inequalities and then differentiated health behaviors and outcomes. However, most previous studies have focused on individual-level factors, such as age [6], education [7], and revenue [8,9], with minimal attention paid to the social resources that individuals may have access to [10]. In fact, a weak ability to mobilize social resources, that is, social capital, could often lead to a disadvantage in health information seeking [11,12] and health behaviors [13].

Furthermore, the communication inequality could manifest in multiple aspects, such as health message search frequency [14], message type [15,16], and media exposure preferences [17]. As media technology advances, disparities in the use of and exposure to traditional and new media, on the one hand, have become visible, resulting in suboptimal health behaviors among certain populations. Several studies support this perspective. For instance, Viswanath and Ackerson [13] explored the phenomenon of mass media exposure in the United States and concluded that social structures, such as race and class, lead to widespread health disparities.

Research on the SIM is an ongoing endeavor. Some scholars believe that more evidence is still needed to establish the relationship among social capital, communication, and health behavior [12,18]. One of the reasons is that current research based on the SIM does not clearly reveal the intrapersonal processes that guide communication decisions [11,19], whereas some other theories state that affective responses to risk, such as anxiety and fear, possibly relate to information seeking [20]. Hovick et al [11] argued that personal perception factors, particularly in the face of risk, can further influence communication inequalities and should be considered. In particular, risk perception can influence information-seeking attitudes and outcomes by shaping affective responses [20]. When individuals perceive threats as serious, it triggers negative emotions, such as fear and anxiety, motivating them to seek and process information to mitigate those emotions [21]. Individuals with a strong sense of risk and self-efficacy are more likely to engage in protective behaviors driven by negative emotions [22]. In this paper, we continue this line of thought and introduce the variable of negative emotions to expand the existing relationship among social capital, media exposure, and protective behavior intention, ultimately forming the theoretical framework.

Variables and Hypotheses

Social Capital
Social capital is defined as a general public resource [23] that is typically found in relationships within kinship, professional, organizational, and neighborhood contexts [24]. Putnam [25,26] developed 2 classification schemes for social capital. The first scheme distinguishes between bonding and bridging components based on the inward and outward social connections of the group. The second scheme divides social capital into cognitive and structural aspects, with the cognitive element encompassing trust and reciprocity, whereas the structural element pertains to social participation and social networks [27].

Social capital has long been recognized as an influential factor in health [28], dating back to the 19th century [29], when Durkheim linked social integration to suicide rates [30]. Many studies have demonstrated a strong correlation between social capital and disease prevalence [31-34] as well as various types of health behavioral habits and lifestyles [35,36]. The SIM further emphasizes the significant role of social capital in the potential relationship among social structural factors, communication inequalities, and health inequalities [11,12].

Different types of social capital may play distinct roles in health outcomes. For example, bridging social capital is believed to provide individuals with new information and resources, enhancing their ability to address health problems [37]. Conversely, bonding social capital may restrict individuals’ access to information, increase psychological stress, and negatively impact health [38]. However, the differential effects of cognitive and structural social capital on communication and health outcomes have not received adequate attention. Kawachi and Berkman [39] identified two mechanisms through which social capital influences health: (1) social networks serve as information channels, facilitating the transfer and exchange of information among members, and (2) trust within the community accelerates the transmission of health behaviors. Building on this perspective, cognitive and structural elements operate through different pathways. Several empirical studies have concluded that cognitive social capital is consistently beneficial for mental health outcomes, whereas the outcomes associated with structural social capital are ambiguous [40,41]. Therefore, this study proposes the following hypothesis: cognitive (hypothesis 1a) and structural (hypothesis 1b) social capital positively predicts the intention to adopt protective behavior.
Media Exposure

The SIM encompasses a broad definition of communication [12], with most studies focusing on information seeking and information scanning [42,43] but paying less attention to preferences for accessing information from different types of media channels. This study specifically examined mass media exposure. It is important to note that social structural differences significantly influence the patterns of media exposure among specific populations. For example, a study on women’s exposure to family planning messages in Nigeria revealed notable disparities between urban and rural women in terms of the likelihood of accessing information through media channels such as print media, radio, and television [44]. This finding underscores the significance of media exposure as an expression of communication inequality in the SIM. Moreover, numerous empirical studies have shown that different forms of media exposure can predict health prevention beliefs and behaviors. For example, Bleakley et al [45] showed that exposure to sexual content in media, including movies, television shows, music, magazines, and video games, influenced adolescents’ beliefs and subsequent behaviors. Sitar-Taut and Mican [46] identified social media exposure as the most expressive driver of attitudes toward vaccination.

Regarding older adults, there is a general belief that they prefer to seek health information from traditional media, such as newspapers and radio [47]. However, with the proliferation of media technologies, internet use among older adults has been on the rise [48]. New media platforms are increasingly playing a vital role in health interventions for older adults. A study conducted in Hong Kong [49] investigated the internet use of older adults and found that the use of media devices, such as phones and computers, can effectively improve the mental health of older adults. Another study by Yi [50] compared the use of paper-based media with the use of electronic media among older adults and discovered that both significantly promote healthy lifestyles. However, the mechanism through which paper-based media use influences older adults is associated with selective socioeconomic characteristics, whereas the effect of electronic media use plays a more direct role. Accordingly, this study proposes the following hypothesis: newspaper and radio exposure (hypothesis 2a), television exposure (hypothesis 2b), and cell phone exposure (hypothesis 2c) mediate the effects of cognitive social capital on protective behavior intention, and newspaper and radio exposure (hypothesis 2d), television exposure (hypothesis 2e), and cell phone exposure (hypothesis 2f) mediate the effects of structural social capital on protective behavior intention.

Negative Emotions

Negative emotions are emotional responses that arise in the face of health risks [51,52]. A strong emotional response can drive individuals to adopt adaptive behaviors and, ultimately, facilitate the reception of information. Although fear has often been considered a mediating factor in the process from risk perception to behavior adoption, it is important to note that fear is not the sole emotion triggered when individuals encounter threats. Other emotions such as worry, anger, and guilt also come into play [21,53]. Although some studies have indicated that both positive and negative emotional reactions affect people’s health behaviors [54,55], most studies focus on negative emotions and highlight their importance as predictors of individuals’ health behaviors among various factors [56].

According to the planned risk information-seeking model, negative emotions can underscore the need for risk information and initiate information-seeking behaviors [20]. When individuals experience feelings of worry, anger, or fear, they are motivated to actively search for relevant risk information to regain a sense of control over the situation [52]. The selective exposure theory suggests that emotional states prompt individuals to actively choose and use media that align with their emotional needs, with stressed people tending to opt for calming media and bored people more inclined to select uplifting media [57]. The study by Huang and Yang [58] of the Chinese public’s perception of air pollution noted that negative effects positively influenced people’s information-seeking behavior regarding air pollution. The appraisal theory of emotion posits that emotions stem from automatic and subjective evaluations of events [59]. Individuals’ assessments of their surroundings give rise to various emotional states [60]. For instance, those experiencing fear or anger are more inclined to seek information aligned with the action tendencies of their respective emotions while being less likely to pursue information incongruent with their current emotional state [61]. Given this framework, negative emotions, prevalent among individuals confronting health risks, are likely to exert a moderating influence on subsequent information-seeking and health behaviors. Furthermore, focusing on different media channels, negative emotions have been shown to increase individuals’ intention to seek health information from various sources, such as interpersonal, traditional, and web sources [54]. Consequently, this study proposes the following hypothesis: negative emotions play a moderating role between social capital and media exposure (hypothesis 3).

This study aimed to shed light on the influences of multiple social, personal, and interpersonal elements on individual health issues within the context of a larger social-ecological system. By integrating the SIM and planned risk information-seeking model, this study investigated the impact of social capital, media exposure, and negative emotions on the health promotion intention of older adults, a group that faces health disadvantages, in a postepidemic era. Specifically, this study measured the association between older adults’ cognitive and structural social capital and their intention to protect themselves from COVID-19 and highlighted the role of media exposure in this process of influence. Furthermore, the variable of negative emotions was also introduced to assess whether it moderates the relationship between social capital and media exposure.

Methods

Design and Participants

This study collected data in March 2023. We developed a web-based questionnaire through Wenjuan.com (Zhongyan Technology), a Chinese platform akin to Google Forms (Google LLC). Snowball sampling method was used in this study. We disseminated the web-based survey link on the WeChat (Tencent
Inc) platform and encouraged our friends to share it with their older family members. The participants were limited to individuals aged ≥60 years. They were requested to complete the questionnaire, either in person or with the assistance of their family members. Each respondent was rewarded with a monetary incentive (RMB ¥5 [US $0.7]) after completing the survey.

Before the formal distribution of the questionnaire, a small-scale pretest was conducted over the web, and 46 completed questionnaires were collected. On the basis of the feedback received from the respondents, certain questions were modified. The final version of the questionnaire was then distributed on a large scale in March 2023, and 514 completed questionnaires were collected. However, 115 (22.4%) questionnaires were deemed invalid owing to incomplete answers or a completion time of <150 seconds. Therefore, a total of 399 valid questionnaires were included in the analysis, representing a completion rate of 77.6%. The age of the participants varied from 60 to 93 years, with an average age of 64.93 (SD 4.91) years. Table 1 presents the characteristics of the respondents.

**Table 1.** Demographic attributes of the participants.

<table>
<thead>
<tr>
<th>Variables and categories</th>
<th>Pretest sample (N=46), n (%)</th>
<th>Formal sample (N=399), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>29 (63)</td>
<td>198 (49.6)</td>
</tr>
<tr>
<td>Female</td>
<td>17 (37)</td>
<td>201 (50.4)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school and below</td>
<td>0 (0)</td>
<td>29 (7.3)</td>
</tr>
<tr>
<td>Junior school</td>
<td>3 (6.5)</td>
<td>82 (20.6)</td>
</tr>
<tr>
<td>High school</td>
<td>5 (10.9)</td>
<td>99 (24.8)</td>
</tr>
<tr>
<td>Professional training college</td>
<td>19 (41.3)</td>
<td>95 (23.8)</td>
</tr>
<tr>
<td>Bachelor and above</td>
<td>19 (41.3)</td>
<td>94 (23.6)</td>
</tr>
<tr>
<td><strong>Income (per month; RMB(^a))</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2500</td>
<td>2 (4.3)</td>
<td>74 (18.5)</td>
</tr>
<tr>
<td>2500-5000</td>
<td>14 (30.4)</td>
<td>215 (53.9)</td>
</tr>
<tr>
<td>5000-10,000</td>
<td>26 (56.5)</td>
<td>90 (22.6)</td>
</tr>
<tr>
<td>&gt;10,000</td>
<td>4 (8.7)</td>
<td>20 (5)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>0 (0)</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td>Married</td>
<td>43 (93.5)</td>
<td>330 (82.7)</td>
</tr>
<tr>
<td>Divorced</td>
<td>2 (4.4)</td>
<td>24 (6)</td>
</tr>
<tr>
<td>Widowed</td>
<td>1 (2.2)</td>
<td>42 (10.5)</td>
</tr>
<tr>
<td><strong>Residence status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living with family</td>
<td>41 (89.1)</td>
<td>319 (79.9)</td>
</tr>
<tr>
<td>Living without family</td>
<td>5 (10.9)</td>
<td>80 (20.1)</td>
</tr>
<tr>
<td><strong>Number of chronic diseases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One or more</td>
<td>19 (41.3)</td>
<td>177 (44.4)</td>
</tr>
<tr>
<td>None</td>
<td>27 (48.7)</td>
<td>222 (55.6)</td>
</tr>
</tbody>
</table>

\(^a\)RMB: Renminbi, the official currency of China; RMB ¥1=US $0.14.

**Ethical Considerations**

The study received ethics approval from the University of Southern California-Shanghai Jiao Tong University Institute of Cultural and Creative Industry Ethics Board. The institutional policy does not require the College to issue a separate approval document with a reference number for this type of humanities and social sciences research currently. Participants were provided with a written description of the study before completing the web-based questionnaire, ensuring the confidentiality of their private information solely for the purpose of this study. The completion of the questionnaire indicated approved consent from each participant.

**Measures**

**Social Capital**

The measurement of social capital in this study encompassed 2 dimensions: cognitive social capital (trust and reciprocity) and structural social capital (social network and social
In this study, we collected demographic information from the participants, including sex, age, education level, income level, marital status, residence status, and chronic disease status. Following independent variables 2-tailed t test and one-way ANOVA in SPSS (version 26; IBM Corp), we identified 3 significant variables that had a notable impact on protective behavior intention, namely gender (P<.001), educational level (P=.047), and income level (P=.002). These variables were used as control variables in the subsequent analyses.

**Analysis**

In this study, SPSS and MPLUS (version 8.3; Muthén & Muthén) were used to test the reliability of the measurement models. Construct reliability was evaluated based on certain conditions, including a Cronbach α value of >.7 and a composite reliability value >0.7 [68,69]. Average variance extracted is considered to be >0.5 [69], but we can also accept 0.4 [70-73]. Then, a confirmatory factor analysis was conducted to test the model fit. The following fit indices, as suggested by Browne and Cudeck [74], were considered for a reasonably fitting model: a root-mean-square error of approximation of <0.08, Tucker-Lewis index and comparative fit index of >0.9, and ratio of chi-square to df χ²/df of <5.

To test the research hypotheses, a linear regression analysis was performed. Before conducting the regression analysis, a validated one-way analysis was performed in MPLUS to assess the presence of common method bias. In addition, multicollinearity was evaluated using variance inflation factors, and variance inflation factor values of <10 were considered acceptable, indicating that all predictor variables were available for a multivariate analysis [69]. This analysis was conducted in SPSS. Next, this study tested the mediating and moderating effects using models 4 and 7 of the SPSS macro. Indirect effects were calculated using a bootstrapping technique [75], which involved generating a 95% bias-corrected CI (BC) for indirect effects based on random samples of data. If the BC did not include 0, mediation could be inferred.

**Results**

**Means, SDs, and Pearson r Correlations**

Means, SDs, and Pearson r correlations between the main variables are reported in Table 2. It can be seen that the mean value of cognitive social capital was much higher than that of structural social capital. Moreover, the mean value of cell phone exposure exceeded that of the 2 traditional media exposures. Notably, older adults demonstrated a strong overall intention to adopt protective behavior, with negative emotions at a moderate level. In addition, cognitive social capital was significantly correlated with protective intention (P<.01), whereas there was a differential correlation between various types of media exposure and the 2 forms of social capital and between various types of media exposure and protective intention.
Table 2. Means, SDs, and Pearson $r$ correlations of the main variables.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>1.000</td>
<td>0.17</td>
<td>0.37</td>
<td>0.07</td>
<td>0.16</td>
<td>0.28</td>
<td>0.002</td>
</tr>
<tr>
<td>$P$ value</td>
<td>__&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.001</td>
<td>&lt;.001</td>
<td>.14</td>
<td>.002</td>
<td>&lt;.001</td>
<td>.97</td>
</tr>
<tr>
<td>2. Structural social capital (mean 2.71, SD 0.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.17</td>
<td>1.000</td>
<td>0.06</td>
<td>0.39</td>
<td>0.24</td>
<td>0.02</td>
<td>0.16</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.001</td>
<td>__</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.76</td>
<td>.001</td>
</tr>
<tr>
<td>3. Protective behavior intention (mean 3.97, SD 0.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.37</td>
<td>0.06</td>
<td>1.000</td>
<td>0.19</td>
<td>0.21</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.21</td>
<td>__</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.37</td>
</tr>
<tr>
<td>4. Newspaper and radio exposure (mean 2.40, SD 0.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.07</td>
<td>0.39</td>
<td>0.19</td>
<td>1.000</td>
<td>0.41</td>
<td>__</td>
<td>–0.06</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.14</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>__</td>
<td>&lt;.001</td>
<td>.90</td>
<td>.002</td>
</tr>
<tr>
<td>5. Television exposure (mean 2.98, SD 0.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.16</td>
<td>0.24</td>
<td>0.21</td>
<td>0.41</td>
<td>1.000</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.002</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>__</td>
<td>.31</td>
<td>.02</td>
</tr>
<tr>
<td>6. Cell phone exposure (mean 3.59, SD 0.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.28</td>
<td>0.02</td>
<td>0.23</td>
<td>–0.06</td>
<td>0.05</td>
<td>1.000</td>
<td>0.15</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.76</td>
<td>&lt;.001</td>
<td>.90</td>
<td>.31</td>
<td>__</td>
<td>.003</td>
</tr>
<tr>
<td>7. Negative emotions (mean 2.35, SD 0.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.002</td>
<td>0.16</td>
<td>0.05</td>
<td>0.16</td>
<td>0.11</td>
<td>0.15</td>
<td>1.000</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.97</td>
<td>.001</td>
<td>.37</td>
<td>.002</td>
<td>.02</td>
<td>.003</td>
<td>__</td>
</tr>
</tbody>
</table>

<sup>a</sup>Not available.

Reliability and Validity

In this study, a model consisting of 7 factors was developed (Figure 1). The construct reliability and validity of the model were evaluated based on the following criteria. As shown in Table 3, all the scales’ Cronbach $\alpha$ value and composite reliability value were >.7, indicating good construct reliability. In addition, the average variance extracted was >0.4, suggesting acceptable construct validity. Furthermore, the fit indices of the 7-factor model, which included all the variables in the study, were within the acceptable range after Modification Index correction ($\chi^2/df=2.230$, root-mean-square error of approximation=0.056, comparative fit index=0.912, and Tucker-Lewis index=0.901), indicating a good model fit and further supporting the validity of the measurement model.
Figure 1. Structural equation model for impacting factors of protective behavior intention among older adults.

Table 3. The scale’s construct reliability and validity.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Items (N=33), n (%)</th>
<th>Cronbach α</th>
<th>CR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>AVE&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive social capital</td>
<td>6 (18)</td>
<td>.80</td>
<td>0.82</td>
<td>0.46</td>
</tr>
<tr>
<td>Structural social capital</td>
<td>5 (15)</td>
<td>.73</td>
<td>0.74</td>
<td>0.41</td>
</tr>
<tr>
<td>Newspaper and radio exposure</td>
<td>6 (18)</td>
<td>.83</td>
<td>0.84</td>
<td>0.48</td>
</tr>
<tr>
<td>Television exposure</td>
<td>3 (9)</td>
<td>.70</td>
<td>0.71</td>
<td>0.46</td>
</tr>
<tr>
<td>Cell phone exposure</td>
<td>3 (9)</td>
<td>.70</td>
<td>0.71</td>
<td>0.46</td>
</tr>
<tr>
<td>Protective behavior intention</td>
<td>4 (12)</td>
<td>.84</td>
<td>0.87</td>
<td>0.63</td>
</tr>
<tr>
<td>Negative emotions</td>
<td>6 (18)</td>
<td>.92</td>
<td>0.92</td>
<td>0.65</td>
</tr>
</tbody>
</table>

<sup>a</sup>CR: composite reliability.

<sup>b</sup>AVE: average variance extracted.

Relationship Between Social Capital and Media Exposure

First, this study examined the relationship between social capital and media exposure (Table 4). Controlling for gender, education, and income, older adults with higher levels of cognitive social capital were more likely to use television (B=0.24; P<.001) and cell phone (B=0.36; P<.001) but not newspaper and radio (B=0.10; P=.16). Structural social capital showed a significant positive association with newspaper and radio exposure (B=0.41; P<.001) and television exposure (B=0.28; P<.001). However, there was no statistically significant association between structural social capital and cell phone exposure (B=0.01; P=.92).
Table 4. Summary of relationships between social capital and media exposure (N=399).

<table>
<thead>
<tr>
<th></th>
<th>Newspaper and radio exposure</th>
<th>Television exposure</th>
<th>Cell phone exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Model 2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Model 3&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sex</td>
<td>B 0.09 &lt;i&gt;P&lt;/i&gt; value .25</td>
<td>B 0.09 &lt;i&gt;P&lt;/i&gt; value .24</td>
<td>B 0.07 &lt;i&gt;P&lt;/i&gt; value .43</td>
</tr>
<tr>
<td>Education</td>
<td>−0.02 &lt;i&gt;P&lt;/i&gt; value .59</td>
<td>−0.04 &lt;i&gt;P&lt;/i&gt; value .36</td>
<td>−0.07 &lt;i&gt;P&lt;/i&gt; value .12</td>
</tr>
<tr>
<td>Revenue</td>
<td>0.004 &lt;i&gt;P&lt;/i&gt; value .95</td>
<td>0.02 &lt;i&gt;P&lt;/i&gt; value .82</td>
<td>−0.03 &lt;i&gt;P&lt;/i&gt; value .70</td>
</tr>
<tr>
<td>Cognitive social capital</td>
<td>0.10 &lt;i&gt;P&lt;/i&gt; value .16</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>0.24 &lt;i&gt;P&lt;/i&gt; value .001</td>
</tr>
<tr>
<td>Structural social capital</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>0.41 &lt;i&gt;P&lt;/i&gt; value &lt;.001</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
</tr>
</tbody>
</table>

<sup>a</sup>R<sup>2</sup>=0.01; 4<sub>3,394</sub>=1.06; <i>P</i> value for the F test>.05.
<sup>b</sup>R<sup>2</sup>=0.15; 4<sub>3,394</sub>=17.96; <i>P</i> value for the F test<.001.
<sup>c</sup>R<sup>2</sup>=0.04; 4<sub>3,394</sub>=4.35; <i>P</i> value for the F test<.01.
<sup>d</sup>R<sup>2</sup>=0.08; 4<sub>3,394</sub>=8.08; <i>P</i> value for the F test<.001.
<sup>e</sup>R<sup>2</sup>=0.12; 4<sub>3,394</sub>=13.72; <i>P</i> value for the F test<.001.
<sup>f</sup>R<sup>2</sup>=0.05; 4<sub>3,394</sub>=5.13; <i>P</i> value for the F test<.001.
<sup>g</sup>The values are not available.

Relationship Between Social Capital and Protective Behavior Intention

Second, the effect of social capital on protective behavior intention was assessed. The results are presented in Table 5. Cognitive social capital positively predicted protective behavior intention (B=0.52; <i>P</i> value<.001; model 7), so hypothesis 1a was supported. However, there was no significant association between structural social capital and protective behavior intention (B=0.07; <i>P</i> value=.21; model 8), so hypothesis 1b was not supported.

Table 5. Summary of regressions predicting protective behavior intention (N=399).

<table>
<thead>
<tr>
<th></th>
<th>Model 7&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model 8&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Model 9&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Model 10&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>B 0.20 &lt;i&gt;P&lt;/i&gt; value .007</td>
<td>B 0.24 &lt;i&gt;P&lt;/i&gt; value .003</td>
<td>B 0.17 &lt;i&gt;P&lt;/i&gt; value .03</td>
<td>B 0.18 &lt;i&gt;P&lt;/i&gt; value .02</td>
</tr>
<tr>
<td>Education</td>
<td>−0.01 &lt;i&gt;P&lt;/i&gt; value .85</td>
<td>−0.004 &lt;i&gt;P&lt;/i&gt; value .92</td>
<td>−0.02 &lt;i&gt;P&lt;/i&gt; value .62</td>
<td>−0.02 &lt;i&gt;P&lt;/i&gt; value .56</td>
</tr>
<tr>
<td>Revenue</td>
<td>−0.12 &lt;i&gt;P&lt;/i&gt; value .06</td>
<td>−0.10 &lt;i&gt;P&lt;/i&gt; value .14</td>
<td>−0.10 &lt;i&gt;P&lt;/i&gt; value .12</td>
<td>−0.07 &lt;i&gt;P&lt;/i&gt; value .25</td>
</tr>
<tr>
<td>Cognitive social capital</td>
<td>0.52 &lt;i&gt;P&lt;/i&gt; value &lt;.001</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>0.43 &lt;i&gt;P&lt;/i&gt; value &lt;.001</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
</tr>
<tr>
<td>Structural social capital</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>0.07 &lt;i&gt;P&lt;/i&gt; value .21</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
</tr>
<tr>
<td>Newspaper and radio exposure</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>0.13 &lt;i&gt;P&lt;/i&gt; value .01</td>
<td>0.14 &lt;i&gt;P&lt;/i&gt; value .01</td>
</tr>
<tr>
<td>Television exposure</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>0.08 &lt;i&gt;P&lt;/i&gt; value .08</td>
<td>0.12 &lt;i&gt;P&lt;/i&gt; value .01</td>
</tr>
<tr>
<td>Cell phone exposure</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>— &lt;i&gt;P&lt;/i&gt; value —</td>
<td>0.14 &lt;i&gt;P&lt;/i&gt; value .005</td>
<td>0.23 &lt;i&gt;P&lt;/i&gt; value &lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>R<sup>2</sup>=0.18; 4<sub>3,394</sub>=21.34; <i>P</i> value for the F test<.001.
<sup>b</sup>R<sup>2</sup>=0.05; 4<sub>3,394</sub>=4.91; <i>P</i> value for the F test<.001.
<sup>c</sup>R<sup>2</sup>=0.22; 7<sub>3,391</sub>=16.13; <i>P</i> value for the F test<.001.
<sup>d</sup>R<sup>2</sup>=0.14; 7<sub>3,391</sub>=9.03; <i>P</i> value for the F test<.001.
<sup>e</sup>The values are not available.

Mediating Effects of Media Exposure

Next, this study evaluated the mediating effects of 3 types of media exposure between social capital and protective behavior intention (Table 5). According to model 9, the positive effect of cognitive social capital on protective behavior intention remained significant after the mediating variable cell phone exposure was added (B=0.43, <i>P</i> value<.001). By contrast, after the mediating variables were added to the regression model of structural social capital and protective behavior intention, newspaper and radio exposure showed a positive effect on protective behavior intention (B=0.14; <i>P</i> value=.01), television exposure (B=0.12; <i>P</i> value=.01), and cell phone exposure (B=0.23; <i>P</i> value<.001; model 10).
Bootstrap repeated sampling analysis (Table 6) examined the indirect effects and provided further insights into the mediating effect of media exposure between social capital and protective behavior intention. Mediated by cell phone exposure, the effect value of cognitive social capital on protective behavior intention was 0.05, with a 95% BC of 0.01-0.11, which does not contain 0, so hypothesis 2c was verified. By contrast, when newspaper and radio exposure acted as a mediator, the indirect effect value was 0.01, with a 95% BC of –0.003 to 0.04, so hypothesis 2a was not supported. Similarly, television exposure was not confirmed as a mediating variable between cognitive social capital and protective behavior intention, so hypothesis 2b was rejected.

Table 6. Direct and indirect effects of social capital on protective behavior intention via media exposure.

<table>
<thead>
<tr>
<th>Path</th>
<th>Direct effect</th>
<th>Indirect effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effect size (SE)</td>
<td>95% BC&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.43 (0.06)</td>
<td>0.30 to 0.57&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.43 (0.06)</td>
<td>0.30 to 0.57</td>
</tr>
<tr>
<td>3&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.43 (0.06)</td>
<td>0.30 to 0.57</td>
</tr>
<tr>
<td>4&lt;sup&gt;f&lt;/sup&gt;</td>
<td>–0.03 (0.06)</td>
<td>–0.14 to 0.08</td>
</tr>
<tr>
<td>5&lt;sup&gt;g&lt;/sup&gt;</td>
<td>–0.03 (0.06)</td>
<td>–0.14 to 0.08</td>
</tr>
<tr>
<td>6&lt;sup&gt;h&lt;/sup&gt;</td>
<td>–0.03 (0.06)</td>
<td>–0.14 to 0.08</td>
</tr>
</tbody>
</table>

<sup>a</sup>BC: bias-corrected and accelerated CI.

<sup>b</sup>When newspaper and radio exposure acts as a mediating variable between cognitive social capital and protective behavior intention.

<sup>c</sup>Mediation is assumed because bias-corrected and accelerated CI does not exceed 0.

<sup>d</sup>When television exposure acts as a mediating variable between cognitive social capital and protective behavior intention.

<sup>e</sup>When cell phone exposure acts as a mediating variable between cognitive social capital and protective behavior intention.

<sup>f</sup>When newspaper and radio exposure acts as a mediating variable between structural social capital and protective behavior intention.

<sup>g</sup>When television exposure acts as a mediating variable between structural social capital and protective behavior intention.

<sup>h</sup>When cell phone exposure acts as a mediating variable between structural social capital and protective behavior intention.

In addition, when newspaper and radio exposure and television exposure were seen as mediating variables for the effect of structural social capital on protective behavior intention, the upper and lower limits of bootstrap 95% CIs did not contain 0, indicating that structural social capital was significantly associated with protective behavior intention through both mediating paths of newspaper and radio exposure and television exposure; thus, hypotheses 2d and 2e were supported. By contrast, cell phone exposure was not proved as a mediator between structural social capital and protective behavior intention, so hypothesis 2f was not supported.

**Moderating Effect of Negative Emotions**

Finally, this study tested the moderating effect of negative emotions between social capital and media exposure (Table 7). According to model 11, the interaction term between cognitive social capital and negative emotions had a significant positive effect on cell phone exposure (B=0.14; P=.03). However, according to models 12 and 13, the interaction term of structural social capital and negative emotions was not statistically significant for newspaper and radio exposure or television exposure. A further simple slope analysis of the moderating effect of negative emotions between social capital and media exposure (Figure 2) showed that, on the one hand, for research participants with higher levels of negative emotions (Mean+1SD), cognitive social capital had a significant positive predictive effect on cell phone exposure (simple slope=0.52; t<sub>395</sub>=4.36; P<.001). On the other hand, for participants with lower levels of negative emotions, cognitive social capital positively influenced cell phone exposure to a lesser extent (simple slope=0.37; t<sub>395</sub>=5.95; P<.001). In addition, the mediating effect of cell phone exposure on the relationship between cognitive social capital and protective behavior intention tended to increase at all 3 levels of negative emotions. Thus, cognitive social capital was more likely to enhance older adults’ protective behavior intention by increasing their cell phone exposure as older adults’ levels of negative emotions rose. Therefore, hypothesis 3 was partially supported.
Table 7. The moderating effect of negative emotions between social capital and media exposure.

<table>
<thead>
<tr>
<th></th>
<th>Cell phone exposure</th>
<th>Newspaper and radio exposure</th>
<th>Television exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 11&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Model 12&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Model 13&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>P value</td>
<td>B</td>
</tr>
<tr>
<td>Sex</td>
<td>0.13</td>
<td>.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Education</td>
<td>0.14</td>
<td>&lt;.001</td>
<td>−0.04</td>
</tr>
<tr>
<td>Revenue</td>
<td>−0.12</td>
<td>.05</td>
<td>0.02</td>
</tr>
<tr>
<td>Cognitive social capital</td>
<td>0.37</td>
<td>&lt;.001</td>
<td><em>d</em></td>
</tr>
<tr>
<td>Structural social capital</td>
<td>___</td>
<td>___</td>
<td>0.40</td>
</tr>
<tr>
<td>Negative emotions</td>
<td>0.10</td>
<td>.01</td>
<td>0.09</td>
</tr>
<tr>
<td>int&lt;sub&gt;1&lt;/sub&gt;&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.14</td>
<td>.03</td>
<td>_</td>
</tr>
<tr>
<td>int&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;f&lt;/sup&gt;</td>
<td>_</td>
<td>_</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<sup>a</sup>R<sup>2</sup>=0.39; F<sub>6,392</sub>=11.45; P value for the F test <.001.
<sup>b</sup>R<sup>2</sup>=0.16; F<sub>6,392</sub>=12.74; P value for the F test <.001.
<sup>c</sup>R<sup>2</sup>=0.29; F<sub>6,392</sub>=5.82; P value for the F test <.001.
<sup>d</sup>The values are not available.
<sup>e</sup>Interaction items of cognitive social capital and negative emotions.
<sup>f</sup>Interaction items of structural social capital and negative emotions.

Figure 2. Moderating effect of negative emotions between cognitive social capital and cell phone exposure.

Figure 3 illustrates the correlations and effect values among the variables used in this study. Notably, cognitive social capital exhibited a significant correlation with protective behavior intention (P<.001), with the mediating effect of cell phone exposure established between the 2 variables (P<.001). In addition, the mediating effect of newspaper and radio exposure and television exposure between structural social capital and protective behavior intention was established (P<.001), whereas negative emotions played a moderating role in the relationship between cognitive social capital and protective behavior intention (P<.001).
Discussion

Principal Findings

This study aimed to investigate the applicability of the SIM in the social context of the post–COVID-19 epidemic era in China. The study focused on assessing the reliability of social capital in influencing health protection intentions through media exposure, with a specific emphasis on the older Chinese population. This study categorized social capital into cognitive and structural dimensions and examined media exposure from the perspective of 3 main media types: newspaper and radio, television, and cell phone. Considering the risk characteristics of COVID-19 messages and the influence of selective exposure, this study also tested whether negative emotions play a moderating role between social capital and media exposure.

The results of this study support the SIM by indicating that potential factors at the personal, interpersonal, and social levels can explain individuals’ intention to adopt health behaviors. First, cognitive social capital was shown to promote health behavior intention, whereas structural social capital did not have a significant effect on health behavior intention, which is consistent with the findings of De Silva et al [40] on social capital and mental health outcomes. However, the underlying reasons for these differences were not fully elucidated and should continue to be explored in depth in future research.

Second, the mediating role of media exposure between social capital and health behavior intention was also emphasized, validating the SIM pathway. Interestingly, the results revealed that the 2 types of social capital influenced protective behavior intention through different patterns of media exposure. Cognitive social capital positively influenced cell phone exposure, indicating that older adults with higher levels of trust and reciprocity were more inclined to use new media for health information seeking and adopt protective behavior. By contrast, structural social capital predicted health behavior intention through newspaper and radio exposure and television exposure, with a stronger mediating effect observed for newspaper and radio exposure. This implies that older adults with more social engagement and larger social networks prefer traditional media, engaging in more offline information activities. Previous studies have suggested that people tend to communicate information across all (or many) media channels [76]. Therefore, most studies have not categorized media channels when examining the relationship between social capital and health information seeking. However, according to the findings of this study, it may be advantageous for health communicators to target different characteristics of the population using different media channels when disseminating messages rather than solely focusing on increasing people’s social capital through welfare measures.

In addition to the association among social capital, media exposure, and protective behavior intention, the results of this study illustrate the importance of negative emotions. The higher the level of negative emotions, the more pronounced the positive effect of cognitive social capital on cell phone exposure, validating the findings of previous studies that negative emotions have a positive effect on information seeking [77,78]. By contrast, the results of this study indicate that negative emotions did not cause interference during the influence of traditional media exposure, such as newspaper and radio exposure and television exposure, among older adults. As numerous studies have pointed out, the intensity of negative emotions during pandemics interacts with the impact of efficiently disseminated media [79,80]. Consistent with this, this study also suggests that new media exhibit a more pronounced correlation with negative emotions than traditional media. However, the specific direction of influence between negative emotions and media
exposure remains unclarified in this study. Although some studies indicate that negative emotions may prompt individuals to seek information from new media [52,81,82], a substantial body of evidence also suggests that an overload of information can significantly induce panic, anxiety, and frustration [83-85].

Although the threshold of emotion varies from person to person [86] and it is difficult to define an optimal level that can be generalized, future research should explore the differences in the effects of negative emotions in specific social and media environments, as well as the reasons that trigger such differences. As a result, public sentiment can be better appeased during health crises, thus encouraging healthy behavior. Furthermore, the findings also suggest the need to explore other factors influencing health communication and behavior to extend the SIM to improve its applicability in relevant research.

Limitations

First, this study used an web-based questionnaire to recruit the sample, which may have introduced a bias in the sample. It is possible that a significant portion of the sample had access to and preferred using the internet for information seeking. Although efforts were made to involve relatives in completing the questionnaire on behalf of older adults who did not use the internet, the effectiveness of this approach in capturing the perspectives of non–internet users could not be fully controlled or measured. Thus, the generalizability of this study may be limited to this sampling method.

Second, the cross-sectional nature of this study, in which participants’ responses were recorded at a single point in time, restricted the ability to draw causal conclusions.

Data Availability

The data sets generated and analyzed during this study are not publicly available to protect the privacy of participants but are available from the corresponding author upon reasonable request.

Authors' Contributions

HG contributed to idea generation, data collection, data analysis, and manuscript writing. WW provided critical feedback, assisted in the study, and offered suggestions for both the initial and revised drafts. Both authors contributed to the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

References


https://formative.jmir.org/2024/1/e53608 JMIR Form Res 2024 | vol. 8 | e53608 | p.3333 (page number not for citation purposes)


64. #COVID19InfectionPneumonia#. Xinhua News Agency. 2020. URL: https://xhpfmapi.zhongguowangshi.com/vh512/share/8833887?channel=weixin&from=singlemessage&isappinstalled=0 [accessed 2024-01-29]


Abbreviations

BC: bias-corrected CI
SIM: structural influence model of communication
Exploring the Prevalence of Tinnitus and Ear-Related Symptoms in China After the COVID-19 Pandemic: Online Cross-Sectional Survey

Di Wang1*, BS; Peifan Li1,2,3*, MS; Xiaoling Huang4*, BS; Yixuan Liu1,2,3*, MS; Shihang Mao1,2,3*, BS; Haoning Yin5, HS; Na Wang6, MD; Yan Luo1, MS; Shan Sun1,2,3,7,8*, PhD

1Research Service Office, Eye & ENT Hospital, Fudan University, Shanghai, China
2NHC Key Laboratory of Hearing Medicine Research, Fudan University, Shanghai, China
3State Key Laboratory of Medical Neurobiology, Fudan University, Shanghai, China
4Department of Otolaryngology Head and Neck Surgery, Second Affiliated Hospital of Anhui Medical University, Hefei, China
5No.2 High School Of East China Normal University, Shanghai, China
6School of Public Health, Fudan University, Shanghai, China
7Tinnitus Hyperacusis Center, Otolaryngology Research Institute, Eye & ENT Hospital, Fudan University, Shanghai, China
8Clinical Research Unit of the Eye & ENT Hospital, Fudan University, Shanghai, China
*these authors contributed equally

Corresponding Author:
Shan Sun, PhD
Research Service Office
Eye & ENT Hospital
Fudan University
83 Fenyang Road
Shanghai, 200031
China
Phone: 86 18917786102
Fax: 86 2164377151
Email: shansun@fudan.edu.cn

Abstract

Background: Tinnitus is a complex and heterogeneous disease that has been identified as a common manifestation of COVID-19. To gain a comprehensive understanding of tinnitus symptoms in individuals following COVID-19 infection, we conducted an online survey called the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS) among the Chinese population.

Objective: Our objective was to investigate tinnitus and ear-related symptoms after COVID-19 infection in the Chinese population, with the aim of providing a solid empirical foundation for improved health care. The findings from CENTSS can contribute to the development of enhanced management strategies for tinnitus in the context of long COVID. By gaining a better understanding of the factors contributing to tinnitus in individuals with COVID-19, health care providers can tailor interventions to address the specific needs of affected patients. Furthermore, this study serves as a basis for research on the long-term consequences of COVID-19 infection and its associated tinnitus symptoms.

Methods: A quantitative, online, cross-sectional survey study design was used to explore the impact of the COVID-19 pandemic on experiences with tinnitus in China. Data were collected through an online questionnaire designed to identify the presence of tinnitus and its impacts. Descriptive statistics were used to analyze individuals’ demographic characteristics, COVID-19 infection–related ear symptoms, and the cognitive and emotional implications of tinnitus. Univariable and multivariable logistic regression analyses were used to model the cross-sectional baseline associations between demographic characteristics, noise exposure, educational level, health and lifestyle factors, and the occurrence of tinnitus.

Results: Between December 19, 2022, and February 1, 2023, we obtained responses from 1262 Chinese participants representing 24 regions, with an average age of 37 years. Among them, 540 patients (42.8%) reported experiencing ear-related symptoms after COVID-19 infection. Only 114 (9%) of these patients sought medical attention specifically for their ear symptoms, while 426...
(33.8%) did not seek hospital care. Tinnitus emerged as the most prevalent and impactful symptom among all ear-related symptoms experienced after COVID-19 infection. Of the respondents, female participants (688/888, 77.78%), younger individuals (<30 years), individuals with lower education levels, participants residing in western China, and those with a history of otolaryngology diseases were more likely to develop tinnitus following COVID-19 infection.

**Conclusions:** In summary, tinnitus was identified as the most common ear-related symptom during COVID-19 infection. Individuals experiencing tinnitus after COVID-19 infection were found to have poorer cognitive and emotional well-being. Different ear-related symptoms in patients post–COVID-19 infection may suggest viral invasion of various parts of the ear. It is therefore crucial to monitor and manage hearing-related changes resulting from COVID-19 as clinical services resume.

(JMIr Form Res 2024;8:e54326) doi:10.2196/54326

**KEYWORDS**

COVID-19 pandemic; tinnitus; ear-related symptoms; online survey; prevalence; ear-related; China; cross-sectional; complex; heterogeneous; symptom; symptoms; Chinese; population; investigate; health care; exploratory; teen; teens; teenager; teenagers; older adult; older adults; elder; elderly; older person; older people; COVID-19; regression analysis

**Introduction**

Auditory-related conditions such as dizziness, tinnitus, and earache have been identified as common symptoms of COVID-19 [1]. These symptoms can vary in duration, ranging from acute (lasting up to 4 weeks) to ongoing (lasting 4-12 weeks) to lasting more than 12 weeks, which is commonly referred to as “long COVID” [2]. The National Institute for Health and Care Excellence (NICE) [3] has also recognized tinnitus, dizziness, and earache as common symptoms of long COVID. According to a systematic literature review, approximately 8% of individuals with COVID-19 reported experiencing tinnitus [4].

Tinnitus is characterized by the perception of sound in the absence of an external sound source [5], and individuals with tinnitus may experience buzzing, ringing, or hissing sounds. It is one of the most prevalent inner ear disorders worldwide, affecting 5% to 43% of adults [6]. Tinnitus often leads to irritation, anxiety, depression, and insomnia and can also impair concentration, negatively impacting quality of life [7]. In severe cases, tinnitus can even contribute to suicidal tendencies [8]. Additionally, because tinnitus frequently coexists with hearing loss, it can result in communication difficulties, particularly among older adults, leading to social isolation, cognitive impairment, and possibly dementia [9]. Given its significant impact on health and quality of life, it is crucial to develop methods for early diagnosis and treatment of tinnitus.

COVID-19 triggers an immune response that can cause damage to cells and tissues [10]. As a result, the virus causing COVID-19 has the potential to affect the cochlea, auditory nerve, and central nervous system [11]. Inflammatory responses induced by the virus are speculated to be a potential cause of tinnitus in patients with long COVID. Direct viral infection may also contribute to the development of tinnitus [12]. Moreover, COVID-19 has been associated with disruptions in daily routines and social isolation, which can lead to elevated levels of depression, anxiety, and perceived stress, potentially exacerbating the perception of tinnitus [13]. Understanding the impact of long COVID on the auditory system and psychological well-being is crucial, and it can provide insights into the characteristics of patients experiencing tinnitus after COVID-19 infection and can inform the development of effective treatment strategies.

Furthermore, in an effort to break the chain of transmission, numerous regional lockdowns were implemented between 2019 and 2022. China recently relaxed certain restrictions, including quarantine rules for travelers and mask-wearing mandates, in December 2022, despite maintaining a comprehensive zero-COVID policy. This study focused on the specific characteristics of ear-related symptoms, particularly tinnitus, that may arise following the lifting of lockdown measures in the context of mass COVID-19 infections. The potential for tinnitus to become more severe due to long COVID presents a concerning challenge, necessitating further research to determine the potential exacerbation of tinnitus following prolonged COVID-19.

To gain a better understanding of the characteristics of patients experiencing tinnitus symptoms following COVID-19 infection, we conducted an online survey called the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS) among the Chinese population. This study aimed to assess the impact of COVID-19 infection on ear-related symptoms, with a specific focus on patients reporting tinnitus as a consequence of COVID-19 exposure. The findings from this study may contribute to improving management strategies for tinnitus in the context of long COVID. By enhancing our understanding of the factors contributing to tinnitus in individuals with COVID-19, health care providers can better tailor interventions to meet the specific needs of affected patients. Furthermore, this study serves as a foundation for future research on the long-term consequences of COVID-19 infection and associated tinnitus symptoms. Considering various risk factors for tinnitus, including infection, eustachian tube dysfunction, and hearing impairment [14], the outbreak of COVID-19, primarily a respiratory infection caused by the novel coronavirus SARS-CoV-2 [15], provides a substantial data set for analyzing the unique attributes of virus-induced tinnitus. Conducting extensive sample analyses would be highly beneficial in gaining comprehensive insights into the characteristics of individuals experiencing tinnitus after COVID-19 infection.
Methods

Study Design
A quantitative, online, cross-sectional survey study design was used to explore the impact of the COVID-19 pandemic on experiences with tinnitus in China. From December 19, 2022, to February 1, 2023, data were collected using wjx.cn, a widely used online questionnaire survey platform in China. The primary source of the samples in this study was from specific online platforms, namely “Guoke Patients,” and social media channels. To ensure sample diversity, the platform is characterized by a wide user base and a diverse user population. Simultaneously, social media channels were selected for their ability to reach user groups covering various ages, genders, professions, and educational backgrounds. Before completing the questionnaire, participants were required to read and confirm their voluntary agreement to participate. In the process of questionnaire data collection, we implemented a validation mechanism between questions to prevent duplicate submissions and the generation of false data. Furthermore, they could withdraw their participation at any time while completing the questionnaire. Only completely answered questionnaires were submitted. Each device could only access the questionnaire once, which is a restriction enforced by wjx.cn to ensure that each respondent only submits 1 questionnaire.

The inclusion criteria were as follows: (1) Chinese citizens currently residing in China, (2) age ≥18 years, (3) capable of independently reading and completing the self-administered questionnaire, and (4) willing to participate in the survey. Exclusion criteria included the following: (1) age younger than 18 years, (2) residing abroad, or (3) invalid data. All participants were anonymous adult volunteers. Thus, the survey posed no potential risk to participants’ physical and mental health. Furthermore, this study adhered to the ethical principles of the Measures for the Ethical Review of Biomedical Research Involving Humans of the National Health Commission of the People’s Republic of China.

Ethical Considerations
Ethical approval for international data collection was obtained from the Ethics Panel at Fudan University’s Eye and ENT Hospital (number 2022127). This clinical study used a cross-sectional design, collecting only patient questionnaire information without intervening in the patients’ treatment plans or posing any physiological risks. The researchers were committed to safeguarding the confidentiality of the provided information and therefore requested exemption from obtaining informed consent. Patient questionnaire information was anonymized, and patient identities were not disclosed. All participants gave their informed consent online.

The CHERRIES (Checklist for Reporting Results of Internet E-Surveys) was used to report the methods and results of the survey (Multimedia Appendix 1).

Data were collected via an online survey from 1334 individuals in China, and after excluding responses with missing information, 1288 participants were included. The data for age, height, and weight showed significant variability, and the maximum values for these variables exceeded the means by 3 SDs. Therefore, using the median to describe the overall level may be more appropriate than the mean, which is heavily influenced by outliers.

Survey Development
The design and implementation of the study and the data analysis were completed with the participation of 2 epidemiology professors. The survey captured the following categories: (1) demographic information such as name, age, height, weight, education, place of residence, and occupation (10 questions); (2) date of COVID-19 infection (1 question); (3) ear symptom–related questions (25 questions; eg, ear-related symptoms such as hearing loss, tinnitus, stuffy ears, ear pain, and dizziness; the impact of tinnitus as measured using the Tinnitus Handicap Inventory [3 questions] and the Athens Insomnia Scale [2 questions]; the level of impact on daily life due to ear symptoms after COVID-19 infection [5 questions]); (4) cognitive and emotional state after COVID-19 infection (6 questions); (5) past medical history of ear, nose, throat (ENT) diseases, hypertension, or diabetes (3 questions).

Survey Distribution
Individuals who provided informed consent were considered eligible. Recruitment primarily took place through the social media channels of patient organizations, such as the Weixin and Guoke websites [16]. To participate in the survey, respondents were required to provide online informed consent, and the survey software restricted multiple submissions from the same IP address.

Data Analysis
Initially, the data cleaning process involved the removal of cases that did not meet the eligibility criteria for the study. The data analysis used a mixed approach that included both quantitative and qualitative analyses. Descriptive statistics, such as frequencies, means, and SDs, were calculated using SPSSAU (QingSi Technology Ltd). If the data were normally distributed, they are expressed as the mean (SD), and Student t tests were used for comparisons between 2 groups. For the analysis of composition ratios, the chi-square test or Fisher exact probability method was used, as appropriate. For multivariate dichotomous dependent variable data, logistic regression analysis was used. Differences were considered statistically significant at P<.05. The chi-square test was used to examine the relationships between categorical variables, and adjusted residuals were used for post hoc analysis to determine significant relationships. To account for multiple testing, the P value was adjusted (Bonferroni) and set to be significant at P<.001. Qualitative data obtained from the open-ended questions were analyzed separately using inductive thematic analysis, and the identified themes were used to support the quantitative analysis. To explore the extent of the impact of the independent variables on the dependent variable, a logistic regression model was used to analyze the relationship between continuous variables, and a logistic regression model was used to handle categorical variables.
Results

Demographic Characteristics
A total of 1334 questionnaires were distributed to the registered users on the platform, of which 23 questionnaires were from foreigners, 23 were excluded due to being younger than 18 years, and 26 were considered otherwise invalid. The demographic characteristics of the study respondents are shown in Table 1.

Table 1. Demographic characteristics of the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS 2023).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Tinnitus status</th>
<th>Total sample (n=1262)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No (n=888)</td>
<td>Yes (n=374)</td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>37.26 (9.949)</td>
<td>36.265 (10.799)</td>
<td>.13</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>688 (77.48)</td>
<td>308 (82.35)</td>
<td>996 (78.92)</td>
</tr>
<tr>
<td>Male</td>
<td>200 (22.52)</td>
<td>66 (17.65)</td>
<td>266 (21.08)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>≤30</td>
<td>192 (21.62)</td>
<td>109 (29.14)</td>
<td>301 (23.85)</td>
</tr>
<tr>
<td>31-36</td>
<td>227 (25.56)</td>
<td>89 (23.80)</td>
<td>316 (25.04)</td>
</tr>
<tr>
<td>37-43</td>
<td>239 (26.91)</td>
<td>82 (21.93)</td>
<td>321 (25.44)</td>
</tr>
<tr>
<td>≥44</td>
<td>230 (25.90)</td>
<td>94 (25.13)</td>
<td>324 (25.67)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>166.422 (50.672)</td>
<td>164.533 (7.168)</td>
<td>.47</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>61.202 (12.070)</td>
<td>61.663 (11.980)</td>
<td>.53</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>22.420 (3.596)</td>
<td>22.685 (3.619)</td>
<td>.23</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Junior high school or less</td>
<td>32 (3.60)</td>
<td>33 (8.82)</td>
<td>65 (5.15)</td>
</tr>
<tr>
<td>Bachelor’s or associate degree</td>
<td>556 (62.61)</td>
<td>242 (64.71)</td>
<td>798 (63.23)</td>
</tr>
<tr>
<td>Master’s or higher degree</td>
<td>300 (33.78)</td>
<td>99 (26.47)</td>
<td>399 (31.62)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
<td>.88</td>
</tr>
<tr>
<td>Noise-exposed occupation</td>
<td>124 (13.96)</td>
<td>55 (14.71)</td>
<td>179 (14.18)</td>
</tr>
<tr>
<td>Non–noise-exposed occupation</td>
<td>540 (60.81)</td>
<td>222 (59.36)</td>
<td>762 (60.38)</td>
</tr>
<tr>
<td>Other occupation</td>
<td>224 (25.23)</td>
<td>97 (25.94)</td>
<td>321 (25.44)</td>
</tr>
<tr>
<td>Region in China, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>East</td>
<td>712 (80.18)</td>
<td>250 (66.84)</td>
<td>962 (76.23)</td>
</tr>
<tr>
<td>Middle</td>
<td>99 (11.15)</td>
<td>69 (18.45)</td>
<td>168 (13.31)</td>
</tr>
<tr>
<td>West</td>
<td>77 (8.67)</td>
<td>55 (14.71)</td>
<td>132 (10.46)</td>
</tr>
<tr>
<td>ENTa disease history, n (%)</td>
<td></td>
<td></td>
<td>.003</td>
</tr>
<tr>
<td>No</td>
<td>402 (45.27)</td>
<td>136 (36.36)</td>
<td>538 (42.63)</td>
</tr>
<tr>
<td>Yes</td>
<td>486 (54.73)</td>
<td>238 (63.64)</td>
<td>724 (57.37)</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td></td>
<td></td>
<td>.89</td>
</tr>
<tr>
<td>No</td>
<td>834 (93.92)</td>
<td>352 (94.12)</td>
<td>1186 (93.98)</td>
</tr>
<tr>
<td>Yes</td>
<td>54 (6.08)</td>
<td>22 (5.88)</td>
<td>76 (6.02)</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td></td>
<td></td>
<td>.90</td>
</tr>
<tr>
<td>No</td>
<td>879 (98.99)</td>
<td>361 (96.52)</td>
<td>1240 (98.25)</td>
</tr>
<tr>
<td>Yes</td>
<td>9 (1.01)</td>
<td>13 (3.48)</td>
<td>26 (1.74)</td>
</tr>
</tbody>
</table>

aENT: ear, nose, throat.
**Representation of Individuals With Tinnitus**

After removing the 72 invalid responses, there were 1262 respondents aged from 18 years to 80 years representing 24 areas in China, although some areas were only marginally represented (Figure 1A).

**Figure 1.** The spatiotemporal distribution of patients with COVID-19 who responded to the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS 2023): (A) distribution and (B) timeline of COVID-19 infection.

---

**Timeline of COVID-19 Infection**

As shown in Figure 1B, the survey respondents first reported having COVID-19 infection on December 16, 2021. The number of infected participants increased from early December 2022 to a peak on December 21, 2022, as policies were continuously adjusted. After the peak, the number of infected participants decreased and stabilized on January 20, 2023.
Impact of the Pandemic on Ear Symptoms

During the COVID-19 pandemic, 540 (42.79%) of the 1262 participants experienced some ear-related symptoms after being infected. Based on the self-reported symptoms of each of the 1262 participants, 374 (29.46%) had tinnitus, 185 (14.4%) had ear pain, 278 (21.74%) had ear fullness, 83 (17.99%) had hearing loss, and 309 (24.29%) experienced dizziness. A total of 114 (9%) of the 540 respondents with ear-related symptoms had seen a doctor because of their ear symptoms, but 426 (33.6%) of the 540 respondents did not go to the hospital (Figure 2A).

Figure 2. Based on responses to the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS) cross-sectional study: (A) ear symptoms and visits to the doctor after COVID-19, (B) location of ear symptoms after COVID-19 infection, (C) outcome of ear symptoms after COVID-19 infection.

Impact of Tinnitus After COVID-19 Infection

Because only 374 respondents had bothersome tinnitus, 888 respondents chose to skip the questions about the impact of tinnitus after COVID-19 infection. The respondents reported the impact of tinnitus on their quality of sleep, and they self-assessed the impact of ear symptoms, including tinnitus, earache, ear fullness, dizziness, and hearing loss, on their daily activities using a score ranging from 0 (no impact) to 10 (severe impact). Additionally, the study investigated the progression of ear-related symptoms in respondents who had contracted COVID-19. Table 2 shows that 146 (11.57%) of the 1262 respondents reported a slight impact on their sleep and that 117 (14.03%) were slightly dissatisfied with their quality of overall sleep. Tinnitus and vertigo had the greatest effects on respondents among all ear symptoms experienced after COVID-19 infection (Figure 2C).

After COVID-19 infection, 556 (43.07%) of the 1262 respondents reported that they had ear-related problems, including earache, ear fullness, dizziness or vertigo, and hearing loss. Of the 185 participants experiencing ear pain, 72 (39%) of the respondents had pain in both ears. Of the 278 participants experiencing ear fullness, 134 (48.2%) experienced fullness in both ears. Of the 374 participants experiencing tinnitus, 194 (51.9%) experienced tinnitus in both ears. Of the 83 participants experiencing hearing loss, 128 (55.4%) experienced hearing loss in both ears (Figure 2B). Only a small number of the respondents treated their ear symptoms, and most of the patients healed on their own, but 174 patients still had residual tinnitus symptoms, 143 patients had residual earache, and 112 patients had residual hearing loss.
Table 2. The impact of tinnitus on sleep and ear symptom score after COVID-19 infection according to the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS 2023; n=1262).

<table>
<thead>
<tr>
<th>Questions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the tinnitus affect your sleep? (eg, difficulty falling asleep, waking up at night, early awakening), n (%)</td>
<td></td>
</tr>
<tr>
<td>No effect</td>
<td>163 (12.92)</td>
</tr>
<tr>
<td>Slight impact</td>
<td>146 (11.57)</td>
</tr>
<tr>
<td>Significant impact</td>
<td>54 (4.28)</td>
</tr>
<tr>
<td>Severe impact or no sleep</td>
<td>11 (0.87)</td>
</tr>
<tr>
<td>Quality of overall sleep, n (%)</td>
<td></td>
</tr>
<tr>
<td>Satisfied</td>
<td>73 (5.78)</td>
</tr>
<tr>
<td>Slightly dissatisfied</td>
<td>177 (14.03)</td>
</tr>
<tr>
<td>Significantly dissatisfied</td>
<td>111 (8.08)</td>
</tr>
<tr>
<td>Very dissatisfied or couldn't sleep</td>
<td>13 (1.03)</td>
</tr>
<tr>
<td>Level of impact on daily life due to ear symptoms after COVID-19 infection, mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Tinnitus</td>
<td>1.366 (2.509)</td>
</tr>
<tr>
<td>Earache</td>
<td>0.692 (1.765)</td>
</tr>
<tr>
<td>Stuffy ear</td>
<td>1.062 (2.267)</td>
</tr>
<tr>
<td>Vertigo</td>
<td>1.193 (2.310)</td>
</tr>
<tr>
<td>Hearing loss</td>
<td>0.956 (2.129)</td>
</tr>
</tbody>
</table>

Cognitive and Emotional State After COVID-19 Infection

People had different cognitive and emotional states after COVID-19 infection. Of the 888 participants with no tinnitus, 281 (31.6%) said they were able to remember most things but had some difficulty thinking and solving daily problems, while 137 (36.63%) of the participants with tinnitus reported the same. Of the 374 participants with tinnitus, 60 (16%) said they were somewhat forgetful but clear-minded and able to solve daily problems, while only 120 (13.5%) of the 888 participants without tinnitus reported the same. Most of the participants with tinnitus (163/374, 43.6%) said they were somewhat unhappy, while only 28.7% (255/888) of the participants without tinnitus reported being unhappy (Figure 3). Table 3 shows the results of the cognitive and emotional state after COVID-19 infection. These findings showed that experiencing tinnitus after COVID-19 infection was associated with a worse cognitive and emotional state.

Figure 3. Cognitive and emotional impact of tinnitus for respondents of the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS) cross-sectional study (2023).
Table 3. Cognitive and emotional state of the respondents of the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS 2023).

<table>
<thead>
<tr>
<th>Cognitive and emotional state after COVID-19 infection</th>
<th>Tinnitus status, n (%)</th>
<th>Total sample (n=1262), n (%)</th>
<th>$\chi^2$ (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No (n=888)</td>
<td>Yes (n=374)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cognitive state after COVID-19 infection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Able to remember most things, be clear-minded, and solve daily problems</td>
<td>0</td>
<td>1 (0.27)</td>
<td>1 (0.08)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Able to remember most things but have some difficulty thinking and solving daily problems</td>
<td>281 (31.64)</td>
<td>137 (36.63)</td>
<td>418 (33.12)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somewhat forgetful but clear-minded and able to solve daily problems</td>
<td>74 (8.33)</td>
<td>60 (16.04)</td>
<td>134 (10.62)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somewhat forgetful and have some difficulty thinking and solving daily problems</td>
<td>120 (13.51)</td>
<td>61 (16.31)</td>
<td>181 (14.34)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Very forgetful and have great difficulty thinking and solving daily problems</td>
<td>407 (45.83)</td>
<td>104 (27.81)</td>
<td>511 (40.49)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unable to remember anything and unable to think and solve daily problems</td>
<td>6 (0.68)</td>
<td>11 (2.94)</td>
<td>17 (1.35)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Emotional state after COVID-19 infection</strong></td>
<td></td>
<td></td>
<td>45.116 (8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>So unhappy to the point of feeling life has no value</td>
<td>21 (2.36)</td>
<td>19 (5.08)</td>
<td>40 (3.17)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Happy and interested in life</td>
<td>323 (36.37)</td>
<td>86 (22.99)</td>
<td>409 (32.41)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somewhat unhappy</td>
<td>255 (28.72)</td>
<td>163 (43.58)</td>
<td>418 (33.12)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somewhat happy</td>
<td>252 (28.38)</td>
<td>82 (21.93)</td>
<td>334 (26.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Very unhappy</td>
<td>37 (4.17)</td>
<td>24 (6.42)</td>
<td>61 (4.83)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>In the past two weeks, how often have you felt nervous, anxious, or restless?</strong></td>
<td></td>
<td></td>
<td>58.004 (6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>More than half of the days</td>
<td>123 (13.85)</td>
<td>99 (26.47)</td>
<td>222 (17.59)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Almost every day</td>
<td>45 (5.07)</td>
<td>39 (10.43)</td>
<td>84 (6.66)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>A few days</td>
<td>454 (51.13)</td>
<td>179 (47.86)</td>
<td>633 (50.16)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not at all</td>
<td>266 (29.95)</td>
<td>57 (15.24)</td>
<td>323 (25.59)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>In the past two weeks, how often have you been unable to stop or control worrying?</strong></td>
<td></td>
<td></td>
<td>71.09 (6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>More than half of the days</td>
<td>82 (9.23)</td>
<td>82 (21.93)</td>
<td>164 (13)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Almost every day</td>
<td>37 (4.17)</td>
<td>40 (10.70)</td>
<td>77 (6.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>A few days</td>
<td>383 (43.13)</td>
<td>153 (40.91)</td>
<td>536 (42.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not at all</td>
<td>386 (43.47)</td>
<td>99 (26.47)</td>
<td>485 (38.43)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>In the past two weeks, how often have you felt a lack of energy or interest in doing things?</strong></td>
<td></td>
<td></td>
<td>46.517 (6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>More than half of the days</td>
<td>171 (19.26)</td>
<td>101 (27.01)</td>
<td>272 (21.55)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Almost every day</td>
<td>76 (8.56)</td>
<td>67 (17.91)</td>
<td>143 (11.33)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>A few days</td>
<td>399 (44.93)</td>
<td>151 (40.37)</td>
<td>550 (43.58)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not at all</td>
<td>242 (27.25)</td>
<td>55 (14.71)</td>
<td>297 (23.53)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>In the past two weeks, how often have you felt down, depressed, or hopeless?</strong></td>
<td></td>
<td></td>
<td>45.511 (6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>More than half of the days</td>
<td>101 (11.37)</td>
<td>81 (21.66)</td>
<td>182 (14.42)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Almost every day</td>
<td>41 (4.62)</td>
<td>28 (7.49)</td>
<td>69 (5.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>A few days</td>
<td>398 (44.82)</td>
<td>181 (48.40)</td>
<td>579 (45.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not at all</td>
<td>348 (39.19)</td>
<td>84 (22.46)</td>
<td>432 (34.23)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings: The Incidence of Tinnitus Increased After the COVID-19 Pandemic

Overall, tinnitus was rated as more bothersome during the pandemic than before. During the COVID-19 pandemic, like many case reports, the onset or aggravation of tinnitus was documented [17-20]. Of the 1262 participants, 540 (42.79%) reported experiencing some ear-related symptoms after being infected, and the most common ear-related symptoms were vertigo and tinnitus. Despite this, only 114 (9.03%) of the participants had seen a doctor regarding their ear symptoms, and 426 (33.75%) did not visit the hospital. COVID-19 infection may present with otological manifestations such as hearing loss, tinnitus, vertigo, ear fullness, and earache [18]. Even if some patients experienced ear symptoms, these were mostly transitory, although some studies have shown evidence for the worsening of preexisting tinnitus as a possible long-term effect of COVID-19 [21,22]. In our practice, we administer questionnaires to patients to assess how troublesome the tinnitus is for them, as well as how it affects their daily activity and ability to function. Most of the surveyed participants have experienced the distress caused by tinnitus, including its impact on attention, emotions, anxiety, depression, and sleep quality. Increased levels of anxiety, depression, and irritability have been reported [23]. These emotional factors are associated with exacerbated tinnitus annoyance. Negative emotions were made worse by a variety of worries and frustrations, such as relationship issues that increased because of confinement, concerns about the availability of food, and worries about contracting the virus. Due to loss of employment, furloughs, and a decrease in the value of investment, financial worries have also caused tinnitus to become significantly worse. These results are consistent with a recent systematic review on the effects of COVID-19 on mental health [24], which found reduced psychological well-being and increased levels of anxiety and depression in the general population compared with before COVID-19. However, less than 10% of individuals with tinnitus sought medical attention after COVID-19 infection, and the management and provision of health care for subsequent individuals with tinnitus need attention. Nocini et al [25] reported that the worldwide burden of tinnitus may have increased significantly during the ongoing COVID-19 pandemic. To improve tinnitus care, better awareness and more accessible resources and management are crucial. Tracking and managing hearing-related changes due to having COVID-19 will be important as clinical services resume.

COVID-19 is an infectious respiratory disease caused by severe acute respiratory syndrome and other systematic symptoms [26]. In some cases, the tinnitus resolved on its own after the resolution of the COVID-19 infection, while in other cases, it persisted even after recovery from the virus. In order to understand the prevalence of tinnitus and ear symptoms after COVID-19 infection, we conducted an epidemiological questionnaire-based survey at the end of 2022 and used regression analysis to identify factors affecting patients’ tinnitus. Table 4 shows the association of gender, age group, height, weight, BMI, education level, occupation, region in China, ENT disease history, hypertension history, and diabetes history in relation to tinnitus after COVID-19, and gender, age group, education level, region in China, and ENT disease history were added into the model. The regression model likelihood ratio test resulted in $\chi^2 = 39.310$ and $P < .001$. For the Hosmer–Lemeshow test, the results were $\chi^2 = 5.362$ and $P = .72$. The regression model was $\ln(p/1-p) = -1.364 - (0.211*\text{gender}) - (0.104*\text{age group}) - (0.106*\text{education level}) + (0.416*\text{region in China}) + (0.387*\text{ENT medical history})$. 

https://formative.jmir.org/2024/1/e54326
### Table 4. Regression analysis results of the China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic (CENTSS 2023).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariable analysis</th>
<th>Multivariable analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR^a (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Male</td>
<td>0.737 (0.541-1.004)</td>
<td>0.737 (0.541-1.004)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤30</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>31-36</td>
<td>0.691 (0.492-0.970)</td>
<td>0.682 (0.482-0.966)</td>
</tr>
<tr>
<td>37-43</td>
<td>0.604 (0.429-0.852)</td>
<td>0.599 (0.421-0.852)</td>
</tr>
<tr>
<td>≥44</td>
<td>0.72 (0.515-1.007)</td>
<td>0.698 (0.492-0.991)</td>
</tr>
<tr>
<td>Height</td>
<td>0.996 (0.981-1.011)</td>
<td>0.61</td>
</tr>
<tr>
<td>Weight</td>
<td>1.003 (0.993-1.013)</td>
<td>0.53</td>
</tr>
<tr>
<td>BMI</td>
<td>1.02 (0.987-1.055)</td>
<td>0.23</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior high school (or below)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Bachelor’s (or associate) degrees</td>
<td>0.422 (0.254-0.702)</td>
<td>0.401 (0.236-0.682)</td>
</tr>
<tr>
<td>Master’s (or above) degrees</td>
<td>0.32 (0.187-0.547)</td>
<td>0.308 (0.176-0.538)</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noise exposure occupation</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Non noise exposure occupation</td>
<td>0.927 (0.651-1.320)</td>
<td>0.976 (0.656-1.452)</td>
</tr>
<tr>
<td>Other occupation</td>
<td>0.976 (0.656-1.452)</td>
<td>—</td>
</tr>
<tr>
<td>Region in China</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Middle</td>
<td>1.985 (1.414-2.787)</td>
<td>1.891 (1.336-2.678)</td>
</tr>
<tr>
<td>West</td>
<td>2.034 (1.398-2.959)</td>
<td>2.063 (1.408-3.022)</td>
</tr>
<tr>
<td>ENT^c disease history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>1.448 (1.129-1.856)</td>
<td>1.530 (1.184-1.977)</td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>0.965 (0.579-1.609)</td>
<td>0.965 (0.579-1.609)</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>1.055 (0.455-2.448)</td>
<td>1.055 (0.455-2.448)</td>
</tr>
</tbody>
</table>

^aOR: odds ratio.

^bNot applicable.

^cENT: ear, nose, throat.

In our survey, female participants, younger participants (<30 years), less educated participants, participants from western China, and participants with a history of otolaryngology diseases were more likely to develop tinnitus after COVID-19 infection. There were 688 (688/888, 77.8%) female participants compared with 200 male participants, and this may have caused selective sex-related bias. Interestingly, this conclusion differs from the known risk factors for tinnitus before COVID-19 infection, such as long-term noise exposure, ototoxic drugs, aging, and genetic predispositions, and concomitant neural alterations are considered the initial source of tinnitus. There is also no unified conclusion on the relationship between tinnitus and gender. Several studies have reported a higher incidence of tinnitus among men than women [27] and that women, especially...
menopausal women, are more susceptible to tinnitus and developing annoying tinnitus [27,28]. Similar studies confirmed that women and younger individuals are more likely to complain about tinnitus [29]. Studies on COVID-19 revealed that ENT symptoms are observed more frequently in younger age groups and in women [30-32]. From the explanations provided, it appears that the unified conclusion on the relationship between tinnitus and gender may be partly attributed to significant lifestyle changes in these groups during the pandemic. These include changes in employment, increased childcare, and household responsibilities. A study involving patients with tinnitus at a clinic in Germany found that COVID-19 increased grief, frustration, stress, and nervousness, although there was only a small increase in tinnitus distress compared with 2 years prior to lockdown [31]. As documented in previous studies [33], the COVID-19 pandemic is a potential environmental stressor that might influence individually perceived tinnitus distress. Because not all people have been affected by the pandemic in the same way, the situation allows one to identify environmental factors and personality traits that impact tinnitus distress differently. The risk factors for tinnitus among respondents with a history of ENT problems were 1.53 times higher than in patients without a history of ENT problems, similar to the conclusions of the study by Mui et al [18] involving 365 older adult respondents in Australia and other countries where tinnitus was reported to be more bothersome during the pandemic, by 36% of respondents, whereas 59% reported no change and 5% reported less bothersome tinnitus. As for the regional distribution, people in the western region of China were 2.063 times more likely to report tinnitus compared with those in the eastern region, and the odds ratio in the central region was 1.891 compared with people in the eastern region. This may be because of economic and environmental factors.

Although it is still not entirely clear why some patients with COVID-19 develop tinnitus, it is speculated that this may be related to inflammation or other changes in the body as a result of the virus [34]. Regarding the possible etiologies of tinnitus during the COVID-19 pandemic, one plausible explanation may be attributed to direct viral injury to the ear. However, the elevated levels of anxiety and stress experienced by individuals during this global health crisis have been identified as exacerbating factors that may have significantly contributed to the worsening or propagation of this debilitating condition [35]. Therefore, the increased burden of anxiety and stress during the pandemic may have played a significant role in the worsening of tinnitus symptoms. More research is needed to explore the relationship between tinnitus and the psychosocial impact of the COVID-19 pandemic. Similar to previous research [32,36,37], we also found that tinnitus was more likely to occur in participants in the western region in China and in participants with low education levels. Differences in tinnitus onset between high- and low-income groups may be because the pressure of work and unemployment caused by COVID-19 infection was more readily reflected in people from the western region and in people with low education levels.

To highlight the impact of COVID-19 on the ear, our focus was on tinnitus along with other ear-related symptoms. COVID-19 infection can potentially lead to symptoms such as ear fullness, ear pain, vertigo, hearing loss, and even deafness. Ear fullness and ear pain may arise from the virus invading the external auditory canal, eustachian tube, or nerves innervating the outer ear. Vertigo may be caused by viral invasion of the inner ear, affecting both the vestibule and cochlea or the vestibulocochlear nerve. Hearing loss or deafness can result from the virus damaging the cochlear hair cells while sparing the vestibular system. It is important to note that different symptoms may indicate viral invasion of different parts of the ear, and this should be considered during diagnosis and treatment. Physicians must remain vigilant and take measures to minimize the detrimental effects of COVID-19 in order to protect the patients' auditory function to the greatest extent possible.

On a worldwide level [38-40], lower health development index and total health expenditure or gross domestic product per capita have been associated with a higher prevalence of moderate to severe ear symptoms.

**Limitations and Future Directions**

There are some limitations in the interpretation of this study that should be considered. First, our study has the limitations of a cross-sectional survey such as convenience sampling and possible recall bias. Second, this online survey potentially introduced selection bias because participants were limited to only those who could access the internet. Therefore, the generalizability of the findings needs to be interpreted with caution. Future studies should include longitudinal follow-up periods to identify the trajectory of the symptoms in order to indicate whether the tinnitus resolves or remains and if the severity changes.

**Conclusion**

This study found that tinnitus has been reported to be more distressing during the pandemic than before it, and individuals experiencing tinnitus after a COVID-19 infection were found to have poorer cognitive and emotional well-being. The presence of various ear-related symptoms in patients following a COVID-19 infection may indicate viral involvement in different parts of the ear. Therefore, it is crucial to monitor and address any changes in hearing associated with COVID-19 as clinical services resume.

**Acknowledgments**

This work was supported by the National Natural Science Foundation of China (grants 81970879, 82371146, 82192862), Shanghai Science and Technology Committee (STCSM) Science and Technology Innovation Program (grant 20MC1920200), and Research Projects of Shanghai Municipal Health Committee (grants 2020YJZX0110 and 2022XD059).

It is important to note that ChatGPT was not used in the ideation or writing process.
Data Availability
The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions
All authors made a significant contribution to the work reported in the conception, study design, execution, acquisition of data, analysis, and interpretation or in all these areas; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys).
[PDF File (Adobe PDF File), 216 KB - format_ve_v8i1e54326_app1.pdf ]

References
15. How are your facial features after COVID-19 infection? Weixin. 2023 Jan 20. URL: https://mp.weixin.qq.com/s?__biz=MzA4NTk4NjQ3NzA==&mid=2650678240&idx=2&sn=cbd364ea10905fe112a4ddce8c26f64a&chksm=87c51e25b0b29733d7de8e12167775f769033000b63e0ec0663fd37b2ea8accf9202bd2edf6 [accessed 2024-04-08]


Abbreviations

CENTSS: China Ear Nose and Throat Symptom Survey in the COVID-19 Pandemic
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
ENT: ear, nose, throat
NICE: National Institute for Health and Care Excellence
Changes in the Clinical Practice of Mental Health Service Providers Throughout the COVID-19 Pandemic: Longitudinal Questionnaire Study

Milena Gotra¹, PhD; Katharine Lindberg², PhD; Nicholas Jasinski¹, PsyD; David Scarisbrick¹,³, PhD; Shannon Reilly⁴, PhD; Jonathan Perle¹, PhD; Liv Miller¹,³, PsyD; James Mahoney III¹,³, PhD

¹Department of Behavioral Medicine and Psychiatry, Rockefeller Neuroscience Institute, West Virginia University School of Medicine, Morgantown, WV, United States
²Atrium Health, Charlotte, NC, United States
³Department of Neuroscience, Rockefeller Neuroscience Institute, West Virginia University School of Medicine, Morgantown, WV, United States
⁴Department of Neurology, University of Virginia Health, Charlottesville, VA, United States

Corresponding Author:
Nicholas Jasinski, PsyD
Department of Behavioral Medicine and Psychiatry
Rockefeller Neuroscience Institute
West Virginia University School of Medicine
64 Medical Center Dr
Morgantown, WV, 26506
United States
Phone: 1 304 598 4740
Email: nicholas.jasinski1@hsc.wvu.edu

Abstract

Background: The COVID-19 pandemic impacted the practices of most mental health providers and resulted in a rapid transition to providing telemental health services, changes that were likely related to stay-at-home policies as well as increased need for services.

Objective: The aim of this study was to examine whether these changes to practice have been sustained over time throughout the course of the COVID-19 pandemic and whether there are differences among mental health provider type and setting. We hypothesized that there would be an increase in the number of patients seen in person after the initial surge of the pandemic in spring 2020 and subsequent discontinuation of stay-at-home policies, though with continued implementation of telemental health services across settings.

Methods: This study surveyed 235 of the 903 mental health providers who responded to a survey in spring 2020 (Time point 1) and at a 1-year follow-up in spring 2021 (Time point 2). Differences in practice adjustments, factors related to telemental health, and number of patients seen were examined across provider type (social worker, psychologist, neuropsychologist) and setting (academic medical center [AMC], community mental health, private practice, and Veterans Affairs hospital).

Results: From Time point 1 to Time point 2, there was a small but significant increase in the overall number of providers who were implementing telehealth (191/235, 81% to 204/235, 87%, \(P=.01\)) and there was a significant decline in canceled or rescheduled appointments (25%-50% in 2020 to 3%-7% in 2021, \(P<.001\)). Psychologists and providers working at AMCs reported decreased difficulty with telehealth implementation (\(P<.001\)), and providers working at AMCs and in private practice settings indicated they were more likely to continue telehealth services beyond spring 2021 (\(P<.001\)). The percent of time working remotely decreased overall (78% to 59%, \(P<.001\)), which was most notable among neuropsychologists and providers working at an AMC. There was an overall increase in the average number of patients seen in person per week compared with earlier in the pandemic (mean 4.3 to 8.7, \(P<.001\)), with no change in the number of patients seen via telehealth (mean 9.7 to 9.9, \(P=.66\)).

Conclusions: These results show that the rapid transition to telemental health at the onset of the COVID-19 pandemic in spring 2020 was sustained over the next year, despite an overall increase in the number of patients seen in person. Although more providers reported returning to working on-site, over 50% of providers continued to use a hybrid model, and many providers reported they would be more likely to continue telemental health beyond spring 2021. This suggests the continued importance...
and reliance on telemental health services beyond the acute pandemic phase and has implications for future policies regulating the availability of telemental health services to patients.

(MJMIR Form Res 2024;8:e50303) doi: 10.2196/50303

KEYWORDS
COVID-19 pandemic; mental health; social worker; psychologist; neuropsychologist; academic medical center; community mental health; private practice; Veteran’s Affairs hospital; longitudinal questionnaire study; COVID-19; implementation; telemental health; hybrid model; availability

Introduction
In March 2020, the World Health Organization declared the COVID-19 outbreak, caused by SARS-CoV-2 infection, a global pandemic [1]. Governments around the globe began implementing various strategies to contain the spread of COVID-19 including recommendations for social distancing, restrictions on in-person gatherings and contact, and international travel limitations, among many others [2,3]. In particular, the recommendations for limited interpersonal contact presented mental health service providers with a unique challenge in an attempt to maintain adequate access to services during a time in which mental health care needs were on the rise [4-7].

As a result of these challenges, numerous researchers began exploring the impact of the COVID-19 pandemic on the provision of mental health services [8,9]. Duden et al [9] conducted a systematic review of 29 international studies documenting the COVID-19 pandemic’s impact on mental health services from December 2019 through March 2022. They indicated 9 major topic areas in mental health services impacted by the COVID-19 pandemic, including (1) lack of preparedness versus timely response and flexible solutions; (2) changes in access, referral, and admission; (3) impacts on outpatient, community, and psychosocial services; (4) inpatient reorganization of hospital psychiatric units and acute wards; (5) diagnostic and therapeutic adaptations (including rapid deployment of telemental health services); (6) effects on medication; (7) infection control measures; (8) changes in patients’ demands, engagement, and mental health; and (9) impacts on staff and team.

Although there have been longitudinal assessments of the impact of the pandemic on patients [10-16], there has been a paucity of published longitudinal data to examine how provider practices have changed throughout the course of the pandemic. Although various studies have collected data at different time points during the pandemic, these cross-sectional evaluations offer little in the way of understanding how providers’ perceptions and activities were modified over time, a limitation which has previously been recognized [8,17-22]. Given the ebb and flow of different COVID-19 variants and changing guidelines in dealing with the ongoing pandemic (eg, reduced distancing requirements, lifting of travel restrictions), it is possible, and likely probable, that provider practice adjustments may subsequently change over time.

This manuscript represents a follow-up to the study conducted by Reilly et al [23] in 2020. Briefly, the results from the original investigation found that over 97% of a sample of US mental health providers (n=903) made practice adjustments as a result of the developing pandemic. The most consistent change identified was the increasing use of telemental health appointments, with nearly 80% of respondents adopting this service delivery method. Other findings of note included providers’ generally positive perceptions of telemental health preparedness, the desire of providers to continue providing telemental health services in the future, reduced rate of patient contacts, and the variations in practice modifications made by different types of US mental health professionals. The broader mental health practice literature has suggested that US mental health providers across disciplines are moving to a hybrid model of mixed telemental health and in-person services [8,9,24-26].

This study sought to address the impact of the continued COVID-19 pandemic by examining longitudinal data related to changes in the clinical practice of US mental health providers from December 2019 through March 2021. We predicted that, given the progressive relaxation of stay-at-home policies and social restrictions, there would be an overall increase in the number of patients seen and fewer cancellations of appointments from spring 2020 to spring 2021. We further predicted there would be continued high adoption of telemental health services and that providers would have easier access to telemental health infrastructure (eg, IT services, technology). Among provider type, we predicted that neuropsychologists would see fewer patients virtually and would have a greater increase in in-person services, likely related to greater difficulty performing services such as neuropsychological testing via telehealth.

Methods
Recruitment
In March 2021, participants who completed the initial survey [23] and indicated at that time that they were interested in participating in possible follow-up surveys (and provided their email address) were contacted via email. The recruitment email included a Qualtrics survey link. All questions were optional, and participants were informed they could discontinue participation at any time. Data collection was completed, and the survey was closed in April 2021.

Data Collection
The follow-up survey was similar in structure and content to the initial survey [23]. Participants were asked to provide information about their demographics, patient populations, practice adjustments in response to COVID-19, perceptions of their employer’s response, and their emotional response and perceptions regarding the COVID-19 pandemic (see Multimedia...
Appendix 1). Additionally, participants were asked about changes in their status since the initial study (eg, household income, employment, training) as well as questions regarding coping strategies and vaccination status. For the Time point 2 (T2) study, precautionary measures were not included as a separate practice adjustment.

For applicable questions, participants were asked about their practices in December 2020 (to match the initial survey that asked about practice in December 2019) and their “current” practices (ie, when they completed the survey between March 2021 and April 2021). Of the 626 initial survey participants who provided their email for follow-up work and were contacted regarding the follow-up survey, 235 participants were included in the final sample. Data were excluded if participants completed less than 66% of the follow-up survey (n=37). We excluded 3 participants because they did not have complete survey data for both time points.

Data Preparation

Less than 5% of the data were missing for each variable of interest, with the exception of the number of patients seen via telehealth in December 2019 (21/235, 8.9%) and percent of time working remotely during the initial survey (18/235, 7.7%) and this follow-up survey (34/235, 14.5%). Missing data were addressed using pairwise deletion. Similar to the initial study, 3 participants who identified as marriage and family therapists were recoded as a master’s-level provider given the small sample size for this category. Participants with a discrepancy between their provider type and highest education (eg, provider with a master’s degree who identified as a psychologist or doctoral-level provider, n=6) were recoded to reflect their level of education. Square root transformations were conducted on continuous variables to address normality and reduce outliers.

Statistical Analysis

Analyses were conducted in SPSS Version 26 (IBM Corp). Outcome variables were compared between the initial survey (Time point 1 [T1]) and follow-up survey (T2) for the overall sample and across provider type (social worker/master’s-level provider vs psychologist/doctoral-level provider vs neuropsychologist) and setting (academic medical center [AMC] vs community mental health vs private practice [PP] vs Veterans Affairs hospital). We chose to limit our analyses to these settings and provider types as these were the largest sample sizes that were also examined in the T1 study [23]. The related-sample McNemar test was used to compare binomial variables among groups (eg, practice adjustments, whether participants were implementing telehealth, whether participants had easy access to IT services). The Wilcoxon signed rank test was used to compare differences in variables ranked on a Likert scale (eg, difficulty of implementing telehealth appointments, likelihood of providing telehealth services in the future). Paired sample t tests were used to compare changes from T1 to T2 for continuous variables for the overall sample. Differences across provider type and setting from T1 to T2 were evaluated using a mixed ANOVA (time x group) for continuous variables, with Bonferroni-corrected post hoc tests. Continuous variables included the percent of time working remotely and the number of in-person, remote, and total patients seen weekly. The number of patients seen was examined across 4 time points, including December 2019 (T1), March 2020 to April 2020 (T1), December 2020 (T2), and March 2021 to April 2021 (T2). Percent of time working remotely was calculated only for those who reported a value greater than 0% (n=135). Analyses were evaluated as significant with a false discovery rate correction for multiple comparisons [27].

Ethical Considerations

This study was determined to be exempt following the research ethics review by the Institutional Review Board of West Virginia University. Participants consented to participate by submitting their survey responses. Participant data were deidentified and matched with the unique identifier from the first survey. Participants were not compensated for completing the survey.

Results

Demographic Characteristics

Participant demographic characteristics are presented in Tables 1 and 2. There were no significant differences in demographic characteristics of participants who responded to the follow-up study compared with the full sample for the initial study [23]. There was a slightly higher proportion of psychologists who responded to the follow-up study (111/235, 47.2% compared with 367/903, 40.64%), though other professional characteristics were comparable. At the time of the T2 study, 15% (34/235) of respondents reported a change of work status since April 2020, due to change of jobs, fewer overall hours, starting new training (ie, new practicum placement, starting internship and postdoctoral fellowship), starting a new position, being laid off, and changing practice settings. Of the respondents, 17% (41/235) reported a change to their provider type since April 2020, and 9% (22/235) of respondents reported moving or changing the state they are employed in since April 2020.
Table 1. Demographic characteristics of participants who completed the follow-up survey (n=235).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>38.7 (10.4)</td>
</tr>
<tr>
<td><strong>Sex/gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>205 (87.2)</td>
</tr>
<tr>
<td>Man</td>
<td>30 (12.8)</td>
</tr>
<tr>
<td><strong>Race/ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>212 (90.2)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>7 (3.0)</td>
</tr>
<tr>
<td>Asian/Asian American</td>
<td>6 (2.6)</td>
</tr>
<tr>
<td>Hispanic/Latinx</td>
<td>6 (2.6)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>4 (1.7)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>199 (84.7)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>17 (7.2)</td>
</tr>
<tr>
<td>Lesbian/gay</td>
<td>8 (3.4)</td>
</tr>
<tr>
<td>Other (ie, queer, fluid, pansexual)</td>
<td>11 (4.7)</td>
</tr>
<tr>
<td><strong>Region, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>106 (45.1)</td>
</tr>
<tr>
<td>Midwest</td>
<td>56 (23.8)</td>
</tr>
<tr>
<td>West</td>
<td>43 (18.3)</td>
</tr>
<tr>
<td>Northeast</td>
<td>30 (12.8)</td>
</tr>
<tr>
<td><strong>Work status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>179 (76.2)</td>
</tr>
<tr>
<td>Part-time</td>
<td>20 (8.5)</td>
</tr>
<tr>
<td>Trainee</td>
<td>35 (14.9)</td>
</tr>
<tr>
<td>Not currently employed</td>
<td>1 (0.4)</td>
</tr>
</tbody>
</table>
Table 2. Professional characteristics of respondents who completed the follow-up survey (n=235).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Results, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provider type</strong></td>
<td></td>
</tr>
<tr>
<td>Psychologist, doctoral-level therapist, counselor</td>
<td>111 (47.2)</td>
</tr>
<tr>
<td>Neuropsychologist</td>
<td>42 (17.9)</td>
</tr>
<tr>
<td>Trainee (ie, practicum student, predoctoral intern, postdoctoral fellow)</td>
<td>35 (14.9)</td>
</tr>
<tr>
<td>Social worker, master’s-level therapist, counselor</td>
<td>32 (13.6)</td>
</tr>
<tr>
<td>Other (eg, support staff, bachelor-level therapist, physician)</td>
<td>15 (6.4)</td>
</tr>
<tr>
<td><strong>Provider level</strong></td>
<td></td>
</tr>
<tr>
<td>Licensed practitioner</td>
<td>146 (62.1)</td>
</tr>
<tr>
<td>Licensed practitioner and board certified in specialty area</td>
<td>38 (16.2)</td>
</tr>
<tr>
<td>Unlicensed practitioner</td>
<td>16 (6.8)</td>
</tr>
<tr>
<td><strong>Current practice setting</strong></td>
<td></td>
</tr>
<tr>
<td>Private practice</td>
<td>55 (23.4)</td>
</tr>
<tr>
<td>Academic medical center</td>
<td>52 (22.1)</td>
</tr>
<tr>
<td>Multiple practice settings</td>
<td>34 (14.5)</td>
</tr>
<tr>
<td>Veterans Affairs hospital or military hospital/clinic</td>
<td>25 (10.6)</td>
</tr>
<tr>
<td>Community mental health setting</td>
<td>18 (7.7)</td>
</tr>
<tr>
<td>Psychiatric hospital or facility</td>
<td>10 (4.3)</td>
</tr>
<tr>
<td>General hospital</td>
<td>8 (3.4)</td>
</tr>
<tr>
<td>Rehabilitation hospital or setting</td>
<td>8 (3.4)</td>
</tr>
<tr>
<td>Other (eg, counseling center, training clinic, prison, school, primary care, specialty clinic, outpatient program, nonprofit organization, online program, insurance company)</td>
<td>24 (10.2)</td>
</tr>
<tr>
<td><strong>Age specialty</strong></td>
<td></td>
</tr>
<tr>
<td>Adults only (ie, 18 years and older)</td>
<td>122 (51.9)</td>
</tr>
<tr>
<td>Lifespan (ie, pediatrics and adults)</td>
<td>92 (39.1)</td>
</tr>
<tr>
<td>Pediatric only (ie, younger than 18 years)</td>
<td>21 (8.9)</td>
</tr>
</tbody>
</table>

a1 respondent did not provide information about their practice setting.
bTime 2 results (follow-up survey) were based on those the participants reported at the Time 1 study (initial survey).

**Overall Sample**

Characteristics of changes in practice across the total sample are presented in Table 3. Slightly more participants reported implementing telehealth services overall, though the percent of participants who endorsed scheduling virtual instead of in-person visits as a practice adjustment decreased by approximately 10% (169/235) from T1 to T2. Additionally, the percent of time working remotely decreased by approximately 20%. Significantly fewer participants were rescheduling and canceling appointments, though participants endorsed continued use of appointment restrictions (eg, by patient age, medical comorbidity, or recent travel). More participants endorsed no change in practice and reported using a greater variety of “other” practice adjustments, including increasing precautions, adding more patients to their schedule, offering reduced fees, using hybrid virtual and in-person models, and adjusting their test batteries. Although there was no change in ease of access to IT services, participants reported a decrease in difficulty with telehealth implementation from a median of 3 (not easy or difficult) at T1 to 2 (somewhat easy) at T2. Participants reported an increase in the likelihood of continuing telemental health services in the future from a median of 4 (somewhat likely) at T1 to 5 (very likely) at T2.

The total number of patients seen weekly significantly decreased from December 2019 to March 2020/April 2020 ($F_{2.01,463.37}=20.94$, $P<.001$; partial $\eta^2=0.08$; mean difference $[M_{diff}]=3.03$, $P<.001$; Figure 1) though returned to pre-COVID-19 levels by December 2020 ($M_{diff}=-3.66$, $P<.001$) and continued to increase by March 2021/April 2021 ($M_{diff}=-0.91$, $P<.001$). The number of patients seen in person showed a similar pattern of initial decline from December 2019 to March 2020/April 2020 ($F_{2.25,481.36}=113.56$, $P<.001$; partial $\eta^2=0.35$; $M_{diff}=11.54$, $P<.001$). Although the number of in-person patients subsequently increased (March 2020/April 2020 to December 2020: $M_{diff}=-3.24$, $P<.001$; December 2020 to March 2021/April 2021: $M_{diff}=-1.36$, $P<.001$), they did not return to pre-COVID-19 levels. The number of patients seen...
virtually initially increased from December 2019 to March 2020/April 2020 ($F_{2,31,476.13}=113.62, P<.001$; partial $\eta^2=0.36$; $M_{\text{diff}}=-9.04, P<.001$) and remained stable through March 2021/April 2021.

### Table 3. Descriptive statistics of changes in clinical practice from Time 1 to Time 2.

<table>
<thead>
<tr>
<th>Clinical practice</th>
<th>Time 1 study</th>
<th>Time 2 study</th>
<th>Statistic (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Practice adjustments, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telemental health or virtual appointments (vs in-person)</td>
<td>194 (82.6)</td>
<td>169 (71.9)</td>
<td>$10.9 (1)^b$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Rescheduling or postponing appointments</td>
<td>118 (50.2)</td>
<td>16 (6.8)</td>
<td>$96.2 (1)^b$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Canceling appointments</td>
<td>60 (25.5)</td>
<td>7 (3.0)</td>
<td>$47.4 (1)^b$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Restrictions on appointments</td>
<td>32 (13.6)</td>
<td>35 (14.9)</td>
<td>0.08 (1)$^b$</td>
<td>.78</td>
</tr>
<tr>
<td>Other adjustment</td>
<td>6 (2.6)</td>
<td>52 (22.6)</td>
<td>$37.5 (1)^b$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>N/A$^a$ (no change in practice)</td>
<td>6 (2.6)</td>
<td>29 (12.3)</td>
<td>$19.4 (1)^b$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Telemental health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported currently not implementing telemental health, n (%)</td>
<td>44 (19.3)</td>
<td>31 (13.2)</td>
<td>$7.5 (1)^c$</td>
<td>.01</td>
</tr>
<tr>
<td>Amount of the week working remotely (%), mean (SD)</td>
<td>77.92 (33.6)</td>
<td>58.46 (37.2)</td>
<td>5.1 (134)$^c$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reported easy access to IT services, n (%)</td>
<td>180 (76.6)</td>
<td>177 (75.3)</td>
<td>0.10 (1)$^c$</td>
<td>.75</td>
</tr>
<tr>
<td>Difficulty with telemental health implementation$^d$, mean (SD)</td>
<td>3.1 (1.2)</td>
<td>2.7 (1.1)</td>
<td>$-5.0^e$</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Likelihood of continuing telemental health services$^f$, mean (SD)</td>
<td>3.6 (1.4)</td>
<td>4.2 (1.1)</td>
<td>$-5.4^e$</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

$^a$N/A: not applicable.
$^b$Chi-squared.
$^c$t test.
$^d$5-point Likert scale (1=easy or not at all difficult to 5=very difficult).
$^e$Z score.
$^f$5-point Likert scale (1=very unlikely to 5=very likely).

### Figure 1. Changes in the number of patients seen weekly for the overall sample.

![Graph showing changes in the number of patients seen weekly for the overall sample.](https://formative.jmir.org/2024/1/e50303)

### Provider Type

Characteristics of changes in practice by provider type are presented in Table 4. Only psychologist/doctoral-level providers worked significantly less time per week remotely and had a significant decrease in the percent of participants who reported implementing virtual instead of in-person visits as a practice adjustment. All providers had a decrease in the percent of
appointments being rescheduled or postponed, and psychologist/doctoral-level providers and neuropsychologists had a decrease in the percent of appointments being canceled. Psychologist/doctoral-level providers and neuropsychologists reported implementing an increase in “other” practice adjustments. “Other” practice adjustments for psychologist/doctoral-level providers included safety precautions, adding extra appointments, smaller therapy groups, hybrid models, and reducing caseloads. Adjustments for neuropsychologists included safety precautions, limits to the number of family members who can attend appointments, and adjustments to test batteries, and hybrid models. Only psychologist/doctoral-level providers had a significant increase in the percent of providers who were not implementing any changes to their current practice. Only psychologist/doctoral-level providers had a significant decrease in difficulty with telemental health implementation and indicated an increased likelihood of continuing telemental health services in the future.

Table 4. Descriptive statistics of the changes in clinical practice from Time 1 to Time 2 across provider type.

<table>
<thead>
<tr>
<th>Clinical practice</th>
<th>Social workers, master’s degree therapists (n=32)</th>
<th>Psychologists, doctoral therapists (n=111)</th>
<th>Neuropsychologists (n=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Practice adjustments, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telemental health or virtual appointments (vs in person)</td>
<td>29 (90.6) 26 (81.3) .25</td>
<td>98 (88.3) 84 (75.7) .002</td>
<td>29 (69.0) 26 (61.9) .63</td>
</tr>
<tr>
<td>Rescheduling or postponing appointments</td>
<td>13 (40.6) 0 &lt;.001</td>
<td>42 (37.8) 5 (4.5) &lt;.001</td>
<td>36 (85.7) 8 (19.0) &lt;.001</td>
</tr>
<tr>
<td>Canceling appointments</td>
<td>6 (18.8) 0 .04a</td>
<td>20 (18.0) 4 (3.6) &lt;.001</td>
<td>14 (33.3) 2 (4.8) .003</td>
</tr>
<tr>
<td>Restrictions on appointments</td>
<td>3 (9.4) 2 (6.3) ≥.99</td>
<td>12 (10.8) 12 (10.8) ≥.99</td>
<td>11 (26.2) 10 (23.8) ≥.99</td>
</tr>
<tr>
<td>Other adjustment</td>
<td>0 5 (15.6) .06</td>
<td>1 (2.7) 21 (19.4) &lt;.001</td>
<td>0 18 (43.9) &lt;.001</td>
</tr>
<tr>
<td>N/Ab (no change in practice)</td>
<td>0 3 (9.4) .25</td>
<td>2 (1.8) 12 (10.8) .002</td>
<td>2 (4.8) 6 (14.3) .13</td>
</tr>
<tr>
<td><strong>Telemental health</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported not implementing telemental health currently, n (%)</td>
<td>1 (3.3) 2 (6.3) ≥.99</td>
<td>10 (9.2) 9 (8.1) .69</td>
<td>15 (36.6) 9 (21.4) .15</td>
</tr>
<tr>
<td>Percent of week working remotely, mean (SD)</td>
<td>81.4 (34.8) 57.7 (40.1) .04a</td>
<td>83.1 (29.8) 62.9 (38.3) &lt;.001</td>
<td>56.9 (38.5) 43.3 (32.1) .08</td>
</tr>
<tr>
<td>Easy access to IT services, n (%)</td>
<td>25 (78.1) 24 (75.0) ≥.99</td>
<td>80 (72.1) 80 (72.1) ≥.99</td>
<td>33 (78.6) 35 (83.3) .69</td>
</tr>
<tr>
<td>Difficulty with telemental health implementationc, mean (SD)</td>
<td>2.9 (1.1) 2.5 (1.1) .04a</td>
<td>3.0 (1.2) 2.5 (1.0) &lt;.001</td>
<td>3.4 (1.3) 3.0 (1.1) .46</td>
</tr>
<tr>
<td>Likelihood of continuing to provide telemental health servicesd, mean (SD)</td>
<td>3.8 (1.3) 4.12 (1.1) .03a</td>
<td>3.9 (1.4) 4.4 (1.1) .001</td>
<td>3.4 (1.5) 4.1 (1.0) .02a</td>
</tr>
</tbody>
</table>

aNot significant after false discovery rate correction.
bN/A: not applicable.
c5-point Likert scale (1=easy or not at all difficult to 5=very difficult).
d5-point Likert scale (1=very unlikely to 5=very likely).

The mixed ANOVA showed that the total number of patients seen for all providers decreased at the start of the COVID-19 pandemic (March 2020/April 2020) though increased by December 2020 ($F_{1.99,357.86}=8.09$, $P<.001$; partial $\eta^2=0.04$; Figure 2). Neuropsychologists saw fewer patients than psychologist/doctoral-level providers and social worker/master’s-level providers ($F_{2.186}=16.19$, $P<.001$; partial $\eta^2=0.15$), and the time x provider type interaction was not significant ($F_{3.98,357.86}=1.46$, $P=.22$). For the number of patients seen in person, there was an initial decrease at the start of the COVID-19 pandemic, with an increase by December 2020, though the number of in-person patients did not return to prepandemic levels ($F_{2.26,372.02}=64.10$, $P<.001$; partial $\eta^2=0.28$). The time x provider type interaction was significant ($F_{4.51,372.02}=7.30$, $P<.001$; partial $\eta^2=0.08$) and showed that neuropsychologists saw fewer patients than other providers in December 2019, though this did not differ from other providers at subsequent time points. For the number of patients seen virtually, there was an initial decrease at the start of the COVID-19 pandemic, which remained stable over time ($F_{2.37,376.04}=73.45$, $P<.001$; partial $\eta^2=0.32$). Neuropsychologists
saw fewer patients overall than other providers \( (F_{2,159} = 28.93, \ P < .001; \text{partial } \eta^2 = 0.27) \). The time x provider type interaction was significant \( (F_{4,73,376.04} = 11.50, \ P < .001; \text{partial } \eta^2 = 0.13) \) and showed that, although there was no difference among providers in the number of patients seen virtually in December 2019, neuropsychologists saw fewer patients virtually than other providers across all other time points.

**Figure 2.** Changes in the number of patients seen weekly by provider type and location: (A) total patient population, (B) patients seen in person, (C) patients seen virtually.

**Setting**

Characteristics of the changes in practice by clinical practice setting are presented in [Table 5](#). Only AMC providers had a significant decrease in the amount of time per week working remotely and in the percent of participants who reported implementing virtual instead of in-person visits as a practice adjustment. All settings had a decrease in the percent of appointments being rescheduled or postponed, though only AMCs and PP had a decrease in appointments being canceled. AMCs and PP reported an increase in implementation of “other” practice adjustments. “Other” practice adjustments for AMCs included safety precautions, hybrid models, offering telehealth as an option, and adjusting test batteries. Adjustments for PP included safety precautions, adding extra appointments, hybrid models, adjustments to test batteries, and reducing caseloads. Only AMCs had a significant increase in the percent of providers who were not implementing any changes to their current practice. Only AMCs had a significant decrease in difficulty with telemental health implementation. AMCs and PPs indicated an increased likelihood of continuing telemental health services in the future.

The mixed ANOVA showed that the total number of patients seen across settings decreased at the start of the COVID-19 pandemic (March 2020/April 2020), though it increased by December 2020 \( (F_{2,08,298.79} = 14.22, \ P < .001; \text{partial } \eta^2 = 0.09) \). There were no differences across settings, and the time x setting interaction was not significant \( (F_{6,23,298.79} = 1.63, \ P = .14) \). For the number of patients seen in person, there was an initial decrease across settings at the start of the COVID-19 pandemic, with an increase by December 2020, though the number of in-person patients seen did not return to prepandemic levels \( (F_{2.27,297.96} = 64.39, \ P < .001; \text{partial } \eta^2 = 0.33) \). The time x setting interaction was significant \( (F_{6.82,297.90} = 2.28, \ P = .03); \)
partial $\eta^2=0.05$), and the follow-up tests of simple main effects indicated that all settings had an increase in the number of patients seen in person after March 2020/April 2020 except for PPs. For the number of patients seen virtually, there was an initial increase at the start of the COVID-19 pandemic, which remained stable over time ($F_{2.40,304.15}=79.72$, $P<.001$; partial $\eta^2=0.39$). There were no differences across settings, and the time x setting interaction was not significant ($F_{7.19,304.15}=1.15$, $P=.33$).

Table 5. Descriptive statistics of changes in clinical practice from Time 1 to Time 2 across settings.

<table>
<thead>
<tr>
<th>Clinical practice</th>
<th>Academic medical center (n=52)</th>
<th>Community mental health (n=18)</th>
<th>Private practice (n=55)</th>
<th>Veterans Affairs (n=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
<td>$P$ value</td>
<td>Time 1</td>
</tr>
<tr>
<td><strong>Practice adjustments, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telemental health or virtual appointments (vs in person)</td>
<td>48 (91.3)</td>
<td>31 (59.6)</td>
<td>&lt;.001</td>
<td>14 (77.8)</td>
</tr>
<tr>
<td>Rescheduling or postponing appointments</td>
<td>29 (55.8)</td>
<td>3 (5.8)</td>
<td>&lt;.001</td>
<td>0 (55.6)</td>
</tr>
<tr>
<td>Canceling appointments</td>
<td>16 (30.8)</td>
<td>3 (5.8)</td>
<td>.001</td>
<td>3 (16.7)</td>
</tr>
<tr>
<td>Restrictions on appointments</td>
<td>8 (15.4)</td>
<td>7 (13.5)</td>
<td>$\geq$.99</td>
<td>2 (11.1)</td>
</tr>
<tr>
<td>Other adjustment</td>
<td>1 (1.9)</td>
<td>10 (19.6)</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td>N/A$^b$ (no change in practice)</td>
<td>1 (1.9)</td>
<td>11 (21.2)</td>
<td>.002</td>
<td>0</td>
</tr>
<tr>
<td><strong>Telemental health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported not implementing telemental health currently, n (%)</td>
<td>5 (10.0)</td>
<td>6 (11.5)</td>
<td>$\geq$.99</td>
<td>5 (27.8)</td>
</tr>
<tr>
<td>Amount of week working remotely (%), mean (SD)</td>
<td>84.8 (27.6)</td>
<td>45.7 (39.1)</td>
<td>.002</td>
<td>77.9 (36.8)</td>
</tr>
<tr>
<td>Easy access to IT services, n (%)</td>
<td>49 (94.2)</td>
<td>51 (98.1)</td>
<td>.63</td>
<td>14 (77.8)</td>
</tr>
<tr>
<td>Difficulty with telemental health implementation$^c$, mean (SD)</td>
<td>3.3 (1.1)</td>
<td>2.7 (1.0)</td>
<td>&lt;.001</td>
<td>3.2 (1.2)</td>
</tr>
<tr>
<td>Likelihood of continuing to provide telemental health services$^d$, mean (SD)</td>
<td>3.5 (1.2)</td>
<td>4.4 (0.8)</td>
<td>.001</td>
<td>3.3 (1.4)</td>
</tr>
</tbody>
</table>

$^a$Not significant after false discovery rate correction.

$^b$N/A: not applicable.

$^c$5-point Likert scale (1=easy or not at all difficult to 5=very difficult).

$^d$5-point Likert scale (1=very unlikely to 5=very likely).
Discussion

This Study

The COVID-19 health crisis was far-reaching, and adjustments in the way clinicians and patients interfaced with health care were necessary. Mental health services were highlighted as crucial throughout the pandemic given the broad day-to-day life changes required to reduce transmission of the virus. During our initial survey [23], we identified significant mental health practice changes and highlighted the implementation of telehealth, which was utilized sporadically prior to the pandemic. With this study, we hoped to create a longitudinal understanding of mental health practices as the pandemic became an everyday part of life. Others have also discovered that the hybrid model of mixed telehealth and in-person services is well accepted by mental health practitioners and their patients [28]. Given the initial findings, we predicted that there would be continued acceptance of adopting telemental health services, fewer canceled appointments, easier access to telehealth technology, increased access to mental health services, and an increase in in-person visits as pandemic restrictions were lifted.

Principal Findings

A year into the COVID-19 pandemic, mental health practitioners reported a significant decrease in the amount of time working remotely, and fewer providers reported offering virtual instead of in-person visits. However, despite some decrease in reliance on telemental health services, the overall number of providers using telemental health services continued to increase, and nearly all providers reported using telemental health in some capacity. Providers continued to work an average of 3 days per week remotely, which is consistent with prior work suggesting that mental health services are moving to a hybrid model of practice [8,9,24-26]. Hybrid models can offer greater flexibility for both patients and providers, though they can pose challenges regarding security of health information, reduced informal social interaction, and access to reliable technology. Although participants in our study reported decreased difficulty with telemental health implementation, this decrease was most notable for psychologists and AMC providers, who may have additional resources that are not readily available to providers in a PP setting. Importantly, providers overall endorsed an increased likelihood of continuing to provide telemental health services in the future, which highlights that continued refinement
of infrastructure in the workplace and home to support these hybrid models is crucial.

The change in patients seen throughout the first year of the COVID-19 pandemic was consistent with predictions. There was an initial decline in the total number of patients and number of in-person patients seen from shortly prior to the start of the pandemic (December 2019) to shortly after most lockdown protocols were implemented (March 2020/April 2020). This coincided with an increase in the number of patients seen virtually. Although the total number of patients seen weekly increased back to prepandemic levels, this was primarily driven by the increase in the number of patients seen virtually, as the number of patients seen in person did not yet return to prepandemic levels by March 2021/April 2021. This increase was consistent with providers’ reporting canceling and rescheduling fewer in-person appointments compared with the start of the pandemic, which was likely related to relaxation of initial stay-at-home orders and implementation of clinic procedures to reduce the risk of transmission. Providers continued to implement restrictions on appointments and other methods of infection control, such as policies regarding recent travel, sanitization practices, and symptom screening, which may have resulted in greater comfort with increasing in-person visits. It is important to note that the number of patients seen virtually remained stable from March 2020/April 2020 to March 2021/April 2021, which resulted in a general trend of more patients being seen overall as the number of in-person patients increased. This finding could be related to increased provider availability due to improved efficiency and comfort with telehealth services as the pandemic progressed [18]. It is also possible that this trend may be due to greater prevalence of mood symptoms resulting in higher demand for mental health services throughout the pandemic [29,30]. It will be important to continue monitoring whether this increase in the total number of patients continues to increase, as increased burden on the health care workers due to the COVID-19 pandemic has contributed to significant burnout and may reduce quality of care provided [31,32].

There were also several findings that were unique to certain settings and provider types. Neuropsychologists saw fewer total patients and virtual patients when compared with social workers and psychologists, though the number of patients seen in person after the start of the pandemic in March 2020/April 2020 did not differ. This was a somewhat expected finding given the time each provider might devote to a single patient (eg, 2-8 hours of neuropsychological assessment for 1 patient versus 1-hour sessions with numerous cases per day), though this highlights unique challenges that neuropsychologists may have had regarding utilizing telemental health in their practice. Patient interviews and feedback sessions are readily adaptable to a virtual format, though procedures for conducting testing in a virtual format are still being developed and are not widely implemented [33]. Among settings, providers who worked at AMCs initially reported the highest amount of time worked remotely (85% per week), though this decreased to below the overall sample mean by March 2021/April 2021. It is possible that certain rules and regulations within settings may contribute to greater pressure to return to a prepandemic baseline. It is also possible that greater resources within a larger system allowed providers to continue the prior standard of care, as AMC providers also reported greater likelihood of continuing to provide telemental health services despite a significant decrease in offering virtual instead of in-person appointments. Additionally, providers in PP did not show the expected increase in patients seen in person between March 2020/April 2020 and March 2021/April 2021. This could be due to less of an initial decline from December 2019 to March 2020/April 2020, as the salary for PP providers is typically more dependent on the number of patients seen compared with other settings with a more fixed salary schedule. Last, although providers in most settings reported a reduction in canceling or rescheduling appointments, all providers in Veterans Affairs hospitals and community mental health settings reported that no appointments were rescheduled nor canceled by March 2021/April 2021. These settings also both had no change in continuing to implement virtual instead of in-person visits as a practice adjustment, which could have contributed to greater flexibility in the type of appointments offered if a patient reported a COVID-19 exposure or onset of symptoms.

Limitations

Longitudinal survey data have a number of common limitations. Given the longitudinal nature of the data in this study, 37.5% of the individuals who provided their email addresses during the initial survey responded to the current follow-up survey. Additionally, some individuals changed practice setting, which we accounted for in our analyses but also impacts the quality of the longitudinal data. An important limitation of survey data is that the data are based on self-report, and we asked for retrospective estimates of the number of patients seen in December 2019 and December 2020. As such, the increase in the total number of patients with stable numbers of telehealth services could also suggest that some of the initial estimates were low or a reflection of changes in the overall practices within mental health specialties (eg, current billing rates compared with current inflation rates). Although our survey had similar characteristics to other surveys and was representative of the original sampling distribution, the demographics of participants who responded to the survey may not be representative of all mental health providers. Participants in our study were predominantly female (87%) and White (90%), and nearly one-half of participants were from the southern United States. As such, these results may not reflect practice changes in other regions of the United States, such as the Northeast, which had varying patterns of COVID-19 mortality and vaccination compared with the South [34].

Conclusions

Hybrid in-person and telemental health services appear to be here to stay. These results show that the rapid transition to telemental health at the start of the COVID-19 pandemic in spring 2020 was sustained over the next year, despite an overall increase in the number of patients seen in person. Although more providers reported returning to working on-site, over 50% of providers continued to use a hybrid model, and many providers reported they would be more likely to continue telemental health beyond spring 2021. Our findings suggest that
it will be important to continue to provide support and resources to successfully implement a hybrid telehealth model. This hybridization of mental health is potentially allowing for an overall increase in access to services, which is particularly important given the mental health impacts of the pandemic. Additionally, easy access to IT infrastructure and education may facilitate adoption of hybrid practice models across settings.

Acknowledgments
We would like to thank all of the participants who took the time to complete the follow-up survey (as well as the initial survey). DS and JM received research support from the National Institute of General Medical Sciences of the National Institutes of Health under Award Number 2U54GM104942-07. The funding source had no role other than financial support. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Survey.
[DOCX File, 42 KB - formative_v8i1e50303_app1.docx ]

References


Abbreviations
- AMC: academic medical center
- PP: private practice
- T1: Time point 1
- T2: Time point 2

Alessandra N Bazzano¹, MPH, PhD; Lesley-Ann Noel², PhD; Tejal Patel¹, MA; C Chantel Dominique³, AA; Catherine Haywood⁴, BSW; Shenitta Moore⁵, MD; Andrea Mantsios⁶, PhD; Patricia A Davis¹, BSc

¹Department of Social, Behavioral, and Population Sciences, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, United States
²College of Design, North Carolina State University, Raleigh, NC, United States
³Global Impact Board, New Orleans, LA, United States
⁴Louisiana Community Health Outreach Network, New Orleans, LA, United States
⁵Louisiana State University Health Sciences Center, New Orleans, LA, United States
⁶Public Health Innovation & Action, New York, NY, United States

Corresponding Author:
Alessandra N Bazzano, MPH, PhD
Department of Social, Behavioral, and Population Sciences
Tulane University School of Public Health and Tropical Medicine
1440 Canal St
New Orleans, LA, 70112
United States
Phone: 1 5049882338
Email: abazzano@tulane.edu

Related Article:
Correction of: https://formative.jmir.org/2023/1/e43101

(JMIR Form Res 2024;8:e58397) doi: 10.2196/58397

In “Improving the Engagement of Underrepresented People in Health Research Through Equity-Centered Design Thinking: Qualitative Study and Process Evaluation for the Development of the Grounding Health Research in Design Toolkit” (JMIR Form Res 2023;7:e43101) the authors noted one error.

In Reference 28, the URL:
https://pcornet.org/about/
has been changed to:
https://www.pcori.org/engagement/engagement-resources/grid-toolkit

The correction will appear in the online version of the paper on the JMIR Publications website on April 4, 2024 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Correction: Implementation Documentation and Process Assessment of the PharmNet Intervention: Observational Report

Lori Ann Eldridge1, PhD; Beth E Meyerson2,3, PhD; Jon Agley4, MPH, PhD

1Department of Health Education and Promotion, College of Health and Human Performance, East Carolina University, Greenville, NC, United States
2Harm Reduction Research Lab, Family and Community Medicine, College of Medicine-Tucson, University of Arizona, Tucson, AZ, United States
3Comprehensive Center for Pain and Addiction, University of Arizona Health Sciences, University of Arizona, Tucson, AZ, United States
4Prevention Insights, Department of Applied Health Science, School of Public Health Bloomington, Indiana University Bloomington, Bloomington, IN, United States

Corresponding Author:
Jon Agley, MPH, PhD
Prevention Insights, Department of Applied Health Science
School of Public Health Bloomington
Indiana University Bloomington
809 E 9th Street
Bloomington, IN, 47405
United States
Phone: 1 812 855 3123
Email: jagle@indiana.edu

Related Article:
Correction of: https://formative.jmir.org/2024/1/e54077
doi:10.2196/59427

In “Implementation Documentation and Process Assessment of the PharmNet Intervention: Observational Report” (JMIR Form Res 2024;8:e54077) two corrections were made.

The peer reviewer information has been updated. The first line in the publication information box has been changed from:

...peer-reviewed by D Carpenter...

to:

...peer-reviewed by SH Linder, D Carpenter...

Additionally, the following sentence appeared within the “Conclusions” section:

As one peer reviewer aptly noted, “the optimal means of naloxone community distribution remains to be determined.”

This has been changed to:

As one peer reviewer (Steven H Linder MD) aptly noted, “the optimal means of naloxone community distribution remains to be determined.”

The correction will appear in the online version of the paper on the JMIR Publications website on April 11, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Corrigenda and Addenda

Correction: Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study

Christina Popescu1*, BSc, MSc; Grace Golden2*, David Benrimoh1*, MSc, CM, MD; Myriam Tanguay-Sela1, BASc; Dominique Slowey3, BSc; Eryn Lundrigan3, BSc; Jérôme Williams3, BSc; Bennet Desormeau3, BSc, MSc; Divyesh Kardani3, BSCE; Tamara Perez3, BSc, MSc; Colleen Rollins4, BSc; Sonia Israel1, BSc; Kelly Perlman1,3, BSc; Caitrin Armstrong1, BASc, MSc; Jacob Baxter3, BSc; Kate Whitmore3, BA; Marie-Jeanne Fradette3, BSc; Kaelan Felcarek-Hope5; Ghassen Soufi3, BSc; Robert Fratila1, BSc; Joseph Mehltretter3, MSc; Karl Looper3, MD; Warren Steiner3, MD; Soham Rej3, MSc, MD; Jordan F Karp5, MD; Katherine Heller6, BSc, MSc, PhD; Sagar V Parikh7, MD; Rebecca McGuire-Snieckus6, BA, MSc, PhD; Manuela Ferrari9, PhD; Howard Margolese3, MSc, CM, MD; Gustavo Turecki9, PhD, MD

1Aifred Health Inc., Montreal, QC, Canada
2University of Waterloo, Waterloo, ON, Canada
3McGill University, Montreal, QC, Canada
4University of Cambridge, London, United Kingdom
5University of Arizona, Tucson, AZ, United States
6Duke University, Durham, NC, United States
7University of Michigan, Ann Arbor, MI, United States
8Barts and the London School of Medicine, London, United Kingdom
9Douglas Mental Health University Institute, McGill University, Montreal, QC, Canada
*these authors contributed equally

Corresponding Author:
David Benrimoh, MSc, CM, MD
Aifred Health Inc.
1250 Rue Guy Suite #600
Montreal, QC, H3H 2T4
Canada
Phone: 1 5144637813
Email: david.benrimoh@mail.mcgill.com

Related Article:
Correction of: https://formative.jmir.org/2021/10/e31862
doi:10.2196/31862

Jmir Form Res 2024;8:e56570 doi:10.2196/56570

In “Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study” (JMIR Form Res 2021;5(10):e31862), the authors noted one error.

In the section “Assessing Physician and Patient Trust in the CDSS and Its Effect on the Clinician-Patient Relationship”, the STAR-P and STAR-C total scores were listed as:

The mean STAR-P and STAR-C scores were 33.62 (SD 2.90) and 31.14 (SD 2.63) comparable with 38.4 (SD 12.0) and 31.5 (SD 6.9) in the original Scale to Assess Therapeutic Relationships in Community Mental Health Care study.

They have been changed to read in the following manner:

The mean STAR-P and STAR-C scores were 42.69 (SD 5.57) and 40.29 (SD 5.65), comparable with 38.4 (SD 12.0) and 31.5 (SD 6.9) in the original Scale to Assess Therapeutic Relationships in Community Mental Health Care study.
The correction will appear in the online version of the paper on the JMIR Publications website on January 24, 2024 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 19.01.24; this is a non–peer-reviewed article; accepted 19.01.24; published 24.01.24.

Please cite as:
Correction: Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study
JMIR Form Res 2024;8:e56570
URL: https://formative.jmir.org/2024/1/e56570
doi:10.2196/56570
PMID:38266244

©Christina Popescu, Grace Golden, David Benrimoh, Myriam Tanguay-Sela, Dominique Slowey, Eryn Lundrigan, Jérôme Williams, Bennet Desormeau, Divyesh Kardani, Tamara Perez, Colleen Rollins, Sonia Israel, Kelly Perlman, Caitrin Armstrong, Jacob Baxter, Kate Whitmore, Marie-Jeanne Fradette, Kaelan Felcarek-Hope, Ghassen Soufi, Robert Fratila, Joseph Mehlretter, Karl Looper, Warren Steiner, Soham Rej, Jordan F Karp, Katherine Heller, Sagar V Parikh, Rebecca McGuire-Snieckus, Manuela Ferrari, Howard Margolese, Gustavo Turecki. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
In “The Influence of Human Connections and Collaboration on Research Grant Success at Various Career Stages: Regression Analysis” (JMIR Form Res 2024;8:e49905) the authors noted one error.

In the “Results” section of the Abstract, the following sentence:

Tracking the process of research development, we found that collaboration during the periods of 10 to 14 years and 15 to 19 years after completing a doctorate degree determined the size of the project that the participant would obtain—interactions with peer researchers and subordinates during the 10- to 14-year postdegree period had positive effects on ≥2 large-program acquisitions (OR 1.51, 95% CI 1.09-2.09 and OR 1.31, 95% CI 1.10-1.57, respectively), whereas interactions with subordinates during the 15- to 19-year postdegree period also had positive effects (OR 1.25, 95% CI 1.06-1.47).

Has been changed to read as follows:

Tracking the process of research development, we found that collaboration during the periods of 10 to 14 years and 15 to 19 years after completing a doctorate degree determined the size of the project that the participant would obtain—interactions with peer researchers and subordinates during the 10- to 14-year postdegree period had positive effects on ≥2 large-program acquisitions (OR 1.51, 95% CI 1.09-2.09 and OR 1.31, 95% CI 1.10-1.57, respectively), whereas interactions with subordinates during the 15- to 19-year postdegree period also had positive effects (OR 1.25, 95% CI 1.06-1.47).

The correction will appear in the online version of the paper on the JMIR Publications website on April 5, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Hashiguchi A, Asashima M, Takahashi S
Correction: The Influence of Human Connections and Collaboration on Research Grant Success at Various Career Stages: Regression Analysis
JMIR Form Res 2024;8:e58733
URL: https://formative.jmir.org/2024/1/e58733
doi: 10.2196/58733
PMID: 38579278

© Akiko Hashiguchi, Makoto Asashima, Satoru Takahashi. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Correction: Tumor Immunotherapy–Related Information on Internet-Based Videos Commonly Used by the Chinese Population: Content Quality Analysis

Chen-xu Ni¹, MM, MS; Yi-bo Fei¹, MD, PhD; Ran Wu¹, MM, MS; Wen-xiang Cao¹, MM, MS; Wenhao Liu¹, MM, MS; Fang Huang¹, MM, MS; Fu-ming Shen¹, MD, PhD; Dong-jie Li¹, MD, PhD

Shanghai Tenth People's Hospital, Shanghai, China

Corresponding Author:
Dong-jie Li, MD, PhD
Shanghai Tenth People's Hospital
301 Middle Yanchang Road
Shanghai, 200072
China
Phone: 86 021 66302570
Email: djli@tongji.edu.cn

Related Article:
Correction of: https://formative.jmir.org/2024/1/e50561/
doi:10.2196/59671

In “Tumor Immunotherapy–Related Information on Internet-Based Videos Commonly Used by the Chinese Population: Content Quality Analysis” (JMIR Form Res 2024;8:e50561) the authors made one addition. The following sentence was added to the “Ethical Considerations” section under “Methods”:

This study was registered in the Chinese Clinical Trial Registry (ChiCTR2400081071).

The correction will appear in the online version of the paper on the JMIR Publications website on May 2, 2024 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submit 21.04.24; this is a non–peer-reviewed article; accepted 25.04.24; published 02.05.24.

Please cite as:
Ni CX, Fei YB, Wu R, Cao WX, Liu W, Huang F, Shen FM, Li DJ
Correction: Tumor Immunotherapy–Related Information on Internet-Based Videos Commonly Used by the Chinese Population: Content Quality Analysis
JMIR Form Res 2024;8:e59671
URL: https://formative.jmir.org/2024/1/e59671
doi:10.2196/59671
PMID:38696774

©Chen-xu Ni, Yi-bo Fei, Ran Wu, Wen-xiang Cao, Wenhao Liu, Fang Huang, Fu-ming Shen, Dong-jie Li. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Correction: Development of Therapeutic Alliance and Social Presence in a Digital Intervention for Pediatric Concussion: Qualitative Exploratory Study

Kiarah M K O’Kane¹, BA; Thalia Otamendi², PhD; Noah D Silverberg¹, PhD; Esther Choi¹, BA; Veronik Sicard³, PhD; Roger Zemek³,⁴, MD; Katherine Healey³, MSc; Olivier Brown³,⁵, BA; Lauren Butterfield³,⁶, BSc; Andra Smith³, PhD; Gary Goldfield³, PhD; Rachel Kardish³, MSc; Bechara J Saab⁷, PhD; Andrée-Anne Ledoux³,⁵,⁶,⁸*, PhD; Molly Cairncross⁹*, PhD

¹Department of Psychology, University of British Columbia, Vancouver, BC, Canada
²Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada
³Children’s Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
⁴Department of Pediatrics and Emergency Medicine, University of Ottawa, Ottawa, ON, Canada
⁵School of Psychology, Faculty of Social Sciences, University of Ottawa, Ottawa, ON, Canada
⁶Department of Neuroscience, Carleton University, Ottawa, ON, Canada
⁷Mobio Interactive, Singapore, Singapore
⁸Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada
⁹Department of Psychology, Simon Fraser University, Burnaby, BC, Canada

*these authors contributed equally

Corresponding Author:
Molly Cairncross, PhD
Department of Psychology
Simon Fraser University
8888 University Drive
Robert C. Brown Hall 7321
Burnaby, BC, V5A 1S6
Canada
Phone: 1 778 782 7458
Email: molly_cairncross@sfu.ca

Related Article:
Correction of: https://formativ.e.jmir.org/2024/1/e49133/
doi:10.2196/59722

(JMIR Form Res 2024;8:e59722) doi:10.2196/59722

In “Development of Therapeutic Alliance and Social Presence in a Digital Intervention for Pediatric Concussion: Qualitative Exploratory Study” (JMIR Form Res 2024 Mar 22:8:e49133) the authors noted several errors.

In the originally published article, the equal contribution note indicating that Andrée-Anne Ledoux and Molly Cairncross had contributed equally as co-senior authors was missed. In the corrected version, these authors are now listed as Andrée-Anne Ledoux*; Molly Cairncross*

*these authors contributed equally

Then, in the originally published article, affiliation 5 was erroneously listed as

School of Psychology, Faculty of Social Sciences, University of Ottawa, Ottawa, BC, Canada

This has been corrected to

School of Psychology, Faculty of Social Sciences, University of Ottawa, Ottawa, ON, Canada

In addition, one author’s name was listed as Bechara Saab

This has been changed to

Bechara J Saab

The correction will appear in the online version of the paper on the JMIR Publications website on May 7, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Web-Based Search Volume for HIV Tests and HIV-Testing Preferences During the COVID-19 Pandemic in Japan: Infodemiology Study

Rie Kanamori¹, MMEdSc; Futaba Umemura², MMEdSc; Kosuke Uemura¹, MD; Taiju Miyagami¹, MD, PhD; Simon Valenti¹, MD; Nobuyuki Fukui³, BA; Mayumi Yuda³, BE; Mizue Saita¹, MD, PhD; Hirotake Mori¹, MD, PhD; Toshio Naito¹, MD, PhD

¹Department of General Medicine, Faculty of Medicine, Juntendo University, Tokyo, Japan
²Department of Sports Medicine and Sportology, Graduate School of Medicine, Juntendo University, Tokyo, Japan
³Center for Promotion of Data Science, Graduate School of Medicine, Juntendo University, Tokyo, Japan

Corresponding Author:
Toshio Naito, MD, PhD
Department of General Medicine
Faculty of Medicine
Juntendo University
3-1-3 Hongo Bunkyo-ku
Tokyo, 113-8421
Japan
Phone: 81 3 5802 1190
Email: naito@juntendo.ac.jp

Abstract

Background: Research has found a COVID-19 pandemic–related impact on HIV medical services, including clinic visits, testing, and antiviral therapy initiation in countries including Japan. However, the change in trend for HIV/AIDS testing during the COVID-19 pandemic has not been explored extensively in the Japanese population.

Objective: This infodemiology study examines the web-based search interest for two types of HIV tests, self-test kits and facility-based tests, before and during the COVID-19 pandemic in Japan.

Methods: The monthly search volume of queried search terms was obtained from Yahoo! JAPAN. Search volumes for the following terms were collected from November 2017 to October 2018: “HIV test,” “HIV test kit,” and “HIV test health center.” The search term “Corona PCR” and the number of new COVID-19 cases by month were used as a control for the search trends. The number of new HIV cases in the corresponding study period was obtained from the AIDS Trend Committee Quarterly Report from the AIDS Prevention Foundation.

Results: Compared to the search volume of “corona-PCR,” which roughly fluctuated corresponding to the number of new COVID-19 cases in Japan, the search volume of “HIV test” was relatively stable from 2019 to 2022. When we further stratified by the type of HIV test, the respective web-based search interest in HIV self-testing and facility-based testing showed distinct patterns from 2018 to 2022. While the search volume of “HIV test kit” remained stable, that of “HIV test health center” displayed a decreasing trend starting in 2018 and has remained low since the beginning of the COVID-19 pandemic. Around 66%-71% of the search volume of “HIV test kits” was attributable to searches made by male internet users from 2018 to 2022, and the top three contributing age groups were those aged 30-39 (27%-32%), 20-29 (19%-32%), and 40-49 (19%-25%) years. On the other hand, the search volume of “HIV test health centers” by male users decreased from more than 500 from 2018 to 2019 to fewer than 300 from 2020 to 2022.

Conclusions: Our study found a notable decrease in the search volume of “HIV test health center” during the pandemic, while the search volume for HIV self-testing kits remained stable before and during the COVID-19 crisis in Japan. This suggests that the previously reported COVID-19–related decrease in the number of HIV tests mostly likely referred to facility-based testing. This sheds light on the change in HIV-testing preferences in Japan, calling for a more comprehensive application and regulatory acceptance of HIV self-instructed tests.
KEYWORDS
HIV test; infodemiology; self-test; COVID-19; search engine; Japan

Introduction
HIV prevalence in Japan is <0.1% among adults aged 15-49 years, and the cumulative total of people diagnosed and living with HIV at the end of 2018 was approximately 30,000 [1]. A recent study using a national database found that 81.5% of people living with HIV who were on antiretroviral therapy (ART) had at least one chronic comorbidity [2]. According to recent national statistics, the proportion of people diagnosed and living with HIV on ART was 86% and 80% as of 2018 [3], lower than the first two 90-90-90 Joint United Nations Programme on HIV and AIDS (UNAIDS) goals, while the proportion of people living with HIV on ART who are virally suppressed was as high as 99% as of 2020 [4], reaching the UNAIDS goal. Evidence has suggested that earlier HIV diagnosis and linkage to treatment should be the core strategy in controlling the HIV epidemic [5]; however, the sudden and ongoing COVID-19 pandemic has overwhelmed the health care system globally and caused severe disruptions in medical service provision, ranging from elective procedures to routine surveillance for manageable diseases such as HIV infections [6-11]. Additionally, social distancing measures implemented early in the pandemic affected people’s access or willingness to visit health care facilities.

There are concerns that disruptions caused by the COVID-19 pandemic may lead to increased HIV incidence and mortality and pose challenges to the international community’s goal to eliminate the HIV or AIDS epidemic by 2030 [1]. Reports from Africa, Asia, Europe, Japan, and the United States have found a pandemic-related impact on HIV medical services, including clinic visits, testing, and antiviral therapy initiation [1-6]. In particular, the number of HIV tests conducted in health centers in Japan reduced from 142,000 in 2019 to 69,000 in 2020, with the difference amounting to a 50% decline in the first year of the COVID-19 pandemic [3]. The number of HIV consultations performed in public health centers also significantly declined in the second quarter of 2020 (32,565 tests) compared to the same period the year before (11,689 tests) [2].

In Japan, voluntary HIV testing is free of charge and anonymous, and offered as a package service bundled with pre- and posttest counseling at appointed public health facilities by law. Although the number of health center–conducted HIV tests performed in public health centers also significantly declined in the second quarter of 2020 (32,565 tests) compared to the same period the year before (11,689 tests) [2].

The internet is a common source of disease- and health-related information, and internet use influences care initiation and the treatment decisions of people living with HIV [13]. During the COVID-19 pandemic, infodemiology studies found that web-based search trends for symptoms associated with COVID-19 coincided with the disease outbreak [14]. Conversely, the global web-based interest in information on HIV/AIDS care services decreased [15]. However, the search engine trend for HIV/AIDS testing during the COVID-19 pandemic has not been explored extensively in the Japanese population. This study aims to characterize and compare the trend for web-based search interest in HIV testing before and during the pandemic to investigate whether there is a change in HIV/AIDS health-seeking behavior in Japan.

Data from the search engine provided by the Yahoo Japan Corporation will be used for this study because Yahoo! JAPAN is the most visited website in the country, and compared to most other countries, Japan uses Google less frequently.

Methods

Data Source
To examine the pattern of web-based search interests in HIV testing before and during the COVID-19 pandemic, the monthly search volume of selected search terms was determined based on the number of searches over a specified period extracted from Yahoo! JAPAN, which is one of the most used digital services in Japan, and compared to the rest of the world, Google is used less. According to the data, 68.9% of the population (aged >2 years) used Yahoo! JAPAN at least once a month between January and November 2021, while 65.1% used Google in the same period [16]. Search volume data were retrieved with authorized access from the Yahoo Japan Corporation server via the DS.INSIGHT People portal. It has been used in studies focusing on transition or trends in search behavior over time among different demographics (eg, gender, age, or prefecture) [17]. The user manual is available on the web [18].

The internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications.

The quarterly number of new COVID-19 cases between January 2019 and December 2022 was obtained from an excerpt from the Japan Ministry of Health, Labour and Welfare [19].

The number of newly infected people with HIV between December 31, 2018, and June 26, 2022, was obtained from the AIDS Trend Committee Quarterly Report from the AIDS Prevention Foundation, API-Net AIDS Prevention Information Net [20].

Search Queries Used in the Analysis
The search term “HIV検査” (ie, “HIV test” in Japanese) was used to assess the web-based interest in general HIV testing from January 2020 to October 2021. The result is presented together with the search volume of “コロナPCR” (ie,
“corona-PCR” in Japanese) and the number of new COVID-19 cases by month. The search volumes of “HIV 検査キット” (ie, “HIV test kit” in Japanese) and “HIV 検査 健康所” (ie, “HIV test health center” in Japanese) were compared to analyze the difference in web-based interest in self-instructed/postal and facility-based HIV testing from November 2018 to October 2022.

Data Standardization
The monthly number of searches per prefecture for each search query was obtained, adjusted by sex, and converted to standardized $z$ scores according to the following formula:

\[
\text{Standardized } z = \frac{\text{Query } A - \text{Mean of Query } A}{\text{Standard Deviation of Query } A}
\]

“Query A” refers to the queried search term.

Statistical Analysis
Internet users were stratified according to age on the day of the search, sex, and year. Demographics were summarized as the number and percentage of patients for categorical variables. Descriptive statistics were summarized using Excel (Microsoft Corporation).

Ethical Considerations
This infodemiology study adhered to chapter 1, section 3, part 1, subsection (C), item 3 of the ethical guidelines for Medical and Health Research Involving Human Subjects of the Ministry of Health, Labour and Welfare of Japan [21]. In accordance with this guideline, since this study used previously anonymized and deidentified data, an ethical review was waived, and patient informed consent was not required.

Results
Between 2018 and 2022, 50% to 51% of overall adult internet users were male, and the user population was evenly distributed across different age groups by decade (11%-18% across different age groups in 2018, 12%-17% in 2019, 12%-18% in 2020, 13%-17% in 2021, and 13%-17% in 2022; Tables 1-3). The first nationwide state of emergency concerning COVID-19 was from April 7 to May 25, 2020, in Japan; since then, the increase in web-based interest in COVID-19 polymerase chain reaction testing (represented by the search volume of “corona-PCR”) roughly corresponded to the peaks and troughs of new COVID-19 infection cases in recent waves of the COVID-19 pandemic in Japan during mid 2021, early 2022, and late 2022. In contrast, the web-based interest in general HIV testing represented by the search term “HIV test” appeared to be relatively stable (Figure 1). Although the sex distribution of the “corona-PCR” search volume varied between 2020 and 2022, people aged 40-49 years consistently showed the highest web-based interest (26%-27%) for COVID-19 polymerase chain reaction testing, followed by those aged 50-59 years (21%-23%) and 30-39 years (15%-18%). In line with the characteristics of people with an increased risk of HIV infection, 57%-60% of the “HIV test” search volume between 2020 and 2022 were from male internet users, and people aged between 20 and 29 years (30%-31%) showed the most interest in general HIV testing within this period, followed by those aged between 30 and 39 years (24%-27%) and 40 and 49 years (17%-20%; Table 1).
Table 1. Search volumes of “HIV test” and “corona-PCR” by the sex and age of internet users in the years 2020, 2021, and 2022.

<table>
<thead>
<tr>
<th>Age group (year)</th>
<th>Search term “HIV検査”</th>
<th>Search term “コロナPCR”</th>
<th>Search term “HIV検査”</th>
<th>Search term “コロナPCR”</th>
<th>Search term “HIV検査”</th>
<th>Search term “コロナPCR”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Internet user&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2020, n (%)</td>
<td>Internet user&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2021, n (%)</td>
<td>Internet user&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2022 (January to October), n (%)</td>
</tr>
<tr>
<td>Age group (year)</td>
<td>Overall</td>
<td>Male</td>
<td>Female</td>
<td>Overall</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>&lt;20</td>
<td>13,396,518 (14)</td>
<td>170 (7)</td>
<td>20 (2)</td>
<td>13,787,372 (14)</td>
<td>280 (10)</td>
<td>100 (5)</td>
</tr>
<tr>
<td>20-29</td>
<td>11,777,729 (12)</td>
<td>770 (30)</td>
<td>130 (10)</td>
<td>12,432,946 (13)</td>
<td>840 (31)</td>
<td>280 (14)</td>
</tr>
<tr>
<td>30-39</td>
<td>13,547,255 (14)</td>
<td>670 (27)</td>
<td>190 (15)</td>
<td>13,615,333 (14)</td>
<td>680 (25)</td>
<td>310 (16)</td>
</tr>
<tr>
<td>40-49</td>
<td>17,429,796 (18)</td>
<td>470 (19)</td>
<td>340 (26)</td>
<td>17,491,018 (17)</td>
<td>460 (17)</td>
<td>540 (27)</td>
</tr>
<tr>
<td>50-59</td>
<td>15,451,255 (16)</td>
<td>230 (9)</td>
<td>300 (23)</td>
<td>16,258,923 (16)</td>
<td>250 (9)</td>
<td>440 (22)</td>
</tr>
<tr>
<td>60-69</td>
<td>12,716,666 (13)</td>
<td>130 (5)</td>
<td>190 (15)</td>
<td>12,887,755 (13)</td>
<td>130 (5)</td>
<td>230 (12)</td>
</tr>
<tr>
<td>≥70</td>
<td>12,378,860 (13)</td>
<td>70 (3)</td>
<td>110 (9)</td>
<td>13,026,322 (13)</td>
<td>80 (3)</td>
<td>80 (4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications [21].

<sup>b</sup>HIV検査: HIV test (in Japanese).

<sup>c</sup>コロナPCR: corona-PCR (in Japanese).

<sup>d</sup>The internet user population in 2022 uses the figures for 2021.
Table 2. Search volumes for “HIV test kit” and “HIV test health center” by the sex and age of the internet users by year (2018-2020).

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>Year 2018, n (%)</th>
<th>Year 2019, n (%)</th>
<th>Year 2020, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Search term “hiv検査キット”</td>
<td>Search term “hiv検査 保健所”</td>
<td>Search term “hiv検査キット”</td>
</tr>
<tr>
<td></td>
<td>Internet user population&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Internet user population&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Internet user population&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Overall</td>
<td>95,590,219 (100)</td>
<td>940 (100)</td>
<td>107,480,629 (100)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>48,932,966 (51)</td>
<td>530 (56)</td>
<td>53,862,486 (50)</td>
</tr>
<tr>
<td>Female</td>
<td>46,657,253 (49)</td>
<td>410 (44)</td>
<td>53,618,143 (50)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>12,760,379 (13)</td>
<td>30 (3)</td>
<td>13,710,441 (13)</td>
</tr>
<tr>
<td>20-29</td>
<td>12,381,162 (13)</td>
<td>280 (30)</td>
<td>12,514,936 (12)</td>
</tr>
<tr>
<td>30-39</td>
<td>14,321,086 (15)</td>
<td>300 (32)</td>
<td>14,167,245 (13)</td>
</tr>
<tr>
<td>40-49</td>
<td>18,127,686 (18)</td>
<td>180 (19)</td>
<td>18,214,081 (16)</td>
</tr>
<tr>
<td>50-59</td>
<td>14,891,256 (16)</td>
<td>90 (12)</td>
<td>15,903,690 (15)</td>
</tr>
<tr>
<td>60-69</td>
<td>12,991,502 (14)</td>
<td>40 (4)</td>
<td>14,688,962 (14)</td>
</tr>
<tr>
<td>≥70</td>
<td>10,117,148 (11)</td>
<td>0 (0)</td>
<td>18,281,274 (17)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications [22].

<sup>b</sup>hiv検査キット: HIV test kit (in Japanese).

<sup>c</sup>hiv検査保健所: HIV test health center (in Japanese).
Table 3. Search volumes for “HIV test kit” and “HIV test health center” by the sex and age of the internet users by year (2021 and 2022).

<table>
<thead>
<tr>
<th>Year 2021, n (%)</th>
<th>Year 2022 (January to October), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet user population&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Search term “hiv検査キット”&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Overall</td>
<td>99,499,669 (100)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50,286,908 (51)</td>
</tr>
<tr>
<td>Female</td>
<td>49,212,761 (49)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>13,787,372 (14)</td>
</tr>
<tr>
<td>20-29</td>
<td>12,432,946 (13)</td>
</tr>
<tr>
<td>30-39</td>
<td>13,615,333 (14)</td>
</tr>
<tr>
<td>40-49</td>
<td>17,491,018 (17)</td>
</tr>
<tr>
<td>50-59</td>
<td>16,258,923 (16)</td>
</tr>
<tr>
<td>60-69</td>
<td>12,887,755 (13)</td>
</tr>
<tr>
<td>≥70</td>
<td>13,026,322 (13)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications [22].

<sup>b</sup>hiv検査キット: HIV test kit (in Japanese).

<sup>c</sup>hiv検査保健所: HIV test health center (in Japanese).

<sup>d</sup>The internet user population in 2022 uses the figures for 2021.

Figure 1. Monthly web-based search interest in HIV testing and COVID-19 genetic testing and the number of new COVID-19 cases from 2020 to 2022. The monthly search volumes of search terms “HIV検査” (HIV test in Japanese, blue line) and “コロナPCR” (corona-PCR in Japanese, red line) from January 2020 to October 2022 according to the search engine Yahoo! JAPAN are shown. The number of new COVID-19 cases in the corresponding months, according to the Ministry of Health, Labour and Welfare in Japan, is also shown in green bars in this figure.
Web-based interest in HIV testing was further refined as either facility-based HIV tests or self-instructed HIV-testing kits. The search volume for HIV facility-based testing was high in late 2018 and decreased in several instances over the next 3 years, especially in 2020 and 2022 (Figure 2). The search volume of “HIV test kit” (used as a representative search term for HIV self-testing) was lower than that for HIV facility-based testing at the end of 2018. However, the interest remained high over the next 4 years relative to facility-based testing, particularly during the COVID-19 pandemic (Figure 2). Around 66%-71% of the search volume of “HIV test kit” was attributable to searches made by male internet users between 2018 and 2022, and the top three contributing age groups were those 30-39 years (27%-32%), 20-29 years (19%-32%), and 40-49 years (19%-25%). The overall search volume of HIV test health centers decreased from 950 in 2018/2019 to 580 in 2020. The actual search volume for “HIV test kit” versus “HIV test health center” was 560 versus 280 among male internet users in 2020 and 590 versus 290 in 2021; while among female internet users the actual search volume for “HIV test kit” versus “HIV test health center” were similar: 230 versus 300 in 2020 and 310 versus 350 in 2021. Hence, the main population searching for “HIV test health center” shifted from male to female users (n=300, 52%) in 2020 and the following years (n=350, 55% in 2021 and n=180, 64% in 2022). Internet users aged 20-29 years (27%-35%), 30-39 years (24%-32%), and 40-49 years (14%-20%) were the top three representative age groups for the search volume of “HIV test health center” (Tables 2 and 3). The quarterly number of newly infected HIV cases remained relatively stable since the declaration of a nationwide state of emergency concerning COVID-19, and the search interest remained high for self-testing relative to facility-based testing (Figure 3).

Figure 2. Monthly web-based search interest in HIV self-instructed and facility-based tests and the number of new COVID-19 cases from 2018 to 2022. The monthly search volumes of search terms “HIV検査キット” (HIV test kit in Japanese, orange line) and “HIV検査保健所” (HIV test health center in Japanese, purple line) from November 2018 to October 2022 according to the search engine Yahoo! JAPAN are shown. The number of new COVID-19 cases in the corresponding months, according to the Ministry of Health, Labour and Welfare in Japan, is shown with green bars.
Figure 3. Monthly web-based search interest in HIV self-instructed and facility-based tests and quarterly number of new infected HIV cases from 2019 to 2022. The monthly search volumes of search terms “HIV検査キット” (HIV test kit in Japanese, orange line) and “HIV検査保健所” (HIV test health center in Japanese, purple line) from January 2019 to December 2022 according to the search engine provided by Yahoo! JAPAN are shown. The number of new HIV cases in the corresponding quarter, according to the AIDS Trend Committee Quarterly Report from the AIDS Prevention Foundation, API-Net AIDS Prevention Information Net, is shown with blue bars.

Discussion

The analysis in our study illustrated that the web-based search interest in general HIV testing remained stable in Japan during the ongoing pandemic. Furthermore, compared to a decrease in HIV facility-based testing, the web-based interest in HIV self-testing has not changed in Japan. The findings suggest that a proportion of the population interested in getting an HIV test or obtaining relevant information may have a lower preference for HIV facility-based testing. This could imply that individuals might have opted for HIV self-testing during the COVID-19 pandemic.

A prolonged decreasing trend in the search volume of “HIV test health center” from late 2018 was observed in this study. This change in search volume was related to the real-world situation reported by recent publications that underscored the situation of facility-based HIV testing during the pandemic [7,8,23]. A trend analysis study exploring data from 2015 to the second quarter of 2020 (during the nationwide state of emergency in Japan) reported a significant decline in the number of HIV tests performed by public health centers in the second quarter of 2020 (n=9584 vs n=35,908 in Q2 2019) [7] and that this decrease coincided with the increase in the number of new HIV cases with an AIDS diagnosis in Japan [7]. Similar situations were reported in other countries, where an approximately 31% to 50% reduction in the number of HIV tests performed by public health centers was observed between 2019 and 2020 due to difficulty accessing testing facilities, shortage of medical staff for HIV-testing services, or closure of HIV-testing facilities [23,24].

Our search trend showed that the web-based search volume for HIV self-tests was relatively unaffected in Japan. Furthermore, we noted the trend for HIV-testing preference was more evident among male internet users. In Japan, men who have sex with men (MSM) constitute more than 70% of newly diagnosed people living with HIV [3,25]; therefore, the difference in search volume for different HIV-testing services most likely reflects the changing preferences among male internet users. This is consistent with the real-world situation reported by other countries. Investigators focusing on the high-risk population’s (MSM) HIV-testing behavior in China during the COVID-19 pandemic found the use of HIV self-testing increased compared to facility-based HIV testing, which decreased by more than 50% overall [26,27]. Earlier investigations in Sweden and Japan showed the interest and demand for HIV self-testing in high-risk populations such as MSM even before the COVID-19 pandemic [12,28]. A study in France also reported that the number of HIV self-testing kits sold in 2021 increased by 3% compared to 2020 [29]. Together, these publications further emphasized the

In this study, interest in HIV self-testing has remained relatively high during the COVID-19 pandemic. Even though web-based search interest only indicates clinical information-seeking behavior, recent infodemiology studies, including that by Ornos et al [15], showed that search volume indices correlated positively with HIV prevalence and negatively with financial and health care service status. Another recent study in Japan by Ishimaru et al [30] also observed a positive correlation between internet search frequency for HIV/AIDS–related terms and the number of voluntary tests.

The change in search volume on HIV-testing preferences in Japan, which highlights the availability, accessibility, and regulatory approval status of HIV self-tests, may be imperative to reduce new HIV infection [7,8]. The pros and cons of alternative HIV-testing approaches such as the use of dried blood spot (DBS) test cards delivered by postal service have been investigated in Japan [8]. So far, DBS-based tests have not been approved as clinical samples for HIV testing in Japan; however, their use has steadily increased in the last two decades, most likely due to convenient and easy self-preparation of blood spots without the need to visit medical facilities [8]. Although the use of DBS is less sensitive than with a plasma sample, the feasibility and reliability of postal DBS have been preliminarily demonstrated in an outreach study in Japan [12]. Other HIV testing to be considered should include the introduction of HIV self-testing, DBS, or oral swabs as an alternative to plasma or serum specimens. However, since HIV self-testing is not yet approved by health authorities, individuals with positive test results may not have been referred to medical facilities for consultations after the test or received appropriate HIV care and treatment. This concern was illustrated in the study by Ejima et al [7], which reported that both the number of HIV tests and consultations performed by public health centers declined during the COVID-19 pandemic in Japan. We believe relevant supporting services, for instance, the provision of web-based counseling services and referral services after HIV self-testing, would be necessary to mitigate the potential concerns and negative impact of using HIV self-test kits without counseling and medical follow-up. Soon, the development of artificial intelligence chatbots may be used to provide real-time instruction and counseling for HIV self-testing users, which may offer potential solutions to this problem [31].

Several factors may limit the generalizability of the preliminary findings presented in this report. First, the data source was solely from Yahoo! JAPAN, and the internet searches conducted using other search engines were not included. Second, the data obtained from a web-based search engine may be subject to the nonrepresentative sampling or methodology bias inherent to the search platform. Third, there could be considerable differences in clinical information-seeking behavior between the internet user population and actual men and women who are affected by HIV; direct and causal relationships cannot be inferred from this study. Nonetheless, most previous studies using search engines have used Google Trends data, but in Japan, Yahoo! JAPAN is the most visited digital service in the country. In addition, sex and prefecture adjustment is possible with Yahoo! JAPAN data, making it appropriate to research topics likely affected by sex differences. Therefore, our preliminary infodemiology using the search engine provided by Yahoo! Japan Corporation is likely representative of the general trend in Japan.

Our infodemiology study indicated that there was a notable decrease in search volume for HIV facility-based testing during the COVID-19 pandemic. Further, a change in HIV-testing preference and interest in HIV self-testing was noted in Japan. To fully delineate and comprehend the changes in HIV-testing behavior, the situation should be continuously monitored and validated by clinical studies.

Acknowledgments
The authors would like to thank 4DIN Ltd, Japan, for the data analysis and for assisting with preparing the manuscript and other materials for journal submission. This work was supported by the Ministry of Health, Labour and Welfare Research on HIV/AIDS (program grant JPMH21HB1005).

Data Availability
The data sets generated or analyzed during this study are not publicly available, complying with the Yahoo! Japan Corporation regulation, but they are available from the corresponding author upon reasonable request.

Authors’ Contributions
FU and TN conceptualized this study. RK and KU are responsible for the methodology, NF and MY collected the data. RK, FU, KU, TM, SV, MS, and HM analyzed the data. RK, NF, MY, and TN prepared the initial draft. RK, FU, KU, TM, SV, MS, HM, and TN critically revised the manuscript. HM and TN obtained the funding. MS was responsible for project administration. All authors approved this version of the manuscript to be submitted.

Conflicts of Interest
None declared.

References
https://formative.jmir.org/2024/1/e52306

JMIR Form Res 2024 | vol. 8 | e52306 | p.3385
(page number not for citation purposes)

Abbreviations
- ART: antiretroviral therapy
- DBS: dried blood spot
- MSM: men who have sex with men
- UNAIDS: Joint United Nations Programme on HIV and AIDS

Edited by A Mavragani; submitted 31.08.23; peer-reviewed by YXiong, S Chen; comments to author 23.10.23; revised version received 27.11.23; accepted 28.11.23; published 18.01.24.

Please cite as:
Kanamori R, Umemura F, Uemura K, Miyagami T, Valenti S, Fukui N, Yuda M, Saita M, Mori H, Naito T
Web-Based Search Volume for HIV Tests and HIV-Testing Preferences During the COVID-19 Pandemic in Japan: Infodemiology Study
JIMIR Form Res 2024;8:e52306
URL: https://formative.jmir.org/2024/1/e52306
doi:10.2196/52306
PMID: 38236022

©Rie Kanamori, Futaba Umemura, Kosuke Uemura, Taiju Miyagami, Simon Valenti, Nobuyuki Fukui, Mayumi Yuda, Mizue Saita, Hirotake Mori, Toshio Naito. Originally published in JIMIR Formative Research (https://formative.jmir.org), 18.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JIMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
A Novel Approach for the Early Detection of Medical Resource Demand Surges During Health Care Emergencies: Infodemiology Study of Tweets

Mahakprit Kaur¹,², BSc; Taylor Cargill¹,², BSc; Kevin Hui²,³, BSc; Minh Vu²,³, BSc; Nicola Luigi Bragazzi⁴,⁵, MPH, MD, PhD; Jude Dzevela Kong²,⁴, PhD

¹Department of Biology, Faculty of Science, York University, Toronto, ON, Canada
²Africa-Canada Artificial Intelligence and Data Innovation Consortium, Toronto, ON, Canada
³Department of Computer Science, Lassonde School of Engineering, York University, Toronto, ON, Canada
⁴Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
⁵Laboratory for Industrial and Applied Mathematics, York University, Toronto, ON, Canada

Corresponding Author:
Jude Dzevela Kong, PhD
Dalla Lana School of Public Health
University of Toronto
155 College St, Room 500
Toronto, ON, M5T 3M7
Canada
Phone: 1 416 978 0901
Email: jdkong@yorku.ca

Abstract

Background: The COVID-19 pandemic has highlighted gaps in the current handling of medical resource demand surges and the need for prioritizing scarce medical resources to mitigate the risk of health care facilities becoming overwhelmed.

Objective: During a health care emergency, such as the COVID-19 pandemic, the public often uses social media to express negative sentiment (e.g., urgency, fear, and frustration) as a real-time response to the evolving crisis. The sentiment expressed in COVID-19 posts may provide valuable real-time information about the relative severity of medical resource demand in different regions of a country. In this study, Twitter (subsequently rebranded as X) sentiment analysis was used to investigate whether an increase in negative sentiment COVID-19 tweets corresponded to a greater demand for hospital intensive care unit (ICU) beds in specific regions of the United States, Brazil, and India.

Methods: Tweets were collected from a publicly available data set containing COVID-19 tweets with sentiment labels and geolocation information posted between February 1, 2020, and March 31, 2021. Regional medical resource shortage data were gathered from publicly available data sets reporting a time series of ICU bed demand across each country. Negative sentiment tweets were analyzed using the Granger causality test and convergent cross-mapping (CCM) analysis to assess the utility of the time series of negative sentiment tweets in forecasting ICU bed shortages.

Results: For the United States (30,742,934 negative sentiment tweets), the results of the Granger causality test (for whether negative sentiment COVID-19 tweets forecast ICU bed shortage, assuming a stochastic system) were significant ($P<.05$) for 14 (28%) of the 50 states that passed the augmented Dickey-Fuller test at lag 2, and the results of the CCM analysis (for whether negative sentiment COVID-19 tweets forecast ICU bed shortage, assuming a dynamic system) were significant ($P<.05$) for 46 (92%) of the 50 states. For Brazil (3,004,039 negative sentiment tweets), the results of the Granger causality test were significant ($P<.05$) for 6 (22%) of the 27 federative units, and the results of the CCM analysis were significant ($P<.05$) for 26 (96%) of the 27 federative units. For India (4,199,151 negative sentiment tweets), the results of the Granger causality test were significant ($P<.05$) for 6 (23%) of the 26 included regions (25 states and the national capital region of Delhi), and the results of the CCM analysis were significant ($P<.05$) for 26 (100%) of the 26 included regions.

Conclusions: This study provides a novel approach for identifying the regions of high hospital bed demand during a health care emergency scenario by analyzing Twitter sentiment data. Leveraging analyses that take advantage of natural language processing, we demonstrate the potential of infodemiology to complement traditional data sources in real-time health care resource management.
processing–driven tweet extraction systems has the potential to be an effective method for the early detection of medical resource demand surges.

**KEYWORDS**
COVID-19; Twitter; social media; medical supply shortage; pandemic; global health; Granger; convergent cross-mapping; causal analysis; intensive care unit bed; ICU bed

**Introduction**

**Background**

The World Health Organization declared COVID-19 a pandemic on March 11, 2020 [1]. The emergence of this pandemic, caused by SARS-CoV-2, led to an unprecedented disruption in the global health care system that exposed and exacerbated existing vulnerabilities in health infrastructure around the world. In particular, the COVID-19 pandemic has had a profound impact on the global medical supply chain, leading to people struggling desperately to access crucial medical resources in the face of case surges and high resource demand [2].

The unprecedented nature of the pandemic and the limited availability of resources, no matter the country, will inevitably lead to the need for prioritizing scarce medical resources to different extents [3,4]. Wealthy countries, such as the United States, experienced shortages of personal protective equipment (PPE) and ventilators [5,6]. This led to the Centers for Disease Control and Prevention developing guidelines for the optimal sourcing of COVID-19 mitigation equipment such as face masks [7]. The pandemic also resulted in an increased strain on hospital capacity around the world. This was especially true for low- and middle-income countries, where health care systems are likely to already be underresourced and stretched thin, making them particularly vulnerable to becoming overwhelmed [8,9]. Considering the potential for future pandemic scenarios and for the recurrence of existing disease outbreaks as new virus variants emerge, the development of accurate real-time methodologies for detecting and forecasting disease impacts is critical for an effective global health response [10,11]. For hospitals that experience volatile demand surges in the face of a finite medical resource supply, timely solutions are required that can allow for rapid and precise decisions to be made regarding resource allocation.

Given that social media are an emerging source for real-time and easily accessible information, there is potential to leverage data from social media for forecasting real-world outcomes [12-15]. Social media platforms and web search data host a wealth of real-time data that broadly reflect the current state of affairs in a particular region [16]. Although the standards of validation for these new data streams are still being validated because they do not have a track record of use, these unconventional data sources have the potential to aid in short-and long-term surveillance, although the surveillance goals must be clearly defined. Studies have found that leveraging social media to identify shortages has the potential to be a cost-effective solution that can be used in real time [17,18]; for instance, Get Us PPE is a grassroots organization that leveraged Twitter (subsequently rebranded as X) to address medical supply shortages in US health care facilities during the first year of the COVID-19 pandemic [19]. Its success in garnering both public and governmental attention to the PPE shortage crisis has demonstrated that Twitter can be a useful tool for mobilizing efforts to address gaps, identifying regional PPE shortages, and informing decision-making in the health care supply chain. In other countries too, such as India, people used Twitter during the pandemic to amplify demands for medical oxygen and intensive care unit (ICU) beds during periods when health care facilities were overwhelmed by case surges [20].

Studies examining the role of social media to glean information about the characteristics of the pandemic note that data derived from social media and search engine data were used to predict new cases in countries such as South Korea [21], the United States [22-24], China [25-27], and Iran [28]. Twitter data have been analyzed to understand the population-level spread of disease [29-31]. Furthermore, forecasting models have been created to track demand for ICU capacity planning in countries such as Chile [32,33], Brazil [34], Colombia [35], the United States [36], India [37], and China [38]. Previous studies have applied convergent cross-mapping (CCM) analysis to explore possible relationships involving antiepidemic measure–related tweets [39], the dynamics of misleading news on Twitter [40], and the identification of the global drivers of influenza [41]. However, to our knowledge, there are limited studies examining the potential of social media, particularly Twitter, to better understand hospital bed demand.

**Objectives**

This study aimed to investigate the potential for social media, a relatively novel data stream, to be leveraged as an early warning and detection system for forecasting medical resource shortages. Specifically, this study sought to determine whether the COVID-19 discourse on Twitter could be linked to real-world ICU bed demand. We applied the Granger causality test and CCM analysis to explore whether a causal relationship exists between the volume of negative sentiment COVID-19 tweets and the proportion of ICU bed occupancy in real time in the United States, Brazil, and India. If social media can be successfully leveraged to develop an effective early warning system for forecasting medical resource demand, health care workers and governments may receive real-time insights into pandemic scenarios to inform urgent resource allocation decisions and gain a head start in preparing for demand surges.

**Methods**

**Overview**

For our analyses, the volume of negative sentiment COVID-19 tweets was compared with ICU bed demand data for each
subregion in the United States, Brazil, and India. These 3 countries were selected for this study because they have high cumulative COVID-19 death tolls [42], and they are among the top 4 nations in terms of Twitter users [43]; in addition, publicly accessible validation data on ICU bed demand are available for each country. Three main restrictions in terms of how many patients can be treated at a hospital during the pandemic are available PPE, available ICU beds, and available health care professionals per shift [44]. The number of available ICU beds was selected as the validation parameter for our model.

Data Sets

For tweets, we used the publicly available Twitter data set Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels (TBCOV), which contains >2 billion COVID-19 multilingual tweets, including geographic location and positive, negative, and neutral sentiment labels [45]. From this data set, negative sentiment tweets were selected to capture the volume of negative Twitter discourse surrounding COVID-19 for the United States, Brazil, and India. From this TBCOV data set, for the period from February 1, 2020, to March 31, 2021, a total of 59,832,393 tweets were extracted for the United States, of which 30,742,934 (51.38%) contained negative sentiment. For Brazil, there were 5,343,723 tweets, of which 3,004,039 (56.22%) contained negative sentiment. For India, there were 9,509,766 tweets, of which 4,199,151 (44.16%) contained negative sentiment.

Real-world hospital bed demand was defined as inpatient_beds_used_covid_coverage from the US Health Data COVID-19 Reported Patient Impact and Hospital Capacity by State Timeseries (RAW) data set [46] for the United States and as ICU beds needed from the Institute for Health Metrics and Evaluation COVID-19 Projections data set for Brazil and India [47]. The COVID-19 Projections data set contains daily information about each region’s need and capacity for hospital beds overall, including ICU beds. In India, similar to Brazil, the greatest medical supply demand during the pandemic was for oxygen cylinders and ICU beds [48]. Amid the pandemic, the insufficient oxygen-manufacturing capacity and the fragmented nature of the Indian health care system made it extremely difficult for people to obtain the supplies they needed in time [49]. Hashtags and sample tweets posted by the Indian public during the pandemic to secure oxygen cylinders and express the urgent need for ICU beds in specific regions have been documented [50, 51].

Data collection and analysis were conducted using Python (Python Software Foundation).

Granger Causality Test Analysis

A time series of each region was generated using the number of negative sentiment tweets per week. This time series was standardized with mean and SD calculated from historical tweet data. Another time series, the ground truth frequency of ICU bed demand, was generated from our preprocessed medical data. In general, all time-series data were binned in intervals of 1 week.

The Granger causality test was used to determine whether past negative tweet frequency contains information that can help forecast ICU bed demand, in addition to the information contained in the past values of ICU bed demand alone [52]. In theory, this test can be applied to a stationary time series. For a nonstationary time series, first or higher difference can be used instead [53, 54]. To see whether the time series could satisfy the requirement for the Granger causality test, the augmented Dickey-Fuller (ADF) test, which determines whether a time series is stationary or nonstationary, was used. In our implementation, the functions grangercausalitytests and adfuller from the statsmodels package for Python were used.

CCM Analysis

The CCM analysis workflow consisted of embedding, cross-mapping, and convergence analysis as well as validation and performance testing. In embedding, the negative sentiment tweets and ICU bed demand for each region were embedded into higher dimensional spaces to capture their underlying dynamics. In cross-mapping, the embedded time series were compared to identify their relationship. In convergence analysis, the results were assessed using statistical measures to determine whether there is a robust relationship between negative sentiment tweets and ICU bed demand.

Put another way, given 2 time series X and Y, their data point entries can be considered to exist in a vector space with x and y axes, and the points over time form a trajectory in the space. Likewise, one can include the time-delayed values of X as new axes, where the vectors can be \( <X(t), X(t-3), X(t-6),...>, <X(t-1), X(t-4), X(t-7),...> \), etc.

If the values of X over time do indeed influence or are linked to the values of Y, then a distance-weighted \( k \)-nearest neighbor mode in the \( X, X \text{ with delay } 1, X \text{ with delay } 2, \text{ etc vector space} \) applied to the same Y (and delay axes) space can have its output converge to the actual observed values of Y, that is, predict the value of Y. If the convergence between modeled Y from \( X \text{ with delays} \) and the actual observed values of Y is close, we can say that the model constructed represents the causality relation between X and Y.

In our implementation, the causal_ccm package for Python was used.

Ethical Considerations

Ethics approval was not required for our study because all data and information are publicly available. In addition, all user-identifiable information was excluded from the study results.

Results

Overview

From the TBCOV data set, for the period from February 1, 2020, to March 31, 2021, a total of 30,742,934 tweets containing negative sentiment were extracted for the United States; 3,004,039 tweets containing negative sentiment were extracted for Brazil; and 4,199,151 tweets containing negative sentiment were extracted for India. Our results can be categorized into (1) Granger causality test analysis and (2) CCM analysis.
Granger Causality Test Analysis

United States

Figure 1 shows time series graphs comparing negative sentiment COVID-19 tweets with real-world ICU bed demand data for each of the 50 US states.

Before performing the Granger causality test, the 2 time series were checked to determine whether they were stationary or nonstationary using ADF tests. After taking the second difference of the 2 time series, the $P$ values of the ADF tests for negative sentiment COVID-19 tweets and ICU bed demand were found to be <.05 for all US states, meaning that we were able to reject the null hypothesis ($H_0$) that a unit root was present in the time series samples; in other words, the 2 time series were stationary. The results are summarized in Multimedia Appendix 1.

The results for the Granger causality test with $H_0$, that is, negative sentiment COVID-19 tweets do not Granger-cause ICU bed demand in US states, are presented in Table 1. At lag 2, $H_0$ was rejected for 14 (28%) of the 50 US states ($P<.05$).
Figure 1. Time series with a comparison of trends for intensive care unit (ICU) bed use with trends for the volume of negative sentiment COVID-19 tweets across all 50 US states. TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels.
Table 1. Granger causality test for all 50 US states.

<table>
<thead>
<tr>
<th>State</th>
<th>P value</th>
<th>Number of lags</th>
<th>Reject null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>.37</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Alabama</td>
<td>.04</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Arkansas</td>
<td>.34</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Arizona</td>
<td>&lt;.001</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>California</td>
<td>.91</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Colorado</td>
<td>.61</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Connecticut</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Delaware</td>
<td>.40</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Florida</td>
<td>.04</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Georgia</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Hawaii</td>
<td>.93</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Iowa</td>
<td>.97</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Idaho</td>
<td>.45</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Illinois</td>
<td>.04</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Indiana</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Kansas</td>
<td>.25</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Kentucky</td>
<td>.97</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Louisiana</td>
<td>.10</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>.001</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Maryland</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Maine</td>
<td>.95</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Michigan</td>
<td>.02</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>Minnesota</td>
<td>.59</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Missouri</td>
<td>.43</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Mississippi</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Montana</td>
<td>.32</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>North Carolina</td>
<td>.02</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>North Dakota</td>
<td>.004</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>Nebraska</td>
<td>.93</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>.01</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>New Jersey</td>
<td>.93</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>New Mexico</td>
<td>.80</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Nevada</td>
<td>&lt;.001</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>New York</td>
<td>.001</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Ohio</td>
<td>.28</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>.047</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Oregon</td>
<td>.54</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>.53</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>.002</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>South Carolina</td>
<td>.78</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>South Dakota</td>
<td>.19</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Tennessee</td>
<td>.94</td>
<td>10</td>
<td>No</td>
</tr>
</tbody>
</table>
Before performing the Granger causality test, the 2 time series were checked to determine whether they were stationary or nonstationary using ADF tests, and \( H_0 \) was rejected for all Brazilian federative units \((P<.05)\), qualifying all of them for the Granger causality test. The results are summarized in Multimedia Appendix 2.

The results for the Granger causality test with \( H_0 \), that is, negative sentiment COVID-19 tweets do not Granger-cause ICU bed demand in Brazilian federative units, are presented in Table 2. At lag 2, \( H_0 \) was rejected for 6 (22\%) of the 27 Brazilian federative units \((P<.05)\).

<table>
<thead>
<tr>
<th>State</th>
<th>( P ) value</th>
<th>Number of lags</th>
<th>Reject null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texas</td>
<td>.32</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Utah</td>
<td>.97</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Virginia</td>
<td>.02</td>
<td>7</td>
<td>Yes</td>
</tr>
<tr>
<td>Vermont</td>
<td>.70</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Washington</td>
<td>.69</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>.95</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>West Virginia</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Wyoming</td>
<td>.24</td>
<td>10</td>
<td>No</td>
</tr>
</tbody>
</table>

**Brazil**

Figure 2 shows time series graphs comparing negative sentiment COVID-19 tweets with real-world ICU bed demand data for each of the 27 Brazilian federative units.

Before performing the Granger causality test, the 2 time series were checked to determine whether they were stationary or nonstationary using ADF tests, and \( H_0 \) was rejected for all Brazilian federative units \((P<.05)\), qualifying all of them for the Granger causality test. The results are summarized in Multimedia Appendix 2.

The results for the Granger causality test with \( H_0 \), that is, negative sentiment COVID-19 tweets do not Granger-cause ICU bed demand in Brazilian federative units, are presented in Table 2. At lag 2, \( H_0 \) was rejected for 6 (22\%) of the 27 Brazilian federative units \((P<.05)\).
Figure 2. Time series with a comparison of trends for intensive care unit (ICU) bed use with trends for the volume of negative sentiment COVID-19 tweets across Brazil’s 27 subdivisions (states and administrative divisions). TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels.
### Table 2. Granger causality test for Brazilian federative units.

<table>
<thead>
<tr>
<th>Federative unit</th>
<th>P value</th>
<th>Number of lags</th>
<th>Reject null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acre</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Alagoas</td>
<td>.90</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Amazonas</td>
<td>.28</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Amapá</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Bahia</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Ceará</td>
<td>&lt;.001</td>
<td>9</td>
<td>Yes</td>
</tr>
<tr>
<td>Federal District</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Espírito Santo</td>
<td>.55</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Goiás</td>
<td>.60</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Maranhão</td>
<td>.04</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>Minas Gerais</td>
<td>.85</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Mato Grosso do Sul</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Mato Grosso</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Pará</td>
<td>.006</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Paraíba</td>
<td>.04</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Paraná</td>
<td>.60</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Pernambuco</td>
<td>.01</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Piauí</td>
<td>.90</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Rio de Janeiro</td>
<td>.16</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Rio Grande do Norte</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Rondônia</td>
<td>.91</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Roraima</td>
<td>.002</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>Rio Grande do Sul</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Santa Catarina</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Sergipe</td>
<td>.48</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>São Paulo</td>
<td>.91</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Tocantins</td>
<td>.50</td>
<td>10</td>
<td>No</td>
</tr>
</tbody>
</table>

**India**

Figure 3 shows time series graphs comparing negative sentiment COVID-19 tweets with real-world ICU bed demand data for the 25 Indian states included in the analysis (Assam, Meghalaya, and Tamil Nadu were excluded owing to lack of data) and the national capital region of Delhi. Before performing the Granger causality test, the 2 time series were checked to determine whether they were stationary or nonstationary using ADF tests, and H₀ was rejected for all Indian states and the national capital region of Delhi (P<.05), qualifying all for the Granger causality test. The results are summarized in Multimedia Appendix 3.

The results for the Granger causality test with H₀, that is, negative sentiment tweets do not Granger-cause ICU bed demand in Indian states and the national capital region, are presented in Table 3. At lag 2, H₀ was rejected for 6 (23%) of the 26 included regions (25 Indian states and the national capital region; P<.05).
Figure 3. Time series with a comparison of trends for intensive care unit (ICU) bed use with trends for the volume of negative sentiment COVID-19 tweets across Indian states. TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels.
Table 3. Granger causality test for Indian states and the national capital region of Delhi.

<table>
<thead>
<tr>
<th>State or national capital region</th>
<th>P value</th>
<th>Number of lags</th>
<th>Reject null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haryana</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Madhya Pradesh</td>
<td>.60</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Andhra Pradesh</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Uttarakhand</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Gujarat</td>
<td>.86</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Manipur</td>
<td>.003</td>
<td>10</td>
<td>Yes</td>
</tr>
<tr>
<td>Himachal Pradesh</td>
<td>.83</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Punjab</td>
<td>.97</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Karnataka</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Jharkhand</td>
<td>.32</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Bihar</td>
<td>.45</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Arunachal Pradesh</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Sikkim</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Mizoram</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Goa</td>
<td>.81</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Kerala</td>
<td>&lt;.001</td>
<td>8</td>
<td>Yes</td>
</tr>
<tr>
<td>West Bengal</td>
<td>.28</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Maharashtra</td>
<td>.002</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>Tripura</td>
<td>.64</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Delhi</td>
<td>.91</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Uttar Pradesh</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Rajasthan</td>
<td>.88</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Nagaland</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Odisha</td>
<td>.98</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Telangana</td>
<td>.99</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>Chhattisgarh</td>
<td>.89</td>
<td>10</td>
<td>No</td>
</tr>
</tbody>
</table>

CCM Analysis

United States

Figure 4 illustrates the results of the CCM analysis for the United States, where a significant causal relationship was found between ICU bed demand and negative sentiment COVID-19 tweets for 46 (92%) of the 50 states (P<.05). The full list of correlation coefficients (r) and P values for the CCM analysis across US states is presented in Table 4.
Figure 4. Graphs showing the convergence of the correlation coefficient ($r$) in convergent cross-mapping analysis as the time series length ($L$) approaches the maximum possible value for each US state. Series $X =$ negative sentiment COVID-19 tweet proportion and series $Y =$ intensive care unit bed demand. The blue graph represents the correlation coefficient when modeling $X$ causing $Y$ ($X \rightarrow Y$). The orange graph represents the correlation coefficient when modeling $Y$ causing $X$ ($Y \rightarrow X$). The correlation coefficients ranged from 0.0405 (Vermont) to 0.7670 (New York). The $P$ values ranged from .44 (Vermont) to <.001 (New York). Of the 50 US states, 46 (92%) had $P$ values <.05.
Table 4. Results of the convergent cross-mapping analysis for each US state. The correlation coefficients ($r$) ranged from 0.0405 (Vermont) to 0.7670 (New York). Series $X$=negative sentiment COVID-19 tweet proportion and series $Y$=intensive care unit bed demand. $X\rightarrow Y$ refers to the correlation coefficient when modeling $X$ causing $Y$; $Y\rightarrow X$ refers to the correlation coefficient when modeling $Y$ causing $X$.

<table>
<thead>
<tr>
<th>State</th>
<th>Correlation coefficient, $r$</th>
<th>$P$ value</th>
<th>$X\rightarrow Y$</th>
<th>$Y\rightarrow X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>0.1080</td>
<td>.04</td>
<td>0.1080</td>
<td>-0.1520</td>
</tr>
<tr>
<td>Alabama</td>
<td>0.2280</td>
<td>&lt;.001</td>
<td>0.2280</td>
<td>0.0211</td>
</tr>
<tr>
<td>Arkansas</td>
<td>0.2320</td>
<td>&lt;.001</td>
<td>0.2340</td>
<td>-0.0166</td>
</tr>
<tr>
<td>Arizona</td>
<td>0.5740</td>
<td>&lt;.001</td>
<td>0.5730</td>
<td>0.5460</td>
</tr>
<tr>
<td>California</td>
<td>0.6270</td>
<td>&lt;.001</td>
<td>0.6260</td>
<td>0.1100</td>
</tr>
<tr>
<td>Colorado</td>
<td>0.5750</td>
<td>&lt;.001</td>
<td>0.5760</td>
<td>0.3770</td>
</tr>
<tr>
<td>Connecticut</td>
<td>0.7010</td>
<td>&lt;.001</td>
<td>0.7000</td>
<td>0.2770</td>
</tr>
<tr>
<td>Delaware</td>
<td>0.4810</td>
<td>&lt;.001</td>
<td>0.4800</td>
<td>0.5290</td>
</tr>
<tr>
<td>Florida</td>
<td>0.7430</td>
<td>&lt;.001</td>
<td>0.7430</td>
<td>0.5600</td>
</tr>
<tr>
<td>Georgia</td>
<td>0.4050</td>
<td>&lt;.001</td>
<td>0.4040</td>
<td>0.0673</td>
</tr>
<tr>
<td>Hawaii</td>
<td>0.1650</td>
<td>&lt;.001</td>
<td>0.1650</td>
<td>0.3960</td>
</tr>
<tr>
<td>Iowa</td>
<td>0.3930</td>
<td>&lt;.001</td>
<td>0.3870</td>
<td>0.2240</td>
</tr>
<tr>
<td>Idaho</td>
<td>0.2080</td>
<td>&lt;.001</td>
<td>0.2070</td>
<td>-0.0386</td>
</tr>
<tr>
<td>Illinois</td>
<td>0.5700</td>
<td>&lt;.001</td>
<td>0.5700</td>
<td>0.4190</td>
</tr>
<tr>
<td>Indiana</td>
<td>0.2150</td>
<td>&lt;.001</td>
<td>0.2190</td>
<td>0.3560</td>
</tr>
<tr>
<td>Kansas</td>
<td>0.3710</td>
<td>&lt;.001</td>
<td>0.3750</td>
<td>0.3380</td>
</tr>
<tr>
<td>Kentucky</td>
<td>0.2790</td>
<td>&lt;.001</td>
<td>0.2790</td>
<td>0.1840</td>
</tr>
<tr>
<td>Louisiana</td>
<td>0.3880</td>
<td>&lt;.001</td>
<td>0.3880</td>
<td>0.5810</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>0.3590</td>
<td>&lt;.001</td>
<td>0.3570</td>
<td>0.4900</td>
</tr>
<tr>
<td>Maryland</td>
<td>0.3980</td>
<td>&lt;.001</td>
<td>0.3980</td>
<td>0.5700</td>
</tr>
<tr>
<td>Maine</td>
<td>0.0526</td>
<td>.31</td>
<td>0.0518</td>
<td>0.2660</td>
</tr>
<tr>
<td>Michigan</td>
<td>0.5480</td>
<td>&lt;.001</td>
<td>0.5500</td>
<td>0.3720</td>
</tr>
<tr>
<td>Minnesota</td>
<td>0.2960</td>
<td>&lt;.001</td>
<td>0.2930</td>
<td>0.2920</td>
</tr>
<tr>
<td>Missouri</td>
<td>0.2920</td>
<td>&lt;.001</td>
<td>0.2890</td>
<td>0.2170</td>
</tr>
<tr>
<td>Mississippi</td>
<td>0.5950</td>
<td>&lt;.001</td>
<td>0.5950</td>
<td>0.2920</td>
</tr>
<tr>
<td>Montana</td>
<td>0.2600</td>
<td>&lt;.001</td>
<td>0.2600</td>
<td>0.2020</td>
</tr>
<tr>
<td>North Carolina</td>
<td>0.5470</td>
<td>&lt;.001</td>
<td>0.5400</td>
<td>0.5850</td>
</tr>
<tr>
<td>North Dakota</td>
<td>0.6000</td>
<td>&lt;.001</td>
<td>0.5990</td>
<td>0.6830</td>
</tr>
<tr>
<td>Nebraska</td>
<td>0.1580</td>
<td>.002</td>
<td>0.1570</td>
<td>0.2520</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>0.4080</td>
<td>&lt;.001</td>
<td>0.4080</td>
<td>0.0395</td>
</tr>
<tr>
<td>New Jersey</td>
<td>0.6830</td>
<td>&lt;.001</td>
<td>0.6770</td>
<td>0.4970</td>
</tr>
<tr>
<td>New Mexico</td>
<td>0.4810</td>
<td>&lt;.001</td>
<td>0.4800</td>
<td>0.3220</td>
</tr>
<tr>
<td>Nevada</td>
<td>0.1800</td>
<td>&lt;.001</td>
<td>0.1800</td>
<td>0.2380</td>
</tr>
<tr>
<td>New York</td>
<td>0.7670</td>
<td>&lt;.001</td>
<td>0.7660</td>
<td>0.4700</td>
</tr>
<tr>
<td>Ohio</td>
<td>0.4840</td>
<td>&lt;.001</td>
<td>0.4830</td>
<td>0.1210</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>0.0565</td>
<td>.27</td>
<td>0.0559</td>
<td>0.0657</td>
</tr>
<tr>
<td>Oregon</td>
<td>0.2590</td>
<td>&lt;.001</td>
<td>0.2570</td>
<td>0.1350</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>0.6290</td>
<td>&lt;.001</td>
<td>0.6280</td>
<td>0.4440</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>0.3900</td>
<td>&lt;.001</td>
<td>0.3990</td>
<td>0.3960</td>
</tr>
<tr>
<td>South Carolina</td>
<td>0.2550</td>
<td>&lt;.001</td>
<td>0.2540</td>
<td>0.2780</td>
</tr>
<tr>
<td>South Dakota</td>
<td>0.4140</td>
<td>&lt;.001</td>
<td>0.4130</td>
<td>0.1500</td>
</tr>
<tr>
<td>State</td>
<td>Correlation coefficient, r</td>
<td>P value</td>
<td>X→Y</td>
<td>Y→X</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Tennessee</td>
<td>0.4920</td>
<td>&lt;.001</td>
<td>0.4920</td>
<td>0.0250</td>
</tr>
<tr>
<td>Texas</td>
<td>0.4430</td>
<td>&lt;.001</td>
<td>0.4430</td>
<td>0.5220</td>
</tr>
<tr>
<td>Utah</td>
<td>0.2970</td>
<td>&lt;.001</td>
<td>0.2950</td>
<td>−0.0438</td>
</tr>
<tr>
<td>Virginia</td>
<td>0.3760</td>
<td>&lt;.001</td>
<td>0.3740</td>
<td>0.1550</td>
</tr>
<tr>
<td>Vermont</td>
<td>0.0405</td>
<td>.44</td>
<td>0.0341</td>
<td>0.0602</td>
</tr>
<tr>
<td>Washington</td>
<td>0.5930</td>
<td>&lt;.001</td>
<td>0.5930</td>
<td>0.2970</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>0.6140</td>
<td>&lt;.001</td>
<td>0.6140</td>
<td>0.2920</td>
</tr>
<tr>
<td>West Virginia</td>
<td>0.0746</td>
<td>.14</td>
<td>0.0752</td>
<td>−0.0510</td>
</tr>
<tr>
<td>Wyoming</td>
<td>0.2700</td>
<td>&lt;.001</td>
<td>0.2690</td>
<td>0.3740</td>
</tr>
</tbody>
</table>

**Brazil**

Figure 5 illustrates the results of the CCM analysis for Brazil, where a significant causal relationship was found between ICU bed demand and negative sentiment COVID-19 tweets for 26 (96%) of the 27 Brazilian federative units (P<.05). The full list of correlation coefficients (r) and P values for the CCM analysis across Brazil is presented in Table 5.
Figure 5. Graphs showing the convergence of the correlation coefficient (r) in convergent cross-mapping analysis as the time series length (L) approaches the maximum possible value for each Brazil subregion. Series X=negative sentiment COVID-19 tweet proportion and series Y= intensive care unit bed demand. The blue graph represents the correlation coefficient when modeling X causing Y (\(X \rightarrow Y\)). The orange graph represents the correlation coefficient when modeling Y causing X (\(Y \rightarrow X\)). The correlation coefficients ranged from 0.0751 (Amapá) to 0.8010 (Amazonas). The \(P\) values ranged from .14 (Amapá) to <.001 (Amazonas). Of the 27 Brazilian federative units, 26 (96%) had \(P\) values <.05.
Table 5. Results of the convergent cross-mapping analysis for each Brazilian federative unit. The correlation coefficients ($r$) ranged from 0.0751 (Amapá) to 0.8010 (Amazonas). Series X=negative sentiment COVID-19 tweet proportion and series Y=intensive care unit bed demand. $X \rightarrow Y$ refers to the correlation coefficient when modeling X causing Y; $Y \rightarrow X$ refers to the correlation coefficient when modeling Y causing X.

<table>
<thead>
<tr>
<th>Federative unit</th>
<th>Correlation coefficient, $r$</th>
<th>$P$ value</th>
<th>$X \rightarrow Y$</th>
<th>$Y \rightarrow X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acre</td>
<td>0.5800</td>
<td>&lt;.001</td>
<td>0.5800</td>
<td>0.7230</td>
</tr>
<tr>
<td>Alagoas</td>
<td>0.6010</td>
<td>&lt;.001</td>
<td>0.6010</td>
<td>0.6250</td>
</tr>
<tr>
<td>Amazonas</td>
<td>0.8010</td>
<td>&lt;.001</td>
<td>0.8010</td>
<td>0.8500</td>
</tr>
<tr>
<td>Amapá</td>
<td>0.0751</td>
<td>.14</td>
<td>0.0758</td>
<td>0.6270</td>
</tr>
<tr>
<td>Bahia</td>
<td>0.5940</td>
<td>&lt;.001</td>
<td>0.5940</td>
<td>0.4860</td>
</tr>
<tr>
<td>Ceará</td>
<td>0.3140</td>
<td>&lt;.001</td>
<td>0.3170</td>
<td>0.7180</td>
</tr>
<tr>
<td>Federal District</td>
<td>0.2720</td>
<td>&lt;.001</td>
<td>0.2600</td>
<td>0.1720</td>
</tr>
<tr>
<td>Espírito Santo</td>
<td>0.1940</td>
<td>&lt;.001</td>
<td>0.1940</td>
<td>0.4150</td>
</tr>
<tr>
<td>Goiás</td>
<td>0.3290</td>
<td>&lt;.001</td>
<td>0.3290</td>
<td>0.2490</td>
</tr>
<tr>
<td>Maranhão</td>
<td>0.3970</td>
<td>&lt;.001</td>
<td>0.3960</td>
<td>0.7280</td>
</tr>
<tr>
<td>Minas Gerais</td>
<td>0.5370</td>
<td>&lt;.001</td>
<td>0.5360</td>
<td>0.3490</td>
</tr>
<tr>
<td>Mato Grosso do Sul</td>
<td>0.1910</td>
<td>&lt;.001</td>
<td>0.1910</td>
<td>0.3350</td>
</tr>
<tr>
<td>Mato Grosso</td>
<td>0.3400</td>
<td>&lt;.001</td>
<td>0.3400</td>
<td>0.3070</td>
</tr>
<tr>
<td>Pará</td>
<td>0.4750</td>
<td>&lt;.001</td>
<td>0.4830</td>
<td>0.6790</td>
</tr>
<tr>
<td>Paraíba</td>
<td>0.3780</td>
<td>&lt;.001</td>
<td>0.3780</td>
<td>0.5710</td>
</tr>
<tr>
<td>Parana</td>
<td>0.6110</td>
<td>&lt;.001</td>
<td>0.6150</td>
<td>0.2810</td>
</tr>
<tr>
<td>Pernambuco</td>
<td>0.4350</td>
<td>&lt;.001</td>
<td>0.4350</td>
<td>0.7090</td>
</tr>
<tr>
<td>Piaui</td>
<td>0.2990</td>
<td>&lt;.001</td>
<td>0.2990</td>
<td>0.3390</td>
</tr>
<tr>
<td>Rio de Janeiro</td>
<td>0.4760</td>
<td>&lt;.001</td>
<td>0.4760</td>
<td>0.5930</td>
</tr>
<tr>
<td>Rio Grande do Norte</td>
<td>0.2810</td>
<td>&lt;.001</td>
<td>0.2780</td>
<td>0.5440</td>
</tr>
<tr>
<td>Rondônia</td>
<td>0.2760</td>
<td>&lt;.001</td>
<td>0.2770</td>
<td>0.4130</td>
</tr>
<tr>
<td>Roraima</td>
<td>0.5170</td>
<td>&lt;.001</td>
<td>0.5160</td>
<td>0.3440</td>
</tr>
<tr>
<td>Rio Grande do Sul</td>
<td>0.5040</td>
<td>&lt;.001</td>
<td>0.5040</td>
<td>0.6780</td>
</tr>
<tr>
<td>Santa Catarina</td>
<td>0.5400</td>
<td>&lt;.001</td>
<td>0.5490</td>
<td>0.2040</td>
</tr>
<tr>
<td>Sergipe</td>
<td>0.3230</td>
<td>&lt;.001</td>
<td>0.3220</td>
<td>0.3400</td>
</tr>
<tr>
<td>São Paulo</td>
<td>0.6240</td>
<td>&lt;.001</td>
<td>0.6250</td>
<td>0.5690</td>
</tr>
<tr>
<td>Tocantins</td>
<td>0.3530</td>
<td>&lt;.001</td>
<td>0.3530</td>
<td>0.2430</td>
</tr>
</tbody>
</table>

India

Figure 6 illustrates the result of the CCM analysis for India, where a significant causal relationship was found between ICU bed demand and negative sentiment COVID-19 tweets for 26 (100%) of the 26 included regions (25 states and the national capital region; $P<.05$). The full list of correlation coefficients ($r$) and $P$ values for the CCM analysis across Indian states and the national capital region is presented in Table 6.
Figure 6. Graphs showing the convergence of the correlation coefficient (r) in convergent cross-mapping analysis as the time series length (L) approaches the maximum possible value for each of the 26 Indian states included in the analysis. Series X = negative sentiment COVID-19 tweet proportion and series Y = intensive care unit bed demand. The blue graph represents the correlation coefficient when modeling X causing Y (X → Y). The orange graph represents the correlation coefficient when modeling Y causing X (Y → X). The correlation coefficients ranged from 0.1630 (Sikkim) to 0.8060 (Delhi). The P values ranged from .001 (Sikkim) to <.001 (Delhi). All 25 states and the national capital region of Delhi had P values <.05.
Table 6. Results of the convergent cross-mapping analysis for each of the 25 Indian states included in the analysis and the national capital region. Correlation coefficients (r) ranged from 0.163 (Sikkim) to 0.806 (Delhi). Series X= negative sentiment COVID-19 tweet proportion and series Y= intensive care unit bed demand. X→Y refers to the correlation coefficient when modeling X causing Y; Y→X refers to the correlation coefficient when modeling Y causing X.

<table>
<thead>
<tr>
<th>State or national capital region</th>
<th>Correlation coefficient, r</th>
<th>P value</th>
<th>X→Y</th>
<th>Y→X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haryana</td>
<td>0.486</td>
<td>&lt;.001</td>
<td>0.4860</td>
<td>0.4840</td>
</tr>
<tr>
<td>Madhya Pradesh</td>
<td>0.417</td>
<td>&lt;.001</td>
<td>0.4170</td>
<td>0.3950</td>
</tr>
<tr>
<td>Andhra Pradesh</td>
<td>0.258</td>
<td>&lt;.001</td>
<td>0.2580</td>
<td>0.2940</td>
</tr>
<tr>
<td>Uttarakhand</td>
<td>0.489</td>
<td>&lt;.001</td>
<td>0.4890</td>
<td>0.5920</td>
</tr>
<tr>
<td>Gujarat</td>
<td>0.669</td>
<td>&lt;.001</td>
<td>0.6680</td>
<td>0.6120</td>
</tr>
<tr>
<td>Manipur</td>
<td>0.511</td>
<td>&lt;.001</td>
<td>0.5110</td>
<td>0.3720</td>
</tr>
<tr>
<td>Himachal Pradesh</td>
<td>0.470</td>
<td>&lt;.001</td>
<td>0.4700</td>
<td>0.3000</td>
</tr>
<tr>
<td>Punjab</td>
<td>0.244</td>
<td>&lt;.001</td>
<td>0.2440</td>
<td>0.2260</td>
</tr>
<tr>
<td>Karnataka</td>
<td>0.646</td>
<td>&lt;.001</td>
<td>0.6480</td>
<td>0.3830</td>
</tr>
<tr>
<td>Jharkhand</td>
<td>0.517</td>
<td>&lt;.001</td>
<td>0.5170</td>
<td>0.4900</td>
</tr>
<tr>
<td>Bihar</td>
<td>0.685</td>
<td>&lt;.001</td>
<td>0.6850</td>
<td>0.3870</td>
</tr>
<tr>
<td>Arunachal Pradesh</td>
<td>0.232</td>
<td>&lt;.001</td>
<td>0.2320</td>
<td>0.2700</td>
</tr>
<tr>
<td>Sikkim</td>
<td>0.163</td>
<td>&lt;.001</td>
<td>0.1630</td>
<td>0.4870</td>
</tr>
<tr>
<td>Mizoram</td>
<td>0.321</td>
<td>&lt;.001</td>
<td>0.3180</td>
<td>0.3080</td>
</tr>
<tr>
<td>Goa</td>
<td>0.493</td>
<td>&lt;.001</td>
<td>0.4920</td>
<td>0.1650</td>
</tr>
<tr>
<td>Kerala</td>
<td>0.612</td>
<td>&lt;.001</td>
<td>0.6120</td>
<td>0.3420</td>
</tr>
<tr>
<td>West Bengal</td>
<td>0.631</td>
<td>&lt;.001</td>
<td>0.6360</td>
<td>0.4420</td>
</tr>
<tr>
<td>Maharashtra</td>
<td>0.739</td>
<td>&lt;.001</td>
<td>0.7420</td>
<td>0.5630</td>
</tr>
<tr>
<td>Tripura</td>
<td>0.470</td>
<td>&lt;.001</td>
<td>0.4680</td>
<td>0.0105</td>
</tr>
<tr>
<td>Delhi</td>
<td>0.806</td>
<td>&lt;.001</td>
<td>0.8090</td>
<td>0.4340</td>
</tr>
<tr>
<td>Uttar Pradesh</td>
<td>0.707</td>
<td>&lt;.001</td>
<td>0.7070</td>
<td>0.6370</td>
</tr>
<tr>
<td>Rajasthan</td>
<td>0.582</td>
<td>&lt;.001</td>
<td>0.5820</td>
<td>0.5200</td>
</tr>
<tr>
<td>Nagaland</td>
<td>0.315</td>
<td>&lt;.001</td>
<td>0.2920</td>
<td>0.1790</td>
</tr>
<tr>
<td>Odisha</td>
<td>0.614</td>
<td>&lt;.001</td>
<td>0.6140</td>
<td>0.6770</td>
</tr>
<tr>
<td>Telangana</td>
<td>0.573</td>
<td>&lt;.001</td>
<td>0.5730</td>
<td>0.2660</td>
</tr>
<tr>
<td>Chhattisgarh</td>
<td>0.390</td>
<td>&lt;.001</td>
<td>0.3900</td>
<td>0.1740</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

Given the need to prioritize the use of limited medical resources during a health care emergency scenario such as the COVID-19 pandemic, social media hold promise in identifying shortages and can be a cost-effective tool for the proper allocation of medical resources, particularly ICU beds, because when governments and organizations are well informed with real-time shortage data, they have the capacity to adequately address the immediate funding and supply needs of health care facilities. This strategy may help mitigate immediate risk to the public until a more systematic solution is possible.

This study sought to determine which patterns existed between negative sentiment COVID-19 tweets and real-world ICU bed shortages during the pandemic, with the aim of leveraging social media to pinpoint regional surges in ICU bed demand in the United States, Brazil, and India. Two statistical tests were conducted to investigate this: the Granger causality test and CCM analysis.

The Granger causality test aims to identify causalities where, in a stochastic system, 1 separable variable is useful for forecasting another. The results of the Granger causality test for this analysis (Figures 1-3 and Tables 1-3) indicate that negative sentiment COVID-19 tweets Granger-caused ICU bed shortage (P<.05) for 14 (28%) of the 50 US states, 6 (22%) of the 27 Brazilian federative unit, and 6 (23%) of the 26 Indian regions included in the analysis (25 states and the national capital region). By contrast, the CCM analysis aims to identify causalities for nonseparable variables that are linked in a dynamic system and can identify and quantify weak to moderate causalities that may be missed by the Granger causality test. For the 3 countries, nearly all subregions—46 (92%) of the 50 US states, 26 (96%) of the 27 Brazilian federative units, and 26...
For the United States (Tables 1 and 4), of the 50 states, 13 (26%) had a significant result for both the Granger causality test and the CCM analysis (Alabama, Arizona, Florida, Illinois, Massachusetts, Michigan, North Carolina, North Dakota, New Hampshire, Nevada, New York, Rhode Island, and Virginia), 33 (66%) passed the CCM test but not the Granger causality test (Alaska, Arkansas, California, Colorado, Connecticut, Delaware, Georgia, Hawaii, Iowa, Idaho, Indiana, Kansas, Kentucky, Louisiana, Maryland, Minnesota, Missouri, Mississippi, Montana, Nebraska, New Jersey, New Mexico, Ohio, Oregon, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Utah, Washington, Wisconsin, Wisconsin, and Wyoming), 1 (2%) passed the Granger causality test only (Oklahoma), and 3 (6%) passed neither the Granger causality test nor the CCM test (Maine, Vermont, and West Virginia). Considering that the majority of the US states (33/50, 66%) passed the CCM test but not the Granger causality test, it can be inferred that the causal relationship between negative sentiment COVID-19 tweets and ICU bed shortage is weak to moderate for US states because CCM analysis is better at detecting weak to moderate causalities than the Granger test. This also implies that the relationship between the 2 variables (social media sentiment and ICU bed shortage) is dynamic and influenced by a number of complex interacting factors such that CCM analysis may be the more appropriate method for detecting and modeling this relationship.

For Brazil (Tables 2 and 5), a similar pattern occurred such that nearly all federative units (26/27, 96%) passed the CCM test, but only a few (6/27, 22%) passed the Granger causality test. Of the 27 federative units, 6 (22%) passed both the Granger causality test and the CCM test (Ceará, Maranhão, Pará, Paraíba, Pernambuco, and Roraima), 20 (74%) passed the CCM test but not the Granger causality test (Acre, Alagoas, Amazonas, Bahia, Federal District, Espírito Santo, Goiás, Minas Gerais, Mato Grosso do Sul, Mato Grosso, Paraná, Piauí, Rio de Janeiro, Rio Grande do Norte, Rondônia, Rio Grande do Sul, Santa Catarina, Sergipe, São Paulo, and Tocantins), none passed the Granger causality test only, and 1 (4%) state passed neither test (Amapá). This result again supports the idea of a dynamic and complex causal relationship being detected between negative sentiment COVID-19 tweets and ICU bed shortage.

For India (Tables 3 and 6), a similar pattern emerged. Of the 26 regions included in the analysis (25 states and the national capital region), 3 (12%) passed both the Granger causality test and the CCM test (Manipur, Kerala, and Maharashtra), whereas the remaining 23 (88%) passed the CCM test only (Haryana, Madhya Pradesh, Andhra Pradesh, Uttarakhand, Gujarat, Himachal Pradesh, Punjab, Karnatak, Jharkhand, Bihar, Arunachal Pradesh, Sikkim, Mizoram, Goa, West Bengal, Tripura, Delhi, Uttarakhand, Rajasthan, Nagaland, Odisha, Telangana, and Chhattisgarh), meaning that all 25 states and the national capital region of Delhi passed at least the CCM test. This further demonstrates that CCM analysis is capable of successfully detecting a causal pattern between negative sentiment pandemic tweets and real-world medical resource shortage and that this can potentially be used to pinpoint specific regions that are expected to face surges in medical resource demand at a given time.

Overall, these results suggest that a significant relationship exists between negative sentiment COVID-19 tweets and real-world ICU bed demand at subnational scales and that this relationship can be effectively detected and modeled using CCM analysis. These findings also indicate that the data contained within social media discourse regarding the COVID-19 pandemic can indeed be leveraged to identify and forecast the real-world impacts of the pandemic in the form of ICU bed demand surges. Further optimization of methods for identifying patterns between X sentiment and real-world medical emergencies can support the development of an early warning system for the robust real-time prediction of health care demand surges. Such a system may give health care workers and government decision makers a critical head start when deciding how to most effectively allocate medical resources in a crisis.

**Future Directions**

These results open up the possibility to develop tools that can forecast hospital bed demand in certain regions, although further research is required. This modeling approach can have a significant impact in the context of the health care supply chain. Forecasting ability is a potential factor affecting supply chain performance [27-29]. One study found that tweets related to food insecurity were strongly correlated with real food insufficiencies [16]; the authors noted that there is potential for tweet sentiment analysis to be used as a cost-effective early warning system that can help direct food-related interventions. Similarly, our results suggest that there is potential for negative sentiment COVID-19 posts to relate to actual medical resource shortage in regions where people use public discourse platforms such as Twitter (since rebranded as X). Although the current iteration of the methods described in this paper is only relevant to the time before a peak infection in the region, such methods have the potential to advance preparedness measures for future pandemics as they become more robust [31,32].

Digital data sources can aid in the identification of changes in disease activity, and it is worth exploring whether they show better performance in this regard than traditional COVID-19 metrics such as confirmed cases [33]. In low- and middle-income countries, there is potential for social media to act as a cost-effective early warning system to identify priority regions for medical resource allocation in real time. User behavior data can be extracted, given the unique social network parameters of a region, including the language spoken and the preference for 1 social media platform over another [34]. Similar to how a multilingual data set was used in the analyses in this study, Lopreote et al [55] analyzed a data set of tweets posted in various European languages and found that there were early warning signals of COVID-19 outbreaks before public announcements about an outbreak were made. Social networks that are particular to a region can provide user behavior data that can inform early warning detection systems specific to that region; for example, using Baidu search data, Qin et al [56] were able to predict the number of new COVID-19 cases such...
as fever, coronavirus, and pneumonia. Their study, along with similar social media–based early warning detection efforts [37-40], shows potential for the creation of a more effective yet affordable model to forecast new cases.

As noted in the previous subsection, there are complex interacting factors that may explain the results observed. Thus, further investigation is needed to unveil the dynamics underpinning the relationship between X negative sentiment pandemic tweets and medical resource demand; for instance, sociopolitical and economic challenges may have had varying influences on the X discourse by affecting public perceptions of COVID-19 management measures across countries. For Brazil (Figure 2), it can be seen that an increased proportion of negative sentiment preceded the ICU bed demand surge. Concern about preventing the proliferation of COVID-19 was among the major emergent X topics in Brazil, and politics also influenced the X landscape [57]. Brazil has a decentralized health system where the federal government finances the states and municipalities that provide health care services. Different levels of government must be able to coordinate effectively, and a failure to do so can lead to a disjunction in the care provided. A study by Neves et al [58] found that key government stakeholders underestimated the magnitude of the pandemic in the early weeks; for example, the Minas Gerais State Health Department posted messages on social media about not suspending the carnival, which is Brazil’s largest festival. When questionable COVID-19 management decisions are made that increase public health risk, such as decisions not to suspend major public events (eg, Brazil’s annual carnival), it is possible for the X discourse to reflect negative public perceptions of these government decisions before their negative impacts on case count and ICU bed demand. Further research is required to better understand the reliability of social media discourse in reflecting current pandemic management landscapes so that web-based public sentiment can accurately forecast pandemic impacts.

Incongruencies among different levels of government may have also contributed to the results observed. The then President of Brazil, Jair Messias Bolsonaro, issued messages that conflicted with those issued by states and municipalities, such as defending hydroxychloroquine as a COVID-19 treatment and countering mask use and physical distancing [59]. At a state-specific level, the governor of Rio de Janeiro state faced charges regarding irregularities in contracts for building emergency field hospitals, which prevented efforts to relieve strain on hospitals [60]. In addition, the state government of São Paulo was being investigated for ventilator fraud, which led to a shortage of ventilators that prevented citizens from accessing life-saving equipment [61]. Overall, the lack of coordinated effort and strategy for dealing with the pandemic contributed to the inconsistent implementation of preventive measures across the states and resulted in confusion for the public. State-specific challenges may have also contributed to the rise in negative sentiment and increased medical resource demand, which can help explain the results observed.

A similar exploration of potential factors for India can shed light on how they contributed to the nuanced and complex relationship observed between negative sentiment COVID-19 tweets and medical resource demand. India emerged as an interesting case study during the pandemic because of the novel use of X by citizens as a way of sourcing ICU beds and ventilators for their loved ones. As citizens found out that there was a shortage of equipment in the hospitals, they took to X in an effort to source medical supplies that were desperately needed. Adherence to government policy is another factor to consider in the Indian X discourse; for example, both COVID-19 waves were associated with nationwide shutdowns, but as the months wore on, people started to adhere less to measures such as masking and physical distancing. This may have contributed to an increase in cases, overwhelming hospitals [62].

Further exploration into region-specific factors as well as social and political contexts will be important for refining our forecasting models and gaining a better understanding of the complex relationship between X negative sentiment pandemic tweets and medical resource demand.

Limitations

One process in our methodology involved using tweets that already had sentiment labels and analyzing them. However, we are aware that there may not be data sets containing pandemic-related tweets in future contexts. Therefore, I recommendation would be to first extract relevant keyword–related tweets and increase accuracy with the help of supervised natural language processing models. A similar study with the aim of predicting medical resource shortages based on tweets was conducted for the state of New York [63]. The method consisted of using supervised learning to find related tweets, which is a more robust method.

One other limitation is that the ICU bed data sets for Brazil and India that were used to validate the model have decreased reliability because they only provide an estimate of the actual data [18], and they may not have accounted for fluctuations in ICU bed supply. However, because there were no other publicly available data sets, it was reasonable to use these data sets.

The use of X data has some important ethical considerations that lie at the intersection of privacy and data collection; for example, social media data collection can be considered a double-edged sword. On the one hand, it provides very valuable data that can increase the possibility of creating important solutions such as a more cost-effective and faster method of gauging PPE shortages and medical resource demand. On the other hand, current artificial intelligence and data collection practices have raised concerns about privacy and the selling of personal data.

Conclusions

The COVID-19 pandemic has made it clear that adequate demand-based allocation of medical resources and adequate preparedness for surges in hospital admissions are paramount to reduce cases and deaths at the onset of a pandemic. Between the time that a pandemic hits and the time that a vaccine is developed and distributed, it is vital that the medical supply is carefully managed to ensure that all health care facilities have adequate capacity and proper plans for meeting unprecedented medical resource demand surges. This study analyzed negative sentiment COVID-19 tweets that were compared with real-world data.
ICU bed use data. Our results show promise that leveraging social media, particularly X, has the potential to provide a cost-effective relatively rapid method that can inform resource allocation to facilities that need it most.

Further investigation into the potential of X data in the modeling of medical supply shortages may lead to the development of powerful tools that can inform health care decision-making in pandemic scenarios. X causal analysis in shortage forecasting has the potential to be applied broadly in a global context for identifying medical resource demand and informing health care supply chain decisions.

Acknowledgments

The authors would like to thank the STEM Fellowship for the opportunity to participate in the Inter-University Big Data Challenge 2022 during which this project was developed. JDK acknowledges support from Canada’s International Development Research Centre (grant 109981) and the International Development Research Centre and the Swedish International Development Cooperation Agency (grant 109559-001). JDK equally acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (grant RGPIN-2022-04559), the NSERC Discovery Launch Supplements (grant DGECR-2022-00454), and the New Frontiers in Research Fund–Exploration Stream (grant NFRFE-2021-00879).

Data Availability

The data sets generated and analyzed during this study are available in the Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels repository [46]; the Institute for Health Metrics and EvaluationCOVID-19 Projections repository [48]; and the COVID-19 Reported Patient Impact and Hospital Capacity by State Timeseries (RAW) [47].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Augmented Dickey-Fuller Test for US states.
[DOCX File , 22 KB - formative_v8i1e46087_app1.docx ]

Multimedia Appendix 2
Augmented Dickey-Fuller Test for Brazil subregions.
[DOCX File , 18 KB - formative_v8i1e46087_app2.docx ]

Multimedia Appendix 3
Augmented Dickey-Fuller Test for India states.
[DOCX File , 17 KB - formative_v8i1e46087_app3.docx ]

References


Abbreviations

ADF: augmented Dickey-Fuller
CCM: convergent cross-mapping
ICU: intensive care unit
PPE: personal protective equipment
TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels

©Mahakprit Kaur, Taylor Cargill, Kevin Hui, Minh Vu, Nicola Luigi Bragazzi, Jude Dzevela Kong. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Exploring the Perspectives of Patients Living With Lupus: 
Retrospective Social Listening Study

Erica Spies¹, PhD; Thomas Andreu², PhD; Matthias Hartung², PhD; Josephine Park¹, PhD; Paul Kamudoni³, PhD

¹EMD Serono, Billerica, MA, United States
²Semalytix GmbH, Bielefeld, Germany
³The Healthcare Business of Merck KGaA, Darmstadt, Germany

Corresponding Author:
Paul Kamudoni, PhD
The Healthcare Business of Merck KGaA
Frankfurter Strasse 250
Darmstadt, 64293
Germany
Phone: 49 15114543257
Email: paul.kamudoni@emdgroup.com

Abstract

Background: Systemic lupus erythematosus (SLE) is a chronic autoimmune inflammatory disease affecting various organs with a wide range of clinical manifestations. Cutaneous lupus erythematosus (CLE) can manifest as a feature of SLE or an independent skin ailment. Health-related quality of life (HRQoL) is frequently compromised in individuals living with lupus. Understanding patients’ perspectives when living with a disease is crucial for effectively meeting their unmet needs. Social listening is a promising new method that can provide insights into the experiences of patients living with their disease (lupus) and leverage these insights to inform drug development strategies for addressing their unmet needs.

Objective: The objective of this study is to explore the experience of patients living with SLE and CLE, including their disease and treatment experiences, HRQoL, and unmet needs, as discussed in web-based social media platforms such as blogs and forums.

Methods: A retrospective exploratory social listening study was conducted across 13 publicly available English-language social media platforms from October 2019 to January 2022. Data were processed using natural language processing and knowledge graph tagging technology to clean, format, anonymize, and annotate them algorithmically before feeding them to Pharos, a Semalytix proprietary data visualization and analysis platform, for further analysis. Pharos was used to generate descriptive data statistics, providing insights into the magnitude of individual patient experience variables, their differences in the magnitude of variables, and the associations between algorithmically tagged variables.

Results: A total of 45,554 posts from 3834 individuals who were algorithmically identified as patients with lupus were included in this study. Among them, 1925 (authoring 5636 posts) and 106 (authoring 243 posts) patients were identified as having SLE and CLE, respectively. Patients frequently mentioned various symptoms in relation to SLE and CLE including pain, fatigue, and rashes; pain and fatigue were identified as the main drivers of HRQoL impairment. The most affected aspects of HRQoL included “mobility,” “cognitive capabilities,” “recreation and leisure,” and “sleep and rest.” Existing pharmacological interventions poorly managed the most burdensome symptoms of lupus. Conversely, nonpharmacological treatments, such as exercise and meditation, were frequently associated with HRQoL improvement.

Conclusions: Patients with lupus reported a complex interplay of symptoms and HRQoL aspects that negatively influenced one another. This study demonstrates that social listening is an effective method to gather insights into patients’ experiences, preferences, and unmet needs, which can be considered during the drug development process to develop effective therapies and improve disease management.

(JMIR Form Res 2024;8:e52768) doi:10.2196/52768

KEYWORDS
systemic lupus erythematosus; SLE; cutaneous lupus erythematosus; CLE; quality of life; health-related quality of life; HRQoL; social media listening; lupus; rare; cutaneous; social media; infodemiology; infoveillance; social listening; natural language

https://formative.jmir.org/2024/1/e52768

JMIR Form Res 2024 | vol. 8 | e52768 | p.3412
(page number not for citation purposes)
Introduction

Systemic lupus erythematosus (SLE) is an autoimmune inflammatory disease affecting multiple systems in the body; it is characterized by fluctuating symptoms and periods of exacerbation and remission [1-3]. The most common symptoms of SLE include fatigue, skin rashes, fever, and joint pain or swelling [3]. SLE has several phenotypes and clinical manifestations involving various organs, including the joints, skin, kidneys, and organs of the neurological or hematological systems [1,2]. Cutaneous lupus erythematosus (CLE) can manifest as either a feature of SLE or an independent skin ailment [4,5]. The most prevalent symptoms of CLE include rashes, hair loss, blood vessel inflammation, ulcers, and increased sensitivity to light [6,7]. Living with these symptoms can be physically debilitating for patients and can disrupt their family, social, and professional life, thereby negatively impacting their health-related quality of life (HRQoL) [8,9].

Globally, SLE and CLE are estimated to have incidence rates of 5.1 and 4.3 per 100,000 person-years, respectively [10,11]. These rates vary widely according to demographic factors such as ethnicity, sex, and socioeconomic status, with a higher incidence in women, African Americans and other non-White populations [10,11]. The current treatment options for SLE include immunosuppressive agents, new biological therapies, and combination therapies of biologicals with immunosuppressive and immunomodulating agents [12]. Despite recent advances in therapeutic strategies, most of the current medications available for the treatment of lupus provide only symptomatic relief and are frequently associated with undesirable side effects [13,14].

Considering the high clinical variability, inadequate treatment options, and poor HRQoL of patients, understanding the perspectives and experiences of patients with SLE and CLE may be critical for developing effective therapies and improving disease management [13-15]. In the context of drug development, regulatory bodies and health care decision makers are emphasizing patient-focused drug development, which involves actively seeking and incorporating patients’ perspectives when designing interventions that can meet their needs, improve outcomes, and enhance the overall patient experience [16]. The increased presence and active engagement of patients on digital platforms, specifically social media, can serve as a source for understanding their needs, treatment experiences, and the factors affecting treatment decisions in real-world scenarios [15]. Additionally, data regarding the experiences of patients living with the disease can potentially influence key decisions and activities during drug development [16].

Data regarding the experiences of individuals living with diseases can be obtained through various sources, such as observational studies, interviews, focus groups, and patient advisory boards. Social listening is a promising newer approach that complements traditional methods of data collection by providing additional insights into patients’ perspectives beyond clinical settings [17]. Furthermore, this method captures the voices of vulnerable and difficult-to-reach populations who may otherwise not participate in clinical or epidemiological studies, thereby providing an opportunity for data triangulation [17,18]. In this social listening study, we explored the experience of patients living with lupus (SLE and CLE), as discussed in web-based social media platforms, such as blogs and forums, by searching for posts regarding disease burden, HRQoL impacts, treatment experience, and unmet needs.

Methods

Study Design

Overview

This retrospective exploratory social listening study was conducted between October 2019 and January 2022 across 13 publicly available English-language social media platforms (Multimedia Appendix 1). The processes of data identification, collection, and analysis involved have been previously presented [19-22]. The process involved algorithmic processing steps (1-3) and research steps (4 and 5) performed by human analysts (Figure 1). A glossary of terms are provided in Multimedia Appendix 2.
Data Source Identification and Data Collection (Step 1)
Identifying data sources was the first step, followed by data collection. Lupus-related keywords, defined and revised by human domain experts, were searched in SocialGist, a third-party search engine providing access to social media sites through an application programming interface search engine, to detect websites hosting relevant lupus-specific content. The following keywords were used for the search: “lupus,” “lupus” AND “systemic” AND “cutaneous” AND “subacute” AND “erythematosus” AND “autoimmune” AND “disease,” “SLE,” and “CLE.” All posts from relevant websites found through the searches were retrieved and collected in Pharos, a proprietary data visualization and analysis platform from Semalytix GmbH.

Filtering and Aggregating Content (Step 2)
Filtering and aggregating the content based on algorithmically determined inclusion criteria was the second step. Posts collected in step 1 were aggregated by unique authors, and all posts by the same author were collapsed into 1 author-specific record. Based on algorithmically determined inclusion criteria, only records classified as having been authored by a patient with lupus were retained. Posts authored by nonpatients, such as caregivers and health care providers, and those involving other types of documents, such as journal papers and press releases, were excluded. Throughout this process, human language engineers and data labeling specialists carefully monitored the patient classifications to ensure accuracy and reliability.

Algorithmic Coding of Patient Experience Concepts (Step 3)
Patient experience themes were algorithmically coded in the study’s third step. All lupus-specific patient records from step 2 were algorithmically annotated with patient experience tags using natural language processing. Patient experience themes mentioned in texts related to disease burden, HRQoL, and treatment experiences were added to patient records as semantic tags. Individual aspects (facets) of HRQoL were investigated based on the taxonomy provided by the World Health Organization Quality of Life instrument [23], which served as a guide for developing the inventory of patient experience tags used in this study. Human language engineers and data labeling specialists monitored, cross-validated, and adapted the algorithms as needed. A subject matter expert reviewed the study’s overall approach and medical content.

Quantitative Analysis (Step 4)
Quantitative data analysis was involved in the fourth step, wherein analysts used the Pharos Patient Experience Platform to generate descriptive statistics regarding patient experience tags. The statistics provided insight into the magnitude of individual patient experience variables, the differences between them, and the associations among the algorithmically tagged variables.

Qualitative Analysis (Step 5)
Qualitative data analysis was involved in the fifth step. Human analysts used Pharos to investigate the relationships between tags, such as symptoms and outcomes or symptoms and impacts, as well as the language patients used to describe their experiences on social media to identify any emerging themes. The results of the qualitative analysis were used to identify themes in patient language, connections to tagged variables, potential connections between themes, emerging themes, and reorganized themes.
**Study Population**

Authors of posts who self-identified as patients living with lupus or any of its subtypes (SLE and CLE; algorithmically identified) were included in this study. All posts by the same user were aggregated into a unique patient record for further data analysis. A machine learning algorithm classified authors as patients based on the language used in their self-reported posts. The algorithm was specifically designed to differentiate expressions that indicated that a patient had lupus (“I have lupus” or “I was diagnosed with SLE in 2020”) from those that were ambiguous.

**Study Outcomes**

Posts from social media platforms were analyzed to investigate symptom burden, impact of lupus on HRQoL, treatment experience, and unmet needs. The related outcomes are described in Table 1. These outcomes were investigated in the SLE as well as CLE-related posts, depending on the robustness of the data.

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom burden</td>
<td>Patient-reported symptoms</td>
</tr>
<tr>
<td></td>
<td>Most burdensome symptoms(^a)</td>
</tr>
<tr>
<td></td>
<td>Areas of involvement (body parts)</td>
</tr>
<tr>
<td></td>
<td>Photosensitivity and body image(^a)</td>
</tr>
<tr>
<td></td>
<td>Sleep disorders</td>
</tr>
<tr>
<td></td>
<td>Comorbidities(^a)</td>
</tr>
<tr>
<td>Overall HRQoL(^b) impact</td>
<td>Psychological well-being (positive feelings and negative feelings)</td>
</tr>
<tr>
<td></td>
<td>Physical well-being(^a) (mobility(^a), activities of daily living(^a), recreation and leisure, and role participation(^a))</td>
</tr>
<tr>
<td></td>
<td>Impact and functioning (including parental care, burden to or on others, and economic or working capability impacts)(^a)</td>
</tr>
<tr>
<td>Treatment experience</td>
<td>Current treatment options (nonpharmacological and pharmacological interventions)</td>
</tr>
<tr>
<td></td>
<td>Treatment options and HRQoL aspects based on treatment(^a)</td>
</tr>
<tr>
<td></td>
<td>Level of satisfaction with current treatment options(^a)</td>
</tr>
<tr>
<td></td>
<td>Diagnostic delay and health care availability</td>
</tr>
<tr>
<td></td>
<td>Coping mechanisms</td>
</tr>
</tbody>
</table>

\(^a\)Owing to low sample size or insufficient data, these outcomes were not analyzed in the cutaneous lupus erythematosus population.

\(^b\)HRQoL: health-related quality of life.

**Ethical Considerations**

Research using data from social media can present ethical challenges owing to the limited guidance on the participants’ consent and anonymity [24]. The data analyzed in this study were obtained from publicly available sources. Identities of the patient posts were appropriately anonymized while ensuring that the data answered specific research questions. To ensure personal data protection, a strict General Data Protection Regulation–compliant process was adopted. Please refer to Multimedia Appendix 3 for further details on the appropriate measures implemented to ensure personal data protection. This study was exempt from ethical review because it examined retrospective publicly available data from a sizeable number of patients.

**Results**

**Patients**

During the initial data collection (step 1), a total of 76,538 lupus-related posts were collected, of which 45,554 posts from 3834 patients were included in the lupus-specific data set based on the algorithmically determined inclusion criteria. Among these patients, 1925 (with 5636 posts) and 106 (with 243 posts) patients were identified as having SLE and CLE, respectively. In the population with SLE, sex information was available for 583 patients, and the female-to-male ratio was 9:1. Age data were available for 402 patients, with 76.7% (n=308) aged <60 years.

In the following sections, we report the findings regarding the population with SLE followed by the population with CLE. The study results are structured under 3 primary sections: disease burden as reported by patients, overall HRQoL, and treatment experiences reported for both patient populations separately.

**Disease Burden in Patients With SLE**

**Overview**

Disease burden was evident among patients with SLE as indicated by a total of 2029 patient mentions across different social media platforms. Disease burden was quantified in terms of the symptoms that were patient-reported, most burdensome, their severity, areas of involvement (body parts), and comorbidities.

**Patient-Reported Symptoms**

The most frequently reported symptoms (n=2029 patients) were pain, fatigue, and rashes. Some patients were more specific regarding the type of pain they experienced, including arthralgia, cephalgia, and arthritis. Other less frequently (<10% [203 patients]) reported symptoms included anxiety, clinical depression, alopecia, and pyrexia (Table 2). In total, 20 patients (in approximately 30 posts) reported experiencing...
photosensitivity, with excessive sun exposure leading to flare-ups and worsening of symptoms such as joint pain, weakness, and fatigue.

Table 2. Symptoms most frequently reported by patients with SLE\(^a\) (n=2029\(^b\) patient mentions from 5636 posts) and CLE\(^c\) (n=118\(^b\) patient mentions from 244 posts).

<table>
<thead>
<tr>
<th>Symptoms or signs</th>
<th>Proportion of patients, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SLE</td>
</tr>
<tr>
<td>Pain</td>
<td>610 (30.1)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>386 (19.2)</td>
</tr>
<tr>
<td>Rash</td>
<td>239 (11.8)</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>184 (9.1)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>150 (7.4)</td>
</tr>
<tr>
<td>Cephalgia</td>
<td>143 (7.1)</td>
</tr>
<tr>
<td>Clinical depression</td>
<td>98 (4.8)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>78 (3.8)</td>
</tr>
<tr>
<td>Alopecia</td>
<td>71 (3.5)</td>
</tr>
<tr>
<td>Pyrexia</td>
<td>70 (3.4)</td>
</tr>
<tr>
<td>Scar</td>
<td>—</td>
</tr>
<tr>
<td>Photosensitivity</td>
<td>—</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)SLE: systemic lupus erythematosus.
\(^b\)Indicates that patients were counted more than once if they discussed more than one symptom in a post.
\(^c\)CLE: cutaneous lupus erythematosus.
\(^d\)Not available.

Although sleep disturbances were not among the top 10 patient-reported symptoms, it was found to significantly impact the HRQoL of patients with SLE. Among the 114 patients discussing “sleep and rest” topics, over half of the patients (n=70) reported experiencing recurrent insomnia since many years. Nighttime sleep disturbances and daytime napping were common, with some patients stating that insomnia or a hyperactive mood was an early sign of disease flare-ups. Patients reported that insomnia was also linked to the side effects of certain medications taken for their lupus, such as hydroxychloroquine, methotrexate, and prednisone, whereas medications such as cannabidiol and nonpharmacological interventions including exercise, meditation, and acupuncture had a positive impact on sleep maintenance disorders. Quotations of patients with SLE describing their symptoms are provided in Multimedia Appendix 4.

Most Burdensome Symptoms

The most burdensome symptoms were identified by analyzing the degree of severity with which a patient with SLE was affected based on linguistic cues provided by them. The most burdensome symptoms included pain (general pain: 289/870, 33.2%; arthralgia: 59/228, 25.9%; and cephalgia: 65/154, 42.2%), fatigue (192/474, 40.5%), and anxiety (51/193, 26.4%). Other symptoms that were perceived burdensome included clinical depression (30/128, 23.4%), photosensitivity (5/33, 15%), arthritis (12/99, 12.1%), hypothyroidism (2/36, 5.6%), and antiphospholipid syndrome (4/79, 5.1%). Of note, it is possible that the same patient may have reported more than one symptom or sign.

Areas of Involvement

The most frequently reported areas of involvement (n=1778 patients) were the skin (n=224, 12.6%), face (n=224, 12.6%), and eyes (n=197, 11.1%; Multimedia Appendix 5).

Comorbidities

In addition to the symptoms from the main indication, patients with SLE reported experiencing multiple comorbidities (n=488 patients). The most frequently (>5% [24 patients]) reported comorbidities were major depressive disorder (n=98, 20.1%), fibromyalgia (n=66, 13.5%), infection (n=62, 12.7%), dry eye syndrome (n=50, 10.2%), rheumatoid arthritis (n=40, 8.2%), and hypothyroidism (n=34, 7%).

Overall HRQoL in Patients With SLE

Overview

To investigate the HRQoL of patients, the frequency of mentions and the perceived importance of issues related to HRQoL facets were analyzed. Social networking communities discussed several aspects of HRQoL, as evidenced by 2199 patient mentions and 2138 posts accounting to 37.9% (2138/5636) of all posts analyzed in this study. These posts on HRQoL by patients with SLE contained descriptions of how they perceived the most burdensome symptoms and the impact on their daily lives. The most commonly discussed HRQoL aspects based on the total
number of patients mentioning any HRQoL facet (n=2199 patient mentions) were “negative feelings” (n=521, 23.7%) and “recreation and leisure” (n=314, 14.3%). Other commonly discussed HRQoL aspects included “health care availability,” “positive feelings,” “mobility,” “energy and motivation,” “cognitive capabilities,” “sleep and rest,” “work capacity,” and others (Table 3).

Table 3. Distribution of HRQoL aspects among patients with SLE\(^b\) (n=2199\(^c\) patient mentions from 5636 posts) and CLE\(^d\) (n=78\(^c\) posts from n=47 patients).

<table>
<thead>
<tr>
<th>HRQoL aspects</th>
<th>Proportion of patients, n (%)</th>
<th>SLE</th>
<th>CLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative feelings</td>
<td>521 (23.7)</td>
<td>20  (25.6)</td>
<td></td>
</tr>
<tr>
<td>Recreation and leisure</td>
<td>314 (14.3)</td>
<td>14  (17.9)</td>
<td></td>
</tr>
<tr>
<td>Health care availability</td>
<td>205 (9.3)</td>
<td>8   (10.3)</td>
<td></td>
</tr>
<tr>
<td>Positive feelings</td>
<td>183 (8.3)</td>
<td>8   (10.3)</td>
<td></td>
</tr>
<tr>
<td>Mobility</td>
<td>176 (8)</td>
<td>7   (8.9)</td>
<td></td>
</tr>
<tr>
<td>Energy and motivation</td>
<td>155 (7)</td>
<td>4   (5.1)</td>
<td></td>
</tr>
<tr>
<td>Cognitive capabilities</td>
<td>154 (7)</td>
<td>5   (6.4)</td>
<td></td>
</tr>
<tr>
<td>Sleep and rest</td>
<td>114 (5.2)</td>
<td>3   (3.8)</td>
<td></td>
</tr>
<tr>
<td>Work capacity</td>
<td>113 (5.1)</td>
<td>3   (3.8)</td>
<td></td>
</tr>
<tr>
<td>Activities of daily living</td>
<td>106 (4.8)</td>
<td>—</td>
<td>e</td>
</tr>
<tr>
<td>Financial resources</td>
<td>104 (4.7)</td>
<td>2   (2.6)</td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>49 (2.2)</td>
<td>3   (3.8)</td>
<td></td>
</tr>
<tr>
<td>Intimacy and sex</td>
<td>5 (0.2)</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)HRQoL: health-related quality of life.  
\(^b\)SLE: systemic lupus erythematosus.  
\(^c\)Indicates that patients were counted more than once if they discussed more than one HRQoL aspect in a post.  
\(^d\)CLE: cutaneous lupus erythematosus.  
\(^e\)Not available.

Patients frequently reported impairments in areas such as “mobility,” “cognitive capabilities,” and “recreation and leisure,” for which pain and fatigue were the leading causes. While pain primarily affected “mobility,” “recreation and leisure,” and “sleep and rest,” fatigue affected multiple HRQoL facets. Overall, the impact of pain on “mobility” and fatigue on “cognitive capabilities” was the most substantial factor impairing HRQoL (Multimedia Appendix 6).

**Psychological Well-Being**

**Negative Feelings**

Posts from 1371 patients revealed “negative feelings” (339/521, 65.1%) as the highest-ranked HRQoL facet in terms of high importance to the patient. Further examination of these posts from 521 patients revealed that severe symptoms or side effects such as pain, fatigue, rashes, and depression affected the HRQoL of these patients, thereby leading to frequent mentions of “negative feelings.” Patients were still considered as highly important HRQoL by 63.2% (116/183) of patients, which is comparable to the high importance of “negative feelings” (339/521, 65.1%), suggesting that patients viewed both dimensions as equally important. Among the positive emotions mentioned, “feeling better,” “thankfulness,” and “hopefulness” were the most commonly reported. Notable patient quotations on “negative feelings” and “positive feelings” are presented in Multimedia Appendix 7.

**Positive Feelings**

“Positive feelings” were expressed by the patients who experienced feelings of satisfaction or wellness and emotions, such as excitement, interest, pride, love, and optimism. “Positive feelings” were mentioned less frequently than “negative feelings” (183/2199, 8.3% vs 521/2199, 23.7%); nevertheless, they were considered as highly important HRQoL by 63.2% (116/183) of patients, which is comparable to the high importance of “negative feelings” (339/521, 65.1%), suggesting that patients viewed both dimensions as equally important. Among the positive emotions mentioned, “feeling better,” “thankfulness,” and “hopefulness” were the most commonly reported. Notable patient quotations on “negative feelings” and “positive feelings” are presented in Multimedia Appendix 7.

**Physical Well-Being**

**Mobility**

Patients reported that the severity of their symptoms limited physical activity and mobility. “Mobility” was mentioned in 176 patient posts corresponding to 8% of all HRQoL patient mentions (n=2199). Within this facet, approximately 58% (103/176) of patients considered this HRQoL aspect as of high importance. Primary factors impeding “mobility” (Multimedia Appendix 6) were attributed to the pain and fatigue they experienced. Moreover, patients noted that reduced mobility negatively impacted their overall health, social life, and mental well-being. One patient described “immobility” as analogous to “poor-performing robot” (Multimedia Appendix 7).
Activities of Daily Living
This HRQoL facet explored a person's ability to perform usual daily living activities. This facet was identified as highly important by 50.4% (53/106) of patients. Most patients described problems performing (routine) household activities, personal hygiene, going (outside) for a walk, or even getting out of bed due to pain or lack of energy. Patients’ quotations on their struggles with “activities of daily living” are provided in Multimedia Appendix 7.

Recreation and Leisure
For most patients (129/314, 41.3%), the HRQoL facet “recreation and leisure” was described as being highly important. Owing to the severity of the symptoms, this aspect was frequently described as being hindered, limited, or achievable only with appropriate treatments. Patients reported that joint pain, bleeding, or exhaustion frequently limited their physical activity. Additionally, patients with photosensitivity found it challenging to perform leisure activities such as outdoor sports, watching television, or just relaxation. Patients’ quotations on “recreation and leisure” are provided in Multimedia Appendix 7.

Role Participation
Symptoms, especially when combined with flare-ups, limited the everyday functions of patients with SLE and impacted their role participation in multiple ways. Patients reported a decrease in self-esteem and restricted social relationships. Several patients felt that they could not fulfill their parenting duties and were a burden to their partners. Moreover, patients reported that symptoms of SLE compromised their ability and capacity to work.

Treatment Experience of Patients With SLE
Overview
To evaluate the treatment experiences and unmet needs of patients with SLE, the most frequently used current treatment options (nonpharmacological and pharmacological interventions) mentioned by patients were quantitatively analyzed. The level of satisfaction among patients with SLE with existing treatment options and coping mechanisms is described below.

Current Treatment Options
Exercise, sun protection measures, meditation, and massage were the most commonly reported nonpharmacological interventions (Table 4). Patients found that nonpharmacological interventions, particularly exercise activities such as walking, e-biking, swimming, yoga, or Pilates, and other activities such as meditation and massages to be effective coping mechanisms for dealing with the disease and its symptoms. Additionally, based on 1566 posts, hydroxychloroquine, prednisone, and methotrexate were the most frequently reported pharmacological interventions (Table 5). Negative feedback for pharmacological interventions was mostly related to the drug’s side effects (Multimedia Appendix 8).

### Table 4. Nondrug treatments most frequently reported by patients with SLE\(^a\) (n=482\(^b\) patient mentions and n=267 patients) and CLE\(^c\) (n=24\(^b\) patient mentions from n=16 patients).

<table>
<thead>
<tr>
<th>Nondrug treatments</th>
<th>Mentions in posts, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SLE</td>
</tr>
<tr>
<td><strong>Exercise</strong></td>
<td>222 (46.1)</td>
</tr>
<tr>
<td><strong>Sun protection</strong></td>
<td>82 (17)</td>
</tr>
<tr>
<td><strong>Meditation</strong></td>
<td>43 (8.9)</td>
</tr>
<tr>
<td><strong>Massage</strong></td>
<td>36 (7.5)</td>
</tr>
<tr>
<td><strong>Herbal medicine</strong></td>
<td>32 (6.6)</td>
</tr>
<tr>
<td><strong>Nutrition therapy</strong></td>
<td>30 (6.2)</td>
</tr>
<tr>
<td><strong>Acupuncture</strong></td>
<td>18 (3.7)</td>
</tr>
<tr>
<td><strong>Cognitive behavioral therapy</strong></td>
<td>11 (2.3)</td>
</tr>
<tr>
<td><strong>Vitamin D supplements</strong></td>
<td>8 (1.7)</td>
</tr>
</tbody>
</table>

\(^a\) SLE: systemic lupus erythematosus.

\(^b\) Indicates that patients were counted more than once if they discussed multiple nondrug treatments in a post.

\(^c\) CLE: cutaneous lupus erythematosus.

\(^d\) Not available.
Table 5. Drug treatments most frequently reported by patients with SLEa (n=1566b posts) and CLEc (n=82b posts).

<table>
<thead>
<tr>
<th>Drug treatments</th>
<th>Mentions in posts, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SLE</td>
</tr>
<tr>
<td>Hydroxychloroquine</td>
<td>538 (34.4)</td>
</tr>
<tr>
<td>Prednisone</td>
<td>322 (20.6)</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>303 (19.3)</td>
</tr>
<tr>
<td>Azathioprine</td>
<td>107 (6.8)</td>
</tr>
<tr>
<td>Prednisolone</td>
<td>100 (6.4)</td>
</tr>
<tr>
<td>Aspirin</td>
<td>49 (3.1)</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>44 (2.8)</td>
</tr>
<tr>
<td>Warfarin</td>
<td>42 (2.7)</td>
</tr>
<tr>
<td>Paracetamol</td>
<td>35 (2.2)</td>
</tr>
<tr>
<td>Quinacrine</td>
<td>26 (1.7)</td>
</tr>
<tr>
<td>Cyclosporine</td>
<td>—</td>
</tr>
<tr>
<td>Tacrolimus</td>
<td>—</td>
</tr>
<tr>
<td>Naproxen</td>
<td>—</td>
</tr>
</tbody>
</table>

aSLE: systemic lupus erythematosus.
bIndicates that patients were counted more than once they discussed more than one drug treatment in a post.
cCLE: cutaneous lupus erythematosus.
dNot available.

Treatment Options and HRQoL Aspects

Many relevant aspects of HRQoL, such as “recreation and leisure,” “energy and motivation,” “activities of daily living,” and “mobility,” were improved by nonpharmacological interventions, particularly exercise or meditation. Despite the undesirable side effects, patients frequently used medications, such as hydroxychloroquine, azathioprine, prednisolone, or mycophenolate mofetil, which improved SLE symptoms, although to a lesser extent than that achieved using nonpharmacological interventions. Regarding treatment effects on different aspects of HRQoL (n=5336 posts), 473 (8.9%) reported worsening, while 286 (5.4%) reported improvement of an HRQoL aspect. The most burdensome symptoms, such as pain, fatigue, and rashes, were poorly managed by available drug treatment options for patients with SLE.

Level of Satisfaction With Current Treatment Options

Both positive (65/286 posts) and negative (56/473 posts) feedback were documented when analyzing the level of satisfaction. The reasons for a low level of satisfaction included the lack of treatment options, failure of previously successful treatments due to side effects, or no treatment at all. A high level of satisfaction was primarily attributed to the management of symptoms, initiation of treatments following a successful diagnosis, knowledge level of the doctors treating the patients, and availability of options in case of treatment failure. The most effective pharmacological interventions included prednisone and hydroxychloroquine. However, side effects, such as allergic reactions, insomnia, and increased appetite, caused some patients to discontinue treatment. Discontinuation of treatment and patient satisfaction with drug treatments were frequently attributed to insurance coverage (such as expensive medications). Patient quotations on the levels of satisfaction with current treatment options are presented in Multimedia Appendix 8.

Diagnosis and Access to Care

Health care availability was frequently mentioned in 285 posts authored by 205 patients with SLE. Health care availability was considered as the most important aspect of HRQoL by 45.8% (n=94) of patients with SLE. Some patients with SLE (n=10) reported having received a misdiagnosis or a correct diagnosis only after several years of experiencing symptoms. Several patients with SLE experienced delays in diagnosis and did not receive proper treatment until they found a doctor with enough expertise to diagnose or investigate their symptoms or until they were referred to a specialist. Patients with SLE often managed and advocated for themselves by switching doctors and actively seeking specialists.

Coping Mechanisms

Despite their symptom-related restrictions, patients with SLE reported attempting to stay active and perform daily tasks, even if they could manage only a few of them. Reducing the number of their daily activities helped patients prevent negative side effects and flare-ups. Patients shared coping strategies such as taking medications for side effects and establishing healthy routines involving adequate sleep, proper nutrition, exercise, and hydration. A total of 28 patients (n=40 posts) reported that physical activities or exercise positively affected their physical and psychological well-being. Exercise, meditation, Tai Chi, and yoga were all reported to reduce the burden of symptoms. Some patients found that following an anti-inflammatory diet...
helped reduce skin and joint pain, thereby allowing them to engage in more physical activities. Patient quotations on coping mechanisms are presented in Multimedia Appendix 8.

**Disease Burden in Patients With CLE**

**Overview**

The investigation of disease burden in patients with CLE was limited owing to the small sample size (106 patients with CLE and 243 patient posts). Disease burden could be quantified only for the most prevalent symptoms, areas of involvement (body parts), and sleep disturbances.

**Patient-Reported Symptoms**

Based on an analysis of 244 posts from 118 patients, the most frequently (>5% [6 patients]) reported symptoms were rash (n=27, 23%), pain (n=25, 21%), and fatigue (n=19, 16%) alopecia (n=11, 9%), arthralgia (n=10, 8%), and anxiety and scar (each n=7, 6%; Table 2). Sleep disorders were reported by a total of 8 patients in 10 posts, with insomnia (n=2, 25%) being the most common side effect of lupus medications. Notable quotations from patients describing their symptoms are presented in Multimedia Appendix 4.

**Areas of Involvement**

Based on an analysis of 169 posts from 148 patients, the skin (n=38, 26%), face (n=20, 14%), legs (n=15, 10%), and hairs (n=15, 10%) were the most frequently reported areas of involvement affected by symptoms (Multimedia Appendix 5).

**Overall HRQoL of Patients With CLE**

HRQoL facets were discussed by 47 patients with CLE in 78 of 243 (32.1%) posts. The distribution of HRQoL facets of this population (n=78 posts) was similar to that of patients with SLE, with “negative feelings” (n=20, 26%) and “recreation and leisure” (n=14, 18%) being the most frequently reported dimensions of HRQoL based on the total number of posts (Table 3). Few patients mentioned that their condition negatively affected their ability to work. Although “positive feelings,” “health care access,” and “mobility” were frequently mentioned, differences in ranking could not be identified because of the small sample size (n<10) of individual HRQoL facets. Quotations of patients with CLE regarding their HRQoL are available in Multimedia Appendix 7.

**Treatment Experience of Patients With CLE**

**Overview**

The treatment experiences and unmet needs of patients with CLE were quantitatively analyzed by assessing the most frequently used treatment options (pharmacological and nonpharmacological interventions) as mentioned by the patients and exploring patient experiences related to diagnostic delay, health care availability, and coping mechanisms.

**Nonpharmacological Interventions**

From 24 posts, sun protection (n=12, 50%) and exercise (n=7, 29%; Table 4) were the most frequently reported nondrug treatments. Patients found dietary changes, massages, and exercise to be helpful in managing their symptoms and flare-ups.

**Drug Agents**

The most frequently reported medications (n=119 posts) were hydroxychloroquine (n=42, 35%), methotrexate (n=29, 24%), and prednisolone (n=17, 14%; Table 5). Negative statements regarding treatment experiences were mostly related to side effects of the medications (Multimedia Appendix 8). Methotrexate was associated with side effects, such as fatigue, brain fog, and lack of sleep, which led to patients having to take time off work. Positive statements regarding pharmacological interventions were infrequent (Multimedia Appendix 8).

**Diagnosis and Access to Care**

Patients with CLE reported difficulties in receiving the correct diagnosis because health care providers did not give sufficient consideration to their skin lesion complaints. They reported being confused or misled by the recommendations of dermatologists and rheumatologists, and they received little support with an understanding of their diagnosis and the symptoms to expect from their health care professionals (HCPs). Patients reported feeling frustrated because dermatologists referred them to rheumatologists who prescribed different medications which only made their symptoms worse (Multimedia Appendix 8). Despite showing clear symptoms and positive diagnostic markers, HCPs often advised patients that they “do not have lupus” or had gone “dormant,” highlighting a concerning trend of misdiagnosis and lack of belief by HCPs when it comes to lupus and related conditions.

**Discussion**

**Principal Findings**

This retrospective study investigated the experiences of patients living with SLE or CLE, who publicly discussed their experiences in web-based social media platforms. Patients who were active on these platforms frequently shared and discussed their HRQoL, their perceptions of the disease burden, and treatment experiences affecting their daily lives. Pain and fatigue were identified as symptoms that most impaired HRQoL, negatively impacting physical and psychological well-being. The most frequently used pharmacological interventions included hydroxychloroquine, prednisone, and methotrexate, while nonpharmacological interventions included exercise, meditation, and proper nutrition (diet). Nonpharmacological interventions were more frequently associated with improved HRQoL.

Several studies have explored the disease burden and treatment experiences among patients diagnosed with SLE and CLE [25-31]. Qualitative studies focusing on outcomes reported by patients with SLE suggest that HRQoL is significantly impaired in these patients [26-29]. A qualitative literature analysis of 58 studies relating to the burden of SLE found that in all of these studies, the factors that appeared most frequently affecting the HRQoL in patients with SLE (based on the number of citations) were advanced age, fatigue, the coexistence of neurological or psychiatric conditions (especially depression or anxiety), limited educational achievement, and financial challenges such as poverty or low household income [26]. This is mostly consistent with the findings of this study, wherein pain and fatigue were...
identified as key factors negatively impacting HRQoL in patients with SLE. The severity of these symptoms affected patients’ mobility, cognitive abilities, recreation and leisure activities, and sleep and rest. Patients with SLE often found it challenging to engage in work or recreational activities such as sports, biking, or running. Moreover, these impairments had a negative impact on psychological aspects such as low self-esteem, social relationships (such as feeling like a burden to others), and role functions (such as parental responsibilities and the capacity to work).

Lupus symptoms often result in sleep disturbances, which may in turn aggravate fatigue and exhaustion. Sleep disorders in patients with SLE are often linked to disease activity, pain, and fatigue and are influenced by psychosocial and psychological factors [30]. This is consistent with the findings of this study, wherein patients with SLE or CLE reported symptoms such as pain, fatigue, and anxiety, which primarily impacted the “sleep and rest” HRQoL facet. In this study, the correlation between sleep and lupus symptoms was found to be bidirectional, that is, disrupted sleep was associated with increased fatigue and may also worsen symptoms such as pain and discomfort leading to poor sleep quality, thereby, negatively impacting the overall HRQoL. Additionally, some patients regularly experienced insomnia-related difficulties, which even lasted for years in some cases.

In this study, patients with SLE discussed worsening of HRQoL twice as much in the context of pharmacological interventions, suggesting that the most burdensome symptoms, such as pain, fatigue, and rash, were still inadequately controlled under the current treatment options. Notably, patients receiving pharmacological interventions reported positive experiences with HCPs, who considered their symptoms seriously and had adequate knowledge of lupus for providing effective treatments. They appreciated receiving treatments that helped manage their symptoms or the availability of other options if treatments failed, particularly following a successful diagnosis. In contrast, a low satisfaction level with current treatment was attributed to the lack of treatment options available, previously successful treatments that later failed, or not receiving any treatment at all. This finding highlights that effective patient-physician communication and improving patients’ knowledge about disease and treatment through patient education strategies are important for improving patients’ adherence to therapy in SLE [31].

Medications that are effective in treating lupus are frequently associated with undesirable side effects, thereby leading to patient dissatisfaction [31]. This can also be emotionally challenging for patients as it can negatively impact their HRQoL. To overcome such difficulties, patients may seek alternative nonpharmacological treatment options and associated coping strategies [32-34]. Notwithstanding the symptom-related limitations, the patients with SLE and CLE included in this study reported that they attempted to remain active and complete daily tasks despite being able to manage only a few tasks.

As observed in patients with SLE, patients with CLE experienced the same impacts on their HRQoL: “negative feelings,” “recreation and leisure,” and “health care availability.” The HRQoL of these patients with CLE was also impacted by the visibility of lesions regardless of the levels of disease activity. Skin lesions or scars that were constantly visible, alopecia, photosensitivity, and the chronic nature of the disease significantly impaired HRQoL. Patients with CLE felt that they were neglected and did not receive the necessary attention in health care settings. They highlighted that they experienced more difficulties than patients with SLE owing to misdiagnosis, delayed diagnosis, and contradictory treatment approaches between dermatologists and rheumatologists.

This study demonstrated that social listening could be useful for the generation and analysis of large amounts of data from web-based platforms, such as social media, and patient forums. The approach offers an opportunity to obtain insights from a large patient population by capturing their real-time conversations and contextual understanding of their perspectives [17]. Thus, social listening complemented with traditional methods, such as interviews, surveys, focus groups, and advisory boards or panels, can facilitate a deeper understanding of patients’ perspectives and their unmet needs. Additionally, social listening is fast, eliminates the need for patient recruitment, reduces recall bias through instant platforms, does not burden patients, and offers anonymity in reporting socially embarrassing symptoms, thereby reducing reporting bias [17,18]. We departed from traditional methods and used natural language processing and artificial intelligence in this study to apply and enhance both qualitative and quantitative methods. This facilitated more rigorous analyses of large amounts of unsolicited patient-reported narratives from multiple web-based patient forums to better understand patients’ perspectives and disease burden and identify unmet medical needs.

In the context of drug development, identifying and integrating unmet patient needs and experiences in the decision-making process and evidence-generation strategies are important to ensure that patients’ perspectives and needs are considered and truly represented across the entire continuum of the process [16]. Social media offers access to difficult-to-reach populations or help to concentrate on groups with specific conditions or disorders [35,36]. By actively listening to patients’ descriptions of their challenges on social media platforms, pharmaceutical organizations could gain insight into their daily struggles and identify the most relevant and impactful factors [17]. The increasing presence of patients on social media represents an opportunity for pharmaceutical organizations to identify relevant knowledge for different stages of drug development, with a focus on patients’ HRQoL and beyond.

Limitations

This study had several limitations. First, this study relied on publicly available social media data, raising concerns regarding the accuracy of self-diagnosis or self-reported patient experience. Patients who opted not to publicly share their profiles (and were therefore excluded from the study) may have different opinions, thereby resulting in potential bias. Second, the data set may contain duplicate author profiles, wherein similar authors could have been active on multiple social media platforms. Third, this study only included English-speaking countries and younger populations, which could be a source of...
bias. Fourth, this study did not examine differences across subgroups (patient populations with SLE and CLE) owing to the inability to identify race or ethnicity or other sociodemographic variables. Fifth, the analysis of patients with CLE was limited by a small sample size, which is why several outcomes reported for SLE could not be reported for CLE; therefore, the results regarding CLE should be interpreted cautiously. The authors’ opinion is that similar type of analyses should be conducted in the future for other novel pharmacological agents to gather valuable insights into the experiences of patients while living with the disease.

Conclusions

This social listening study sheds light on the experiences of patients living with SLE and CLE active on social media, providing valuable insight into their experiences, which have not been extensively investigated. SLE and CLE affect all aspects of patients’ lives owing to their wide-ranging manifestations. This study showed varying severity and frequency of symptoms that were reported as burdensome by patients. Moreover, several aspects of HRQoL and factors contributing to HRQoL improvement or worsening were discussed by patients. The results of this study suggest that current treatment options provide insufficient relief, thereby warranting the development of more effective treatments with good tolerability and safety to address the heavy disease burden and unmet needs of patients with SLE and CLE. Our findings can serve as a valuable resource to inform and shape activities and decisions in drug development to meet patients’ needs and improve their HRQoL.

Acknowledgments

Medical writing assistance was provided by Vimal Kumar Muthyala and Pritorthi Bhattacharjee of Merck Specialties Pvt Ltd, an affiliate of the Healthcare Business of Merck KGaA, Darmstadt, Germany. This work was supported by the Healthcare Business of Merck KGaA, Darmstadt, Germany (CrossRef Funder ID 10.13039/100009945).

Authors’ Contributions

ES, TA, and MH contributed to the study concept and design, data acquisition, data analysis, and data interpretation. PK and JP contributed to the study concept and design, data analysis, and data interpretation.

Conflicts of Interest

ES was affiliated with EMD Serono, Billerica, MA, United States at the time of the study and paper preparation and is currently affiliated with Modus Outcomes. TA and MH are employees of Semalytix GmbH. MH is a shareholder of Semalytix GmbH. JP is an employee of EMD Serono, Billerica, MA, United States. PK is an employee of the Healthcare Business of Merck KGaA, Darmstadt, Germany.

Multimedia Appendix 1

List of relevant web-based social media sources retrieved from third-party social media data provider “SocialGist”.
[DOC File, 118 KB - formative_v8i1e52768_app1.doc ]

Multimedia Appendix 2

Glossary of terms.
[DOC File, 110 KB - formative_v8i1e52768_app2.doc ]

Multimedia Appendix 3

Personal data protection measures.
[DOC File, 115 KB - formative_v8i1e52768_app3.doc ]

Multimedia Appendix 4

Quotations of patients with systemic lupus erythematosus or cutaneous lupus erythematosus describing their symptoms.
[DOC File, 123 KB - formative_v8i1e52768_app4.doc ]

Multimedia Appendix 5

Most frequently mentioned body parts by patients with systemic lupus erythematosus or cutaneous lupus erythematosus.
[DOCX File, 93 KB - formative_v8i1e52768_app5.docx ]

Multimedia Appendix 6

Heatmap showing the impairment (negative impact) of a symptom on an aspect of health-related quality of life in patients with systemic lupus erythematosus (n=159 patients from 716 posts).
[PNG File, 43 KB - formative_v8i1e52768_app6.png ]
Multimedia Appendix 7
Quotations of patients with systemic lupus erythematosus or cutaneous lupus erythematosus on their health-related quality of life.
[DOC File, 121 KB - formative_v8i1e52768_app7.doc]

Multimedia Appendix 8
Quotations of patients with systemic lupus erythematosus and cutaneous lupus erythematosus on their drug treatment experiences.
[DOC File, 127 KB - formative_v8i1e52768_app8.doc]

References
16. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Patient-focused drug development: methods to identify what is important to patients: guidance for industry, food and drug administration staff, and other stakeholders. U.S. Food and Drug Administration. 2022 Feb. URL: https://www.fda.gov/media/131230/download [accessed 2023-11-06]


Abbreviations

- CLE: cutaneous lupus erythematosus
- HCP: health care professional
- HRQoL: health-related quality of life
- SLE: systemic lupus erythematosus
Please cite as:
Spies E, Andreu T, Hartung M, Park J, Kamudoni P
Exploring the Perspectives of Patients Living With Lupus: Retrospective Social Listening Study
JMIR Form Res 2024;8:e52768
URL: https://formative.jmir.org/2024/1/e52768
doi: 10.2196/52768
PMID: 38306157

© Erica Spies, Thomas Andreu, Matthias Hartung, Josephine Park, Paul Kamudoni. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Patient Experiences and Insights on Chronic Ocular Pain: Social Media Listening Study

Brigitte Sloesen1, PharmD; Paul O’Brien2*, Grad Dip; Himanshu Verma3, BDentS, MBA, PGDBM; Sathyaraj Asaithambi3*, MSc; Nikita Parashar3*, MPharm; Raj Kumar Mothe3*, MPharm; Javed Shaikh3*, MS; Annie Syntosi4*, MSc

1Novartis Pharma NV, Vilvoorde, Belgium
2Novartis Ireland Limited, Dublin, Ireland
3Novartis Healthcare Pvt Ltd, Hyderabad, India
4Novartis Pharma AG, Basel, Switzerland
*these authors contributed equally

Corresponding Author:
Brigitte Sloesen, PharmD
Novartis Pharma NV
Medialaan 40
B-1800
Vilvoorde
Belgium
Phone: 32 478881453
Email: brigitte.sloesen@novartis.com

Abstract

Background: Ocular pain has multifactorial etiologies that affect activities of daily life, psychological well-being, and health-related quality of life (QoL). Chronic ocular surface pain (COSP) is a persistent eye pain symptom lasting for a period longer than 3 months.

Objective: The objective of this social media listening study was to better understand COSP and related symptoms and identify its perceived causes, comorbidities, and impact on QoL from social media posts.

Methods: A search from February 2020 to February 2021 was performed on social media platforms (Twitter, Facebook, blogs, and forums) for English-language content posted on the web. Social media platforms that did not provide public access to information or posts were excluded. Social media posts from Australia, Canada, the United Kingdom, and the United States were retrieved using the Social Studio platform—a web-based aggregator tool.

Results: Of the 25,590 posts identified initially, 464 posts about COSP were considered relevant; the majority of conversations (98.3%, n=456) were posted by adults (aged >18 years). Work status was mentioned in 52 conversations. Patients’ or caregivers’ discussions across social media platforms were centered around the symptoms (61.9%, n=287) and causes (58%, n=269) of ocular pain. Patients mentioned having symptoms associated with COSP, including headache or head pressure, dry or gritty eyes, light sensitivity, etc. Patients posted that their COSP impacts day-to-day activities such as reading, driving, sleeping, and their social, mental, and functional well-being.

Conclusions: Insights from this study reported patients’ experiences, concerns, and the adverse impact on overall QoL. COSP imposes a significant burden on patients, which spans multiple aspects of daily life.

(JMIR Form Res 2024;8:e47245) doi:10.2196/47245

KEYWORDS
chronic ocular surface pain, patients’ experiences; quality of life; social media; Twitter; unmet needs; ocular pain; ophthalmology; ocular; listening; experience; experiences; tweet; eye pain; eye condition; social media platforms; social media use; patient experience; chronic pain; pain; internet; eye; retina; online health; digital health; web; vision; optical
**Introduction**

Chronic ocular surface pain (COSP) is defined as moderate to severe corneal-induced chronic pain lasting more than 3 months. COSP distracts from, or interferes with, regular daily activities, which may result in poor quality of life (QoL) and health-related QoL, thus impacting the individuals’ ability to carry out daily activities and their psychological well-being [1]. Common pain-related symptoms include ocular irritation, burning, dryness, and other pain descriptors such as corneal sensitivity, shooting, stabbing, grittiness, sharp pain, a hot burning sensation, light sensitivity, and the sensation of a foreign body on the ocular surface. The intensity of ocular pain ranges from simple discomfort to intense unbearable pain; however, evidence elaborating a classification system for COSP is limited [2,3]. COSP can be considered an umbrella term due to the wide spectrum of clinical manifestations, which can result from multiple factors [4].

Patient-reported outcomes (PROs) and patients’ experiences are most often analyzed via conventional methods by using cross-sectional surveys (digital, face-to-face interviews, questionnaires, etc) in different research settings (at hospital sites, home, on the web, through patient-supported organizations, etc). In countries with high internet penetration, access to electronic devices and social media platforms has significantly influenced the health care landscape [5,6]. Social media platforms enable patients to share their perceptions about diseases, treatment patterns, satisfaction with outcomes, and other factors affecting their lives. Therefore, these platforms act as a source to obtain disease-related information, identify and access health care resources, network with fellow patients, and communicate problems and experiences on the web [5-7].

Social media listening (SML) is a new approach to gather information from social media platforms and can be useful in generating insights from users’ experiences. SML has been used to monitor and analyze discussions on health-related topics in diverse diseases [8-13]. Social media facilitates efficient capture of patients’ understanding of a disease and their coping strategies. In addition, SML provides a clear understanding of the patients’ perspectives of a disease, their barriers to health behavior change, and disease-related symptoms [14]. To date, there is no published literature on the use of SML to investigate the needs and experiences of patients with COSP. This study explores SML as a research tool to provide insights on disease burden, diagnosis, treatment patterns, and QoL in patients with COSP.

The objectives of this study are to (1) understand the burden of COSP; its symptoms; its impact on daily life, activities, and QoL, including social well-being and mental health; its management; and the hurdles, gaps, and needs from a patient’s perspective; and (2) capture information directly from patients or caregivers by using their own words and in their familiar environment regarding when and how they feel at that moment, attempting to understand related emotions as well.

The results of this study could help define COSP-related gaps and needs from patients’ perspectives and set priorities for drug development to address them. Our results would also help to develop adequate tools (PRO questionnaires) to quantify the impact of COSP on patients and to develop new PRO end points in clinical studies assessing the potential improvement of new treatments.

**Methods**

**Study Design and Data Source**

A comprehensive search from February 2020 to February 2021 was performed on social media platforms (Twitter, Facebook, blogs, and forums) for English-language content posted on the web. Social media platforms that did not provide public access to information or posts were excluded. The following predefined search terms were used to retrieve the relevant web content: “chronic ocular pain” OR “chronic eye pain” OR “ocular surface pain” OR “ocular pain” OR “eye pain” OR “eye aching” OR “eye ache” OR “eye surface pain” OR “keratoconus pain” OR “conjunctivitis pain” OR “keratitis pain” OR “blepharitis pain” OR “Iritis pain” OR “corneal neuropathy pain.” Social media posts from Australia, Canada, the United Kingdom, and the United States were retrieved using the Social Studio platform [15]—a web-based aggregator tool. In the first step (wave 1), the study was limited to English-speaking countries; non–English-speaking countries were assessed later (in wave 2) and will be the subject of a separate publication.

**Selection of Posts and Text Data Cleaning**

We used the Social Studio tool to download the links of posts with the date stamp and geographic location of users from web-based social media platforms on the basis of the abovementioned keywords. The tool assigned a unique article ID for each downloadable link. These links were used to retrieve the content and remove duplicates (based on unique article IDs and content snippets), irrelevant comments, and out-of-scope content. Irrelevant posts including job postings, aggregator, or junk websites (ie, websites leading to unsolicited advertisements or unrelated misleading links), nonfunctional links, promotional content, and pharmaceutical or e-retailer market reports were removed. Spelling correction was applied while evaluating the content of the post. Additional secondary research including hand searching for posts was performed for patients’ or caregivers’ conversations on forums to increase the robustness of insights and volume of data. The term “mention” indicates the number of times a symptom, treatment, diagnostic test, or another parameter is mentioned in each post. The number of mentions were independent of the number of posts, and due to data anonymization, these users could not be tracked across different platforms. The terms “conversation,” “post,” and “discussion” were synonymously used.

**Categorization and Indexing of Social Media Posts**

The downloaded links were further indexed using the WordNet Lemmatizer natural language processing (NLP) model to arrive at a sample of possible patient posts using patient lexicons and disease-related keywords. A relevancy check was carried out through a two-step process: (1) NLP was used to identify the relevancy and (2) a manual evaluation of relevant posts identified using NLP was carried out. Gender, age, work status, and other demographic information about the users was recorded.
when available or categorized as unknown if the information was not available at the time. Final analysis was conducted manually to generate patient or caregiver insights. A 4-level hierarchy of the decision-making process was followed to ensure the creation of a robust model of consensus-based analysis and tagging of posts. A self-review, followed by a peer review by analysts, a review by senior analysts, and finally a team review were performed during the study (Figure 1).

**Figure 1.** Decision-making process for including and tagging of posts.

![Decision-making process](image)

The key themes identified were patient demographics, perceived causes of COSP, symptoms, disease management (medical consultations, diagnosis, and treatments), and QoL aspects (impact on physical, functional, emotional and social well-being, and work productivity). Posts were further analyzed for the “number of mentions” of a particular theme in the posts. Posts with a mention of COSP-related symptoms (itching, pain, irritation, etc) were used to provide a qualitative description of the patients’ self-reported experience with the disease. The impact of COSP on patients’ QoL was analyzed using posts describing work productivity loss or inefficiency and problems in performing activities of daily living. One post can have multiple mentions (themes) across the patient journey.

**Ethical Considerations**

This paper does not contain any studies with identifiable participants. All web-based content was anonymized and was in accordance with the HIPAA (Health Insurance Portability and Accountability Act) search strategy and data collection. Approval was obtained from the Novartis safety registry 1P1R—the governing body that holds oversight on the use of social media by Novartis (1P1R ID DE005899). All relevant local and global laws affecting and relating to the use of social media were aligned with and, as reflected in Novartis processes, followed in the conduct of this study. The authors of this manuscript consent to the publication of the submitted manuscript and declare that no individual patient data requiring consent has been presented.

**Data Analysis**

All data were analyzed using descriptive statistics and are presented as the number of posts, number of mentions, or percentages. All posts were analyzed by (1) social media channel (Twitter, Facebook, forums, and blogs), (2) country (Australia, Canada, the United Kingdom, and the United States), (3) tone (positive, neutral, or negative), and (4) key themes of discussion.

**Results**

**Characteristics of Analyzed Posts**

A total of 25,590 outputs were initially extracted from social media platforms using COSP-related keywords during the study period. In all, 5100 patient posts were identified by NLP, which facilitated an automated process of generating unique patient posts. The 5100 posts were then filtered for relevance to the disease area by analysts by manually identifying the content that explicitly stated information about COSP. A total of 464 posts determined to be written by patients or caregivers were thus identified and included for further analysis. Figure 2 summarizes the filtering of posts.
Forums and Twitter were the key channels for discussions across all geographies, contributing to ~59% and ~39% of the total conversations, respectively. A detailed list of domains from where the posts were obtained and demographic information are provided in Multimedia Appendix 1. Most of the social media posts were from the United Kingdom (n=208), followed by the United States (n=175), Canada (n=65), and Australia (n=16; Figure 2). Of the 130 conversations mentioning the duration of their ocular pain, the majority of patients had experienced ocular pain for several years (53.1%, n=69), followed by 7 to 12 months (18.5%, n=24), 4 to 6 months (14.6%, n=19), and 3 months (13.8%, n=18).

The majority of conversations were from patients (95%, n=441) as compared to 3.9% (n=18) from caregivers. The gender of the patient was mentioned in 118 posts, of which 57.6% (n=68) of posts were from women and 42.4% (n=50) of them were from men. Only 42 posts specified the patients’ ages; 40.5% (n=17) of patients were 21-30 years of age and 26.2% (n=11) of them were >60 years of age. Approximately 214 posts mentioned whether the patients were adults (99.1%, n=212) or children or teenagers (<18 years of age; 0.9%, n=2). Regarding the work status of the patients determined from among 52 posts, the majority of patients were working (51.9%, n=27), 19.2% (n=10) of them were students, and 19.2% (n=10) of them were retired.

**Key Themes of Discussion**

Of the 464 conversations, symptoms (61.9%, n=287) and the causes of COSP (58%, n=269) were the most common themes that emerged, along with the impact on QoL (20%, n=93). Patients or caregivers also discussed their experiences with health care professionals (HCPs; 25%, n=116), diagnosis (8%, n=37), triggers of their COSP (19%, n=88), treatments (18.1%, n=84), and coping strategies and lifestyle modifications (15.9%, n=74).

Symptoms described by patients varied widely and could be categorized as eye pain–related, such as burning and irritation; vision-related, such as blurred vision and light sensitivity; and physical, such as insomnia and headache (Figure 3). The severity of the symptoms was ranked (high and moderate to low) on the basis of key words and phrases used in the posts. Of the 261 conversations, 85.8% (n=224) of posts were rated as high severity and 14.2% (n=37) as moderate to low severity. Some phrases for high severity mentioned by patients included “eyes burn like hell,” “excruciating (eye) pain,” “eye pain is the
absolute worse,” “unbearable pain,” etc. Moderate to low–severity phrases included “mild redness in my eye occasionally,” “discomfort to mild pain,” and “the pain is pretty manageable.” In 154 conversations, patients also cited emotions that were related to their symptoms, including being upset (35.1%, n=54), worried (26%, n=40), angry (23.4%, n=36), and confused (5.8%, n=9; Figure 3).

**Figure 3.** Patient-reported symptoms on social media by the number of mentions in the posts.

There were 282 mentions of the causes of COSP, of which 46.1% (n=130) mentioned underlying ocular medical conditions such as dry eye disease and ocular surgery, and 53.9% (n=152) of them mentioned nonocular conditions contributing to COSP, such as migraine and COVID-19 (Multimedia Appendix 2). Commonly reported environmental triggers of eye pain, such as exposure to bright light or sun and long screen time use (Multimedia Appendix 3), reflect the burden on daily activities.

Overall, 158 patients mentioned the type of HCP they consulted. Of them, 27 consulted an ophthalmologist; 25 consulted an eye, retina, or cornea specialist; 21 consulted a doctor (unspecified specialty); 15 consulted an optometrist; and 14 consulted an optician. Others consulted a neurologist, neuro-ophthalmologist, eye clinic or eye surgeon, otolaryngologists, etc. When the HCP consulted was mentioned (158 patients), the majority of the related comments (n=123) were negative. Many patients expressed that they were not able to get a proper diagnosis and hence expressed negative sentiments in various conversations, citing emotions such as confusion and worry. Pharmaceutical drugs (140/247, 56.7% treatment-related mentions) were cited as the major treatment option for managing COSP symptoms such as topical or oral drugs, followed by alternative remedies (83/247, 33.6%) such as hot compresses, lenses, etc, and lifestyle modifications (24/247, 9.7%) such as sleep, diet changes etc (Multimedia Appendix 4).

**Impact of COSP on Patients’ QoL**

A total of 159 posts by patients, ascertained to describe their QoL, were grouped in 4 domains: physical, emotional, functional, and social impact (Figure 4). COSP symptoms significantly impacted all aspects of patients’ QoL. Of the 159 posts, various aspects reported were difficulty watching TV or using a phone or computer (8.8%, n=14), executive dysfunction (cognitive, behavioral, and emotional difficulties; 5%, n=8), difficulty performing day-to-day activities such as reading (5%, n=8), and difficulty driving (4.4%, n=7). Emotional well-being, including feelings of depression or hopelessness (10.1%, n=16), frustration or anger (6.9%, n=11), fear (6.3%, n=10), and even suicidal thoughts (1.9%, n=3), were also mentioned.

Conversations also mentioned the impact of COSP on functional well-being such as difficulty at the workplace or at the place of study (6.9%, n=11), reduced productivity or having to quit their job (5.7%, n=9), as well as social impacts such as being irritated around people (1.9%, n=3) and having a less active social life (2.5%, n=4). Overall, COSP greatly affected the patients’ functional and psychological well-being, as illustrated through patients’ example quotes in Figure 5.
Patients’ Perspectives on Unmet Needs

Patients’ perceptions of unmet needs were grouped under 4 broad categories: diagnosis, HCP-related, treatment, and price- or access-related. A summary of patients’ perceptions of unmet need is presented in Figure 6. Failure to obtain a diagnosis of the underlying cause of COSP, even after undergoing multiple tests, emerged as the primary concern; this also led to a loss of confidence in HCP consultations in most cases.
Although not assessed in detail, no major differences were observed among the 4 English-speaking countries. Of note, however, the eye pain and related symptoms described by patients most often referred to the eye in general without specification of the ocular surface, so the study assessed chronic ocular pain rather than COSP.

Discussion

Principal Findings

This SML study attempts to use relevant data from social media platforms to understand patients’ problems and concerns related to COSP. This is the first study to use an SML approach to analyze the COSP disease burden and patients’ perspectives on COSP-related symptoms, causes and triggers, treatment options, and QoL aspects. The study provides a useful insight of the significant burden of COSP on patients’ daily activities, such as reading and driving and the negative impact on their emotional and social well-being and their work productivity. The strong wording that patients are using in describing their pain and its impact on their daily life illustrates patients’ suffering as well as their frustration in desperately seeking effective relief for their chronic ocular pain. This highlights the importance and difficulty of interpreting patients’ own wording expressed in spontaneous SML quotes to understand their real suffering. The results also illustrate how the lack of disease awareness to confirm a diagnosis, a dearth of effective treatment options, and the need to increase awareness of COSP among patients and HCPs are the key unmet needs.

Two broad aspects of patient experience that this study identify are negative impact on QoL and patient perception of unmet needs. Almost 1 in 5 COSP patients described the negative impacts of COSP on work performance. Treatment strategies could be improved by understanding the issues related to patients’ QoL from social media posts. A common theme among patients’ conversations was the delay in the diagnosis of COSP, which resulted in multiple clinic visits and suboptimal management of COSP symptoms. In this study, patients reported that delayed diagnosis and a lack of proper treatments are associated with a negative perception of COSP. Additionally, even after being diagnosed with COSP, patients cited a lack of pain relief despite using multiple pain management options.

Social media interactions on the web present an inherent opportunity to identify and analyze unfiltered experiences of COSP directly from the own words of patients and caregivers. SML has been shown to be a valuable approach to gain an understanding of the patients’ experience of diseases and treatments with a particular advantage that it can allow data to be gathered from a very large, representative sample that is geographically dispersed (ie, from a wide range of countries or locations) [17]. SML has been used as a tool to investigate patients’ perceptions for different indications, including cancer screening [9], dry eye disease [18], HIV [11], inflammatory bowel disease [10], multiple sclerosis [8], presbyopia [19], total joint arthroplasty [13], and Zika virus [12]. Insights and summaries derived from such conversations can be a vital resource in enriching treatment outcomes associated with chronic diseases. It is, however, critical to carefully detect and assess different aspects included in the sometimes lengthy web-based conversations, ensuring the correct interpretation of patients’ posts, including the underlying emotional aspects. In our study, the high severity of COSP symptoms and its lasting effect resulted in negative sentiments in most patients. The variety in terms and the often strong wording used to describe their pain symptoms, their feelings, and the impact on their daily lives reflects the importance of SML in collecting first-line patient information that has been spontaneously expressed. Careful attention must be paid to capture the different aspects addressed in multiple conversations and interpret them correctly.
to adequately understand individuals’ actual suffering. Correct categorization and indexing of the SML posts was critical. Multiple analysts, senior analysts, and experts reviewed the posts during the whole process in order to obtain robust data. This highlights the importance and difficulty of interpreting patients’ own wording expressed in spontaneous SML conversations.

NLP has been widely used in SML studies, and the resulting data have been reported for a variety of diseases [10,12]. In this SML review, NLP has been used for its high-throughput capability, which was then paired with manual assessment to provide the necessary focus and specificity of the outputs. Inherent biases that may affect the accuracy (representativeness of the collected sample and linguistic selection in posts), reliability (consistency of reports from individual patients across time points or descriptors), and quality of information (self-selection bias) obtained from any social media platform may be present in this study as well [20]. One of the limitations of this study was that only publicly available information on digital platforms has been accessed and used; all the personal identifiers were anonymized in the report. Further, owing to the unstructured nature of social media data, it was not possible to obtain information on every research question; therefore, we objectively determined whether the available information supported the research question and, accordingly, our findings, which may be considered a limitation of this study.

Conclusions

Insights from this study reported the experiences and concerns and the adverse impact on overall QoL among patients with COSP. Assessment of patients’ own wording demonstrated that COSP imposes a significant burden of impaired QoL on patients, which spans multiple aspects of daily life, including physical and functional impacts, as well as impacts on emotional and social well-being. The lack of disease awareness among HCPs with long diagnostic delays and the inefficacy of prescribed treatments were cited as concerns. The variety in terms patients used on the web to describe their suffering, issues, and needs as well as the different aspects addressed herein contribute to a better understanding of the patients’ perspectives and highlight the value of SML studies. It is indeed most critical to obtain patients’ insights and address their unmet needs when considering disease management including drug development.

Acknowledgments

The study and article processing charges were funded by Novartis Pharma AG (Basel, Switzerland). The authors acknowledge Gurpreet Kaur Virya (Novartis Healthcare Pvt Ltd, India, at the time of study conduct) and Pradeep Basa (Novartis Healthcare Pvt Ltd, India) for editorial and writing support, in accordance with the Good Publication Practice (GPP3) guidelines of the International Society for Medical Publication Professionals.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

BS, PO, HV, SA, and AS conceptualized the study design. NP, HV, and SA contributed to the conduct of study and data collection. NP, HV, and SA analyzed and interpreted the data, along with other authors. RKM drafted the manuscript, along with JS, PO, and BS. All authors reviewed the manuscript critically and provided final approval for submission. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this paper and take responsibility for the integrity of the work as a whole.

Conflicts of Interest

BS, NP, SA, and AS are full-time employees at the Novartis group of companies and declare that they have no competing interests. HV, JS, PO, and RKM were full-time employees at the Novartis group of companies at the time this study was conducted and declare that they have no competing interests.

Multimedia Appendix 1
Demographic and clinical characteristics as represented by patient posts.

[DOCX File, 15 KB - formative_v8i1e47245_app1.docx]

Multimedia Appendix 2
Patients’ and caregivers’ example quotes regarding ocular causes of chronic ocular surface pain.

[PDF File (Adobe PDF File), 556 KB - formative_v8i1e47245_app2.pdf]

Multimedia Appendix 3
Patients’ and caregivers’ example quotes regarding nonocular causes of chronic ocular surface pain.

[PDF File (Adobe PDF File), 591 KB - formative_v8i1e47245_app3.pdf]
Multimedia Appendix 4

Patients' and caregivers' discussion on social media regarding treatment options for chronic ocular surface pain.

[PDF File (Adobe PDF File), 602 KB - formative_v8i1e47245_app4.pdf ]

References


Abbreviations

COSP: chronic ocular surface pain

https://formative.jmir.org/2024/1/e47245 JMIR Form Res 2024 | vol. 8 | e47245 | p.3434 (page number not for citation purposes)
Identification of Myths and Misinformation About Treatment for Opioid Use Disorder on Social Media: Infodemiology Study

Mai ElSherief¹, PhD; Steven Sumner², MD; Vikram Krishnasamy², MD; Christopher Jones², DrPH; Royal Law², PhD; Akadia Kacha-Ochana², MPH; Lyna Schieber², DPhil; Munmun De Choudhury³, PhD

¹Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States
²Centers for Disease Control and Prevention, Atlanta, GA, United States
³Georgia Institute of Technology, Atlanta, GA, United States

Corresponding Author:
Mai ElSherief, PhD
Khoury College of Computer Sciences
Northeastern University
360 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 (617) 373 2462
Email: m.elsherif@northeastern.edu

Abstract

Background: Health misinformation and myths about treatment for opioid use disorder (OUD) are present on social media and contribute to challenges in preventing drug overdose deaths. However, no systematic, quantitative methodology exists to identify what types of misinformation are being shared and discussed.

Objective: We developed a multistage analytic pipeline to assess social media posts from Twitter (subsequently rebranded as X), YouTube, Reddit, and Drugs-Forum for the presence of health misinformation about treatment for OUD.

Methods: Our approach first used document embeddings to identify potential new statements of misinformation from known myths. These statements were grouped into themes using hierarchical agglomerative clustering, and public health experts then reviewed the results for misinformation.

Results: We collected a total of 19,953,599 posts discussing opioid-related content across the aforementioned platforms. Our multistage analytic pipeline identified 7 main clusters or discussion themes. Among a high-yield data set of posts (n=303) for further public health expert review, these included discussion about potential treatments for OUD (90/303, 29.8%), the nature of addiction (68/303, 22.5%), pharmacologic properties of substances (52/303, 16.9%), injection drug use (36/303, 11.9%), pain and opioids (28/303, 9.3%), physical dependence of medications (22/303, 7.2%), and tramadol use (7/303, 2.3%). A public health expert review of the content within each cluster identified the presence of misinformation and myths beyond those used as seed myths to initialize the algorithm.

Conclusions: Identifying and addressing misinformation through appropriate communication strategies could be an increasingly important component of preventing overdose deaths. To further this goal, we developed and tested an approach to aid in the identification of myths and misinformation about OUD from large-scale social media content.

(JMIR Form Res 2024;8:e44726) doi:10.2196/44726

KEYWORDS
addiction treatment; machine learning; misinformation; natural language processing; opioid use disorder; social media; substance use

Introduction

In the United States, more than 100,000 drug overdose deaths occurred in 2021 [1]. Beyond lives lost, the economic costs of both fatal opioid overdose and opioid use disorder (OUD) are estimated to be greater than 1 trillion US dollars per year [2,3]. Furthermore, the extent of OUD is significant; in 2020, about 2.7 million people in the United States aged 12 years or older met the diagnostic criteria for an OUD in the past year [4].

https://formative.jmir.org/2024/1/e44726
The American Psychiatric Association’s *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) defines OUD as a “problematic pattern of opioid use leading to clinically significant impairment or distress” [5]. OUD is characterized by 11 defining criteria, including taking opioids in larger amounts or over a longer period of time than intended [5]. Medications for opioid use disorder (MOUD), including methadone, buprenorphine, and extended-release naltrexone, are effective treatments for OUD [6]. MOUD increase treatment retention and reduce opioid use and overdose mortality, among other public health benefits [7]. Yet in 2020, it is estimated that only 11.2% of people with OUD received treatment with MOUD in the past year, based on data from the National Survey on Drug Use and Health [8]. Despite its effectiveness, MOUD are underused, in part due to stigma, financial constraints, treatment availability, and a lack of perceived treatment need [4].

When individuals do seek information on MOUD—often on the internet or through social media—they may access inaccurate and potentially harmful health misinformation [9]. A recent review pointed to several areas of misinformation regarding MOUD, including that MOUD may be detrimental to health and can be perceived as simply substituting one addiction with another [10]. Although social media provides a venue for individuals to seek help, advice, and support surrounding their OUD experiences, journeys, and recovery goals [11-13], these platforms also provide a mechanism for misinformation about MOUD to spread [14]. However, most studies to date have focused on general opinions and attitudes regarding MOUD, and less is known about specific misinformation that is shared. For example, Tofighi et al [15] conducted a qualitative analysis of 1010 Twitter (subsequently rebranded as X) posts related to MOUD, assessing general experiences and perceptions. A subsequent study by Chenworth et al [16] assessed tweets that mentioned methadone or suboxone and found that a large percentage expressed negative sentiment about MOUD. Pertaining specifically to misinformation, a study [14] quantified the prevalence of a single myth about MOUD across multiple social media and web-based communication platforms. This myth was drawn from clinical literature; however, it is likely that other myths or misinformation related to MOUD are emerging or being discussed that have not been previously described. Indeed, health care professionals’ understanding of what misinformation and myths may be circulating related to substance use disorder treatment is currently limited to expert opinion, and there is no systematic or large-scale quantitative approach to identify new opioid-related myths from web-based communications and social media platforms. Thus, in this study, we developed and evaluated an approach for identifying potentially novel myths that may exist regarding MOUD. This approach can help identify harmful content to inform strategies to educate clinicians and the public about MOUD and counter myths and misinformation related to MOUD.

**Methods**

**Overview**

To accomplish our goal of identifying potentially new myths about MOUD from social media, we used a multistep analytic pipeline (Figure 1), described in detail below. The steps included curating a data set of social media posts across multiple platforms; extracting posts with a high probability of including a myth; using a clustering algorithm to group these posts into themes; and lastly, examining the resulting content for language indicative of misinformation and myths.

**Figure 1.** Pipeline for the identification of new myths pertaining to treatment for opioid use disorder. We start by transforming each post and myth into a mathematical representation. Next, we identify the posts closest to myths in meaning and have public health experts annotate them for accuracy. We then computationally cluster similar themes together and have public health experts assign a clinical theme. KNN: k-nearest neighbor.

**Data Set Curation**

We first developed a lexicon of opioid-related keywords; for this, we adopted a 2-pronged approach that combined insights from the substance use literature and feedback from the substance use expert coauthors of this study. Our lexicon encompassed different types of opioids, such as natural opiates, semisynthetic opioids, and synthetic opioids, and included opioids that were prescription or illicit. For each generic drug name, we also included brand and combination product names. In addition, we also included street names of substances, where useful, from the Drug Enforcement Administration. Our final lexicon of 152 keywords was then used in the ensuing data collection.

Using this lexicon, we constructed a diverse data set from Twitter, YouTube, and online health communities (OHCs) such as Reddit and Drugs-Forum. For all the platforms we
investigated, we focused on public posts and messages created between January 1, 2018, and December 31, 2019. Our data set collection methodology for Twitter included querying for all public posts that contained 1 of the words in our lexicon using the then-available Twitter Academic application programming interface (API). This process yielded 6,365,245 posts. For YouTube, due to rate limitations imposed in the data collection process by the platform’s API, we restricted the keywords to 11 MOUD treatment keywords such as buprenorphine and naltrexone. We used the YouTube API to identify 552 public YouTube videos that contain 1 of the 11 keywords in the title and then collected all of the associated comments (99,386 comments). We relied on expert domain knowledge to identify subforums pertinent to OUD for Reddit and Drugs-Forum. We used data from 22 opioid-specific subreddits (carfentanil, opiates, fentanyl, opiatesmemorial, modquittingkratom, methadone, suboxone, kratom, heroin, quittingkratom, Tianeptine, loperamide, naltrexone, oxycodone, OpiatesRecovery, opiatewithdrawal, lean, heroinaddiction, HeroinHeroesine, OpiateChurch, suboxone, OurOverUsedVeins), resulting in 1,189,590 posts and 12,293,829 comments. Additionally, we collected 5549 messages posted under the various “Opiates and Opioids” subforums on Drugs-Forum. Throughout this paper, we combine Reddit and Drugs-Forum content under the category of OHCs because of their similar affordances.

**Mixed Methods Approach to Identify Social Media Posts Relevant to MOUD Myths**

We began our analytic and data processing efforts by investigating three “seed” myths drawn from the substance use literature [17-20]: (1) agonist therapy or medication-assisted treatment for OUD replaces one drug with another, (2) only patients treated with opioids who have certain characteristics are at higher risk for opioid addiction, and (3) tramadol is a nonaddictive nonopioid alternative. These seed myths were used, in concert with machine learning approaches, to filter the large volume of semantically rich social media content that would be subsequently investigated for the presence or absence of a new MOUD myth.

Specifically, we used InferSent, a sentence embedding method that provides semantic sentence representations [21], to construct document embeddings for the 3 myth statements noted above. Document embeddings are long sequences of numbers that mathematically represent the semantic meaning of each document (ie, a social media post). InferSent embeddings have been shown to outperform unsupervised methods such as SkipThought vectors on a range of natural language processing (NLP) tasks [21]. In our experiments, we evaluated the embedding values of the social media posts most similar to the seed myths and observed that, indeed, the most similar posts express a similar meaning but are expressed differently (eg, the following blockquote from a sample post in our data set was found to be similar to the seed myth “agonist therapy or medication-assisted treatment for OUD replaces one drug with another”).

I'm not getting any support from them being on MMT. They don't see it any different than me doing heroin every day....

After constructing document embeddings for the seed myth text, we constructed document embeddings for all social media posts in our data set. Using the mathematical representation of each post, we were able to identify posts similar to the seed myths and containing additional information useful for understanding MOUD myths. Specifically, we identified the 200 most semantically similar posts per platform for each seed myth using the k-nearest neighbor (KNN) algorithm [22], implemented in Python’s scikit-learn library [23]. The KNN algorithm assumes that similar things exist in close proximity. In other words, KNN uses the idea that similar things are near each other—in our case, it would compare how close posts are to the seed myths. This process provided us with candidate social media posts that were likely to be discussing MOUD myths or related topics. Additional details of our machine learning and NLP approaches are provided in Multimedia Appendix 1 [21,24-34].

Because not all posts identified through the methods described above may be directly discussing a MOUD myth and document embeddings could pick up some noise, we performed a second data processing step, in which we harnessed annotations from public health experts. The experts reviewed and evaluated whether each post in our now-filtered data set of 800 messages (200 per platform) was relevant to 1 of the seed myths, discussing a new myth, or neither. Specifically, a total of 3 public health experts (coauthors of this paper) reviewed each of the 800 social media posts to perform this qualitative assessment and reach a consensus on the topic discussed therein. Public health experts included 2 clinicians and 1 doctoral-level epidemiologist. The experts reviewed all posts and collectively came to a consensus across all posts in our data set. Our rationale for the qualitative annotation approach followed the guidelines given in the seminal research of McDonald et al [35]. The experts leveraged thematic coding, an iterative process that involves multiple coders developing, discussing, and refining codes through continual discussion. Our qualitative annotation, followed by consensus-building discussion, is situated in grounded theory [36,37].

This manual review resulted in a total of 303 posts identified as discussing potentially new myths. The pipeline for identifying these posts is outlined in Figure 1.

**Methods to Understand Discussions of New Myths Arising From the Seeds**

We developed additional machine learning–based techniques to better characterize the content of the 303 annotated posts. We leveraged an unsupervised machine learning technique known as hierarchical clustering [24]. This approach provides a probabilistic mechanism to group items (social media posts) into categories (discussion themes). Multimedia Appendix 1 describes this unsupervised technique in greater detail.

Hierarchical clustering was used to construct 10 categories from the 303 posts that are potentially indicative of new myths. All posts per theme were then presented to the above public health experts.
experts to interpret and name the extracted themes. To give richer context to the experts and help with the generation of theme descriptors, we also provided the linguistic markers (n=1-grams or single words) present in the posts for each theme using a commonly used lexical analytic generative model known as Sparse Additive Generative Models (SAGE) of Text [25]. SAGE identifies distinguishing words in our topic themes, where the SAGE magnitude of a word signals the degree of its uniqueness. Previous work has shown that SAGE outperforms other topic models, such as latent Dirichlet allocation (LDA) models, by focusing on high-frequency terms with accurate counts, thus leading to learning more robust interpretable topics [25]. Using this information, the experts developed descriptions of each category and identified myths related to OUD. After careful inspection, the experts aggregated a few topically similar themes, with a final focus on seven themes.

Ethical Considerations
This study is considered exempt research since there are no human participants involved. As such, the study proceeded without obtaining informed consent from social media users. Moreover, social media users who authored posts in the data set were not compensated because the data were publicly available. Social media posts are anonymized by not including usernames in our analysis. All examples given in this study are slightly paraphrased from different social media platforms to further protect user confidentiality. The potential use of these findings by malicious actors cannot be overstated. Motivated actors perpetrating myths and misinformation surrounding OUD could use machine learning and NLP approaches to target susceptible people with OUD and redirect them toward clinically unverified treatments, leading to misinformation exacerbation. Additionally, individuals with OUD are often stigmatized on multiple levels. Individuals with OUD are perceived as dangerous, of moral failure, and called “addicts” [38]. In light of these possible negative outcomes, although we provided links to all open-source libraries associated with our computational analyses, we have not shared the text data from different social media platforms to minimize the possible identification of OUD social media users.

Results
Our analytics pipeline resulted in the identification of 7 clusters or discussion themes of social media posts related to MOUD. Textbox 1 shows paraphrased example posts that represent new myths or potentially harmful information identified from this data set, and Table 1 displays the salient keywords identified by SAGE for each category to help provide further context.

Textbox 1. Example posts from our data set (n=303) that represent inaccurate or harmful information identified by our human-machine mixed strategies. Themes (in bold font) are labeled by public health experts. The bullet points depict example posts from our data set for each identified type of misinformation. Posts are slightly paraphrased to prevent traceability and author identification.

<table>
<thead>
<tr>
<th>Discussing addiction or addictiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Calling addiction a disease cheapens what I mean.</td>
</tr>
<tr>
<td>• Addiction is just made up in your mind.</td>
</tr>
<tr>
<td>• Buprenorphine is more addictive than opioids.</td>
</tr>
<tr>
<td>• Calling it a disease is just an excuse.</td>
</tr>
<tr>
<td>• Fentanyl is less addictive than marijuana.</td>
</tr>
<tr>
<td>• Withdrawal from buprenorphine is worse than heroin.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taking medication for addiction is not true recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>• You have to be strong. Willpower is the only real way to quit.</td>
</tr>
<tr>
<td>• Who cares if people are addicted to kratom, it’s just like coffee,</td>
</tr>
<tr>
<td>• it’s good for you.</td>
</tr>
<tr>
<td>• Ibogaine cures addiction big time.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alternative or nonrecommended treatment for addiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The only people overdosing are those taking street heroin.</td>
</tr>
<tr>
<td>• People with chronic pain are physically dependent on opioids to function like a normal person.</td>
</tr>
</tbody>
</table>
The most prevalent category described and discussed treatments for opioid use disorder (90/303, 29.8%). Posts described several Food and Drug Administration (FDA)–approved medications for OUD, including buprenorphine and methadone, and nonapproved treatments such as kratom. Additional posts reference a regulated supply of heroin to treat OUD, while other posts discuss the need for long-term use of buprenorphine or methadone. New myths present in this category perpetuated the notion that addiction is not a medical disease and discussed alternative therapies such as kratom.

The second category explored the nature of addiction (68/303, 22.5%). Posts discussed the definition of addiction and the distinction between dependence and addiction and debated whether addiction is a disease. Similar to the first category, misinformation found in this category included statements that addiction is not a medical disease and the promotion of alternative therapies, such as “Ibogaine cures addiction big time.”

The third category described the pharmacologic properties, effects, and addictiveness of substances (52/303, 16.9%). Posts included comments on how addictive substances were compared with each other (eg, nicotine vs heroin or fentanyl vs marijuana). Additional posts discussed drug metabolism and drug synthesis. Misinformation in this category included statements that “Buprenorphine is more addictive than opioids” and that “Fentanyl is less addictive than marijuana.”

The fourth category focused on intravenous or injection drug use (36/303, 11.9%). Posts covered drug injection techniques, advice, and the health consequences of such injections. Misinformation in this category included statements that “withdrawal from buprenorphine is worse than heroin” and that “willpower is the only way to quit.”

The fifth category centered on pain, opioid use, and addiction (28/303, 9.3%). Posts covered the use of opioids for chronic pain, the risk of addiction when using opioids for pain, and the stigma associated with using opioids for pain. Misinformation in this category included diverse statements such as “the only people overdosing are those taking illicit heroin,” “people with chronic pain are physically dependent on opioids to function like a normal person,” and “Who cares if people are addicted to kratom, it’s just like coffee, it’s good for you.”

The sixth category described physical dependence on medications (22/303, 7.2%). Posts commented on the distinction between dependence and addiction and specifically noted kratom as resulting in less addiction. Misinformation in this category included statements promoting alternate therapies.

Finally, the seventh category largely described tramadol use (7/303, 2.3%). Posts commented on tramadol dosing and administration. A discussion of the alternative therapy, ibogaine, was also present.

Each of the 7 categories included posts that contained some form of new misinformation that was not present in the initial seed myths we used to build our detection approach. Textbox 1 presents examples of misinformation or myth text. Further review of the misinformation resulted in classifying the misinformation into three themes, as shown in Textbox 1: (1) discussing addiction or addictiveness, (2) taking medication for addiction is not true recovery, and (3) alternative or nonrecommended treatments for addiction.

### Discussion

#### Overview

The aim of this study was to develop and test a methodology for identifying new myths and misinformation related to MOUD. While health professionals are cognizant of the presence of misinformation and its effects on patient populations [39,40], there is currently no systematic approach for identifying potentially harmful information that individuals who use substances are actually exposed to and discussing. We developed a semiautomated pipeline that uses known myths as seed text and a sophisticated NLP approach to identify other misinformation that is circulating. The approach used a recent

<table>
<thead>
<tr>
<th>Topic or myth</th>
<th>Frequency, n (%)</th>
<th>SAGE keywords (score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments for opioid use disorder</td>
<td>90 (29.8)</td>
<td>therapy (1.36), mat (1.33), primary (1.31), assisted (1.28), treatment (1.24), care (1.22), replacement (1.18), buprenorphine (1.17), medication (1.13), and methadone (1.06)</td>
</tr>
<tr>
<td>Exploring the nature of addiction</td>
<td>68 (22.5)</td>
<td>physical (1.93), difference (1.85), dependence (1.83), addiction (1.77), between say (1.68), disease (1.53), not (1.47), no (1.41), and treated (1.4)</td>
</tr>
<tr>
<td>Pain, opioid use, and addiction</td>
<td>28 (9.3)</td>
<td>ppl (2.31), please (2.31), severe (2.31), majority (2.31), treating (2.31), function patients (2.27), chronic (2.27), people (2.23), and addicted (2.21)</td>
</tr>
<tr>
<td>Use of tramadol</td>
<td>7 (2.3)</td>
<td>study (3.47), extended (3.47), 200 (3.47), 30 (3.47), studies (3.29), mg tramadol (2.82), ibogaine (2.74), safe (2.54), and hard (2.54)</td>
</tr>
<tr>
<td>Physical dependence on medications</td>
<td>22 (7.2)</td>
<td>they’re (2.8), physically (2.57), trading (2.57), you’re (2.48), dependent (2.26), less (2.26), become (2.23), side (2.02), addicted (2.02), and being (1.9)</td>
</tr>
<tr>
<td>Intravenous or injection drug use</td>
<td>36 (11.9)</td>
<td>hit (1.61), nerve (1.61), hands (1.61), veins (1.52), feet (1.46), shoot (1.46), inject (1.46), artery (1.46), foot (1.46), and tying (1.46)</td>
</tr>
<tr>
<td>Pharmacologic properties, effects, and addictiveness of substances</td>
<td>52 (16.9)</td>
<td>sub (1.74), acetaminophen (1.6), targin (1.60), addictive (1.57), codeine (1.37), sometimes (1.36), drug (1.34), seizure (1.32), medical (1.32), and release (1.32)</td>
</tr>
</tbody>
</table>
algorithm released by Facebook Research for measuring text similarity in social media postings [21] and then explored results through automated clustering and human expert review. We found that this approach identified new myths and forms of misinformation beyond those used as seed myths, suggesting that this approach may be useful in identifying new and emerging forms of potentially harmful information that may be circulating.

The posts we identified were grouped into 7 main themes related to MOUD, each revealing inaccuracies upon review by public health experts. The most common misinformation included statements endorsing alternative therapies such as kratom or ibogaine or discouraging medication use, favoring less effective abstinence-only approaches [41]. These are critical areas that require ongoing public health attention since alternative therapies such as kratom have been linked to fatal overdoses, and nonpharmacologic therapy for OUD is associated with higher rates of drug resumption and mortality. Our approach, using automatic data mining techniques, successfully flagged these types of misinformation, even uncovering subtler concerns that have not received sufficient attention, such as misinformed statements about the addictiveness of buprenorphine and unwarranted fears about buprenorphine withdrawal. These findings are significant as buprenorphine, a partial-opioid agonist, stands as one of the most effective treatments for OUD [7]. In light of the rising impact of health misinformation on patient populations, this study addresses the lack of tools for identifying potentially harmful information that spreads. We introduce an approach that quantitatively taps into extensive discussions about addiction across major communication platforms, contributing to misinformation identification. This is pivotal because, although patients seek health information from social media, it is also a breeding ground for false information dissemination [42], and there is a lack of systematic tools to assess health information related to addiction shared on these platforms [43]. Most closely related to this study is the work done by Sarker et al [13], Garett and Young [44], and Johnson et al [45]. Garett and Young [44] conducted a review on how inaccurate and false beliefs by both patients and providers can lead to stigma and serve as barriers to receiving appropriate treatment. This study comments on the consequences of 4 types of stigma, including structural stigma, public stigma, self-stigma, and stigma associated with treatment medications [44]. Similar to this study, Johnson et al [45] conducted a content analysis of 33 YouTube videos to identify and understand the lived experiences of parents and families impacted by the opioid crisis. In contrast to their work, we focused on social media posts and comments to identify new myths surrounding OUD. Most closely related to this study is Sarker et al [13], wherein the authors similarly leveraged NLP methods and built a classifier that identified whether a post’s language promoted 1 of the leading myths challenging addiction treatment: that the use of agonist therapy for MOUD is simply replacing one drug with another [13]. However, this study differs in the sense that it does not focus on 1 particularly known myth but rather on finding new pieces of misinformation.

To our knowledge, no studies report interventions directed at web-based misinformation on MOUD, yet lessons from analogous interventions during the COVID-19 pandemic are informative. One study of US-based Facebook users showed decreased distance traveled among those who viewed video messages from health professionals during the 2020 holiday season [46]. Another study found that journalistic fact-checks may be effective against COVID-19 misinformation [47]. Finally, a third study demonstrated the utility of accuracy nudges in addressing COVID-19 misinformation [48]. Taken together, these interventions suggest initial steps in building the evidence base for infodemic response across public health areas. Further study is needed to understand interventions that address MOUD-related misinformation.

The greatest strength of this study is that it is nonobtrusive and leverages a large-scale data set along with advances in machine learning and NLP to identify new pieces of misinformation that could exacerbate OUD health-related risks. Nevertheless, this study is subject to limitations. First, the qualitative assessment of myths and misinformation was conducted by only a limited number of health experts. As myths can be nuanced, further work should aim to codify definitions and guidance around opioid-related myths, particularly as this field of study grows. Second, while we demonstrate success at identifying misinformation beyond the seed myths we used to initialize the algorithm, the nature and scope of all misinformation related to MOUD are not known, and thus we are unable to assess the sensitivity of our approach to capturing all misinformation that may exist. Nevertheless, the ability to quantitatively and algorithmically identify misinformation related to addiction is still an important advancement. Third, although the approach we use harnesses machine learning and NLP techniques, a component of our pipeline still relies on public health expert review. While this necessitates some labor, human-in-the-loop designs are a leading framework for product development, have certain advantages, and can be particularly useful for complex areas such as health misinformation where expert judgment is required [14]. Finally, we acknowledge that there are a multitude of social media–based communication modalities, and this study is limited to those that are publicly available. It is possible that the nature of health misinformation may differ by platform, and further study of these differences is needed.

Our research demonstrates promise in identifying myths and misinformation related to treatment for OUD included in social media posts. With rapidly rising opioid overdose fatality rates, the initiation and adoption of MOUD are increasingly urgent to prevent additional loss of life. Attention by health and public health professionals to the health misinformation that may be affecting individual decisions related to OUD treatment engagement and retention can be a critical element in enhancing prevention.

Conclusions

Health misinformation regarding treatment for OUD is prevalent, contributing to the challenges of preventing opioid-related overdoses. However, a systematic and quantitative methodology to identify this misinformation is lacking. In response, we developed a multistage analytic pipeline to analyze social media posts from platforms such as Twitter, YouTube, Reddit, and Drugs-Forum for OUD-related misinformation. Our
methodology successfully identified 7 main clusters of misinformation related to MOUD. The most salient topics for these myths include treatments for OUD, the nature of addiction, as well as the pharmacologic properties, effects, and addictiveness of substances. Further understanding of the identified myths and continual monitoring of emerging myths are critical in the battle against the opioid overdose epidemic.

Acknowledgments
This research was supported through a contract by the Centers for Disease Control and Prevention (CDC; through the Department of Health and Human Services) with the Georgia Institute of Technology (Principal Investigator MDC). CDC investigators participated in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; the preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication. ME and MDC had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. We thank Kenneth Kannampully for his help with the Drugs-Forum data collection.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
SS, MDC, and ME designed the research. ME, SS, and MDC conceptualized and developed the analytic techniques. ME gathered and analyzed the data and designed the machine learning models. SS, VK, CJ, RL, AKO, and LS provided expert clinical review and annotations for the models and read, edited, and provided feedback on the paper. ME, SS, and MDC interpreted the results and drafted the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Machine learning and clustering approaches.
[DOCX File , 17 KB - formative_v81e44726_app1.docx ]

References


44. Garett R, Young SD. The role of misinformation and stigma in opioid use disorder treatment uptake. Subst Use Misuse 2022;57(8):1332-1336 [FREE Full text] [doi: 10.1080/10826084.2022.2079133] [Medline: 35611913]

Abbreviations

API: application programming interface
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
FDA: Food and Drug Administration
KNN: k-nearest neighbor
LDA: latent Dirichlet allocation
MOUD: medications for opioid use disorder
NLP: natural language processing
OHC: online health community
OUD: opioid use disorder
SAGE: Sparse Additive Generative Models
Arabic Web-Based Information on Oral Lichen Planus: Content Analysis

Azzam AlMeshrafi1, BDS; Arwa F AlHamad2,3,4, BDS, MSc, PhD; Hamoud AlKuraidees5, BDS; Lubna A AlNasser3,6, BDS, MPH, PhD

1Department of Periodontics, King Saud Medical City, Riyadh, Saudi Arabia
2Dental Services, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia
3King Saud bin Abdulaziz University for Health Sciences, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia
4King Abdullah International Medical Research Center, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia
5Department of Dental Medicine, Maple Clinic, Dammam, Saudi Arabia
6Department of Population Health, King Abdullah International Medical Research Center, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

Corresponding Author:
Arwa F AlHamad, BDS, MSc, PhD
Dental Services
Ministry of National Guard Health Affairs
Prince Mutib bin Abdullah bin Abdulaziz Rd
Riyadh, 11426
Saudi Arabia
Phone: 966 118011111
Email: hamadar@mngha.med.sa

Abstract

Background: The use of web-based health information (WBHI) is on the rise, serving as a valuable tool for educating the public about health concerns and enhancing treatment adherence. Consequently, evaluating the availability and quality of context-specific WBHI is crucial to tackle disparities in health literacy and advance population health outcomes.

Objective: This study aims to explore and assess the quality of the WBHI available and accessible to the public on oral lichen planus (OLP) in Arabic.

Methods: The Arabic translation of the term OLP and its derivatives were searched in three general search platforms, and each platform’s first few hundred results were reviewed for inclusion. We excluded content related to cutaneous LP, content not readily accessible to the public (eg, requiring subscription fees or directed to health care providers), and content not created by health care providers or organizations (ie, community forums, blogs, and social media). We assessed the quality of the Arabic WBHI with three standardized and validated tools: DISCERN, Journal of the American Medical Association (JAMA) benchmarks, and Health On the Net (HON).

Results: Of the 911 resources of WBHI reviewed for eligibility, 49 were included in this study. Most WBHI resources were provided by commercial affiliations (n=28, 57.1%), with the remainder from academic or not-for-profit affiliations. WBHI were often presented with visual aids (ie, images; n=33, 67.4%). DISCERN scores were highest for WBHI resources that explicitly stated their aim, while the lowest scores were for providing the effect of OLP (or OLP treatment) on the quality of life. One-quarter of the resources (n=11, 22.4%) met all 4 JAMA benchmarks, indicating the high quality of the WBHI, while the remainder of the WBHI failed to meet one or more of the JAMA benchmarks. HON scores showed that one-third of WBHI sources had scores above 75%, indicating higher reliability and credibility of the WBHI source, while one-fifth of the sources scored below 50%. Only 1 in 7 WBHI resources scored simultaneously high on all three quality instruments. Generally, WBHI from academic affiliations had higher quality scores than content provided by commercial affiliations.

Conclusions: There are considerable variations in the quality of WBHI on OLP in Arabic. Most WBHI resources were deemed to be of moderate quality at best. Providers of WBHI could benefit from increasing collaboration between commercial and academic institutions in creating WBHI and integrating guidance from international quality assessment tools to improve the quality and, hopefully, the utility of these valuable WBHI resources.
INTRODUCTION

The internet has revolutionized the visibility and accessibility of data [1]. Over half of today’s world population used the internet in 2021 [2]. Searching the internet for web-based health information (WBHI) was the third most frequent internet activity, and 6.75 million searches are performed daily for health information [3,4]. Often, WBHI is considered a “first aid” resource for health information and is used to compare diagnoses or treatment options [5,6], or to supplement insufficient time with a health care provider [6,7]. More seriously, some patients believe that WBHI is trustworthy and might defer or replace medical consultation or treatment from health care providers [6,8,9]. Lastly, recent evidence suggests that seeking WBHI could be associated with changes in health behavior or patient outcomes [6,8,9]. Therefore, ensuring the availability, accessibility, and quality of WBHI is essential to the well-being of individuals and the community.

The format and quality of WBHI vary substantially. The format ranges from health blogs/forums based on personal experiences offering unregulated information to peer-reviewed journal articles that provide complex data addressed to medical professionals [10,11]. Therefore, the quality of WBHI could vary considerably between the resources of WBHI. Some global research indicates that the quality of some WBHI targeting the public could be of low quality [10,11]. Evidence that examined Arabic WBHI often found that a considerable proportion of the WBHI had low quality, including but not limited to not disclosing the authorship of the information, outdated information, and lack of advice that WBHI should not replace a health care consultation [12-16]. Often, high-quality WBHI would require payment (ie, subscription) for access, exacerbating the inequalities in accessing WBHI for some individuals [10]. The evaluation of context-specific WBHI (ie, language) emerges as a public health priority that might address some health literacy inequality.

Oral lichen planus (OLP) is a prevalent chronic inflammatory mucocutaneous disease that frequently affects the oral mucosa. Between 0.5%-2% of the world’s population is affected by OLP [17-19]. In a review of studies conducted in Arab countries, OLP had a potential malignant transformation rate ranging from 0.4% to 6.5% [20]. Although there is no gold standard measure of the quality of WBHI, some international tools like the Journal of the American Medical Association (JAMA) benchmarks and Health On the Net (HON) are frequently used to examine the WBHI in different languages. Literature that evaluated the English content of WBHI regarding OLP using the aforementioned tools reported moderate accuracy and reliability [11]. A comprehensive evaluation of 122 Arabic health websites revealed that these websites varied substantially in meeting some industry benchmarks, like the HON code. For instance, 16% of the websites provided information on their advertising policies, while 73% provided justification for the content included within a website [21,22].

Though accessing high-quality WBHI can effectively increase the public’s knowledge, support health-related decision-making, and improve health-seeking behaviors or outcomes [8,23], no studies focused on evaluating OLP-related WBHI in Arabic. As such, the need arises to scrutinize the content and accuracy of Arabic WBHI related to OLP that is accessible to the public. This study assessed the availability and quality of OLP-related resources in Arabic.

METHODS

Search Strategy

This study was a cross-sectional evaluation of Arabic WBHI on OLP. We searched for the Arabic translation of the keywords “oral lichen planus,” “treatment of oral lichen planus,” and their derivatives ( , , ) in three main search engines, namely, Google, Yahoo, and Bing. We used the Boolean operator (OR) to link these terms but did not use any conditions of filters to mimic electronic research that patients or members of the public might perform on OLP.

We reviewed the first few hundred links on each platform for inclusion until links were no longer relevant to OLP. We included resources (ie, web pages) with information on OLP in Arabic. We excluded WBHI focusing on extra-oral lichen planus, scientific content requiring membership (eg, subscriptions) or directed to professionals (ie, specialty journals), community-based forums without professional guidance, social media posts, and results promoted or advertised by search engines. We also excluded resources that included video or audio content only with no accompanying text and results that were in .doc, .pdf, or .ppt format, as the public might be less likely to seek such resources. Two trained evaluators (AA and HA) simultaneously screened and evaluated the OLP Arabic WBHI resources and deferred to the senior author (AFA) in case of disagreement.

Domains and Tools for Evaluating WBHI

The WBHI resources were assessed for content and quality. The content of the resources was categorized as reported previously by Ni Riordan and McCready [24] and attached in Multimedia Appendix 1 [24]. Briefly, results were grouped according to the affiliation of the web page (commercial, nonprofit, governmental, or university/medical center), specialization (if a web page is either entirely or partially related to the searched topic), content type (medical facts, clinical trials, human interest stories, or question and answer), and content presentation (text, images/graphs, videos, and audio). The quality of the web pages was assessed with the DISCERN
The DISCERN tool aims to empower consumers when making treatment choices by evaluating the quality of written health information (ie, publications). DISCERN is a validated 16-item instrument, with each item rated on a 5-point scale (1=did not fulfill item to 5=complete fulfillment of item) and is arranged in three sections. The first section (items 1-8) addresses the reliability of the publication, the second section (items 9-15) assesses the quality of the information related to treatment choices, and the third section consists of one item (item 16) that gives an overall quality rating of the content. 

JAMA benchmarks assess the accountability of WBHI by examining the following domains of resources: authorship (providing authors and their affiliations), attribution (eg, citations) of the source of information provided, disclosure (revealing any ownership, sponsorship, or conflict of interest in providing the WBHI), and currency (ie, updates of the WBHI).

The HON tool evaluates the reliability and credibility of health websites. HON asks reviewers to evaluate a given web page regarding eight sections: authority, complementarity, confidentiality, attribution, justifiability, transparency, financial disclosure, and advertising policy. The reviewer will score each item from 0% (did not fulfill the specific item) to 100% (total fulfillment of that item), while the HON final score is an average of scores across the eight items.

Statistical Analysis

Data were extracted, coded, and cleaned in Excel (Microsoft Corporation), then analyzed using SPSS Statistics (Version 24.0; IBM Corp). The data were summarized using frequency and percentage distribution for the categorical variables and means with SDs for the continuous variables. We classified WBHI by type of affiliation into commercial and academic (including nonprofit, medical center, and governmental), then compared the quality of WBHI across these groups using bivariate analyses. To facilitate comparison with previous literature, the HON score was categorized into an ordinal variable as such: scores <50% (low credibility), 51%-75% (moderate credibility), and scores >75% (high credibility). Fleiss κ was used to calculate the interrater reliability of the reviewers in assigning the HON score as a categorical variable. Lastly, to synthesize information from all three quality measures, we created a composite variable as follows: high-quality WBHI means the resource satisfied at least 3 of the JAMA benchmarks, had a HON score ≥75, and had a DISCERN overall score ≥3; otherwise, the WBHI was considered low-quality.

Ethical Considerations

This study was exempt from ethical review since it used publicly available data sources that do not include any patient-identifying data.

Results

Overview

The keywords returned the highest number of hits in Google (n=56,770), followed by Yahoo (n=5188) and Bing (n=1910). We reviewed the first 400 results on Google, 300 on Yahoo, and 211 on Bing until the web pages were unrelated to OLP. After removing duplicated results, a total of 49 results were included in this study. The most frequent reason for exclusion was results about oral lesions other than OLP (Figure 1). The interrater reliability measure indicated a high agreement between the evaluators in assigning HON scores (κ=0.826, 95% CI 0.823-0.828).
Content Assessment
More than half of the web pages had commercial affiliations (n=28, 57.1%), that is, WBHI provided by for-profit clinics with clear advertisements for treatment or establishments (Table 1). Conversely, 43% (n=21) of the WBHI resources had academic affiliations (nonprofit, medical centers, and governmental affiliations). Only one-fifth of the web pages (n=11, 22%) were focused exclusively on OLP, rather than lichen planus in general. The vast majority of Arabic OLP WBHI was in the format of medical facts (n=42, 85.7%), and most WBHI included images of OLP (n=31, 63%; Table 1).
Table 1. Content assessment of Arabic web-based health information on oral lichen planus (OLP) using Ni Riordan and McCreary’s [24] method (n=49).

<table>
<thead>
<tr>
<th>Categories</th>
<th>Web pages, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content affiliation</strong></td>
<td></td>
</tr>
<tr>
<td>Commercial</td>
<td>28 (57.1)</td>
</tr>
<tr>
<td>Nonprofit</td>
<td>19 (38.8)</td>
</tr>
<tr>
<td>Medical center</td>
<td>2 (4.1)</td>
</tr>
<tr>
<td>Governmental</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td><strong>Content specialization</strong></td>
<td></td>
</tr>
<tr>
<td>Entirely related to OLP</td>
<td>11 (22.3)</td>
</tr>
<tr>
<td>Partially related to OLP</td>
<td>38 (77.6)</td>
</tr>
<tr>
<td><strong>Content format</strong></td>
<td></td>
</tr>
<tr>
<td>Medical facts</td>
<td>42 (85.7)</td>
</tr>
<tr>
<td>Human stories</td>
<td>2 (4.1)</td>
</tr>
<tr>
<td>Question and answer format</td>
<td>4 (8.2)</td>
</tr>
<tr>
<td>Clinical trials</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td><strong>Content presentation</strong></td>
<td></td>
</tr>
<tr>
<td>Included images</td>
<td>31 (63.3)</td>
</tr>
<tr>
<td>Included videos</td>
<td>2 (4.1)</td>
</tr>
<tr>
<td>Included audio</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>Text only</td>
<td>15 (30.6)</td>
</tr>
</tbody>
</table>

**Quality Assessment by DISCERN**

The overall mean score of DISCERN for the included WBHI was rather low (mean 2.61, SD 1); however, the average rating of single items varied substantially (Table 2). Items with the highest DISCERN scores described the publication’s aims and the alternative OLP treatment options. There was no evidence to indicate meaningful differences in the DISCERN scores between WBHI presented by commercial and academic affiliations.
<table>
<thead>
<tr>
<th>Domain and DISCERN item</th>
<th>All (N=49), mean (SD)</th>
<th>Academic (n=21), mean (SD)</th>
<th>Commercial (n=28), mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1. Explicit aims</td>
<td>3.37 (1.0)</td>
<td>3.61 (0.7)</td>
<td>3.19 (1.1)</td>
<td>.07</td>
</tr>
<tr>
<td>Q2. Attainment of aims</td>
<td>3.27 (1.0)</td>
<td>3.17 (1.1)</td>
<td>3.42 (0.7)</td>
<td>.27</td>
</tr>
<tr>
<td>Q3. Relevance</td>
<td>3.18 (1.1)</td>
<td>3.47 (0.92)</td>
<td>2.96 (1.10)</td>
<td>.07</td>
</tr>
<tr>
<td>Q4. Explicit sources</td>
<td>2.20 (1.8)</td>
<td>2.76 (1.94)</td>
<td>1.78 (1.57)</td>
<td>.06</td>
</tr>
<tr>
<td>Q5. Explicit date</td>
<td>1.73 (1.4)</td>
<td>2.05 (1.5)</td>
<td>1.5 (1.17)</td>
<td>.13</td>
</tr>
<tr>
<td>Q6. Balanced and unbiased</td>
<td>2.69 (1.5)</td>
<td>2.57 (1.50)</td>
<td>2.78 (1.47)</td>
<td>.66</td>
</tr>
<tr>
<td>Q7. Additional sources</td>
<td>1.92 (1.2)</td>
<td>1.90 (1.22)</td>
<td>1.92 (1.27)</td>
<td>.96</td>
</tr>
<tr>
<td>Q8. Areas of uncertainty</td>
<td>2.08 (1.3)</td>
<td>3.09 (1.13)</td>
<td>1.32 (0.77)</td>
<td>&lt;.001&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Treatment options</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9. How treatment works</td>
<td>2.57 (1.5)</td>
<td>2.47 (1.28)</td>
<td>2.64 (1.63)</td>
<td>.84</td>
</tr>
<tr>
<td>Q10. Benefits of treatment</td>
<td>2.08 (1.3)</td>
<td>1.95 (0.86)</td>
<td>2.17 (1.56)</td>
<td>.85</td>
</tr>
<tr>
<td>Q11. Risk of treatment</td>
<td>1.92 (1.2)</td>
<td>1.95 (0.86)</td>
<td>1.89 (1.47)</td>
<td>.24</td>
</tr>
<tr>
<td>Q12. Effect of no treatment</td>
<td>1.67 (1.4)</td>
<td>1.42 (1.20)</td>
<td>1.85 (1.48)</td>
<td>.25</td>
</tr>
<tr>
<td>Q13. Effect on quality of life</td>
<td>1.24 (0.7)</td>
<td>1.09 (0.43)</td>
<td>1.35 (0.78)</td>
<td>.18</td>
</tr>
<tr>
<td>Q14. All alternatives described</td>
<td>3.37 (1.9)</td>
<td>3.38 (1.93)</td>
<td>3.35 (1.88)</td>
<td>.96</td>
</tr>
<tr>
<td>Q15. Shared decision</td>
<td>3.20 (1.8)</td>
<td>2.61 (1.49)</td>
<td>3.64 (1.80)</td>
<td>.04</td>
</tr>
<tr>
<td>Q16. Overall rating of the source</td>
<td>2.61 (1.0)</td>
<td>2.61 (1.16)</td>
<td>2.60 (0.91)</td>
<td>.98</td>
</tr>
</tbody>
</table>

<sup>a</sup>Italics indicate a statistically significant result.

**Quality Assessment by JAMA Benchmarks**

Only 22.4% (n=11) of the WBHI resources met all 4 JAMA benchmarks, indicating high quality, and another 24.5% (n=12) met 3 of the 4 benchmarks. Overall, most WBHI fulfilled the “Disclosure” criteria, while less than half of the web pages fulfilled the “Attribution” criteria (Table 3). Academic affiliations satisfied the authorship and attribution more often than commercial affiliations but were less current than commercial resources (Table 3).
Table 3. Overall and stratified distribution of the fulfillment of Journal of the American Medical Association (JAMA) benchmarks and Health On the Net (HON) scores among reviewed web-based health information (WBHI; n=49).

<table>
<thead>
<tr>
<th></th>
<th>Academic affiliation (n=21), n (%)</th>
<th>Commercial affiliation (n=28), n (%)</th>
<th>All (n=49), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>JAMA benchmarks fulfilled</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authorship</td>
<td>13 (61.9)</td>
<td>14 (50.0)</td>
<td>27 (55.1)</td>
<td>.56</td>
</tr>
<tr>
<td>Attribution</td>
<td>11 (52.4)</td>
<td>7 (25.0)</td>
<td>18 (36.7)</td>
<td>.07</td>
</tr>
<tr>
<td>Currency</td>
<td>13 (61.9)</td>
<td>20 (71.4)</td>
<td>33 (67.3)</td>
<td>.55</td>
</tr>
<tr>
<td>Disclosure</td>
<td>17 (80.9)</td>
<td>23 (82.1)</td>
<td>40 (81.6)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Fulfilling 3 benchmarks</td>
<td>6 (28.6)</td>
<td>21 (43.7)</td>
<td>27 (55.1)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Fulfilling all 4 benchmarks</td>
<td>5 (23.8)</td>
<td>6 (21.4)</td>
<td>11 (22.4)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td><strong>HON score (range 0-100)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>2 (10)</td>
<td>8 (28.6)</td>
<td>10 (20.4)</td>
<td>.09</td>
</tr>
<tr>
<td>51-75</td>
<td>15 (53.6)</td>
<td>5 (17.9)</td>
<td>25 (51.1)</td>
<td></td>
</tr>
<tr>
<td>&gt;75</td>
<td>9 (43.9)</td>
<td>5 (23.8)</td>
<td>14 (28.6)</td>
<td></td>
</tr>
<tr>
<td>High-quality WBHI&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5 (23.8)</td>
<td>2 (7.14)</td>
<td>7 (14.29)</td>
<td>.12</td>
</tr>
</tbody>
</table>

<sup>a</sup>JAMA comparisons denote raw percentages comparing the relative contribution by affiliation type within each JAMA benchmark.

<sup>b</sup>HON comparison denotes column contribution by affiliation type across levels of HON scores.

<sup>c</sup>High-quality WBHI denotes a WBHI resource that satisfied at least 3 of the JAMA benchmarks, had a HON score >75, and had an overall DISCERN score ≥3; otherwise, a resource was labeled as low-quality.

**Quality Assessment by HON**

The average HON score was significantly higher for academic affiliation sources of WBHI than commercial affiliations (mean 71, SD 14 vs mean 56, SD 15; P<.001). When HON scores were examined as an ordinal variable, almost one-third (n=14, 29%) of the WBHI resources scored higher than 75, indicating high credibility of the WBHI, with more contributions from academic rather than commercial affiliations (Table 3).

**Overall Quality Assessment**

When WBHI information was classified by collective score combining all three quality instruments, it was found that 1 in every 7 resources (n=7, 14.3%) had high quality, and most of this high-quality evidence came from academic affiliations (n=5, 23.8% vs n=2, 7.1% from commercial affiliations; Table 3). However, the difference in the distribution of quality levels across affiliation types was not significant (Fisher exact test P=.12).

**Discussion**

**Principal Findings**

This study evaluated the content and quality of the Arabic WBHI on OLP available to the public using prevalidated standardized tools. The results indicated that approximately one-third of the WBHI had high scores, corresponding to higher quality, as indicated by the JAMA and HON tools. The Arabic WBHI on OLP was often presented as a narration of medical facts and frequently included audiovisual aids to enhance the consumer's comprehension of health information, and most were provided by commercial affiliations. To the best of our knowledge, this is the first study to evaluate Arabic WBHI on OLP. As OLP can be a chronic condition with occasional atypical forms and a potential for malignant transformation, patients are likely to supplement their clinical consultations with WBHI [7,17,32,33]. Hence, this evaluation of the quality of WBHI on OLP in Arabic is essential and timely.

The Arabic WBHI on OLP in this study included a higher percentage of resources deemed to be of higher quality compared to Arabic WBHI on other diseases, including periodontal diseases and breast and oral cancers [12,14,34]. For instance, in this study, 1 in every 5 WBHI sources had fulfilled all 4 JAMA benchmarks, while recent results by Halboub et al [16] and Al-Ak'hali et al [34] found the same for 8% and 4%, respectively, of the websites they reviewed. The association of higher quality with resources classified as academic was noted in previous studies evaluating Arabic WBHI [34]. Lastly, DISCERN scores in this study were slightly higher than those reported by Alakhali et al [14] on oral cancer. However, these studies evaluated different diseases, and the higher scores could reflect “better” WBHI on OLP compared to oral cancer or an inherent subjectivity in the DISCERN instrument by virtue of having a Likert scale.

Most of the WBHI resources we reviewed had more commercial affiliations than affiliations to medical or research centers. This generally agrees with the literature where commercial affiliations were the most prolific providers of WBHI [12,13,34]. This is consequential as the web pages with a commercial affiliation often include advertising of certain treatments or care providers. Some evidence suggests that advertising negatively affects the credibility of the information offered and individuals’ trust in the source of WBHI [35]. Moreover, compared to academic affiliations, we observed a lower quality among WBHI by commercial affiliations, as evidenced by the HON, DISCERN,
and JAMA benchmarks scores. While these differences in quality scores did not always reach statistical significance (possibly due to the relatively small sample size), they indicate a consistent trend among commercial affiliations across all the quality tools used.

Of the WBHI we reviewed, the authority and attribution were the least fulfilled JAMA domains—that is, resources rarely mentioned the authors and the references of the WBHI. This was in line with previous studies on Arabic WBHI on oral cancer, periodontal diseases, and epilepsy [13,34]. This highlights an area of potential improvement as the authority of the author (ie, the level of expertise of the person writing the information) was found to increase the credibility and trust in the content of the WBHI [35]. Moreover, other evidence from an Arabic-speaking country showed that participants prioritized information provided by health care professionals over other sources [9]. Our findings, along with conclusions from other Arabic WBHI, are useful for developers of WBHI resources to promote transparency when creating web-based content to maximize the benefit of WBHI to the public.

Enhancing the content and quality of Arabic WBHI is pivotal in optimizing its effectiveness and influence on patients’ well-being. For example, research reported that 49% of the participants in a US-based national survey used WBHI as a first resource to address health concerns [36]. Another study from Saudi Arabia reported that the patients who sought diabetes mellitus–related WBHI, compared to nonseekers, were more health conscious and showed a positive trend of better self-care [8]. Three critical issues must be highlighted when we combine evidence from this study with conclusions from similar research on Arabic WBHI. First, the quality of available resources in Arabic is moderate at best, leaving the public at a disadvantage as most of the evidence is deemed low-quality. Second, developers and providers of WBHI of digital health content must invest in the quality of WBHI in Arabic to empower the public [37]. This could be accomplished by benchmarking the finalized content against the quality tools frequently used in the relevant literature to build more robust, impactful WBHI. Third, it is imperative to establish policies and guidelines that guarantee the quality of Arabic WBHI that are produced, disseminated, or promoted for the benefit of the public.

**Strengths and Limitations**

This study reviewed the content of three major search engines, potentially covering the majority of OLP-related WBHI that the public could find, access, and use. Additionally, we used validated and standardized tools to evaluate the quality and content of OLP health information on the web, enhancing the comparability with research elsewhere. However, this study could not assess Arabic readability due to some challenges related to the available tools and our belief that readability in Arabic is context specific and best explored through research that uses community participatory methods (ie, focus groups). Future efforts should strive to include the readability of the Arabic WBHI from both the providers’ and the consumers’ perspectives. Lastly, the WBHI we evaluated included published web pages up to mid-2020, coinciding with the height of the COVID-19 pandemic. The COVID-19 pandemic increased the demand for digital resources in general, which might have positively affected the quality of WBHI in general in response to the increased demand for web-based information. However, given the comparison of our results to other recent evidence on Arabic WBHI post COVID-19 [14,34,38], we assume that the Arabic WBHI content did not experience a change of a magnitude large enough to substantially alter our conclusions regarding the quality of Arabic WBHI related to OLP available to the public.

**Conclusion**

This study evaluated the quality of Arabic WBHI on OLP available to the public, using several validated and standardized tools, namely, JAMA, DISCERN, and HON. The results indicated that the quality of the WBHI is moderate at best, with only 1 in 7 resources scoring simultaneously high on all three resources. Commercial affiliations of the WBHI provided more content than academic affiliations; however, the content delivered by the former was of lower quality. Therefore, the collaboration between commercial and academic WBHI providers could improve the quality of the OLP resources offered to the public. Lastly, providers of WBHI in Arabic could benefit from integrating guidance from international quality tools to enhance the quality and, hopefully, the utility of these valuable WBHI resources.

**Acknowledgments**

The authors wish to thank Ms Lolwah Almousa for her help with the statistical analysis and creating graphs. The authors also thank the anonymous peer reviewers for their valuable feedback that shaped the final version of this manuscript.

The authors received no financial support for this research. No generative artificial intelligence tools were used to write any portion of this manuscript.

**Data Availability**

Data are available from the corresponding author upon request.

**Authors’ Contributions**

AFA conceptualized the study. AA and HA acquired and curated the data. LAA performed the data analysis. AA, HA, and AFA wrote the initial draft. LAA and AFA edited and critically reviewed the draft. All authors have reviewed and agreed to publish the final version of this manuscript.
Conflicts of Interest

None declared.

Multimedia Appendix 1

Riordan and McCreary content categorization. [DOCX File, 16 KB - formative_v8i1e49198_app1.docx]

References


29. AlMeshrafi A, AlHamad AF, AlKuraidees H, AlNasser LA. Arabic Web-Based Information on Oral Lichen Planus: Content Analysis. JMIR Form Res 2024;8:e49198 [FREE Full text] [PMID:33461516]

Abbreviations

HON: Health On the Net

JAMA: Journal of the American Medical Association

OLP: oral lichen planus

WBHI: web-based health information
Social Network Analysis of e-Cigarette–Related Social Media Influencers on Twitter/X: Observational Study

Runtao Zhou1*, BS; Zidian Xie2*, PhD; Qihang Tang1, BS; Dongmei Li2, PhD

1Goergen Institute for Data Science, University of Rochester, Rochester, NY, United States
2Department of Clinical and Translational Research, University of Rochester Medical Center, Rochester, NY, United States
*these authors contributed equally

Corresponding Author:
Dongmei Li, PhD
Department of Clinical and Translational Research
University of Rochester Medical Center
265 Crittenden Boulevard Cu 420708
Rochester, NY, 14642-0708
United States
Phone: 1 5852767285
Email: Dongmei_Li@urmc.rochester.edu

Abstract

Background: An e-cigarette uses a battery to heat a liquid that generates an aerosol for consumers to inhale. e-Cigarette use (vaping) has been associated with respiratory disease, cardiovascular disease, and cognitive functions. Recently, vaping has become increasingly popular, especially among youth and young adults.

Objective: The aim of this study was to understand the social networks of Twitter (now rebranded as X) influencers related to e-cigarettes through social network analysis.

Methods: Through the Twitter streaming application programming interface, we identified 3,617,766 unique Twitter accounts posting e-cigarette–related tweets from May 3, 2021, to June 10, 2022. Among these, we identified 33 e-cigarette influencers. The followers of these influencers were grouped according to whether or not they post about e-cigarettes themselves; specifically, the former group was defined as having posted at least five e-cigarette–related tweets in the past year, whereas the latter group was defined as followers that had not posted any e-cigarette–related tweets in the past 3 years. We randomly sampled 100 user accounts among each group of e-cigarette influencer followers and created corresponding social networks for each e-cigarette influencer. We compared various network measures (eg, clustering coefficient) between the networks of the two follower groups.

Results: Major topics from e-cigarette–related tweets posted by the 33 e-cigarette influencers included advocating against vaping policy (48.0%), vaping as a method to quit smoking (28.0%), and vaping product promotion (24.0%). The follower networks of these 33 influencers showed more connections for those who also post about e-cigarettes than for followers who do not post about e-cigarettes, with significantly higher clustering coefficients for the former group (0.398 vs 0.098; P=0.005). Further, networks of followers who post about e-cigarettes exhibited substantially more incoming and outgoing connections than those of followers who do not post about e-cigarettes, with significantly higher in-degree (0.273 vs 0.084; P=0.02), closeness (0.452 vs 0.137; P=0.04), betweenness (0.036 vs 0.008; P=0.001), and out-of-degree (0.097 vs 0.014; P=0.02) centrality values. The followers who post about e-cigarettes also had a significantly (P<0.001) higher number of followers (n=322) than that of followers who do not post about e-cigarettes (n=201). The number of tweets in the networks of followers who post about e-cigarettes was significantly higher than that in the networks of followers who do not post about e-cigarettes (93 vs 43; P<0.001). Two major topics discussed in the networks of followers who post about e-cigarettes included promoting e-cigarette products or vaping activity (55.7%) and vaping being a help for smoking cessation and harm reduction (44.3%).

Conclusions: Followers of e-cigarette influencers who also post about e-cigarettes have more closely connected networks than those of followers who do not themselves post about e-cigarettes. These findings provide a potentially practical intervention approach for future antivaping campaigns.
KEYWORDS
social network; social media; influencer; electronic cigarettes; e-cigarette; vaping; vape; Twitter; observational study; aerosol; consumer; influencers; social network analysis; antivaping; campaigns

Introduction

e-Cigarettes, also known as “e-cigs,” “vapes,” “e-hookahs,” “vape pens,” and “electronic nicotine delivery systems,” heat up a liquid containing propylene glycol, vegetable glycerin, flavors, and frequently nicotine via a battery to produce an aerosol for users to inhale [1,2]. In 2018, 4.5% of respondents sampled in the United States self-reported being current e-cigarette users [3]. Based on the annual national youth tobacco survey from the US Centers for Disease Control and Prevention (CDC) in 2023, e-cigarette is the most popular tobacco product among youth, with 10.0% of high school students and 4.6% of middle school students reported as e-cigarette users [4]. While the CDC suggests that e-cigarettes might be less harmful than regular cigarettes, they have also warned that e-cigarettes are not harmless [5]. One study showed that e-cigarette usage has health effects on seven specific organ systems along with other negative effects to the human body [6]. Another study showed that during the vaporizing process of the e-cigarette liquid, formaldehyde-containing compounds appeared to be formed, and the US National Cancer Institute suggested that formaldehyde is one of the cancer-causing substances in e-cigarettes [7]. Many studies have shown that e-cigarette use is significantly associated with various health risks, including respiratory diseases (such as asthma, chronic obstructive pulmonary disease, and wheezing), cardiovascular diseases, and cognitive defects [8-13].

Given the popularity (especially among youth) and the adverse health effects of e-cigarettes, it is critical to understand how e-cigarettes have become so popular and, more importantly, how to educate the public, especially youth, about their health risks.

The tobacco industry has recognized the power of social media. Accordingly, the industry actively promotes tobacco products or tobacco product use, and mounts opposition to regulatory policies on social media platforms such as Twitter (rebranded as X in July 2023), Facebook, Instagram, and YouTube by hiring social media influencers [14-19]. Conversely, to prevent and reduce tobacco use, the US Food and Drug Administration (FDA) launched “The Real Cost” campaign in 2014 to disseminate antitobacco messages on social media [20]. Several studies have shown that exposure to e-cigarette advertising and promoting messages (eg, on social media) can increase interest in e-cigarettes and the likelihood of initiating e-cigarette use [21-29]. For example, McCausland et al [30] showed that conversations about e-cigarettes on Twitter in Australia generally encourage e-cigarette use. Huang et al [14] found that Twitter appears to be a popular marketing platform for e-cigarette products, with most e-cigarette–related tweets being commercial tweets. Since e-cigarette companies cannot directly market their vaping products online, they typically pay social media influencers (users who have a large number of social media followers) to promote their products [31]. While there are no official definitions or hard thresholds for a social media influencer, most previous studies have defined influencers as social media users who have a large number of followers, remain relevantly active, and gain financial rewards from their activities on social media platforms [32-34]. On Instagram, e-cigarette influencers often promote different brands of e-cigarette products [32,35]. Considering the large number of followers, the posts from e-cigarette influencers can reach many social media users. Studies have shown that higher exposure to e-cigarette–promoting content was significantly associated with e-cigarette use [27,36,37]. In addition, social media influencers posted some misinformation about e-cigarettes (eg, e-cigarettes are harmless) on YouTube [38]. To the best of our knowledge, there has been no study focused on understanding the nature of e-cigarette influencers on Twitter.

Most studies in this field have focused on the association between social media exposure and e-cigarette perception and initiation, and although some of the data analyzed may have covered e-cigarette influencers, the social networks among e-cigarette–related social media influencers and general users are relatively less studied [14,30]. On social media, topical influencers play a vital role in information diffusion within the social network [39-41]. One study tried to reveal the interaction between social media influencers and e-cigarette brands on Instagram using social network analysis [32]. However, the social networks of e-cigarette influencers on social media remain elusive, even though they are critical for understanding how the information is disseminated.

Clustering analysis can provide valuable insights into the structure, behavior, and dynamics of social interactions among users within social networks, which can facilitate the exploration of commonalities within each cluster and potential effective targeted marketing [42]. As one of the fundamental measures, the clustering coefficient can help us understand the network’s architecture by measuring how well-connected users are [43], which can reflect information sharing.

In this study, we aimed to identify e-cigarette–related influencers on Twitter and measure the social network features (eg, the clustering coefficient) between different types of e-cigarette influencers. Although e-cigarette influencers also post tweets unrelated to e-cigarettes, it is of great interest to compare the clustering measures for followers who have posted e-cigarette–related tweets and those who have not, which can help us understand how followers of e-cigarette–related posts are connected. Therefore, we compared the clustering measures between the networks of these two types of followers of e-cigarette influencers. Our findings will provide a more comprehensive understanding of the differences between different types of e-cigarette influencers, discern the types of prevailing discussions within each influencer’s network, reveal the social network structure of e-cigarette–related influencers on Twitter, and provide a potentially novel intervention strategy for effectively disseminating health-related information for future antivaping campaigns.
Methods

Data Collection
Twitter data related to e-cigarettes were collected through the Twitter streaming application programming interface using e-cigarette–related keywords such as “e-cigarette” and “vaping” [44]. From November 19, 2019, to June 10, 2022, we collected 15,787,239 English tweets related to e-cigarettes. For each tweet, we collected as much information as possible, such as tweet ID, created date, text, retweets, favorites, user ID, user name, user tweets, followers count, friends count, and hashtags. In this data set, we identified 3,617,766 unique Twitter user accounts.

Identifying e-Cigarette Influencers
To identify e-cigarette influencers on Twitter, we defined the following selection criteria based on previous studies [32-34]: (1) social media influencers must have at least 1000 followers, (2) the follower and following ratio for social media influencers must be at least 2:1, (3) social media e-cigarette influencers have posted at least 10 e-cigarette–related posts within the past year, and (4) social media e-cigarette influencers should have posted at least one promotional tweet related to e-cigarettes within the past year.

We iteratively checked the users and tweets in our database against the above criteria and filtered out the Twitter accounts that did not match the criteria. We used keyword matching to determine the promotional tweets related to e-cigarettes [44]. Toward this end, we created a similar list that contained the most common keywords related to promotional tweets and we compared all of the tweets of the candidate influencers against the keywords in this list. If any words in the tweets matched any of the keywords in the list, we categorized the tweets as promotional tweets. At the end of the process, we collected 33 accounts that completely satisfied the above requirements and were not from the e-cigarette industry, vape shops, or retailers.

Based on the tweet content related to e-cigarettes, through a group discussion of all four authors (RZ, ZX, QT, and DL), we split the 33 identified influencers into three broad influencer categories. The first category is “vape advocates,” who provide educational content about e-cigarettes and aim to persuade smokers to switch from traditional cigarettes to e-cigarettes owing to multiple health benefits. The second category is “vape reviewers,” who tweet about their reviews of certain e-cigarette products/devices. The third category is “other,” encompassing Twitter accounts owned by other vaping-related groups that are not affiliated with any vaping brands, such as vape social network platforms or even vape expos. The tweets from all 33 influencers were carefully examined, and the influencers’ categories were assigned by the coders based on the contents of their tweets and account information.

Social Network Analysis
Since we were particularly interested in e-cigarette–related influencer social networks, we decided to group the followers of each e-cigarette influencer into those posting about e-cigarettes and those not posting about e-cigarettes, defined as those who posted at least five e-cigarette–related tweets in the past year and those who had not posted any e-cigarette–related tweets in the past 3 years, respectively. To maximize the differences between these two follower groups, we did not include followers of e-cigarette influencers who posted 1-4 e-cigarette–related tweets in the analysis. Considering the large number of followers for each influencer, we randomly sampled 100 followers who post about e-cigarettes and 100 followers who do not post about e-cigarettes from each of the 33 e-cigarette influencers for further social network analysis.

For each e-cigarette influencer, we constructed two social networks for each follower type. We then compared key measures of these networks, including the average clustering coefficient, in-degree centrality, out-degree centrality, closeness centrality, and betweenness centrality, for each social network. The clustering coefficient of a network measures the degree to which nodes in the networks tend to cluster or form groups; a higher clustering coefficient suggests that nodes within the network are more likely to be connected, which can impact the overall structure and function of the network. In-degree centrality measures the number of incoming edges or connections of a node in a directed network, representing the popularity, influence, or attention received by one node from other nodes in the network. Similar to degree centrality, out-degree centrality focuses on the number of outgoing edges or connections from a node in a directed network, representing the extent to which a node reaches out to other nodes and influences them. Betweenness centrality measures the extent to which a node lies on the shortest paths between other pairs of nodes in a network, whereas closeness centrality measures how close a node is to all other nodes in a network. Both betweenness centrality and closeness centrality have significant control over the network’s communication dynamics and are crucial for maintaining efficient information flow.

Two-sample t tests and generalized linear models were used to compare the social network metrics between followers who do and do not post about e-cigarettes using R statistical analysis software (R Core Team, 2017) with the significance level set at 5%.

Topic Modeling Analysis
We performed topic modeling analysis using the latent Dirichlet allocation (LDA) model on all e-cigarette–related tweets posted by the 33 influencers [45]. Further, we identified 100 unique sampled e-cigarette followers for each of the 33 influencers and collected 3.48 million e-cigarette–related tweets posted by these followers in the past 12 months. After standard data cleaning procedures were applied to all tweets, such as removing hashtags, URLs, and emojis [46], we used the LDA model to obtain the most popular topics [45]. We chose the number of topics based on the coherence score and the intertopic distance.

Ethical Considerations
This study was reviewed and approved by the Research Subjects Review Board of the Office for Human Subject Protection at the University of Rochester (STUDY00006570). This study is a secondary analysis of publicly available Twitter data. All study data have been deidentified before analysis.
Results

Identification of e-Cigarette Influencers on Twitter

Between November 19, 2019, and June 10, 2022, we identified 3,617,766 unique Twitter user accounts that posted at least one e-cigarette–related tweet. Among these, 126 Twitter accounts were determined to be e-cigarette influencer candidates based on our criteria. Among them, 33 Twitter accounts were ultimately characterized as e-cigarette influencers, while the remaining 93 were vape shop or company accounts.

Among the 33 e-cigarette influencers, 10 belonged to the category of “vape advocates,” 19 were “vape reviewers,” and four were classified in the “other” category. “Vape advocates” (14,748 followers on average) had more followers than found in the “vape reviewers” (7426 followers on average) and “other” (4683 followers on average) categories. By comparison, “vape reviewers” had a higher percentage of followers that also post about e-cigarettes (1152/7426, 15.5%) than found for the influencer categories “vape advocates” (1357/14,748, 9.2%) or “other” (398/4683, 8.5%). E-Cigarette influencers categorized as “vape reviewers” and “vape advocates” had almost the same proportion of e-cigarette–related tweets at 53.1% (8205/15,454) and 53.2% (8,668/16,295), respectively. Influencers from the “other” category had a slightly higher proportion of e-cigarette–related tweets at 55.7% (3035/5449).

Further examination of all 19,908 e-cigarette–related tweets posted by the 33 e-cigarette influencers identified three major topics discussed (Table 1). The most popular topic was “advocating against vaping policy,” followed by “vaping helps to quit smoking” and “vaping product promotion.”

<table>
<thead>
<tr>
<th>Topics</th>
<th>Tweets (N=19,908), n (%)</th>
<th>Top 10 keywords</th>
<th>Example tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advocating against vaping policy</td>
<td>9556 (48)</td>
<td>People, Products, Don't, Ban, THC, Health, Harm, Public, Know</td>
<td>“Years and years from now. When cigarette smoking is finally eliminated. When no one is dying from COPD. You will realize that you wanted JUUL, punished for eliminating youth smoking and #vaping is absolutely a public health benefit.”</td>
</tr>
<tr>
<td>Vaping helps to quit smoking</td>
<td>5774 (28)</td>
<td>Quit, Vewapewevote, World-vapeday, Smokers, Health, Quitlying, Study, Like, Blackmarketthe, Today</td>
<td>“The number of ‘Vaping is illegal in my country but it helped me quit smoking and now I can’t get vape liquids so I had to buy cigarettes’ comments I get on YouTube is far… far too high. Ban-ning #vaping is anti public health. Discouraging people who WANT to quit is shameful.”</td>
</tr>
<tr>
<td>Vaping product promotion</td>
<td>4778 (24)</td>
<td>Vapefam, Vapelifе, New, Eliquid, Vapecommunity, Follow, Vapepen, Online, Dubs, Shops</td>
<td>“Arez 120 mod for a thorough test drive!”</td>
</tr>
</tbody>
</table>

Table 1. Major topics discussed by the 33 e-cigarette influencers on Twitter from November 19, 2019, to June 10, 2022.

As shown in Table 2, the average clustering coefficient for the networks of followers who post about e-cigarettes was significantly higher than that of the networks of followers who do not post about e-cigarettes. Therefore, the networks of followers posting about e-cigarettes have more connections than those of followers not posting about e-cigarettes for these 33 e-cigarette influencers. In addition, other social network metrics were significantly higher for the followers who post about e-cigarettes compared to those of followers who do not post about e-cigarettes, including in-degree centrality, closeness centrality, betweenness centrality, and out-of-degree centrality (Table 2). Therefore, the social networks of followers of e-cigarette influencers who also post about e-cigarettes have much more incoming and outgoing connections than those of the social networks of followers who do not post about e-cigarettes.

e-Cigarette Influencer Social Networks

The followers of the 33 identified e-cigarette Twitter influencers were grouped into those who previously did or did not post e-cigarette–related tweets. Considering the large number of followers (range 1097-49,215) for each influencer, we randomly selected 100 followers from each group (with or without e-cigarette–related posts of their own) for each influencer to create individual social networks based on their interactions (how they follow each other). To better illustrate the density differences between the social networks of the two types of followers, we randomly selected three e-cigarette influencers to demonstrate their sampled followers’ social networks. As shown in Figure 1, for any of the three representative influencers, the social network of followers who post about e-cigarettes had more connections (edges) than that of followers who do not post about e-cigarettes.
In addition, we compared the number of followers and their tweets in the social networks of the two e-cigarette influencer follower groups. The average number of follower accounts for followers who post about e-cigarettes was significantly higher than that for followers who do not post about e-cigarettes (322 vs 201; \(P<.001\); see Multimedia Appendix 1). Furthermore, the average number of tweets posted in the past 36 months by followers who post about e-cigarettes was higher than that of followers who do not post about e-cigarettes (93 vs 43; \(P<.001\); see Multimedia Appendix 2). Considering the higher number of followers and more tweets in the networks of followers who post about e-cigarettes, we decided to fit a generalized linear model to account for these differences. After controlling for the number of followers and tweets, compared to that of the networks of followers who do not post about e-cigarettes, the clustering coefficient of the network of followers who post about e-cigarettes was significantly higher (\(\beta=.26, P<.001\)).

Social network measures of the followers who post about e-cigarettes differed between vape advocate and vape reviewer influencers (Table 3). The followers of vape reviewers had higher clustering coefficients than those of vape advocates, which indicates that the e-cigarette–posting followers of vape reviewers have tighter connections than those of vape advocates. However, the network of followers of vape advocates had significantly higher in-degree centrality and closeness centrality compared to those of the networks of followers of vape reviewers. This finding indicates that, on average, the e-cigarette followers of vape advocates have better information flow.
between each other and a greater amount of incoming information compared to those of followers of vape reviewers. We did not observe a significant difference in the betweenness centrality and out-of-degree centrality between the followers of vape reviewers and vape advocates.

Table 3. Social network measures of follower networks of vape advocate and vape reviewer influencers on Twitter from November 19, 2019, to June 10, 2022.

<table>
<thead>
<tr>
<th>Social network measures</th>
<th>Vape reviewers</th>
<th>Vape advocates</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering coefficient</td>
<td>0.428</td>
<td>0.335</td>
<td>.02</td>
</tr>
<tr>
<td>In-degree centrality</td>
<td>0.061</td>
<td>0.193</td>
<td>.007</td>
</tr>
<tr>
<td>Closeness centrality</td>
<td>0.392</td>
<td>0.459</td>
<td>.008</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>0.007</td>
<td>0.019</td>
<td>.15</td>
</tr>
<tr>
<td>Out-of-degree centrality</td>
<td>0.006</td>
<td>0.004</td>
<td>.52</td>
</tr>
</tbody>
</table>

**Topics Discussed in e-Cigarette Social Networks**

Since we found that the followers who post about e-cigarettes were more likely to follow each other than the followers who do not post about e-cigarettes for the 33 e-cigarettes influencers, we further wanted to understand the types of information that these followers were sharing about e-cigarettes. As shown in Table 4, we identified two major topics from e-cigarette–related tweets posted on e-cigarette social networks. The relatively more popular topic was promoting e-cigarette products or vaping activity (74,572/133,881, 55.7%), followed by discussing the ability of vaping to help with smoking cessation and harm reduction (59,309/133,881, 44.3%).

Table 4. Major e-cigarette–related topics discussed by followers of 33 e-cigarette influencers.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Tweets (N=133,881), n (%)</th>
<th>Top 10 keywords</th>
<th>Example tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaping helping with smoking cessation and harm reduction</td>
<td>59,309 (44.3)</td>
<td>Vapefam, New, Ban, People, Don’t, Like, Know, Need, Time, News</td>
<td>“mother to mother, i’d like to know that vaping is important to us because it helped us stop smoking. Our kid need us to be healthy and able to watch them grow up. My beautiful family deserves a mom who doesn’t smoke.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“only 1 in 5 youth report flavors as a reason they vape. Why don’t you spend your time on policies that will actually be effective rather than trying to ruin the most successful form of tobacco harm reduction to date?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“over two thirds of vapers (68%) said they never thought they would quit smoking until #vaping came along - study of 2000 smokers. It's never too late to make the switch.”</td>
</tr>
<tr>
<td>Promoting e-cigarette products or vaping activity</td>
<td>74,572 (55.7)</td>
<td>Products, People, Smokers, Quit, Ban, Adults, Like, Youth, Market, E cigs</td>
<td>“Have you tried the range with cold shots yet? we recommend that you do! digi vape range is available for wholesale, no moq, further enquiries to shane”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“go to our Facebook page and join in the competition. ‘Let’s play a game, can you spot our new disposable vapes? try your luck’”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“starting my day with the beautiful vintage apple from vapour_distillery check out the fam cloud_legion”</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

In this study, with the identification of 33 influencers related to e-cigarettes on Twitter, we investigated the social network differences between followers of these influencers who do and do not post about e-cigarettes on Twitter. We showed that the social networks of followers who post about e-cigarettes have significantly more connections and higher density than those of followers who do not post about e-cigarettes, as indicated by the higher network clustering coefficient. The social networks of followers who post about e-cigarettes also had higher closeness centrality and betweenness centrality than those of followers who do not post about e-cigarettes, which suggests that the social networks of the former group are more efficient in terms of information dissemination, coordination, and influence. There were two main topics shared in the social networks of the followers who post about e-cigarettes, including vaping helping with smoking cessation and harm reduction and promoting e-cigarette products and vaping activity.

**Comparison With Previous Research**

With thousands or even millions of followers, social media influencers can have a significant influence on their followers, especially with respect to behavior changes. Therefore, it is crucial to understand how these influencers might engage with their followers and how they influence their followers through social network analysis. First, it is important to understand who these influencers are. A previous study investigated the influence of e-cigarette influencers on their followers on Instagram, showing that only 2 of 55 e-cigarette influencers used the word...
“influencer” in their descriptions, instead opting to use words such as "public figure," "brand ambassador," "promoter," "video creator," "artist," "photographer," "blogger," "model," or "fitness lover" to describe themselves [32]. In this study, we characterized e-cigarette influencers into different categories, including vape reviewers, vape advocates, and others. We did not observe any obvious difference in the hashtags used in their e-cigarette–related tweets, and most of them were related to vaping (Multimedia Appendix 3). Our topic modeling results of the tweets posted by the 33 influencers showed they were advocating against vaping policy, supporting vaping to help smoking cessation, and promoting vaping products on Twitter. In addition, based on the social network measures, we showed that influencers that advocate for vaping can spread their information to more general Twitter users, as indicated by high in-degree and closeness centrality values. However, influencers that review certain e-cigarette products attract more followers who post about e-cigarettes themselves, as indicated by a high clustering coefficient. Together, these results show that the vape reviewers’ followers tend to have more connections with each other, while the vape advocates’ followers tend to have more efficient information spread among each other.

To understand how social networks might be associated with user behavior, we grouped the influencers’ followers into those who post and do not post about e-cigarettes. Our results showed that the network of followers who also post about e-cigarettes had a higher clustering coefficient than that of the network of followers who do not post about e-cigarettes, suggesting that followers who post about e-cigarettes form a denser network (more connections) than followers who do not post about e-cigarettes. While we do not know if this tight connection among followers who post about e-cigarettes is due to their shared interest in e-cigarettes, the messages about e-cigarettes are much easier to share among these followers. In addition, besides the connection between the e-cigarette influencers and followers who also post about e-cigarettes, we found connections between the e-cigarette influencers themselves. For the 33 identified e-cigarette influencers, the number of followers who were influencers ranged from 0 to 17 (Multimedia Appendix 3). Therefore, users with similar interests tend to form a more closed network or community on social media. A previous study showed that social networks with active posting of alcohol-related messages were significantly associated with more frequent alcohol use [47]. It is plausible to speculate that the denser network for followers who post about e-cigarettes might indicate a greater influence on vaping behaviors, such as being more likely to vape and more frequent vaping.

Recognizing the significance of social networks on social media in disseminating e-cigarette–related messages, tobacco companies and vape shops have used social media influencers to promote their e-cigarette products. A previous study showed that, on average, each e-cigarette influencer on Instagram was sponsored by more than 10 different e-cigarette brands such as Voopotech, Innokin, and Geekvape [32]. Moreover, despite the FDA’s efforts to reduce e-cigarette advertisements targeting adolescents on social media, only 25% of the sampled e-cigarette influencers indicated age restrictions to access their e-cigarette promotional posts on Instagram [48]. These influencers on Instagram are very impactful, and their promotional posts through interconnected social networks can reach tens of thousands of adolescents. While more attention has focused on the influence of social media influencers on promoting e-cigarettes, it is important to understand how these influencers influence their followers through social networks. In this study, we showed that, on average, the followers who post about e-cigarettes have more connections in the sampled 100-node network compared to those of followers who do not post about e-cigarettes. The dense network might help disseminate e-cigarette–related promoting messages to more users in the social network, which can be a potentially effective strategy to effectively communicate health education messages with the public for future digital tobacco education campaigns.

Provaping messages are dominant on social media [49-51]. One study demonstrated that while the FDA’s sponsored antivaping hashtag “TheRealCost” only appeared 50 times a month on Instagram, provaping hashtags such as “e-juice” or “e-liquid” appeared more than 1000 times [52]. Another study showed that despite the FDA's requirement to add warning labels on e-cigarette advertisements on social media, most e-cigarette promotional images have no warning labels [53]. Our study showed that all e-cigarette–related influencers were promoting vaping as a smoking cessation tool and various vaping products on Twitter/X. Besides posting provaping messages, vaping advocates also posted messages advocating against tobacco regulatory policies, as shown in the major topics discussed by the e-cigarette influencers on Twitter/X. The vaping advocates directly contribute highly negative comments to FDA communication efforts, such as the FDA flavor enforcement policy and the proposed rules on menthol cigarettes [30,54-57]. These antiregulatory posts on social media might have a negative effect on the effectiveness of the FDA regulatory efforts. In addition, we noticed that some tweets from e-cigarette influencers are spreading misinformation about e-cigarettes, such as vaping being harmless, which is consistent with findings from other social media studies [58-60]. The spreading of this misinformation about e-cigarettes on social media, especially within social networks or communities, might undermine public understanding of the health risks associated with e-cigarettes, thereby promoting the initiation or continuation of vaping.

In our study, two topics from the networks of followers who post about e-cigarettes were vaping helping with smoking cessation and harm reduction and promoting e-cigarette products and vaping activity. The most popular hashtag we found in this social network was “WeVapeWeVote” (Multimedia Appendix 3), an organization fighting to protect the rights of adults to access vapor devices and other smokeless alternatives. The Hashtag “Vapefam,” used by someone who views the vaping community as an extended family, has also been mentioned more than 65,000 times in our data set. It is reasonable to speculate that these dominant provaping messages on social media might push back against current efforts of vaping regulation policies (such as the FDA’s flavor enforcement policy) that aim to reduce the current vaping epidemic in youth, which should draw more attention from public health authorities. Further, more educational warning messages about the health risks of e-cigarette use should be encouraged and promoted on
social media, which future tobacco education campaigns should consider. In contrast, as shown in Multimedia Appendix 3, the top 10 hashtags in tweets posted by the followers who do not post about e-cigarettes were more likely to be related to health, such as #health and #selfimprovement.

**Limitations**

There are several limitations of this study. First, the definition of followers who post about e-cigarettes was based on Twitter user accounts that have posted at least 10 e-cigarette–related tweets within the last 12 months. However, it is uncertain whether these posters are also e-cigarette users. In future studies, we will determine if they are actual e-cigarette users through use of a human-guided deep-learning model. Second, other factors (such as the demographics of Twitter users) might partially influence the average clustering coefficients, which need to be controlled in the future. In this study, we did not investigate the impact of social networks on user behaviors, which can be further measured in the future through a longitudinal study. Finally, the social network is dynamically evolving. Therefore, it is important to monitor the social network longitudinally to study how it evolves and influences user behaviors.

**Conclusion**

Through social network analysis, we showed that the social networks of followers of e-cigarette Twitter influencers who themselves also post about e-cigarettes have a denser connection compared to that of followers who do not post about e-cigarettes, which helps us better understand how e-cigarette influencers influence their followers through social networks. More importantly, social networks on social media might provide a novel and effective intervention approach for future health education campaigns to disseminate antivaping messages. For example, considering the dense social networks among e-cigarette social media users, public health advocates can join these social networks by following either the influencer or their followers who post about e-cigarettes, which would enable vaping prevention messages to be disseminated quickly in the networks and to specifically target potential or current e-cigarette users.

**Acknowledgments**

This study was supported by the rapid response project grants from the Western New York Center for Research on Flavored Tobacco Products (CRoFT) under cooperative agreement U54CA228110, funded by the National Cancer Institute and the US Food and Drug Administration (FDA). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the FDA.

**Data Availability**

The Twitter/X data used in the study are publicly available from the Twitter/X website [61]. The Python codes used for data analyses are available upon reasonable request to the corresponding author.

**Authors' Contributions**

DL, ZX, RZ, and QT conceived and designed the study. RZ analyzed the data. RZ, QT, and ZX wrote the manuscript. DL, ZX, RZ, and QT edited the manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Comparison of number of followers in the networks of followers of e-cigarette influencers who post (e-cigarette followers) and do not post (non-e-cigarette followers) about e-cigarettes.

[PNG File, 86 KB - formative_v8i1e53666_app1.png ]

Multimedia Appendix 2
Comparison of the number of tweets posted in the networks of e-cigarette influencer followers who post about e-cigarettes (e-cigarette followers) and do not post about e-cigarettes (non-e-cigarette followers).

[PNG File, 88 KB - formative_v8i1e53666_app2.png ]

Multimedia Appendix 3
Top 10 hashtags observed in e-cigarette-related tweets from e-cigarette influencers (Table S1); Number of followers who themselves are influencers for each e-cigarette influencer (Table S2); Top 10 hashtags in e-cigarette-related tweets from followers (Table S3).

[DOCX File, 18 KB - formative_v8i1e53666_app3.docx ]
References


49. Xie Z, Xue S, Gao Y, Li D. Characterizing e-cigarette-related videos on TikTok: observational study. JMIR Form Res 2023 Apr 05;7:e42346 [FREE Full text] [doi: 10.2196/42346] [Medline: 37018026]


61. Twitter/X. URL: https://twitter.com/home [accessed 2024-03-19]

Abbreviations

CDC: Centers of Disease Control and Prevention

FDA: Food and Drug Administration

LDA: latent Dirichlet allocation
Zhou R, Xie Z, Tang Q, Li D
Social Network Analysis of e-Cigarette–Related Social Media Influencers on Twitter/X: Observational Study
JMIR Form Res 2024;8:e53666
URL: https://formative.jmir.org/2024/1/e53666
doi:10.2196/53666
PMID:38557555

©Runtao Zhou, Zidian Xie, Qihang Tang, Dongmei Li. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Preferences on Governance Models for Mental Health Data: Qualitative Study With Young People

Emma Grace Carey1, PhD; Faith Oluwasemilore Adeyemi2, BSc; Lakshmi Neelakantan3, PhD; Blossom Fernandes4, PhD; Mina Fazel1, DM; Tamsin Ford1, PhD; The MindKind Consortium5; Anne-Marie Burn1, PhD

1Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom
2Department of Psychology, University of Bath, Bath, United Kingdom
3School of Population and Global Health, University of Melbourne, Melbourne, Australia
4Department of Psychiatry, University of Oxford, Oxford, United Kingdom
5See Acknowledgments

Corresponding Author:
Anne-Marie Burn, PhD
Department of Psychiatry
University of Cambridge
Herchel Smith Building for Brain and Mind Sciences
Forvie Site, Robinson Way
Cambridge, CB2 0SZ
United Kingdom
Phone: 44 01223 336961
Email: amb278@cam.ac.uk

Abstract

Background: Improving access to mental health data to accelerate research and improve mental health outcomes is a potentially achievable goal given the substantial data that can now be collected from mobile devices. Smartphones can provide a useful mechanism for collecting mental health data from young people, especially as their use is relatively ubiquitous in high-resource settings such as the United Kingdom and they have a high capacity to collect active and passive data. This raises the interesting opportunity to establish a large bank of mental health data from young people that could be accessed by researchers worldwide, but it is important to clarify how to ensure that this is done in an appropriate manner aligned with the values of young people.

Objective: In this study, we discussed the preferences of young people in the United Kingdom regarding the governance, sharing, and use of their mental health data with the establishment of a global data bank in mind. We aimed to determine whether young people want and feel safe to share their mental health data; if so, with whom; and their preferences in doing so.

Methods: Young people (N=46) were provided with 2 modules of educational material about data governance models and background in scientific research. We then conducted 2-hour web-based group sessions using a deliberative democracy methodology to reach a consensus where possible. Findings were analyzed using the framework method.

Results: Young people were generally enthusiastic about contributing data to mental health research. They believed that broader availability of mental health data could be used to discover what improves or worsens mental health and develop new services to support young people. However, this enthusiasm came with many concerns and caveats, including distributed control of access to ensure appropriate use, distributed power, and data management that included diverse representation and sufficient ethical training for applicants and data managers.

Conclusions: Although it is feasible to use smartphones to collect mental health data from young people in the United Kingdom, it is essential to carefully consider the parameters of such a data bank. Addressing and embedding young people’s preferences, including the need for robust procedures regarding how their data are managed, stored, and accessed, will set a solid foundation for establishing any global data bank.

(JMIR Form Res 2024;8:e50368) doi: 10.2196/50368

KEYWORDS
young people; mental health; data; governance; deliberative democracy; mobile phone
**Introduction**

**Young People’s Mental Health in the United Kingdom**

Worldwide, 1 in 7 individuals aged 10 to 19 years experiences a mental health disorder, which accounts for 13% of the global burden of disease in this age group according to the World Health Organization [1]. In the United Kingdom, the prevalence of probable mental health disorders in children and young people aged 6 to 16 years has increased from 11.6% in 2017 to 17.4% in 2021 [2]. The most common mental illnesses experienced by children and young people are emotional disorders, namely, anxiety and depressive disorders [3]. Emotional disorders are more common in girls than in boys, potentially getting worse with increasing age to the extent that, in a 2018 UK population-based survey, 22% of young women aged 17 to 19 years met the ICD-10 (International Statistical Classification of Diseases, Tenth Revision) criteria for any emotional disorder, including anxiety disorders, depression, and bipolar disorder [3].

Mental health conditions, including anxiety and depression, often continue into adulthood [4], whereas other potential adverse outcomes of poor mental health in childhood and adolescence include lower educational achievement, self-harm, substance abuse, and violence [5-8].

**Use of Mobile Technology to Capture Mental Health Data**

There has been a growing use of web-based resources and mobile technologies to better understand and support mental health. For example, data related to sleep; body movement; or exercise, social connections, and positive activities can be collected via smartphones and wearable devices (smartwatches and sensors). Web-based mental health programs can be used to help remotely monitor patients, send reminders to patients to engage in health-promoting behaviors, or complete electronic standard mental health scales [9,10].

Mobile devices are portable and inexpensive compared to traditional desktop computers and can contain features that provide real-time feedback, although they can drain batteries and use up data quotas [9,10]. Furthermore, new ethical and regulatory issues arise as these technologies develop and evolve, and it is essential to understand the implications of these developments. Because of this, it is more important than ever to research young people’s views on data governance and data sharing.

Previous research has shown that adult study participants are willing to share their data to be used for research by people working to improve health [11-15]. The societal benefit of sharing data arguably outweighs some of the privacy concerns or potential negative consequences, which are described in the following paragraphs [13,15].

Participants’ overall trust in an individual or institution can determine their willingness to share their health data [16]. There have been reservations about government access to data due to issues in the past with governments misusing public data and harming people in vulnerable and minority communities [17].

While one study found participants were accepting of governments accessing their data, it was concluded that this came from a position of resignation—access was considered acceptable because “the government already have all our information” [18].

Previous research has identified discomfort about the sharing of health information with commercial entities [14,15]. A systematic review found that participants were happy to share their data for research as long as the research conducted was not used to discriminate against a specific group of people and was used to help build knowledge, identify issues, and find answers to questions that participants endorsed [18]. Participants suggested that this risk could be mitigated by controlling who has access to health data via a screening process [11,17]. Not all studies have found opposition to government or commercial data access. One study on diabetes research data found that most participants (56%) were happy to allow commercial companies access to their medical history, genetic information, blood test results, and personal information, although it was unclear why this was the case because these findings were not supported with qualitative data [17].

**Rationale and Aims of This Study**

Several studies have examined adults’ attitudes toward sharing their health data or genetic data, covering issues surrounding data storage, access to the data, data use, and anonymity of the data. Adults’ views on sharing their data and data governance have been quite consistent across the studies, but we lack evidence of what young people think.

A global mental health data bank containing information about how various factors affect young people’s mental health would provide valuable resources for mental health researchers. Terms of governance acceptable to young people must be established before such a resource is developed.

This study was embedded within the MindKind Study, which was a large, multinational study investigating young people’s preferences for the collection, storage, and sharing of mental health data [19-22]. In the quantitative part of this study, participating young people downloaded and used a mental health app (henceforth, the *MindKind app*), which was used to inform us of young people’s preferences in developing a data bank of information pertaining to young people’s mental health.

The qualitative study described in this paper involved a deliberative democracy approach with young people who had either used or not used the MindKind app to gain a more detailed insight into their preferences for the use of their mental health data. We aimed to determine the preferences of UK youth for a future global data bank of mental health data and capture their views on the collection, storage, and sharing of mental health data.

**Methods**

**Ethical Considerations**

Ethics approval was granted by the University of Cambridge Department of Psychology Ethics Committee (reference: PRE.2021.031) and the University of Oxford (reference: June 2023). Participants were informed about the study in an online consent form, and their participation was voluntary. Participation consent was anonymized and sent to the data storage system (the *MindBank*). The data were stored securely on the *MindBank* server. The study was conducted in accordance with the Declaration of Helsinki and the Good Clinical Practice Guidelines. Participants were ensured of confidentiality and the right to withdraw from the study at any time without prejudice.
Sampling and Recruitment

Participants in the study were young people aged 16 to 24 years at the time of enrollment recruited from across the United Kingdom. Participants had lived experience of anxiety or depression as ascertained by a positive response to any of the following eligibility questions:

1. Have you ever felt that you could have benefited or did benefit from access to support for anxiety or depression?
2. Have you witnessed or experienced anxiety or depression within your family or close friends?
3. Do you have a strong interest in anxiety or depression?

We advertised the study via paid social media advertisements (eg, on Facebook and Instagram). We also used our own networks, including departmental Twitter (subsequently rebranded X) pages and Instagram accounts, and placed posters in our communities in Cambridge, Oxford, and Kent. Individuals who wished to participate or learn more about the study contacted the UK study team via email. We recruited 2 groups of participants. App-naive participants were those who had not downloaded and used the MindKind app and were recruited primarily via social media. Coenrolled participants were those who had previously downloaded and interacted with the MindKind app, and they received a pop-up notification after 2 weeks in the study asking whether they would like to sign up to join deliberative democracy sessions.

A member of the research team was in regular contact with participants to answer any questions that they had about the study and to arrange the times for the web-based sessions. All participants were provided with a participant information sheet and completed a web-based consent form and a sociodemographic survey to confirm study eligibility and assess the diversity of the sample. Each session’s participants were unique and attended only 1 deliberative democracy session.

Procedure

Educational Material

The MindKind Study developed 2 animated videos, which were coproduced with YPAGs based in the United Kingdom, India, and South Africa and voiced by our professional youth advisor [21,22]. The educational videos were sent to participants 1 to 2 weeks before their scheduled session to ensure that participants had a basic understanding of the topics to be covered.

The first video was approximately 10 minutes long and provided a brief overview of the proposed global mental health data bank and a brief introduction to each of the 7 questions that participants would be asked to deliberate on. The second video was approximately 25 minutes long and described 4 possible data governance models that could apply to a future data bank.

The Deliberative Democracy Process

A topic guide (Multimedia Appendix 1) and a set of facilitation slides were developed in collaboration with the YPAGs to address the 7 key questions as part of the MindKind Study. The questions (Multimedia Appendix 2) were (1) who can access the data? (2) Where are the data hosted? (3) Who controls the data? (4) What do people have to do before they can access the
Deliberative democracy sessions were stratified by previous exposure to the MindKind app (i.e., naïve or coenrolled participants). Where possible, we arranged sessions with different age groups to ensure that younger participants would feel comfortable. The process is further described in the MindKind Study protocol [22] and a paper describing the MindKind adaptations for conducting public deliberation using digital platforms [20].

Each web-based deliberative democracy session was facilitated by 2 trained members of the study team and lasted approximately 2 hours including 2 short breaks. Sessions were held on the Zoom videoconferencing platform (Zoom Video Communications) using Otter AI (Otter.ai, Inc) to provide automated closed captions. Participants were able to contribute to the discussion by speaking and using the chat feature to make comments or ask the cofacilitator any questions. To ensure a smooth discussion and that all voices could be heard, the cofacilitator verbalized comments from the chat box.

At the start of each session, participants were provided with the terms of engagement, which encouraged active and respectful participation. Participants were reminded that their privacy would be protected, they did not have to answer questions if they did not want to, and they could withdraw from the study at any time.

The session facilitator presented the facilitation slides described previously, which addressed the 7 questions in turn. Attendees had previously learned about these questions via the presession educational material. The group discussion addressed each of the 7 questions, with participants giving their views as to whether each option was seen as acceptable, unacceptable, or somewhere in between. Participants were free to speak or raise their hands to contribute, and the facilitator prompted participants who had not spoken on each question.

Immediately, during and after the session, a member of the research team collated the deliberative outputs from the session, creating a table classifying governance options as acceptable, maybe acceptable, or unacceptable. Session audio recordings were transcribed and anonymized, and the written content from the chat box was integrated into the appropriate parts of the transcript.

Analysis
We applied the framework method of thematic analysis [26,27]. Initially, 5 transcripts were allocated to different members of the research team (AMB, BF, EC, FA, and LN). Team members independently read the transcripts; familiarized themselves with the data; and highlighted salient points, which were open coded by each researcher. These excerpts were put into a Excel (Microsoft Corp) spreadsheet. Team members then met to compare the labels applied to their own transcripts and agree on a set of codes to be used for subsequent coding. We added a short summary description to each of the codes generated, forming a working analytic framework.

The transcripts and working framework were entered into NVivo (version 12; QSR International), and this file was placed on a shared drive. In total, 5 members of the team (AMB, BF, EC, FA, and LN) were allocated 2 transcripts each, which they coded line by line to the thematic framework. Researchers created additional codes where they were required during a series of meetings to discuss the coding process.

Once all the transcripts were coded, we generated a matrix framework that was then exported to Microsoft Excel. We charted the data by summarizing verbatim text and making analytical notes. The framework matrix enabled us to identify characteristics of the data and any group differences according to age or app exposure. Regular team meetings were held to ensure charting reliability and interpret the patterns between the themes and subthemes.

Results
Sample
A total of 11 web-based deliberative democracy sessions were held between August 2021 and April 2022 with 46 young people aged 16 to 24 years (median age 19 years). Of the 46 participants, 22 (48%) were naïve participants who had not used the app, and 24 (52%) were coenrolled participants who had previously used the app (see Table 1 for more details). Due to significant difficulties in recruitment, including many who did not attend agreed sessions, some group sizes were smaller than anticipated.
Table 1. Participants who attended each deliberative democracy session.

<table>
<thead>
<tr>
<th>Session number</th>
<th>Participants, n (%)</th>
<th>Age of participants (years)</th>
<th>App exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3 (7)</td>
<td>16-17</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>4 (9)</td>
<td>16-18</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>3 (7)</td>
<td>19-23</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>7 (15)</td>
<td>19-23</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>6 (13)</td>
<td>16-18</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>4 (9)</td>
<td>16-18</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>2 (4)</td>
<td>16</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>2 (4)</td>
<td>18-19</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>4 (9)</td>
<td>19-22</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>4 (9)</td>
<td>19-22</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>7 (15)</td>
<td>18-23</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The sociodemographic characteristics of the coenrolled and naïve samples are shown in Table 2.

Table 2. Sociodemographic characteristics of the samplea.

<table>
<thead>
<tr>
<th></th>
<th>Non–app users</th>
<th>App users</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Womanb</td>
<td>18 (82)</td>
<td>18 (75)</td>
</tr>
<tr>
<td><strong>Ethnic background, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Whitec</td>
<td>11 (50)</td>
<td>10 (42)</td>
</tr>
<tr>
<td>White</td>
<td>10 (45)</td>
<td>13 (54)</td>
</tr>
<tr>
<td><strong>Current stage in life</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary school or college, n (%)</td>
<td>10 (45)</td>
<td>8 (33)</td>
</tr>
<tr>
<td>University, n (%)</td>
<td>7 (32)</td>
<td>10 (42)</td>
</tr>
<tr>
<td>Not in educationd</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td><strong>Illness or disability, n (%)e</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5 (23)</td>
<td>11 (46)</td>
</tr>
<tr>
<td>No</td>
<td>16 (73)</td>
<td>12 (50)</td>
</tr>
</tbody>
</table>

aOne participant chose prefer not to answer and is not included in the table.
bOthers were male or nonbinary.
cNon-White participants were of South Asian, East Asian, Black, and mixed or multiple ethnicity.
dThis group includes those working or not in employment or education combined to avoid accidental disclosure.
eWhere values do not add up, the participants declined to answer.

**Themes**

Three overarching and interrelated themes were identified by participants as key to increasing trust and participation in the establishment of a global mental health data bank: (1) **accessibility and openness**, (2) **risks associated with the data**, and (3) **mitigation of risks**. The themes and subthemes are summarized in Figure 1 and detailed in the following sections; quotes are provided to illustrate the themes.
Figure 1. Themes and subthemes generated from the analysis.

**Theme 1: Accessibility and Openness**

**Data as a Public Good**
Young people believed that establishing a mental health data bank was acceptable. Typically, young people expressed hope that such a resource could be used to empower those with mental health difficulties and contribute to research to improve their lives. Typically, they wanted their data to be as accessible as possible while mitigating potential risks for data misuse and protecting confidentiality. This was driven in part by the belief that sharing data more widely would allow for more research, creativity, and discovery. Furthermore, young people believed that data could be used to empower communities by improving the understanding of mental health difficulties:

*My idea for who can access the data, if it was mental health data, I thought it would be good for teenagers to be able to access it, because then they'll be able to understand mental health.* [FG2; aged 16-18 years]

**Removing Barriers to Ensure Fairness**
Participants talked extensively about removing barriers that could prevent certain groups from being able to access the data. One of the primary perceived barriers to fair data access was cost—both direct and indirect. Young people understood that the costs of storing and managing data would need to be covered, but most were concerned that, if this cost were to be met by those accessing the data, individual researchers with less access to funding might be unable to conduct the research they want:

*I don’t know if classist is the right word to say, but like it makes it so that people with money are able to* access the data and people that don’t have as much money can’t access the data. [FG4; aged 19-22 years]

Some young people preferred that costs be met by those who access the data but with flexible charges to avoid barriers to access. For example, those with less funding could pay less than those with more funding using a sliding scale. Others thought that cost could depend on how much data people needed, how long people wanted to use the data for, or the complexity of the analysis proposed.

In general, charging at a macro level was preferred (ie, an organization, institution, or government) because it was thought not to preclude researchers without the funds from data access. Participants were also concerned about indirect costs—for example, some thought that requiring ID for data access might result in unacceptable costs for individuals in countries where IDs can only be obtained by paying and wanted to ensure that the data could be accessible wherever a researcher might reside. Young people also wanted to avoid nonfinancial barriers to data access, such as limiting access to data to researchers in certain countries:

*I agree, don’t want to gatekeep by ability to pay for ID (as someone who couldn’t afford a passport for many years!).* [FG4; aged 19-22 years]

*I feel like nowhere in the world should the data not be reached just because you’re from a certain country. It’s kind of unfair.* [FG5; aged 16-18 years]

**Theme 2: Risks Associated With the Data**

**Inappropriate Uses and Loss of Confidentiality**
While young people supported the equitable, open access of data, they had strong opinions on mitigating the risk of misuse...
and whom they would trust with their data. For example, many expressed concerns about how countries with strong societal stigma against those with mental health problems or with a poor track record on human rights issues might manage and use the data. Concerns were raised about the risk of identification and the harm that this could cause to individuals:

\[
\text{I think the only place I wouldn’t want my data going is North Korea...because they don’t have the best reputation for being not-harmful.} \quad [\text{FG6; aged 16-18 years}]
\]

**Access by Those With the Wrong Intentions**

The intentions of those accessing the data were important to young people. Private companies were viewed with some skepticism and described by young people as self-serving and profit-oriented and might misuse data to these ends:

\[
\text{There have been instances in research where the way that—even the way that data is interpreted will be biased to give the outcome that the people involve want, especially in for-profit studies. So I guess that would be a concern for me.} \quad [\text{FG1; aged 16-17 years}]
\]

Some voiced concerns about a poor track record for security among private companies and specifically distrusted advertising companies, insurance companies, and “big pharma” to protect their data. The media were also perceived as untrustworthy and likely to exaggerate or misinterpret findings for profit.

While researchers were generally seen as having better intentions for the data, some participants highlighted that researchers can also have bad intentions and their research can have harmful consequences. Specifically, they wished to limit any research that actively harms or has the potential to harm vulnerable and marginalized groups:

\[
\text{One thing I had in mind when I saw this question is there’s a book called “The Bell Curve”...but some researcher took a bunch of data about—I think it was IQ scores, and it was supposed to be a big mental health thing, but it just ended up being a bunch of racist crap. So, he was a researcher and he had proper access to the data...he will have passed the first six questions about all the other controls, but I think if you put question seven to him about what type of analysis, what are you using the data for, I think he would have failed that. So that would probably be most important for me.} \quad [\text{FG8; aged 18-19 years}]
\]

One participant highlighted the potential for genetic data contributing to research that might then be applied in eugenics, giving rise to some young people suggesting that research with potentially damaging consequences should receive a higher level of scrutiny.

**Data Loss or Corruption**

Beyond these concerns about how the data could be used to cause harm, young people were also concerned about the sabotage, corruption, and destruction of their data. For example, if the data were downloadable, they could be altered and false information could be shared, whereas malicious individuals could sell or leak the data or deliberately destroy them entirely.

Young people suggested that having multiple people controlling the data and storing the data in multiple locations could reduce the risk of intentional loss or corruption despite the potential for loss of complete control over the data:

\[
\text{My only concern is how safe the data is, how it’s encrypt[ed]. Because some people could just hack into a server and just corrupt the whole data.} \quad [\text{FG9; aged 19-22 years}]
\]

**Theme 3: Mitigation of Risks**

**Robust Procedures Are Needed to Mitigate Risk**

Participants were keen to ensure that risks involving loss, corruption, or misuse of data were mitigated in various ways. Ethical use of data was seen as fundamentally important, and young people emphasized the need for robust ethics procedures. Ensuring that data are used ethically was seen as one of the most important functions of a review board. Some ethical training was seen by most as crucial. Some young people wanted mandatory ethics training for all applicants, with exemptions issued to those who already have a good knowledge of ethical data use. Young people wanted this training to be robust and not a “tick-box” exercise, with many participants proposing a thorough ethics assessment after training:

\[
\text{There should be an assessment at the end of it that’s like a proper assessment, not just like an online quiz, multiple choice...Like something maybe scenario based or something where they actually have to put themselves in the position of something that could go wrong in the databank and what they’d do or what they’re allowed to do and what they can’t do, and it’s actually not just like closed responses.} \quad [\text{FG4; aged 19-22 years}]
\]

As well as being concerned about the data being used ethically, young people wanted to ensure that the data themselves were reliable. The concept of recreated data was raised by researchers as a method to protect confidentiality. A recreated data set is a data set that reflects key aspects of the original data but has been modified and combined in such a way that data points no longer reflect specific individuals. Many young people expressed concern that use of a recreated data set might iron out nuances and granularity in the data, which could have serious real-life implications by distorting research findings.

Maintaining privacy, anonymity, and confidentiality was seen as fundamentally important. Thorough anonymization enabled young people to feel safer—even if data were to be misused as this would not have an individual, targeted impact. Data that could lead to identification could be used to discriminate or stigmatize a specific person. In addition, answers to other questions about data access were more liberal on the potential for loss of complete control over the data:

\[
\text{My only concern is how safe the data is, how it’s encrypt[ed]. Because some people could just hack into a server and just corrupt the whole data.} \quad [\text{FG9; aged 19-22 years}]
\]

**Theme 3: Mitigation of Risks**

**Robust Procedures Are Needed to Mitigate Risk**

Participants were keen to ensure that risks involving loss, corruption, or misuse of data were mitigated in various ways. Ethical use of data was seen as fundamentally important, and young people emphasized the need for robust ethics procedures. Ensuring that data are used ethically was seen as one of the most important functions of a review board. Some ethical training was seen by most as crucial. Some young people wanted mandatory ethics training for all applicants, with exemptions issued to those who already have a good knowledge of ethical data use. Young people wanted this training to be robust and not a “tick-box” exercise, with many participants proposing a thorough ethics assessment after training:

\[
\text{There should be an assessment at the end of it that’s like a proper assessment, not just like an online quiz, multiple choice...Like something maybe scenario based or something where they actually have to put themselves in the position of something that could go wrong in the databank and what they’d do or what they’re allowed to do and what they can’t do, and it’s actually not just like closed responses.} \quad [\text{FG4; aged 19-22 years}]
\]

As well as being concerned about the data being used ethically, young people wanted to ensure that the data themselves were reliable. The concept of recreated data was raised by researchers as a method to protect confidentiality. A recreated data set is a data set that reflects key aspects of the original data but has been modified and combined in such a way that data points no longer reflect specific individuals. Many young people expressed concern that use of a recreated data set might iron out nuances and granularity in the data, which could have serious real-life implications by distorting research findings.

Maintaining privacy, anonymity, and confidentiality was seen as fundamentally important. Thorough anonymization enabled young people to feel safer—even if data were to be misused as this would not have an individual, targeted impact. Data that could lead to identification could be used to discriminate or stigmatize a specific person. In addition, answers to other questions about data access were more liberal on the potential for loss of complete control over the data:

\[
\text{My only concern is how safe the data is, how it’s encrypt[ed]. Because some people could just hack into a server and just corrupt the whole data.} \quad [\text{FG9; aged 19-22 years}]
\]
If someone has two intersectionalities like they’re black and gay and the dataset is recreated and it only takes into account one of them because it averages it out, it sort of ignores the really small...pockets of people in that dataset. [FG4; aged 19-23 years]

In light of these concerns, some participants proposed flexibility in level of access depending on individual needs, expertise, intentions, and requirements. Under such a system, only properly trained researchers with clearly positive intentions who had a true need to access the most sensitive, granular data would be granted access to it. Other researchers may have access to less sensitive, aggregated, or recreated data without such stringent requirements. Some young people also proposed that the most sensitive, confidential data should be stored in only one, well-protected location, with less sensitive or confidential data stored in other places.

Control by the Community Represented

Participants reported that it is important to retain some control over the data. Most young people did not want the data to be available to just anyone but instead believed that some kind of vetting process was important, such as assessing a person’s intentions for using the data and that they could be trusted to do this, providing them with the required training, and potentially asking them to sign a legal contract. Young people wanted to ensure that a group of people were responsible for this process of determining who could access their data. Having the entire community vote was seen as ideologically desirable but practically infeasible, so young people tended to see a community review panel who would make decisions about who should be able to access their data as a good option:

I think that the only way that you can really represent a community is that if you let that community speak for itself, and that is actually part of the reason why I don’t like community hires a manager either, because I don’t really think it’s feasible a whole group of people to be represented by one person. [FG7; aged 16 years]

Given their desire for vetting, young people tended to view data download as unacceptable because once data were downloaded, they could be freely shared with people who have not gone through this procedure. Some young people proposed a more flexible approach whereby a recreated data set, aggregated data, or a subset of less sensitive data was available for download but a more thorough procedure would be necessary to access the true, granular data. Young people generally approved of data access via a server and proposed a high level of security for the server to avoid hacking:

You’d have to make sure that that server was extremely well protected and able to fend some attacks from people who may wish to hack into it and use that data in a negative way. [FG9; aged 19-22 years]

A single data manager was viewed as unacceptably risky, and young people were keen to distribute power among a group of people with a personal stake in the data. Distributed power over the data bank would avoid any one person’s biases becoming entrenched in the system. Some young people also wanted multiple organizations or governments funding the data as they were concerned that a sole funder could seize control or misuse the data. A few participants also mentioned that storing data in multiple locations would aid the distribution of power:

If it’s just like one person or like one or two people, that’s only one set of biases and their biases sort of control the whole thing, if that makes sense. Whereas if it’s a community or like tens of people, then their biases counteract each other. [FG4; aged 19-23 years]

Young people generally supported procedures to hold anyone who misused the data accountable, such as using a formal contract. Funders and managers of the data should also be held to account to the same standards as anyone who accesses it:

I think that [signing a contract is] important because it will deter people from doing anything they shouldn’t be doing with the data, and then even if they do something bad, you can sue them. So, it’s not perfect—you can still do bad things—but at least they’re held accountable for it. [FG3; aged 19-23 years]

Ensure Skilled Data Management and Funding

Another way to mitigate risks of data misuse was to stipulate that the data be accessed by the “right” people with appropriate skill levels. Participants were concerned that sensitive data should only be accessible to suitably trained professional researchers. However, even sharing completely deidentified data was considered to be risky as harm beyond identification is possible. For example, young people were concerned that the press (in particular tabloid media) could distort findings in a way that is harmful to youth mental health generally. Some young people proposed flexible systems with different levels of access depending on individual expertise and intentions:

I guess people working in a newspaper would have an interest to try and make it controversial or like make a finding out of just generic data. [FG4; aged 19-23 years]

If a normal person that’s not part of the mental health medical field, they should just see maybe an infographic or maybe just some paragraphs that conclude what’s in the data so they’re aware, globally, of mental health. [FG1; aged 16-17 years]

Researchers at universities and other non–profit-motivated organizations were perceived to have an interest in conducting research properly and ensuring that the findings were accurate, as opposed to private companies, who were seen as primarily motivated by profit. Some—but not all—young people placed the government in a similar category with an interest in the public good. Others presented a more nuanced view:

Whether a government should fund the databank depends on whether the government is democratic or not, or even if they are democratic, whether they’re corrupt or not. [FG7; aged 16 years]

Most participants trusted charities and nongovernmental organizations, although others listed specific named charities that they considered not to be trustworthy because of a track record of stigmatization or harm:
I wouldn’t want somebody like [named charity] holding my data, because I don’t like them. [FG6; aged 16-18 years]

Mental health care practitioners were seen as motivated to improve the mental health of young people and having the appropriate skills to use the data well. Young people wanted those with skills in data dissemination to have access so that findings could be relayed back to relevant communities:

Data-analysis-wise, maybe you need some data scientists which may not be very good at psychological stuff. So maybe psychiatrists maybe need to work with data scientists to interpret data. [FG9; aged 19-22 years]

The intentions of those accessing the data were seen as important while acknowledging that even well-intentioned use could cause harm. Young people wanted those who accessed the data to be working for the benefit of society and in particular to improve the health of young people. Profit was not an appropriate intention for most participants, although it was acknowledged that profit-making is not inherently harmful. Particular concern arose about intentions that could harm, for example, developing targeted advertising to encourage harmful coping mechanisms, stigmatizing mental health problems, or charging individuals higher life insurance premiums:

I was also going to say, yeah, getting targeted ads based on the stuff that you look at, the last thing you really want is them being able to have access to private data about things like, for example, your mental health, because the last thing you want is something saying, “Oh, are you struggling from depression? Try Jack Daniel’s whiskey.” [FG11; aged 18-23 years]

Skilled data management was also seen as very important; young people proposed that the data managers should have expertise in confidentiality, with more trust that privacy would be maintained if data were managed by a group of skilled, reliable individuals. Some suggested that the review panel should represent a variety of skills (eg, psychologists, students, and activists), and others thought that different decisions should be made by people with specific expertise in that area:

In the review panel there could be a few psychologists, some students as in college/university students, researchers who don’t necessarily have a degree or any kind of certificate in psychology, but they are known for research. And then...you know activists, like mental health activists. [FG5; aged 16-18 years]

Young people were also keen that the data bank be funded by individuals with the correct intentions. They often questioned why an organization or government would volunteer to spend money on a data bank. Some young people felt that the government has a duty to improve mental health and were therefore obvious funders for a data bank. Universities were generally viewed as trusted institutions with an interest in furthering knowledge as well as having research expertise and thorough ethics review processes:

If the research can help the general public, then the government should be contributing towards that, just like they give money for the NHS. [FG3; aged 19-23 years]

Ensure That Mitigations Are as Feasible and Flexible as Required

Young people acknowledged that accounting for feasibility, practicality, and financial viability was necessary. One frequent concern was the global nature of the proposed data bank. This means that the data bank would include data collected from individuals in countries with varying data protection laws. Some young people considered that the lack of certain laws applying to their data would make them feel unsafe. Others were more concerned about how these differences could be resolved practically. Some participants argued that, if data are held and managed in the United Kingdom, other countries should abide by the UK laws or rules but with agreement that one country having more control than the others could create an unfair power dynamic:

I’d be worried about GDPR [General Data Protection Regulation], especially if one country had it and another country had their GDPR regulations different to the other countries...so they might have different rules of how they store the data. [FG2; aged 16-18 years]

A few young people were concerned enough about differences in global policy and attitudes that they did not think that a truly global data bank was feasible, whereas others believed that a global data bank would be a valuable asset to global mental health research and that challenges could be managed.

Other practical concerns regarding a global data bank revolved around the ability to hold people to account for misuse of data. Variations in professional expertise worldwide were also of concern. Language barriers, currency conversion, and differences in ID worldwide were also flagged as issues that would need to be resolved when establishing a global data bank. Young people felt that having data managed by “real people” could allow for flexible systems that mitigate some of these concerns (eg, those without ID could be vetted more thoroughly in a different way):

If anything did happen and, for some reason, that contract was broken, where would it go to? Would you have a trial in your own country or would it have to go to an international court or the European courts, or somewhere like that? [FG9; aged 19-22 years]

Practicalities were often in tension with young people’s ideological preferences. For example, young people favored the democracy afforded by consultation on data management with the entire community as the most representative option. However, practical concerns here were numerous—it would take a great deal of effort and may become bureaucratic, a vocal minority might be heard over a quieter majority, and the community may become disengaged. For these practical reasons, participants proposed a community review panel as a more viable option that distributed power among a group. Some young people proposed that a community review panel should be...
combined with the community deciding—for example, the panel could be elected by the community, or the community could be given a yes or no vote with nuances decided by the panel. Some participants suggested that the most controversial decisions should be made by the entire community, whereas less potentially harmful uses of data could be approved by a panel. Others argued that less controversial studies could be approved by an algorithm. Young people stipulated that a review panel should be as representative as possible of the entire community to compromise between the practical benefits of a review panel and the ideological desire to represent the entire community well:

I think the community decides is pretty unacceptable is that because in a community, the most vocal people tend to have the largest say and it can end up becoming a vocal minority who have really strong opinions on certain things. [FG10; aged 19-22 years]

Similarly, young people were against the idea of charging for access to data. However, they acknowledged that it might be the most feasible option to cover the operational costs of the data bank. Nevertheless, they believed that institutions, rather than individuals, should primarily bear the cost:

I’m all for having people pay money...I know that there are some health databases that people pay money to access for research purposes, and I think that would help cut down costs. [FG10; aged 19-22 years]

Similarly, young people thought that a secure server would provide the best protection for their data but equally saw that this might limit the ability of researchers to conduct their preferred analysis or be unreliable in parts of the world with inconsistent power or internet connection:

I do believe, for practical reasons, it is better to download the data from the researcher’s point of view. Just from personal experience, doing things in the server can be very tedious and very chaotic. [FG8; aged 18-19 years]

Despite strong support for robust ethics training, some young people were concerned that the practical hurdle might lead people to seek out lower-quality, less scrupulous data sets. Some proposed flexible systems whereby only research with the right intentions and with a high risk of harm or controversy would need thorough ethics training and assessment.

Ensuring that a data bank could continue in the long term was important to participants, with contrasting views on how the data bank should be managed and by whom. While it could be more practical for one organization or group to manage the data, multiple funders could ensure sustainability. Others suggested that having governments fund the data bank could also provide this consistency or resource. Participants were concerned about what would happen to the data in the long term, where data access becomes less stringently monitored as the data age (and lose some of the sensitivity and identifiability of recently collected data). It was seen as practically unviable to maintain a review panel in the longer term:

I was thinking that a review panel might not end up being useful in the long run, because I was thinking, over time, it would be harder to find enough people who are actually willing to review it. So maybe a manager might be easier after a while. [FG10; aged 19-22 years]

Discussion

Principal Findings

We ran 11 deliberative democracy discussions of 7 questions related to the governance of a potential global mental health data bank among young people aged 16 to 22 years in the United Kingdom as the vast majority of research to date has focused on adults. Participants were enthusiastic about data being as accessible as possible to a variety of people because data access was seen as a public good. Our broader, global study found some differences by region, so we report findings specific to the United Kingdom [21].

Young people were also very enthusiastic about access of data determinants being equitable, but this desire for accessible data was in tension with concerns about risks associated with data use. In particular, young people wanted to limit inappropriate or badly intended use of their data, ensure that their privacy was maintained, and prevent the sharing of unreliable data. There was enthusiasm for various risk mitigation strategies to counter these possibilities. Young people wanted robust procedures to ensure ethical use of data and the ability to hold people to account for misuse. Control of the data by the community who contributed to them was important, as were distributed power and diversity of those in charge of the data. Young people wanted to ensure that these mitigations were both practical and flexible; mitigations were a necessary means to ensure that data would be accessed by those who could be trusted to use them skillfully, with the right intentions, and with positive consequences.

Implications for Policy

Our finding that young people were generally positive about the idea of widely sharing mental health data was in keeping with those of previous research suggesting that adults are willing to share their data for research purposes [11-15] and perceived that the societal benefits of this outweighed potential negative consequences [13,15]. While previous research has also suggested that participants want their data to be as useful as possible via increased access, the young people in this study framed fair and equal access to data as a good in itself rather than suggesting that fair access is simply a means to maximize the public good of data availability.

The desire to obtain equitable access and maximize public good led to many interesting discussions about how to remove barriers such as cost. Cost was especially interesting as it affected young people’s approach to all aspects of the discussion—for example, requiring researchers to provide ID was perceived as ideal in terms of controlling data and limiting access by those with poor intentions but might unfairly affect researchers from countries or institutions with less money or those who are not affiliated with an institution. This has broad implications for policy as it
highlights the need to consider both direct and indirect costs as a cause of potential injustice in data access.

The intention behind data access application was an important consideration consistent with the findings of other studies [12]. Participants were especially concerned about the uses of data that could increase stigma or prejudice against certain groups of people, echoing the findings of Breakey and Dipinto [18]. There is an obvious tension here with the desire for research to be shared fairly and freely, especially when young people mentioned specific countries as being untrustworthy. Young people recognized that the presence of an untrustworthy government does not imply that the people of a country are themselves untrustworthy, but in practice, it would be challenging to allow free access to citizens and prevent misuse of data by their government.

Previous studies have shown that people are willing to share their physical health data with commercial entities [17]. In contrast, our participants were cautious about sharing their data with commercial organizations and suspicious of profit-driven interest in using their data. This may be because the participants in our study were younger or because data about physical health conditions are seen as less sensitive than mental health data. Our participants expressed strong views on sharing their data with commercial organizations even though many young people might regularly, and possibly inadvertently, share personal data with for-profit social media companies and health apps. Future research could explore directly this disconnect between principles and practice.

In addition to limiting who could access data by their identity or profession, young people were keen on robust ethical procedures. The level of modular ethical training and assessment proposed by young people in this study was interesting insofar as it was significantly beyond what is typically required by researchers applying to access data. However, the reasoning of young people in this study was sound—researchers from different institutions and worldwide locations may have very different previous training in ethics, and conducting additional training could standardize abilities. Ajuwon and Kass [28] conducted a research ethics training workshop with researchers and reported significantly improved understanding of research ethics in a follow-up test, which indicates that, even among trained researchers, additional training can improve and standardize skills.

Young people understood that data would not be attached to themselves untrustworthy, but in practice, it would be challenging to allow free access to citizens and prevent misuse of data by their government.

Strengths and Limitations

This study uniquely provided insights into the views of young people in the United Kingdom on the governance of mental health data. We acknowledge that this qualitative study used a convenience sample and, therefore, may not represent the views of other young people.

The main limitation of this study was that some of the sessions had a small number of participants; our group sizes ranged from 2 to 7 participants as we considered it important to proceed with planned sessions despite last-minute cancellations. Smaller groups will have limited the range of experience and opinions expressed. Nevertheless, we found that the smaller groups were cohesive and the participants engaged well and contributed their ideas freely. Participants had the time to speak more in depth about their ideas and justify their reasoning, were able to deliberate the questions, and respected each other’s opinions and ideas. In addition, across the entire study, there was a diverse group of participants in terms of ethnicity and age, and therefore, we were able to capture a wide range of life experiences across the sessions. We found that, regardless of session size, participants tended to say similar things between sessions and that the consensus reached did not seem to vary greatly. While this study is a relatively preliminary exploration of data governance preferences among young people, we were struck by the consistency of themes and consensus building, which did suggest that the data captured accurately reflect the opinions of young people in the United Kingdom.
Deliberative democracy sessions are usually conducted face-to-face; however, we conducted our sessions digitally using videoconferencing [20]. We encountered technical challenges such as poor broadband connection or microphone issues that limited some participants’ contributions to the discussions. Because the sessions were web based, this may have made it easier for participants to not attend at the last minute. Grönlund et al [30] compared web-based deliberative democracy with face-to-face deliberative democracy and found that participants were able to engage in high-quality deliberation when technology performed properly.

One strength of conducting the sessions digitally was that participants had a range of methods they could use to communicate; we made use of the chat and reaction emojis throughout the sessions. In addition, this feature was beneficial for those who felt less comfortable speaking out loud as they were still able to share their ideas. Digital sessions also extended our geographical range. Participants could take part in the sessions regardless of where they lived in the United Kingdom and could choose from a number of time options. This was beneficial as it enhanced the findings and made them potentially more representative of the entire population with the addition of diversity and, therefore, life experience to the groups.

The coproduced educational materials were helpful for participants to grasp some difficult scientific concepts. However, there was one concept (a “recreated dataset”) that proved a difficult term to convey to young people, which meant that it was difficult to have a meaningful discussion about it [31].

Conclusions
We found that young people in the United Kingdom are able and willing to successfully prepare for and participate in public deliberation about the sharing of their mental health data. While the web-based setting made this more challenging in some ways, it also brought potential benefits for the diversity of our sample and the accessibility of the research. Young people had many concerns about the possibility of misuse and negative consequences of sharing their data but were able to suggest mitigations to these concerns and were broadly positive about the collection, sharing, and use of mental health data by researchers and other relevant, vetted stakeholders to improve mental health and well-being. A global mental health data bank is acceptable to young people in the United Kingdom and would provide significant resources for future research.

Acknowledgments
The results presented in this paper are in whole or in part based on data obtained from the MindKind Study funded by Wellcome Trust. The MindKind Databank data were contributed by participants of the MindKind Study developed by the MindKind Study Consortium [32] led by lead principal investigator Lara Mangravite (Sage Bionetworks) with Megan Doerr and Solveig Sieberts (Sage Bionetworks; study leads) in collaboration with Pamela Collins and Patricia Areán (University of Washington) and site principal investigators Soumitra Pathare (Indian Law Society, India), Zukiswa Zingela (Walter Sisulu University, South Africa) and Melvyn Freeman (World Health Organization and Higher Health, South Africa), and Mina Fazel (University of Oxford and Oxford University Hospitals, United Kingdom) and Tamsin Ford (University of Cambridge, United Kingdom) and with guidance from the MindKind Youth Group Advisory Panels in India, South Africa, and the United Kingdom; the International Youth Panel; the Global Youth Panel; and the Data Use Advisory Group. The study was commissioned by the Mental Health Priority Area at Wellcome Trust, and data were obtained through Synapse [33].

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
TF’s research group receives funds from Place2Be, a third sector organisation that provides mental health training and interventions in school for research methods consultation.

Multimedia Appendix 1
Topic guide.
[DOCX File, 25 KB - formative_v81e50368_app1.docx ]

Multimedia Appendix 2
Seven key questions.
[DOCX File, 15 KB - formative_v81e50368_app2.docx ]

References


Abbreviations

ICD-10: International Statistical Classification of Diseases, Tenth Revision
YPAG: young people’s advisory group

©Emma Grace Carey, Faith Oluwasemilore Adeyemi, Lakshmi Neelakantan, Blossom Fernandes, Mina Fazel, Tamsin Ford, The MindKind Consortium, Anne-Marie Burn. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: Media narratives can shape public opinion and actions, influencing the uptake of pediatric COVID-19 vaccines. The COVID-19 pandemic has occurred at a time where infodemics, misinformation, and disinformation are present, impacting the COVID-19 response.

Objective: This study aims to investigate how narratives about pediatric COVID-19 vaccines in the media of 4 English-speaking countries: the United States, Australia, Canada, and the United Kingdom.

Methods: The Narrative Policy Framework was used to guide the comparative analyses of the major print and web-based news agencies’ media regarding COVID-19 vaccines for children aged 5 to 11 years. Data were sought using systematic searching on Factiva (Dow Jones) of 4 key phases of pediatric vaccine approval and rollout.

Results: A total of 400 articles (n=287, 71.8% in the United States, n=40, 10% in Australia, n=60, 15% in Canada, and n=13, 3% in the United Kingdom) met the search criteria and were included. Using the Narrative Policy Framework, the following were identified in each article: hero, villain, survivor, and plot. The United States was the earliest country to vaccinate children, and other countries’ media often lauded the United States for this. Australian and Canadian media narratives about vaccines for children aged 5 to 11 years were commonly about protecting susceptible people in society, whereas the US and the UK narratives focused more on the vaccine helping children return to school. All 4 countries focused on the vaccines for children aged 5 to 11 years as being key to “ending” the pandemic. Australian and Canadian narratives frequently compared vaccine rollouts across states or provinces and bemoaned local progress in vaccine delivery compared with other countries globally. Canadian and US narratives highlighted the “infodemic” about the COVID-19 pandemic and disinformation regarding child vaccines as impeding uptake. All 4 countries—the United States, Australia, the United Kingdom, and Canada—used war imagery in reporting about COVID-19 vaccines for children. The advent of the Omicron variant demonstrated that populations were fatigued by the COVID-19 pandemic, and the media reporting increasingly blamed the unvaccinated. The UK media narrative was unique in describing vaccinating children as a distraction from adult COVID-19 vaccination efforts. The United States and Canada had narratives expressing anger about potential vaccine passports for children. In Australia, general practitioners were labelled as heroes. Finally, the Canadian narrative suggested altruistic forgoing of COVID-19 vaccine “boosters” as well as pediatric COVID-19 vaccines to benefit those in poorer nations.
Conclusions: Public health emergencies require clear, compelling and accurate communication. The stories told during this pandemic are compelling because they contain the classic elements of a narrative; however, they can be reductive and inaccurate.

KEYWORDS
COVID-19; SARS-CoV-2; vaccine; mRNA; Pfizer-BioNTech; pediatric; children; media; news; web-based; infodemic; disinformation

Introduction

Background

The COVID-19 pandemic has had an enormous impact on our lives, and COVID-19 vaccines remain key to preventing severe disease for most people. The media has played an important role in the pandemic, informing and educating populations about COVID-19, vaccines, and public health mitigation measures. Previous research has demonstrated that the media influences the beliefs, attitudes, and behaviors of institutions, governments, and individuals [1]. Mainstream and social media have become focal points during the pandemic, to the point where the World Health Organization (WHO) has named the excess information an “infodemic,” a term defined as “excess information including, but not limited to, false or misleading information, in digital and physical environments during an acute public health event” [2]. Disinformation and misinformation have also been features of the COVID-19 response. These are descriptions of false information, the former being deliberate to deceive and the latter not deliberate.

False and misleading information has been present for many years and has played a role in past pandemics. However, the COVID-19 pandemic has accentuated recent changes in the way we find, write, receive, and share information. Emotive misinformation travels quickly, and controversies can affect individuals, organizations, public opinion, and decision-making [3].

Infodemics and disinformation or misinformation can cause confusion and undermine the collective action needed to protect public health. There is a need to increase our understanding of the media, its messages, themes, and interpretations to improve communication in crises and ultimately improve health [4,5].

Media

The influence of various forms of media is increasing but unpredictable. This is in part because of the opaque workings of the information delivery algorithms behind the media that is consumed digitally, as we rush to understand our rapidly changing global information landscape [6]. Current and major events in recent history have demonstrated the influence of the media on public and institutional views and actions [7]. Media can skew perceptions and cause bias [8], influence public opinion [9], and influence policy makers [10]. Notably, social media plays a major role in shaping media narratives. However, the traditional media remains an equally important influence on public opinion.

Media narratives influence people in a multitude of ways. The producers of the narrative decide which issues are salient, what suits their agenda, political leanings, and known audience, and therefore what will generate “clicks” or drive reader attention [11]. Extended coverage on issues lead us to think the issues are important and credible [12,13]. The effects of media narratives can be immediate or slow acting; can be positive or negative; and can affect individuals, institutions, and society as a whole [14].

COVID-19 Vaccines

Vaccines are arguably the most clinically effective and cost-effective interventions in health care [15]. During pandemics, notably Zika, Ebola, Middle East Respiratory Syndrome, and now COVID-19, vaccines have been fast-tracked for approval using emergency use authorization processes for adults and subsequently for children. Emergency use authorization enables a fast track between product development and use where there is an urgent clinical need; evidence exists, but more will be gathered [16].

Public health strategies have aimed for high uptake of the COVID-19 vaccine in all age groups, as a predominantly vaccinated population will help reduce the transmission of COVID-19 and the sequelae of COVID-19 infection [1]. Achieving this goal depends on multiple aspects of the system as well as individual choice. Behavioral research identifies three categories of drivers of vaccine uptake, in addition to people having the necessary knowledge: (1) an enabling environment, (2) social influences, and (3) motivation [1]. The media plays a role in each of these, for instance, by sharing eligibility, vaccine timing, or locations where the vaccine will be offered. During the COVID-19 pandemic, many were confined to their homes or reduced their social networks, meaning perceptions of other people’s behavior—particularly regarding vaccination and also mask wearing and social distancing—were more likely to be inferred from mainstream and social media [17].

Initially, vaccinating children aged 5 to 11 years against COVID-19 was viewed as a dilemma. Although COVID-19 can be severe in adults [18], children with COVID-19 generally have a less severe disease [19]. Although only a small proportion of children with COVID-19 will go on to have severe disease, the widespread prevalence of COVID-19 in this population means that the absolute number is still large, including those at risk of sequential severe outcomes, particularly “long COVID” [19] or multisystem inflammatory syndrome in children [20-29]. Symptomatic or asymptomatic children may infect their parents or other susceptible people with COVID-19 [30-32]. Furthermore, children have been affected by disruptions to daycare, school, and extracurricular activities as well as the economic and mental health impacts of COVID-19 restrictions [33]. The argument for vaccinating...
children is strong for children with neurodisabilities; Down syndrome; immunodeficiencies; malignancies; some cardiac, respiratory, and renal diseases; obesity; and poorly controlled diabetes [34] as well as those living with high-risk household members [32]. The factors for and against vaccinating children have been described in more detail elsewhere [32].

Studies regarding the uptake of COVID-19 vaccines in children have largely focused on parents, their trust in governments, and the information they receive [35-39]. The Kaiser Family Foundation found that before approval of the vaccines for children aged 5 to 11 years in the United States, only 27% of American parents of children aged 5 to 11 years would immunize their children against COVID-19, compared with 95% of kindergarten children being immunized against several other diseases including measles, mumps, and rubella (MMR) [40]. Likewise, vaccination rates for MMR are 95% for children aged 5 years in Australia [41], 95.2% for children aged 5 years in the United Kingdom [42], and 46% to 95% for children aged 5 to 7 years [43] in Canada. In the United States, 91% of children aged 2 years and 92% of children aged 13 to 17 years are vaccinated with MMR [44]. Despite these high rates of MMR vaccination uptake, these figures are not reflected in pediatric COVID-19 vaccination rates. The latest data (July 17, 2022) in Australia for the vaccination rate for children aged 5 to 11 years is just 52.3% for first doses and 40.3% for second doses [45], despite the Pfizer-BioNTech COVID-19 vaccine for children aged 5 to 11 years being available and free since early January 2022. The disparity in vaccination rates between COVID-19 and other vaccine-preventable diseases is likely in part because of the media playing a critical role in opinions and behaviors [46].

To date, COVID-19 and COVID-19 vaccination experiences have been vastly different in Australia, Canada, the United Kingdom, and the United States (Textbox 1). Arguably, Australia has fared the best through the pandemic, with the (until now) lowest case load (Figure 1) [47] and the lowest mortality rate (Figure 2) [47], noting that Australia has had a higher testing rate than Canada, the United Kingdom, and the United States [48]. To date (July 31, 2022), the cumulative number of cases [49] and total deaths [50] from COVID-19 infection per million people are approximately 270,971 and 3056 in the United States, 364,021 and 458 in Australia, 107,265 and 1126 in Canada, and 347,681 and 2737 in the United Kingdom, respectively. These countries have also experienced different pandemics in different health care systems. For instance, the US health expenditure per capita (US $10,921) is approximately twice that of Australia, the United Kingdom, and Canada [51]. The US health system is a complex mix of public and private, with both profit and not-for-profit insurers and providers [52]. Canada, the United Kingdom, and Australia all have universal health care provided through a government fund; however, Australia uses a mixed public or private model [53].

As COVID-19 devastated the globe, the WHO recognized the increasingly detrimental effects of the coinciding infodemic, misinformation, and disinformation on health [5]. This paper contributes new knowledge to the WHO priority of “detecting signals and understanding the spread and risk of infodemics” [2]. Given the importance of the media in shaping public and institutional knowledge, perception, attitudes, and behavior, we aimed to investigate how pediatric COVID-19 vaccine narratives have unfolded in the media of 4 countries: the United States, Australia, Canada, and the United Kingdom.

Textbox 1. Summary of each country’s pediatric vaccine timeline.

Australia
- COVID-19 vaccines became available for children aged 5-11 years during the initial wave of Omicron, when testing capacities became overwhelmed. There were supply chain issues due too many individuals isolating after testing positive for COVID-19 or being listed as close contacts, and first child doses were due when many adults were due for their booster doses. This meant that many vaccine clinics, doctor offices, and pharmacies were inundated with the need for vaccines. Furthermore, at this time, there were supply chain issues with rapid antigen tests. This was also a holiday period for many Australians (December 2021 to January 2022), with Christmas, New Year, and school summer holidays occurring during this time.

Canada
- Canada approved vaccination for children aged 5-11 years shortly after the United States, and rollout started just prior to Thanksgiving (November 2021) and in time for the Christmas or New Year 2021 Holidays, around the time when the spread of Omicron was announced.

United Kingdom
- The United Kingdom was the last of the 4 countries to approve vaccinations for children aged 5-11 years, with the government wanting to focus on vaccinating the adult population. In January 2022, vaccines for susceptible children or children who live with susceptible adults were approved, and in late February 2022, vaccines were approved for all children aged 5-11 years, with rollout at the start of April 2022.

United States
- The United States was the first country to approve vaccines for children aged 5-11 years, after the pediatric society pleaded for the Food and Drugs Administration to speed up the vaccine approval for this age group, stating that the benefits outweighed the risks, and the data gathered from the ongoing trials and adult populations were sufficient. The process from approval to rollout was rapid, and the Omicron variant announcement came later.
Figure 1. Confirmed COVID-19 cases per million during the COVID-19 pandemic.

Figure 2. Confirmed COVID-19 deaths per million during the COVID-19 pandemic.

Methods

Search Criteria
To better understand, compare, and contrast media narratives about the COVID-19 vaccines for children aged 5 to 11 years, we selected 4 countries. These countries were selected for specific reasons. All 4 countries are English-speaking, high-income, founding or first Organization for Economic Cooperation and Development members, democratic, and market-economy countries with multicultural populations. For each of the 4 selected countries, the federal governments have been responsible for vaccine procurement, approval, and scientific advice on their use. COVID-19 vaccines have been available for free in each of the 4 selected countries. All 4 countries had similar health expenditures per capita and, hence, were thought to be good comparators. The United Kingdom, Canada, and Australia spend very similar amounts on health per capita, whereas the United States is an outlier, spending more than twice the amount of the other listed countries per capita [51]. These 4 countries also spent similar amounts on COVID-19 in relation to their gross domestic product [54]. In Australia, Canada, and the United States, state, provincial, or territory governments have been responsible for the administration of vaccines, whereas in the United Kingdom, the deployment of vaccines is a responsibility that sits with the Department of Health and Social Care, working with the...
National Health Service (NHS) England and Public Health England [55]. Vaccination deployment in the rest of the United Kingdom is managed by the health services in each nation: NHS Wales, NHS Scotland, and Health and Social Care Northern Ireland [55].

We sought to better understand the narratives following advisory body statements regarding pediatric COVID-19 vaccines. Therefore, we reviewed official media releases of government advisory bodies regarding COVID-19 vaccinations for children aged 5 to 11 years published on the websites of these organizations. We then aimed to understand the mass media’s responses to these official media releases by advisory bodies.

We searched the Factiva database, a current international news database with >30,000 news sources provided by Dow Jones, to obtain media articles for the 7-day period following 4 different announcements by advisory bodies in the 4 study countries (8 data collection segments). These announcements were (1) approval of the COVID-19 vaccines for children aged 5 to 11 years by the Food and Drug Administration (FDA), the Therapeutic Goods Administration (TGA), Health Canada, and the Medicines and Healthcare products Regulatory Agency (MHRA; United States, Australia, Canada, and the United Kingdom, respectively); (2) when vaccinating children aged 5 to 11 years was recommended or scheduled by the Centers for Disease Control and Prevention (CDC), the Australian Technical Advisory Group on Immunisation, the National Advisory Committee on Immunization (NACI), and the Joint Committee on Vaccination and Immunisation; (3) when countries started vaccinating children aged 5 to 11 years; and (4) the Omicron announcement by the WHO (November 26, 2021; Table 1; Figure 3).

We limited the searches to the top 5 media outlets in each country; refer to Multimedia Appendix 1 [56-69] for further details. The search strategy used for each of the 4 phases is detailed in Textbox 2. Duplicate articles were removed.

### Table 1. Country-specific advisory and regulatory bodies and timeline.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Country</th>
<th>Canada</th>
<th>United Kingdom</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Organization</td>
<td>TGA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NACI&lt;sup&gt;b&lt;/sup&gt;</td>
<td>MHRA&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Approval date</td>
<td>December 5, 2021</td>
<td>November 19, 2021</td>
<td>December 22, 2021</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Organization</td>
<td>ATAGI&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Health Canada</td>
<td>JCVI&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Approval date</td>
<td>December 9, 2021</td>
<td>November 19, 2021</td>
<td>December 22, 2021</td>
</tr>
<tr>
<td>Phase 3</td>
<td>Date of vaccination rollout</td>
<td>January 10, 2022</td>
<td>Last week of November, 2021&lt;sup&gt;i&lt;/sup&gt;</td>
<td>January 31, 2022</td>
</tr>
<tr>
<td>Phase 4</td>
<td>Date of Omicron announced</td>
<td>November 26, 2021</td>
<td>November 26, 2021</td>
<td>November 26, 2021</td>
</tr>
</tbody>
</table>

<sup>a</sup>TGA: Therapeutic Goods Administration.

<sup>b</sup>NACI: National Advisory Committee on Immunization.

<sup>c</sup>MHRA: Medicines and Healthcare products Regulatory Agency.

<sup>d</sup>FDA: Food and Drug Administration.

<sup>e</sup>ATAGI: Australian Technical Advisory Group on Immunisation.

<sup>f</sup>JCVI: Joint Committee on Vaccination and Immunisation.

<sup>g</sup>CDC: Centers for Disease Control and Prevention.

<sup>h</sup>Initially only approved for clinically susceptible children or children living with household members who are immunocompromised. On February 16, 2022, JCVI extended the offer of COVID-19 vaccination to all children aged 5 to 11 years.

<sup>i</sup>Varied by province.
Figure 3. Timeline for the Pfizer-BioNTech COVID-19 vaccination in children aged 5 to 11 years in Australia, Canada, the United Kingdom, and the United States (2021 and 2022). *Only clinically susceptible children.

Textbox 2. Factiva search strategy.

- [Shot* OR vaccine OR immunisation OR immunization OR jab*] AND
- [child OR children OR paediatric OR pediatric OR five year OR eleven year OR 5 OR 11] AND
- [COVID OR COVID-19 OR CORONAVIRUS] AND
- United States
  - [“Food and Drug Administration” OR “FDA” OR “Centers for Disease Control and Prevention” OR “CDC”]
- Australia
  - [“Australian Technical Advisory Group on Immunisation” OR “ATAGI” OR “Therapeutic Goods Administration” OR “TGA”]
- Canada
  - [“NACI” OR “National Advisory Committee on Immunization” OR “Health Canada”]
- United Kingdom
  - [MHRA OR “Medicines and Healthcare Products Regulatory Agency” OR “JCVI” OR “Joint Committee on Vaccination and Immunisation”]

Inclusion and Exclusion Criteria

Media articles eligible for inclusion were those produced by media organization employees. To be included in further analysis, media articles were required to reference COVID-19 vaccines among children aged 5 to 11 years, the country-specific regulatory bodies, at least 1 character (hero, villain, or survivor), and a plot (in accordance with the Narrative Policy Framework [NPF] requirements); to have been published within 7 days after the event of our 4 phases detailed in Textbox 2; and to be written in English. Articles were excluded if they were based on countries other than Australia, Canada, the United Kingdom, and the United States or if they did not mention COVID-19 or vaccines for children aged 5 to 11 years.

Analysis Framework

This study applies the NPF to compare and contrast the narratives deployed by the media as well as by advisory and
regulatory bodies approving and recommending COVID-19 vaccinations for pediatric populations in Australia, Canada, the United Kingdom, and the United States and how narratives regarding the COVID-19 vaccines for children changed upon arrival of the Omicron variant. The NPF is a theoretical framework that specifies common assumptions, concepts, and hypotheses for the study of policy narratives [70] and provides guidance on how to conduct empirical research on the role of said narratives in the policy process [71]. The NPF holds that narratives have distinct components including settings (bounded by time, geography, or other characteristics), characters (heroes, villains, and survivors that may either be humans, places, or things), plots (which connect the characters to the setting), and morals (policy solutions) [46,72].

Ethical Considerations
We did not have ethics approval for this study as no personal data was used.

Results
Overview
In total, 400 articles met the selection criteria (n=287, 71.8% in the United States, n=40, 10% in Australia, n=60, 15% in Canada, and n=13, 3% in the United Kingdom). The NPF was used to guide our qualitative analyses of these data to explore the framing of these media narratives about the COVID-19 mRNA vaccines for children aged 5 to 11 years. These articles were organized according to phases, which we have denoted as follows:

- Phase 1: announcement of approval by the safety governing body (FDA, TGA, NACI, and MHRA)
- Phase 2: announcement of COVID-19 vaccine approval for pediatrics aged 5 to 11 years by the country’s public health advisory body (CDC, Australian Technical Advisory Group on Immunisation, Health Canada, and Joint Committee on Vaccination and Immunisation)
- Phase 3: the rollout of vaccinating children aged 5 to 11 years
- Phase 4: the announcement of the Omicron variant (Table 1).

Phases 1 and 2: Announcement of Approval by the Safety Governing Body (FDA, TGA, NACI, and MHRA)
There was excitement at the prospect of the COVID-19 pandemic “ending” by the US media following the CDC’s decision to approve vaccinating children. “As a mother and as a physician I know parents, caregivers, school staff and children who have been waiting for today’s authorization vaccinating younger children against COVID-19, bring[ing] us closer to returning to a sense of normalcy” [73], with vaccinating children the end to worrying about doing usual activities, “it’s a total game changer for parents and families...so many parents who’ve been worried about going to work because we’re afraid of bringing back COVID to our unvaccinated children...worry about our kids going to school...surrounded by others who are not masked and may not be taking the same precautions in their homes” [74]. The US media narrative repeatedly emphasized that “vaccines will be the nail in the coffin to the pandemic” [75]. There was also repeated reassurance to parents that mandates would not be enforced, “I’m not using the M-word!,” despite acknowledging that “we already mandate vaccines for our kids, if they to go to public school” [75] because of parental concerns, “This is only emergency use. That sounds experimental to me. You’re not going to guinea pig my kid. And we’ve never really done anything like this before” [75].

In Australia, Canada, and the United States, the media accentuated the role of families protecting susceptible individuals from COVID-19 as “vaccinating children can also help reduce community transmission and help prevent children passing the virus on to younger siblings, grandparents and the wider community” [76], and the importance of vaccinating children who are susceptible to diseases, “If your child is obese, has any lung disease, heart disease, immunosuppression, this is a no-brainer. You need to get vaccinated now” [75]. In the United Kingdom, there were conflicting reports about the importance of vaccinating children to offer them protection, while at the same time considering widespread vaccination of children a “distraction” from adult vaccination efforts, “Even if the JCVI had recommended extending the age range this would not be a priority for us right now. There is a danger that if you ask the NHS to do two different things at once you lose focus and that is exactly what we don’t want to happen” [76].

The US media regarding the COVID-19 vaccines for children aged 5 to 11 years emphasized the importance of attending school and being with other children with vaccines helping, “kids can go back to something that’s better than being locked at home on remote schooling, not being able to see their friends” [77], as did Canada’s media, writing that “the virus has disrupted the school year and children’s activities” [78], and the UK media, writing that “We also now have concrete evidence on the harms of children being out of school, which we must balance against the risk of harms from COVID [79],” while maintaining that “The majority of children aged five to 11 are at very low risk of serious illness due to COVID-19” [57], and therefore vaccinating and providing boosters to adults remained the priority.

During phase 1, the US media message changed from needing to get children vaccinated for herd immunity and needing to get a pediatric COVID-19 vaccination approved to end the pandemic to convincing parents that vaccination is safe and that there is “no evidence that the vaccine can lead to loss of fertility” [58]. The message to “trust your regulatory bodies” and the advice that the vaccines are safe for children was also common in Australia, “We understand parents have questions about the vaccines, especially around vaccine safety, and they need to have the opportunity to have these addressed and to access clear, trustworthy information to guide their decision-making” [59]. In particular, the initial concern and barrier to having Pfizer approved for children was “the risk of myocarditis,” but “there were no cases of myocarditis in the children’s trial,” which was key to the vaccine being approved for this age group [60]. Australian media depicted children of antivaccination parents as the survivors, “confronted with that predicament when her ex-partner started to speak out against vaccinations...Before
that, the Canberra mother said she had not known her former partner to be opposed to vaccines” [61]. Similarly, in the UK media, trust in authorities was further reinforced, “Parents and carers can be reassured that no new vaccine for children would have been approved unless the expected standards of safety, quality and effectiveness have been met” [57].

US and Canadian media also focused on the winter weather, “deaths nationally have been ticking up over the past few weeks, some rural hospitals are showing signs of strain, and cold weather is setting in” [80], leading to increased COVID-19 and influenza infections, with this seen as encouragement to vaccinate children early given the expected rise in COVID-19 infections during winter.

**Phase 3: Week After Vaccine Rollout in Children**

Canadian and US media stories shifted to convincing parents (with the help of quoted pediatricians) that children need to be vaccinated to help children get back to normalcy, “beyond the clinical impact...there have been detrimental social and mental health impacts that we are just beginning to fully understand. For almost two full years, school has been fundamentally changed...Tragically, COVID-19 was among the top 10 leading causes of death for children” [81]. Canada rolled out their COVID-19 vaccines for children aged 5 to 11 years a few weeks later than the United States, and the Canadian media often referred to the vaccination success of the United States, “For many Canadians, spring was a dark period marked by surging coronavirus infections, lockdowns – and the envy of watching their American neighbours get vaccinated en masse” [82]. In contrast, the Australian media narrative was predominantly expressing anger against the Morrison (federal) government for the late purchase and arrival of vaccines, and then supply chain issues causing cancelations of appointments for children’s vaccines, “My grandchildren were infected two weeks before they could get a vaccine. This need not have happened. It should not have happened. That it did is the result of the chronic incompetence of Prime Minister Scott Morrison” [62]. In the Australian media, general practitioners were painted as heroes, with media articles showing decorated offices in preparation for children’s vaccinations, but then delays to children’s vaccine appointments because “many GPs [were] unable to access adequate supplies or cope with demand, with some practices reporting receiving only 50 to 100 doses a week” [83]. Similarly, in the United Kingdom, frustration was growing as cases among children peaked owing to Omicron and children faced disruptions to schooling, “There’s a strange disconnect between the lack of mitigations and restrictions, and the lived experiences of families and schools right now” [63].

In countries with different jurisdictions, the media compared the vaccine rollouts of states or provinces. For example, in Canada, the Leader of the Opposition criticized Ontario for being too slow while British Columbia was doing much better, “provinces are taking different approaches,” with the politician noting that in “British Columbia, generally, there will not be in-school clinics during school hours” whereas “in Ontario...vaccines are being given in community clinics, hospitals and pharmacies, as well as school clinics” [64]. This was true in the United States too, with California imposing “the nation’s first coronavirus vaccine mandate for schoolchildren,” whereas “Texas Gov. Greg Abbott issued an executive order stating that no entity in Texas can mandate getting a COVID-19 vaccine,” leading to different vaccination rates across states [84]. The United Kingdom did not have the issue of discordant policies between states, but there was resistance to vaccinations taking place in primary schools, with most vaccinations taking place “at local vaccination centres or community pharmacies outside of school hours” [85]. Requiring vaccine passports was also a heated topic in the United States [86] and Canada [65], further contributing to the politicization of COVID-19.

The US media portrayed an emotive picture of brave young children lining up to get vaccinated and wanting to help the country in its fight against COVID-19, “Thousands got their first kids-size shots yesterday. And luckily they didn’t have to do it alone...comfort dogs were seen working hard to keep kids calm as they rolled up their sleeves” [87]. This de facto militarization of society against COVID-19 was echoed in Australian media, where a “command and control” military-style scale-up of the “beleaguered vaccination rollout” occurred [88]. In Canadian, UK, and US media, children were depicted as the survivors of the COVID-19 pandemic because of school closures, high case numbers, and family disruptions. Parents of susceptible children expressed relief and gratitude that their kids could get vaccinated in the United Kingdom and the United States, “On my Christmas list I said the COVID vaccine,” said Adeline Gieng, aged 11 years. “And then I was just driving in the car and I was like ’Oh my gosh, it’s finally here! We can get the vaccine!’” [89].

Canadian media was less emotive than the other countries and quoted a confused public experiencing information overload, “Health Canada COVID-19 update...I was once again struck by the conflicting information being dispensed by the various authorities. There was a discussion around the intervals between vaccine doses for children; one group advocated three weeks while the other advised eight. The reason given was that each group looked at different research. Issuing these conflicting opinions only reinforces folks to do their own research and come up with their own treatment regimes, which, in many cases, does not include a vaccine” [90]. There was also reassurance provided to parents that “children tend to have robust immune responses, so reducing the dosage ensures kids will have high levels of protection while minimizing the risk of side effects” [91]. The Canadian media repeatedly suggested that parents needed to talk to their pediatrician, “Of course, pediatricians are also happy to answer questions about vaccination” [66], echoing the US media message, despite Canadians often having a family general practitioner instead of a family pediatrician.

**Phase 4: Week After the Omicron Variant Announced**

Although the arrival of the Omicron variant meant that vaccines had reduced effectiveness against COVID-19 infection and against hospitalization and death for adults [92], there was less media coverage regarding the COVID-19 vaccines for children aged 5 to 11 years after Omicron arrived in comparison with the other 3 phases. In all 4 countries, Omicron (and any other new variants to come) was portrayed as the villain, posing an...
ongoing risk to society returning to normal. Nonetheless, reassuring messages were broadcasted, “We have more tools today to fight the variant than we’ve ever had before, from vaccines to boosters to vaccines for children, 5 years and older and much more” [93], with children wanting to go to school and families wanting to celebrate the December holidays together.

The Canadian media, and to a much lesser degree the US media, discussed low- and middle-income countries with limited vaccine availability. “When Cambodia rolled out COVID-19 vaccines, lines stretched down entire streets and people left their shoes out to save their places as they sheltered from the sun. But three months into its campaign, just 11 per cent of the population had received at least one dose” [94], compared with high-income countries with vaccine oversupply and wastage. “An informal survey shows that at least one million doses of Canada’s COVID-19 vaccine supply have gone to waste” [95]. However, the paradox in Canada was noted, “several people who have made a conscious decision on moral or ethical grounds not to get a booster, and many others...who have already foregone third doses given that large swaths of the world remain unvaccinated...[but] nobody will say, ‘I don’t want to have my cancer chemotherapy because somebody in a low or middle-income country isn’t going to get their chemotherapy if I get mine’.” [67].

There was cross-country comparison by the media; in Canada, mostly with the United States; in the United Kingdom, mostly with the United States and Europe; and in Australia, mostly with the United Kingdom. Initially, the United States and Europe had higher vaccination rates in children compared with Canada, the United Kingdom, and Australia, and these countries looked to vaccinated children in the United States as real-world evidence that vaccines are safe, “Dr. Eileen de Villa, the city’s medical officer of health, told reporters ‘more than three million children have been vaccinated in the United States so far, and...there have been no major signals of concern’.” [68]. In Canada and the United States, there was substantial confusion expressed by parents regarding information overload, and in Canada, there were mixed messages about COVID-19 policies, “The vaccine promotion effort would be on considerably firmer ground, surely, if its leading lights hadn’t spent the last 20 months playing the entire country for fools - dressing up ‘operational considerations,’...as science and expecting us not to notice” (Table 2) [69].
<table>
<thead>
<tr>
<th>Phase 1 and 2</th>
<th>Australia</th>
<th>Canada</th>
<th>United States</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Similarities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protect susceptible individuals</td>
<td><em>a</em></td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Children deserve to go to school, and the vaccine will help make this a possibility</td>
<td>—</td>
</tr>
<tr>
<td>Excitement regarding “ending” the pandemic with vaccinating children</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>Focus on winter weather and expected rise in COVID-19 outbreaks</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Trust your regulatory bodies</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Trust your regulatory bodies</td>
</tr>
<tr>
<td><strong>Phase 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of vaccine rollouts in different states or provinces</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Parents of susceptible children expressed relief that their children can get vaccinated</td>
<td>—</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Children portrayed as the survivors of COVID-19 outbreaks because of school closures</td>
<td>—</td>
</tr>
<tr>
<td>Militarization of terms regarding COVID-19</td>
<td>—</td>
<td>—</td>
<td>Militarization of terms regarding COVID-19</td>
<td>—</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Parents encouraged to talk to their pediatrician about their concerns regarding vaccinating children</td>
<td>—</td>
</tr>
<tr>
<td><strong>Phase 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>Discussed low- and middle-income countries with less vaccines available</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Comparison with other countries’ more successful pediatric vaccine uptake</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Comparison with other countries’ pediatric vaccine uptake</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Substantial confusion because of information overload</td>
<td>—</td>
</tr>
<tr>
<td><strong>Differences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phase 1 and 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children of antivaccination parents portrayed as the survivors</td>
<td>—</td>
<td>—</td>
<td>Children of antivaccination parents portrayed as survivors</td>
<td>Vaccinating children described as a distraction from adult COVID-19 vaccination efforts</td>
</tr>
<tr>
<td><strong>Phase 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPs b as heroes</td>
<td>Referenced the pediatric vaccination success in the United States</td>
<td>Anger about the possibility of requiring pediatric vaccine passports</td>
<td>Anger of parents with ongoing schooling disruptions owing to rising COVID-19 cases</td>
<td></td>
</tr>
<tr>
<td>Anger against the Morrison (federal) Government for the late purchase and arrival of vaccines and cancelation of pediatric vaccine appointments</td>
<td>Less emotive than other countries and expressed confusion being felt by society because of information overload</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

Media narratives regarding the COVID-19 vaccine for children aged 5 to 11 years were analyzed for 4 high-income countries using the NPF. Although epidemiological models track the progression of pediatric COVID-19 infection and vaccination rates quantitatively, our analysis captures the richness of the news media portrayals and the stories behind the numbers. The media plays an important role in sharing information, shaping opinions, and ultimately influencing behavior. Analyzing media narratives offers unique clues regarding the uptake of vaccines.

The United States was the first country to vaccinate children against COVID-19, with the American Academy of Pediatrics recommending in August 2021 that the FDA speed up its approval of the Pfizer-BioNTech mRNA vaccine for children, assessing that the benefits of vaccinating this population would outweigh the risks. The primary barrier to the vaccine’s approval was concern regarding the risk of myocarditis; however, the clinical trial investigating the Pfizer-BioNTech COVID-19 vaccine in children aged 5 to 11 years found that no children developed myocarditis [60]. This led to the FDA issuing an Emergency Use Authorization for the mRNA COVID-19 vaccine for children aged 5 to 11 years, followed by Canada, Australia, and the United Kingdom approving the vaccine for this use. All countries emphasized the importance of getting children back to school and herd immunity and that vaccines would help achieve this. Initially, Australia was greatly affected by a lack of vaccine supply for adults and children, and the scarcity of vaccines was a common theme, with much blame directed at the Australian Government for its inability to deliver. However, at the same time, the Australian media (like the US media) voiced patriotism and portrayed COVID-19 as a war to win; imagery that was less prevalent in the Canadian or UK media. In contrast to patriotism, Canada had frequent media discourse regarding disadvantaged countries’ lack of access to vaccines. The United Kingdom was the last country to approve and rollout Pfizer vaccination for children, only starting to vaccinate susceptible children aged 5 to 11 years from January 30, 2022, and then all children from February 16, 2022, despite the UK medical regulator (MHRA) approving the vaccine for this age group on December 22, 2021 [96]. This decision was said to be made to avoid distracting from the adult booster vaccination efforts. Later, the media discussed the narrative of parents in the United Kingdom, particularly parents of susceptible children who were frustrated with this delay.

Television media indicated that “anti-vaxxer” frustration was rife in the United States, and similar frustration was seen in Australia through in-person demonstrations in the streets. Vaccine misinformation was (and remains) widely available. For example, the top Amazon book searches in 2021 were about how vaccines are ineffective [97], and theories on the “great reset” have gone viral [98]. However, the US mainstream print media did not report these messages in any detail. We did not find a single media article in the 4 countries studied during our study period that reported a case of severe outcome or death following the COVID-19 vaccine. We understand that reporting severe outcome or death from COVID-19 vaccines was done on social media and is thought to have had a large influence on vaccine uptake [99]. The relatively low vaccine uptake in the United States has been attributed to the subpopulations being “anti-vax” or against vaccination.

Analyses of less regulated social media (eg, TikTok and Twitter [subsequently rebranded as X]) are likely to yield different narratives. The mainstream media is built upon the premise of reporting verifiable facts and providing sources for these facts. This practice is believed to enhance the credibility of the media outlet, attracting a larger audience and consequently generating income. In contrast, social media have fewer incentives to operate in this manner and exhibit significantly less accountability [100]. According to a 2018 report, 79% of UK adults received their news from television, 64% from the internet, and 44% from social media [101]. Data from the United States in 2021 show that about half of the population receives their news from social media and the other half from traditional news outlets [102]. In Australia, a 2021 study suggests that 23% receive their news from social media [103]. In Canada, in 2021, although 52% said they read news on social media, less than one-third trusted it to be accurate [104]. Features distinct to social media, such as likes, retweets, and shares, build an ideological echo chamber with the same piece of real or fake news recirculating [105]. Other studies are exploring the role of social media narratives and their impact on COVID-19 vaccine uptake [106]. Although research into the impact of social media on COVID-19 vaccine uptake is ongoing, research into mainstream media narratives continues to be important, with the majority, or half, of the population still consuming mainstream media.

By the time Omicron was spreading globally, there was significant public fatigue about COVID-19 restrictions as well as COVID-19 information [107], and this was in part evident through significantly less media coverage of COVID-19 during this period. The media narratives of the United States and Canada also shifted to expressing increasing frustration that some people did not play their part in getting society out of this pandemic, with some saying that a key step to returning to “normal living” included the vaccination of children. The Australian media voiced this sentiment later as well, in part owing to an initial lack of Pfizer vaccine for this age group, and
this sentiment was not evident in the UK media, where the vaccine for children was only recently (February 16, 2022) recommended for all children. There were some questions about whether current vaccines would protect against the Omicron variant in US media; however, this did not seem to be a theme in other countries’ narratives surrounding vaccinating children, likely because the United States authorizing the use of vaccines for children before the knowledge that Omicron was less severe and before data on the vaccines’ effectiveness against Omicron became available.

Comparison With Prior Work
Our study contributes to the COVID-19 social science literature in the following ways: most existing COVID-19 studies on vaccinating children have focused on either the safety or efficacy of the vaccines in 5–11-year-old children or parents’ thoughts and intentions regarding vaccinating their children. This study is the first to examine the media narrative surrounding COVID-19 vaccines for children in 4 different, high-income, English-speaking countries [1]. We also dynamically tracked web-based media narratives for 4 countries following the regulatory approval of the vaccines for children aged 5 to 11 years, the rollout of vaccines, and the announcement of the Omicron variant. Given the fast-moving nature of the pandemic globally and the flurry of evolving government interventions locally, our analysis over 4 months provides a systematic picture of how media narratives in the news have shifted during the pandemic and between countries with respect to vaccination for children.

On the basis of our knowledge about media narratives, different narratives will influence the COVID-19 vaccine uptake. Pediatric vaccinations for COVID-19, as well as other pediatric vaccinations, will continue to be rolled out, and narratives in the media will continue to play an important role in vaccine uptake. The nuanced nature of the benefits of pediatric COVID-19 vaccines compared with other childhood vaccines—where the data on benefits and risks have had more time to accrue—has made the messaging of pediatric COVID-19 vaccines more challenging, and indeed, this was observed in the UK media. To date, Australian media is endorsing vaccine uptake more uniformly than US media, and the Australian pediatric uptake of the COVID-19 vaccine (adjusted for population) was more rapid than the United States’ uptake, albeit now stalled at approximately 50% [108-116].

Limitations
This study has some limitations. First, we did not seek to analyze the media narratives of non–English-speaking countries and focused only on 4 countries. Further studies might explore media narratives on COVID-19 vaccines in non–English-speaking countries and vaccination rates to understand the potential influence of the media on vaccination uptake in different settings. Second, we only included studies that mentioned advisory and regulatory bodies, as we aimed to understand how the media reported on the decisions made by these bodies as well as how the media reported on professional and public opinion following these decisions. Further research could explore other themes regarding pediatric vaccines in the media. Third, we analyzed only traditional media narratives and did not explore social media narratives, which may have offered different narratives. Further research could examine social media narratives in response to advisory and regulatory bodies that approve pediatric vaccinations.

Conclusions
Analysis of the hero, villain, survivor, and plot of 400 articles on COVID-19 vaccination in children shows that the media coverage of this pandemic relies on interpersonal narrative stories. Public health emergencies require clear; compelling; and, above all, accurate communication. The stories told during this pandemic are compelling because they contain the classic elements of a narrative; however, they can be reductive and inaccurate. The media analysis presented in our paper is significant. It assists in informing how health policy and guidelines are presented to the public in the mainstream media, particularly as the science behind those policies and guidelines is unfolding in real time. Understanding the role of the mainstream media in health communications will remain crucial as we continue through the COVID-19 pandemic as well as for future health crises.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Sources for top 5 media outlets per country.
[DOCX File, 14 KB - formative_v8i1e38761_app1.docx]

References
2. Infodemic. World Health Organization. URL: https://www.who.int/health-topics/infodemic#tab=tab_1 [accessed 2022-03-18]


24. Chadwick et alJMIR FORMATIVE RESEARCH

https://jmirformative.research/2021/1/e38761


44. Measles, mumps, and rubella. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/fastats/measles.htm [accessed 2023-08-01]


49. Cumulative confirmed COVID-19 cases per million people. Our World in Data. URL: https://tinyurl.com/vcmdcy5 [accessed 2024-02-26]


63. Devlin H, Adams R. One in eight primary pupils in England had Covid last week, figures show. The Guardian. 2022 Feb 4. URL: https://tinyurl.com/5n895v8i [accessed 2024-02-26]


66. Parents, it’s now up to you. COVID vaccines are right for kids and everyone else. Toronto Star. 2021. URL: https://www.thestar.com/opinion/editorials/parents-it-s-now-up-to-you-covid-vaccines-are-right-for-kids-and-everyone/article_43b4eb64-9493-52d6-88a4-c978363704d0.html [accessed 2024-01-01]


76. Lay K, Blakely R. Covid vaccine to be offered to third of a million vulnerable 5-to-11-year-olds. The Times. 2021 Dec 22. URL: https://www.thetimes.co.uk/article/covid-vaccine-to-be-offered-to-third-of-a-million-vulnerable-5-to-11-year-olds-n0dqns7h8 [accessed 2024-02-26]


82. Cecco L. ‘I knew supply was coming’: how Canada’s push for COVID vaccines paid off. The Guardian. URL: http://www.theguardian.com/lifeandstyle/2021/nov/03/i-knew-supply-was-coming-how-canadas-push-for-covid-vaccines-paid-off [accessed 2024-02-26]


111. Coronavirus (COVID-19) vaccinations. Our World in Data. URL: [https://ourworldindata.org/vaccinations] [accessed 2021-06]


113. COVID-19 vaccination coverage in Canada. Government of Canada. URL: [https://health-infobase.canada.ca/covid-19/vaccination-coverage#a5] [accessed 2024-02-26]


115. Vaccinations in Wales. United Kingdom Government. URL: [https://coronavirus.cymru.gov.wales/details/vaccinations?areaType=nation&areaName=Wales] [accessed 2024-02-26]

116. Vaccinations in Scotland. United Kingdom Government. URL: [https://coronavirus.scottishgovernment.gov.uk/details/vaccinations?areaType=nation&areaName=Scotland] [accessed 2024-02-26]

Abbreviations

CDC: Centers for Disease Control and Prevention
FDA: Food and Drug Administration
MHRA: Medicines and Healthcare products Regulatory Agency
MMR: measles, mumps, and rubella
NACI: National Advisory Committees on Immunization
NHS: National Health Service
Examining the Gateway Hypothesis and Mapping Substance Use Pathways on Social Media: Machine Learning Approach

Yunhao Yuan¹, MSc; Erin Kasson², MSW, MS; Jordan Taylor³, MS; Patricia Cavazos-Rehg², PhD; Munmun De Choudhury⁴, PhD; Talayeh Aledavood¹, PhD

¹Department of Computer Science, Aalto University, Espoo, Finland
²School of Medicine, Washington University in St. Louis, St. Louis, MO, United States
³Carnegie Mellon University, Pittsburgh, PA, United States
⁴Georgia Institute of Technology, Atlanta, GA, United States

Corresponding Author:
Yunhao Yuan, MSc
Department of Computer Science
Aalto University
P.O. Box 11000 (Otakaari 1B)
FI-00076 AALTO
Espoo, FI-00076
Finland
Phone: 358 509113635
Email: Yunhao.Yuan@aalto.fi

Abstract

Background: Substance misuse presents significant global public health challenges. Understanding transitions between substance types and the timing of shifts to polysubstance use is vital to developing effective prevention and recovery strategies. The gateway hypothesis suggests that high-risk substance use is preceded by lower-risk substance use. However, the source of this correlation is hotly contested. While some claim that low-risk substance use causes subsequent, riskier substance use, most people using low-risk substances also do not escalate to higher-risk substances. Social media data hold the potential to shed light on the factors contributing to substance use transitions.

Objective: By leveraging social media data, our study aimed to gain a better understanding of substance use pathways. By identifying and analyzing the transitions of individuals between different risk levels of substance use, our goal was to find specific linguistic cues in individuals' social media posts that could indicate escalating or de-escalating patterns in substance use.

Methods: We conducted a large-scale analysis using data from Reddit, collected between 2015 and 2019, consisting of over 2.29 million posts and approximately 29.37 million comments by around 1.4 million users from subreddits. These data, derived from substance use subreddits, facilitated the creation of a risk transition data set reflecting the substance use behaviors of over 1.4 million users. We deployed deep learning and machine learning techniques to predict the escalation or de-escalation transitions in risk levels, based on initial transition phases documented in posts and comments. We conducted a linguistic analysis to analyze the language patterns associated with transitions in substance use, emphasizing the role of n-gram features in predicting future risk trajectories.

Results: Our results showed promise in predicting the escalation or de-escalation transition in risk levels, based on the historical data of Reddit users created on initial transition phases among drug-related subreddits, with an accuracy of 78.48% and an F₁-score of 79.20%. We highlighted the vital predictive features, such as specific substance names and tools indicative of future risk escalations. Our linguistic analysis showed that terms linked with harm reduction strategies were instrumental in signaling de-escalation, whereas descriptors of frequent substance use were characteristic of escalating transitions.

Conclusions: This study sheds light on the complexities surrounding the gateway hypothesis of substance use through an examination of web-based behavior on Reddit. While certain findings validate the hypothesis, indicating a progression from lower-risk substances such as marijuana to higher-risk ones, a significant number of individuals did not show this transition. The research underscores the potential of using machine learning with social media analysis to predict substance use transitions. Our results point toward future directions for leveraging social media data in substance use research, underlining the importance of continued exploration before suggesting direct implications for interventions.
Introduction

Background

Substance misuse has become a major public health challenge around the globe, leading to a considerable health burden and substantial financial loss. For example, in the United States, drug-involved overdoses driven by fentanyl and other opioids resulted in the loss of nearly 106,000 lives in 2021 [1].

Understanding the dynamic transitions between types of substances used and the timing of transitions to polysubstance use is vital for developing effective prevention, harm reduction, and recovery strategies. Since the 1970s, researchers have explored the gateway hypothesis, a theory suggesting a structured, sequential pattern of drug involvement among individuals, often starting with the use of substances such as alcohol, tobacco, or cannabis and potentially progressing to the use of more potent substances with higher associated risks [2]. Despite the long-standing prevalence of the gateway hypothesis in the literature, recent studies [3,4] have begun to scrutinize its validity. While the gateway hypothesis may explain specific substance use patterns, it does not adequately capture the “causal” or “direct” relationships. Moreover, it is not able to describe complex and multidirectional patterns of substance use. Nevertheless, this does not diminish the significance of the gateway framework. At the macrolevel, patterns have been identified where higher-risk substance use is often preceded by lower-risk substance use [5]. While many individuals might begin their substance use journey with alcohol or marijuana, it does not imply an inevitable progression to substances like opiates. Most users of marijuana do not escalate to using opiates [6]. However, this raises important questions about what microlevel factors differentiate substance use escalation and de-escalation. More information is needed to understand these microlevel risk factors associated with substance use escalation and de-escalation. Identifiable risk factors exist among groups that engage in substance use, potentially increasing the likelihood of transitioning from one substance to another, and the gateway hypothesis framework could be evaluated and expanded upon given new sources of data (ie, social media). Given the complex nature of substance use initiation and people’s highly individualized journeys, understanding common trajectories of substances used is essential for informing substance use intervention programs and for preventing escalation to misuse behaviors and substances associated with higher risk (ie, overdose).

There has been extensive work on developmental stages in substance use involvement using survey data [7-9]. For instance, taking cannabis as a gateway drug case study, recent research from the National Institute on Drug Abuse in the United States confirms that marijuana use typically precedes harder drug use, but most marijuana users do not escalate to harder drugs [6]. Moreover, according to a study by McCarthy [10], 1 in 8 Americans report smoking marijuana [10]. This suggests that the current interventions aiming to prevent substance use escalation targeted at those who use marijuana may not be sufficiently granular. Furthermore, much of the research to date is either cross-sectional or focused on a single substance rather than across substances with variance in perceived risk. To fill this gap, our study aimed to investigate substance use trajectories on social media to improve our understanding of transitions between types of substances, factors associated with the likelihood of progression to higher-risk substances, and the linguistic patterns associated with escalating and de-escalating.

Social media platforms such as X (previously known as Twitter), Reddit, and YouTube have created spaces where individuals can engage with each other and share various parts of their private lives, including their personal experiences and life events [11-13]. This is especially true among individuals who misuse substances, as networking with others with lived experiences has been shown to decrease self-stigma and encourage safer substance use behaviors and recovery efforts [14]. However, it should be noted that the increased visibility and influence of these platforms might inadvertently normalize substance misuse and provide easier access to information related to harmful substances, thus potentially promoting harmful substance misusing behaviors [15,16]. Nevertheless, the data from social media are invaluable for research. The information gathered from social media platforms is timely and relevant for studying substance use characteristics longitudinally [17-19]. As a result, the data obtained from social media offer a unique opportunity to observe substance use behaviors within a broad population, enabling a deeper comprehension of the underlying patterns and factors influencing substance use behaviors.

Several studies have used social media data to examine different aspects of substance-related behaviors, such as adverse drug reactions [20,21], substance misuse [22,23], and substance use–related slang [24]. Lu et al [25] used machine learning techniques to estimate Reddit users’ shifts from casual drug discussion forums to drug recovery forums. They further developed a survival model to estimate the probability of a Reddit member posting in a recovery forum within the upcoming year [26]. Another work developed a language model of opioid consumption and generated alternative words for opioids, routes of administration, and drug tampering [27]. One study [28] used transfer learning techniques to detect Reddit posts that indicate opioid recovery. With a combination of deep learning and human annotation techniques, they identified and analyzed the effectiveness of particular alternative drugs for treating opioid use disorder (eg, kratom). Relatedly, another study [29] used social media content related to the misuse of fentanyl to build a machine learning model to identify risky discourse and develop a vocabulary set consisting of community-specific and colloquial terms associated with fentanyl and its analogs [29]. However, few studies have assessed the gateway hypothesis using social media data. Furthermore, to our knowledge, none of the existing studies have assessed the gateway hypothesis using social media data. Furthermore, to our knowledge, none of the existing studies have assessed the gateway hypothesis using social media data.
studies have focused on identifying or understanding specific risky behaviors associated with the gateway hypothesis.

**Study Objectives**

In this work, we investigated substance use behaviors on social media, particularly the transitions between substances with different risk levels. We specifically addressed the following research questions (RQs):

- **RQ1:** Does substance use–related social media behavior support the gateway hypothesis?
- **RQ2:** How can we use social media to identify transitions in different risk levels of substances?
- **RQ3:** What specific linguistic cues in web-based behaviors can be identified using social media data to distinguish between escalating and de-escalating risk transitions?

To answer these RQs, we analyzed Reddit data to understand the trajectory of substance use: a Reddit user’s entire historical path of posts and comments within targeted substance-related subreddits, encoded by risk levels assigned to these subreddits. This risk level is a numeric indicator assigned to each subreddit based on the potential harm of the substance use behaviors discussed, ranging from discussions of nonharmful substances to those with a high potential for those with a high potential for substance use disorder and harm. We then examined transitions, defined as changes in risk levels from one time bin to another within a user’s trajectory, to identify potential early signs of risk level changes associated with substance use. A transition begins when we observe a series of time bins (groupings of weekly posts and comments) that have the same risk level. The transition is marked when a subsequent time bin displays a different risk level, indicating a change in the Reddit user’s posting and commenting with either a higher-risk (escalating transition) or a lower-risk (de-escalating transition) substance-related subreddit. This shift from one risk level to another defines the transition, capturing the Reddit user’s movement toward potentially different substance use behaviors.

We implemented a logistic regression model, a Bidirectional Encoder Representations from Transformers (BERT) model, and a Robustly Optimized BERT Approach (RoBERTa) model to identify potential escalation or de-escalation transition in risk levels. We show the potential of using deep learning methods with large-scale historical data to predict the escalation or de-escalation transitions of risk levels. Our linguistic analyses showed the most salient linguistic cues between escalating transitions and de-escalating transitions. These findings highlight the importance of language patterns within self-generated content related to substance use trajectories. They suggest the need for further research to refine these methodologies and the future applications to use social media data to analyze substance use behaviors to refine policies and therapeutic strategies targeting substance use disorders and recovery strategies.

**Methods**

**Data Source**

Past research has suggested that Reddit is a popular social media platform well suited for substance-related research due to its large number of users with diverse backgrounds, which facilitates longitudinal data analysis and insights into the evolving nature of substance use and discussions [30]. Various studies have indicated that engaging in substance-related discussions can be a strong marker of an individual’s experience with substance use [26,28,31]. For our data set, we focus on Reddit, a prevalent social media platform where individuals can anonymously participate in topic-centered communities known as “subreddits.” Each subreddit operates under its own set of rules to moderate posts and comments, offering unfiltered discussions with less interference from overarching platform policies, distinguishing Reddit as an ideal platform for analysis of substance-related conversations [30]. On Reddit, there are various subreddits dedicated to discussing different substances, such as cannabis, cocaine, and fentanyl, as well as behaviors associated with these substances, such as active misuse, harm reduction, and recovery.

Grounding our approach in psychiatric theory related to the developmental patterns of substance involvements [8,19,32-35] and the discussions within these substance-subrelated subreddits, we curated a list of relevant subreddits and annotated each subreddit based on the risk level associated with specific substances discussed. The rationale behind annotating at the subreddit level rather than the post level is the structural organization of Reddit itself. Subreddits are often named after or centered on a specific topic, suggesting that a majority of posts within a given subreddit, such as r/fentanyl, likely align with a particular risk profile related to fentanyl. Furthermore, subreddits are typically guided by a set of rules and guidelines enforced by moderators to ensure that the content remains relevant to the subreddit’s central theme. The presence of these regulations provides a level of consistency in the nature and risk associated with the content, making it more practical to assess and annotate at the subreddit level.

To create a comprehensive list of relevant subreddits, we used a methodology similar to the previous research [36]. We began by compiling an extensive lexicon of terms related to substance use, focusing on keywords that encompass various facets of the experience and treatment of substance use disorders. This included a range of drug names, incorporating substances available over the counter or through prescription or those considered illicit. For every generic drug name, we also incorporated the corresponding trade names and combination products, drawing upon insights from substance use research and from collaboration with coauthors specializing in clinical psychology. We found 152 keywords in this process. We then used these 152 keywords to query Reddit posts through BigQuery (Google). Our focus was not limited to the most popular or active subreddits; instead, we aimed to encompass a diverse representation of the substance use spectrum, including active use, harm reduction, and recovery. This process led to the curation of 73 drug-related subreddits.

Using a set of 73 drug-related subreddits, we collected all the posts and comments from Google Big Query [37]. Google Big Query is the publicly available data warehouse containing all the posts and comments made on Reddit. We used Google’s BigQuery application programming interface to query publicly available Reddit data from December 2015 to December 2019. With increasing privacy concerns and changes in data protection...
laws, researchers frequently find themselves limited to using dated historical data to gain insights from social media platforms. In the Discussion section, we revisit the limitation of the temporal restriction of our data set. In total, we collected 2,291,356 posts and 29,372,044 comments created by 1,426,621 Reddit users. We excluded 4321 authors with usernames beginning or ending with “bot” and deleted accounts. We also removed those with <1 year of posting and commenting activities in our collected data set to ensure we had enough historical data to conduct the analysis. After these steps, 313,846 authors remained in our data set.

Ethical Considerations

Although we use publicly accessible historical Reddit data from Google BigQuery that do not qualify for ethics board approval, we are committed to protecting the privacy of the data owners. Following the prior work on sensitive topics on social media 

[28,38,39], we have removed any personally identifiable information and have paraphrased all quotations in this paper to avoid traceability. The authors of this paper have backgrounds in both computer science and data analytics, as well as clinical psychology and addiction medicine. Such a diverse knowledge base enhances our understanding of the privacy considerations of this group and allows an interdisciplinary approach to understanding our findings.

Annotation Approach

The coauthors in this study, who are substance use domain experts, used inductive and deductive approaches to develop a risk level codebook based on substance type. First, substance use literature was reviewed, and a rough structure was developed for the codebook based on respective risks related to each substance (ie, the risk for overdose, dependence, risky routes of administration, etc). Next, coders systematically examined each of the 73 subreddits in our data set to understand the substances and the level of risk discussed in each community and to refine the categories in the codebook. This process involved 2 independent coders (Kevin Davet and Nina Kaiser) searching the subreddit name on Reddit, reviewing the first 25 posts and their associated comments listed in each subreddit, and then assigning a risk category to this subreddit based on the substance and substance use behaviors mentioned in this community. Interrater reliability was 96% for the initial risk category assigned by the 2 independent coders, and the remaining discrepancies were discussed among a group and resolved by a third coder (EK). Once a final codebook was established, coders worked to assign a risk level to each subreddit ranging from 0 to 4, with 0 denoting no risk and 4 indicating the highest level of risk. Multimedia Appendix 1 [40-49] provides the details on risk level definitions by substance used in this study.

The domain expert coauthors determined that 1 subreddit in our data set (r/supplements) did not constitute a risky substance subreddit. This was the only subreddit that was assigned to level 0. Due to the underrepresentation of the risk level 0, we excluded the subreddit (r/supplements) from our subsequent analysis. We also excluded 7 subreddits annotated as risk level mixed (r/drugtesthelp, r/DrugNerds, r/REDDITORSINRECOVERY, r/addiction, r/ReagentTesting, r/Drugs, r/drugsarebeautiful). These subreddits discussed multiple substances and different aspects of substance use behaviors that were placed in the high-risk category. In the end, we had the following data set for the subsequent analysis: risk level 1 had 3 subreddits, risk level 2 had 22 subreddits, risk level 3 had 8 subreddits, and risk level 4 had 32 subreddits (Textbox 1). Among these subreddits, we had 133,220 Reddit users with 9,205,558 posts and comments.

Textbox 1. Categorization of subreddits according to risk levels.

<table>
<thead>
<tr>
<th>Categorization</th>
<th>Risk level 0: r/supplements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk level 1: r/SMARTRecovery, r/secularsoibriety, r/AtheistTwelveSteppers</td>
<td></td>
</tr>
<tr>
<td>Risk level 2: r/TokeSpot, r/vaporents, r/CannabisExtracts, r/EntExchange, r/Kanna, r/GetOutAndVape, r/quittingkramot, r/AthleticEnts, r/Kratom, r/cannabis, r/mod quittingkramot, r/trees, r/Peitoles, r/leaves, r/entshop, r/portablelabs, r/abw, r/Waxpen, r/vapebonging, r/StonerEngineering, r/elderitrees, r/reevidebles</td>
<td></td>
</tr>
<tr>
<td>Risk level 3: r/researchchemicals, r/Tianeptine, r/tryptonaut, r/TripSit, r/pharmas, r/Psychonaut, r/Nootropics, r/afinil</td>
<td></td>
</tr>
<tr>
<td>Risk level 4: r/OpiatesRecovery, r/HeroinHeroes, r/MDMA, r/enan, r/maltrexone, r/heroin, r/suboxone, r/gabapentin, r/opiates, r/opiatesmemorial, r/ObscureDrugs, r/phenibut, r/pillhead, r/pregabalin, r/loperamide, r/Opiatewithdrawal, r/meth, r/Methadone, r/askdrugs, r/fentanyl, r/drugwar, r/benzodiazepines, r/stimupinleitung, r/sims, r/heroinaddiction, r/benzorecovery, r/toxycodeone, r/OurOverusedVeins, r/OpiateChurch, r/Thevanondrugs, r/drugscirclejerk, r/Carbentinal</td>
<td></td>
</tr>
<tr>
<td>Risk level mixed: r/drugtesthelp, r/DrugNerds, r/REDDITORSINRECOVERY, r/addiction, r/ReagentTesting, r/Drugs, r/drugsarebeautiful</td>
<td></td>
</tr>
</tbody>
</table>

Approach Overview

Our study aimed to investigate the gateway hypothesis on Reddit through 3 RQs: analyzing social media behaviors that support the gateway hypothesis (RQ1), predicting transitions between substances (RQ2), and comparing the linguistic differences between escalating and de-escalating risk transitions (RQ3). To tackle these RQs, we began by categorizing Reddit users into groups based on their historical trajectories and analyzed their behaviors to answer RQ1. We then constructed a risk transition data set using the trajectories of Reddit users from our collected data sets. For RQ2, we used the risk transition data set to develop and compare different machine learning and deep learning models for forecasting the future directions of risk transitions. In addition, we extracted coefficient scores for each input feature from our machine learning models to identify the most important features in distinguishing between future
escalating and de-escalating transitions. To address RQ3, we used an unsupervised language modeling technique to measure the linguistic disparities in posts and comments created on initial transition phases between escalating and de-escalating transitions.

**Modeling Trajectories and Transitions**

**Overview**
To evaluate the gateway hypothesis on Reddit, we investigated and analyzed the historical activity of every Reddit user in our collected data set. Each Reddit user’s posts or comments history with substance-related subreddits was compiled into a single trajectory. Our approach involved constructing a trajectory for each Reddit user in our data set, encompassing all the posts and comments they had made across 65 substance-related subreddits. This enabled us to investigate whether a pattern emerged among Reddit users, beginning with discussions or interactions related to specific substances and subsequently transitioning toward others.

For each Reddit user’s trajectory, we replaced the subreddit names that a Reddit user had posted or commented on with the annotated risk levels of that subreddit as assigned by our domain experts. By broadening our analysis beyond specific substances and focusing on risk levels, we aimed to explore Reddit users’ overall progression and movement across different levels of risk within the substance-related subreddit landscape.

**Time Binning**
We grouped posts and comments into bins by each active week to reduce noise in our data. This grouping process involved defining a starting time, denoted as \( t \), which represents the time stamp of the first post or comment created. We then established a week-long time frame by adding the total number of seconds in a week to \( t \). Thus, an active week bin encompasses all posts and comments with time stamps falling within the range of \( t + \text{number of seconds in a week} \). To establish the subsequent active week, we identified the time stamp \( t \) of the next post or comment that occurs after \( t + \text{number of seconds in a week} \). We assigned a majority label within each bin based on the posts and comments in the bin. This involved computing the risk label for the bin by converting the subreddits present within it into their respective risk levels. We then determined the risk level that occurred most frequently within the bin, establishing the majority label for that particular period.

**Breaking Ties**
Following the process of creating time bins for each trajectory and assigning majority labels to those bins, we found that approximately 5% of Reddit users in our data set encountered ties in their posting or commenting history at the subreddit level. A tie occurs when multiple labels have the same frequency of occurrence within a given time bin. In such cases, we used the strategy of breaking ties by selecting the highest risk level as the majority label. The rationale behind this tie-breaking approach lies in the assumption that when Reddit users begin engaging in discussions within higher risk level subreddits, they likely already possess some level of experience or familiarity with those particular drugs [26,28,31]. By assigning the higher risk level as the majority label in cases of ties, we aimed to capture the progression or transition of Reddit users toward subreddits associated with potentially greater risk. Furthermore, we discussed more about the limitation of this approach in assigning majority labels in the Discussion section, acknowledging the complexities and nuances involved in such categorizations.

**Grouping Trajectories**
To understand how Reddit users exhibit different behaviors within their trajectories, we grouped each trajectory into three classes; these classes allowed us to categorize trajectories based on the patterns observed in their risk levels over time. (1) Escalating trajectories: this class includes trajectories demonstrating a clear progression or escalation in risk levels across the time bins within the trajectories. It indicates a pattern where Reddit users gradually shift from lower-risk subreddits to those associated with higher risk levels. (2) De-escalating trajectories: in contrast to escalating trajectories, de-escalating trajectories exhibit a pattern of decreasing risk levels over time. (3) Trajectories with no change: trajectories falling into this class consist of time bins labeled with the same risk level throughout.

**Creating Transitions**
After transforming each Reddit user’s activity into one continuous trajectory by applying time binning and resolving ties, we analyzed the data by creating transitions from these trajectories. While each Reddit user in our data set has one trajectory, within this trajectory, a Reddit user may experience multiple transitions that is defined as shifts from one or multiple consecutive time bins with the same risk level to a time bin with a different risk level.

To identify these transitions, we first located the beginning of a transition (Figure 1). This occurs when a time bin with the same risk level persists consecutively before encountering a time bin with a different risk level. The beginning of a transition signifies the period where the Reddit user’s activity remains within a consistent risk level. Subsequently, the ending of a transition marks the point where the risk level changes within the trajectory, represented by a single time bin with a different risk level than the preceding bins. Once an ending transition is identified, the subsequent beginning transition starts with the time bin following the ending transition. This approach allows us to delineate the risk transitions within a Reddit user’s trajectory, capturing the shifts in risk levels and potentially indicating the escalation transitions or de-escalation transitions in their engagement with substance-related subreddits. We removed all the transitions that contained <100 words at the beginning of the transitions to ensure that we had enough data to conduct the following analyses.
**Predicting the Future Risk Transitions**

**Overview**

For prediction, we used a logistic regression model, a popular interpretable machine learning model, to perform text classification tasks. Using transformer-based pretrained models has become the norm in various natural language processing fields and has achieved state-of-the-art performances [50]. Motivated by this, we used BERT [51] and its extension, RoBERTa [52], as additional methods.

**Logistic Regression Model**

The logistic regression model was developed to predict the direction of transitions between subreddit risk levels by using a binary dependent variable indicative of the transition direction (escalating transitions vs de-escalating transitions). Independent variables, or predictors, were derived from textual features and user engagement metrics extracted at the beginning of identified transitions. Four categories of predictors were extracted from the model:

1. Psycholinguistic attributes: the Linguistic Inquiry and Word Count (LIWC) lexicon [53-55] is a well-validated psycholinguistic lexicon used for extracting psycholinguistic characteristics from texts. It analyzes texts across various dimensions, such as emotions, perceptions, interpersonal focus, and social and personal concerns. For each transition, we used the LIWC lexicon to obtain normalized occurrences of words in each LIWC category for each individual.

2. Open vocabulary (n-grams): the concept of open vocabulary, which refers to the unrestricted and diverse set of words or word combinations a person can use, has been used to investigate the psychological attributes of individuals in multiple previous studies [38,56,57]. We calculated the distribution of the top 500 uni- and bigrams at the beginning of the transition as open vocabulary features.

3. Part-of-speech tagging: the Stanford Log-linear Part-Of-Speech Tagger [58] was used to identify nouns, verbs, and adjectives in posts and comments. This tool calculates the ratios of nouns, verbs, and adjective words in the posts and comments.

4. Activities: for each Reddit user in our collected data set, we measured the average number of Reddit posts or comments made within our data set per bin and the number of words per bin.

After extracting these 4 sets of features from the posts and comments made during the initial stages of the transition, we constructed a logistic regression model. This model predicts whether an individual’s transition is escalating or de-escalating. It considers the independent variables, in this case, the extracted features: psycholinguistic attributes, open vocabulary, part-of-speech tagging, and activities, and outputs a binary prediction of whether the transition will escalate or de-escalate in the future. To ensure the optimal performance of our model and reduce the noisy features, we applied a feature selection strategy, determining to retain the top 100 features that exhibit the highest levels of mutual information with the classification output. This approach allowed us to focus on the most critical features and facilitate a more effective predictive model.

We randomly split our entire transitions data set into training and testing sets in an 8:2 ratio to evaluate our logistic regression model. During the training phase of the model, we used cross-validation to ensure that our model generalizes well to unseen data and does not overfit. Specifically, we adopted a 5-fold cross-validation approach, where is set to 5. The performance across all iterations was then averaged to provide a more accurate measure of the model’s performance.

**Deep Learning Model**

To tackle the multifaceted problem of predicting future risk levels for each substance-related transition, we used BERT [51] and its extension, RoBERTa [52]. As a pioneer in transformer-based models, BERT revolutionized the natural language processing field with its bidirectional training approach. RoBERTa further improved upon it by modifying key hyperparameters in the model architecture, such as removing the next-sentence pretraining objective and training with much larger minibatches [52]. Both models have been adapted to solve
various domain-specific tasks, such as suicidal ideation [53], mental disorders detection [59], web-based social support [60], and social dimensions describing human relationships [61].

We fine-tuned pretrained BERT and RoBERTa models using concatenated texts at the beginning of transitions as inputs and the corresponding escalation directions as outputs. To assess the performance of the fine-tuned models, we randomly split the entire transition data set into training and testing sets in an 8:2 ratio and used a \( k \)-fold cross-validation \((=5)\) approach on the training data set. In our training setting, we set the epoch count to 5, a decision guided by initial tests that observed marginal improvement of the model beyond this epoch number. We set the batch size to 32, considering our computational capacity. The learning rate was chosen as \( 5 \times 10^{-5} \) by default.

### Analyzing the Linguistic Cues

We used an unsupervised language modeling technique known as the Sparse Additive Generative Model (SAGE) [62] to investigate the linguistic indicators at the concatenated texts created during the beginning phase of escalating or de-escalating transitions. This approach enabled us to compare the parameters of 2 multinomial models with the logistic regression, allowing us to identify significant terms. The regularization parameter in SAGE was self-tuned, and this balanced the importance of common and rare terms in the selection process. We applied the SAGE technique specifically to differentiate between n-grams \((n=2, 3)\) present in the posts and comments at the beginning of transitions between escalating and de-escalating transitions.

During our analysis, a positive SAGE value \( (>0)\) suggested that an n-gram was more representative of escalating transitions, indicating its association with an increase in risk levels. Conversely, a negative SAGE value suggested greater representativeness for the absence of escalating transitions, indicating its association with a decrease in risk levels. By examining these linguistic indicators using the SAGE technique, we gained insights into the specific language patterns and characteristics that distinguished escalating and de-escalating transitions within our data set. This analysis helped us better understand the linguistic dynamics and factors that influenced the progression or regression of risk levels on Reddit.

### Results

#### Modeling Trajectories and Transitions

Table 1 provides an overview of the activity of Reddit users within drug-related subreddits, highlighting the number of Reddit users, the frequency of posts and comments, and the instances of risk level transitions. A subset of users displayed at least 1 transition in risk level behavior, with variations in the length and frequency of their posts and comments. These data show the analysis of transitions between subreddits of differing risk levels (Table 1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reddit users, n</td>
<td>133,220</td>
</tr>
<tr>
<td>Reddit users who have at least 1 transition, n (%)</td>
<td>58,279 (43.75)</td>
</tr>
<tr>
<td>Posts and comments, n</td>
<td>9,205,558</td>
</tr>
<tr>
<td>Transitions, n</td>
<td>169,811</td>
</tr>
<tr>
<td><strong>Posts and comments length (by words)</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>36.67 (65.49)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>18 (1-9688)</td>
</tr>
<tr>
<td><strong>Subreddits that Reddit users posted or commented</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>2.69 (2.00)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>2 (1-48)</td>
</tr>
<tr>
<td><strong>Unique risk levels per Reddit users</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.67 (0.68)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>2 (1-4)</td>
</tr>
<tr>
<td><strong>Posts and comments per Reddit user</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>69.10 (280.91)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>15 (1-36,533)</td>
</tr>
<tr>
<td><strong>Posts and comments at the beginning of transition</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>18.12 (97.58)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>3 (1-6650)</td>
</tr>
</tbody>
</table>

Table 2 presents the number of Reddit users based on their historical trajectories in our collected data set, specifically their risk transition patterns. The table categorizes Reddit users into 3 types based on their trajectories: escalation trajectories,
de-escalation trajectories, and no change. For each category, the table shows the starting and ending risk levels of the transition within trajectories, the total number of users who exhibited the corresponding transition in their trajectories in our collected data set, and the average transition time (ATT) duration in weeks. Note that no ATT is recorded for users showing no change in risk levels.

Table 2. Distribution of Reddit users categorized by their historical behavior patterns.

<table>
<thead>
<tr>
<th>Trajectories type and users, n (%)</th>
<th>Risk level transition</th>
<th>ATT (weeks), mean (SD; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>No change (n=39,000)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>26,381 (67.6)</td>
<td>7814 (20)</td>
<td>4</td>
</tr>
<tr>
<td>4805 (12.3)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Escalation (n=8,428)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3706 (44)</td>
<td>3617 (42.9)</td>
<td>2</td>
</tr>
<tr>
<td>1105 (13.1)</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>De-escalation (n=4925)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2065 (41.9)</td>
<td>1936 (39.3)</td>
<td>3</td>
</tr>
<tr>
<td>924 (18.8)</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

aNote that no average transition time is recorded for Reddit users, showing no change in risk levels.
bATT: average transition time.
cNot available.

Regarding Reddit users with trajectories indicating only the escalation in risk levels across the time bins, Reddit users predominantly showed transitions from risk level 2 to 4, followed by transitions from 2 to 3 and 3 to 4. The chi-square test for goodness of fit on the number of Reddit users in escalation trajectories showed significant differences (χ² = 1552.4, P < .001). For example, one Reddit user wrote about their first-time kratom (risk level 2) experience in the risk level 2 subreddit:

Yesterday I had my first experience taking Kratom...
After around 39 weeks, the same user wrote another post disclosing their own experiences with phenibut (risk level 3), varying dosages, frequency, and the effects of combining it with other substances:

I’ve been using phenibut ranging from once every 2 weeks to 3 times a week…I took 2g followed by 2, 1g redoses…I took 2g early on the day and decided to drink like 3 beers probably 12 hours later.

The ATT for these transitions ranged from 28.71 (from risk level 3 to risk level 4) to 39.71 weeks (from risk level 2 to risk level 3). In the group with de-escalating trajectories, Reddit users primarily showed transitions from risk level 4 to 2 (χ² = 475.7, P < .001), with the ATT being 32.73 weeks. Transitions from 3 to 2 and 4 to 3 also occurred, with average active use ranging from 29.51 to 35.79 weeks. For the group with no-change trajectories, most Reddit users in our data set remained at risk level 2 without any transition as indicated by a significant chi-square result (χ² = 21,008.0, P < .001).

Predicting the Future Risk Transitions

We implemented and evaluated multiple classifiers on our transition data set to predict the future transitions of risk level, including the logistic regression, BERT, and RoBERTa models. Table 3 summarizes the performance of our models and includes the accuracy, precision, recall, F1-score, and area under receiver operating characteristic curve score (Figure 2). These metrics are critical for evaluating the performance of machine learning and deep learning models [63,64]. On the basis of these results, we found that the fine-tuned BERT and RoBERTa models had better performance than the logistic regression model. The BERT model, after fine-tuning, achieved an accuracy of 78.48% and an F1-score of 79.2%, while logistic regression had an accuracy of 75.5% and an F1-score of 76.08%. The differences in their performance might suggest the complex language patterns found in substance use–related discussions. Overall, these models showed promise in predicting the future risk level, with the fine-tuned BERT and RoBERTa models demonstrating better performance than logistic regression.
Table 3. Comparative performance metrics of logistic regression, Bidirectional Encoder Representations from Transformers (BERT), and Robustly Optimized BERT Approach (RoBERTa) models on a 5-fold cross-validation and test data set.

<table>
<thead>
<tr>
<th>Model</th>
<th>5-fold cross-validation, mean (SD)</th>
<th>Test data set (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Precision</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>75.29 (0.29)</td>
<td>80.55 (0.30)</td>
</tr>
<tr>
<td>BERT</td>
<td>79.14 (0.27)</td>
<td>82.34 (1.96)</td>
</tr>
<tr>
<td>RoBERTa</td>
<td>79.09 (0.53)</td>
<td>84.97 (1.15)</td>
</tr>
</tbody>
</table>

$^a$ROC-AUC: area under the receiver operating characteristic curve.

Figure 2. Receiver operating characteristic curves comparing the performance of Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Approach (RoBERTa) models on a test data set. AUC: area under the curve.

Although the BERT and RoBERTa models showed better performance metrics than the logistic regression model, it is crucial to recognize the interpretability and explainability of the logistic regression model. As an explainable model, logistic regression allows for the extraction and examination of the most influential features in its decision-making process [65]. Table 4 shows the top 20 predictive features (independent variables) with the $\beta$ coefficients and significance from the logistic regression. A subreddit with a positive $\beta$ coefficient indicates that the occurrence of the features in the posts increases its possibility of future escalation into a higher risk level subreddit (escalating transition from a lower risk level subreddit to a higher risk level). A subreddit with a negative $\beta$ coefficient indicates that the occurrence of the features in the posts increases its possibility of future de-escalating into a lower risk level subreddit (de-escalating transition from a higher risk level subreddit to a lower risk level). In the predictors associated with future risk escalations, we found many n-gram features related to substance names, including kratom, weeds, marijuana, tree, buds, and cannabis, showing high relative importance. We also observed words related to substance-related behaviors or tools, such as bong, vape, smoke, smoking, and smoked, showing high relative importance.
Table 4. Top 20 features in the logistic regression classifier with the $\beta$ coefficients. *P* values reported after Bonferroni correction. Positive features indicate the most predictive features for identifying escalating transitions, and negative features suggest the most predictive features for identifying de-escalating transitions.

<table>
<thead>
<tr>
<th>Feature</th>
<th>$\beta$</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kratom</td>
<td>.510</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>bong</td>
<td>.261</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>weed</td>
<td>.249</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>smoke</td>
<td>.246</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>smoking</td>
<td>.216</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>marijuana</td>
<td>.214</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>vape</td>
<td>.202</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>bud</td>
<td>.190</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>joint</td>
<td>.179</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>strain</td>
<td>.167</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>bowl</td>
<td>.166</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>edible</td>
<td>.158</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>tree</td>
<td>.156</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>oil</td>
<td>.130</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>high</td>
<td>.123</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>cannabis</td>
<td>.099</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>legal</td>
<td>.095</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>glass</td>
<td>.094</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>green</td>
<td>.087</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>smoked</td>
<td>.084</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phenibut</td>
<td>−.351</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>mdma</td>
<td>−.265</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>pill</td>
<td>−.258</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>xanax</td>
<td>−.234</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>meth</td>
<td>−.225</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>opiate</td>
<td>−.220</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>roll</td>
<td>−.191</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LIWC^[b]bio</td>
<td>−.138</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LIWC:insight</td>
<td>−.111</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>heroin</td>
<td>−.110</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>trip</td>
<td>−.105</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LIWC:time</td>
<td>−.097</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>take</td>
<td>−.082</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>benzos</td>
<td>−.081</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LIWC:you</td>
<td>−.079</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LIWC:death</td>
<td>−.075</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>shit</td>
<td>−.064</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>took</td>
<td>−.063</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
In n-gram features associated with future risk de-escalation transitions, the n-gram features indicated possible treatments, such as **phenibut**, **pills**, and **xanax**. We also found that the n-gram features around “harder substances” are reported to have high relative importance in predicting future risk de-escalation transitions such as **mdma**, **meth**, **heroin**, and **opiate**. LIWC features, including biological processes, insight, time, you (second pronouns), and death, contributed significantly to predicting future risk de-escalation transitions. This might be because Reddit users used these words to describe their symptoms or give suggestions or warnings at the beginning of their de-escalating transitions.

### Analyzing the Linguistic Cues

To investigate the most salient linguistic cues at the beginning of escalating and de-escalating transitions, we adopted the SAGE analysis [62] and reported the top salient n-grams in Table 5. A positive SAGE score indicates the terms frequently associated with the beginning of escalation transitions, while a negative SAGE score points to the terms linked with the beginning of de-escalation transitions.

We discovered that keywords such as **taking kratom**, **smoking weed**, and **dry herb** related to substance use frequently appear in individuals’ self-motivated texts before transitioning into higher-risk subreddits. For instance, one person described their experiences with kratom:

> I was taking kratom for over a year or two and had positive experiences. However...

Approximately 31 weeks later, the same author wrote in a risk level 4 subreddit:

> I was in a temporary role, and they offered me a full-time position, contingent on passing a drug test. I usually don’t consume opiates, but I took 6 pills on December 24th...

We noted that keywords describing substance use; purchase locations; and feelings associated with use, such as **head shop**, **smell like**, and **get high**, were more salient in the language of escalating transitions (eg, “I went to the head shop,” “My room smells like cannabis,” and “I get high from it.”).

In contrast, keywords indicating harm reduction and recovery, such as **test kit**, **get clean**, and **stay safe** were prevalent at the beginning of de-escalating transitions. We observed these keywords frequently in texts describing individuals’ recovery experiences from substance misuse and providing suggestions for others. For example, a post reads as follows:

> To get clean, I reduced my intake gradually over two months, decreasing my dosage by 25% at each step.

Roughly 20 weeks later, the same poster sought suggestions and support in a lower-risk subreddit regarding severe withdrawal symptoms:

> I used to struggle with anxiety...and a strong desire for Xanax. While many of these symptoms have subsided, I often find myself without an appetite...I am inquiring whether Xanax withdrawal can lead to this issue.

We also observed many time-related keywords with high salience scores in de-escalating transitions, such as **3 months**, **hour late**, **year clean**, and **3 days**. This suggests that before posting or commenting on lower-risk subreddits, Reddit users might have previously shared their experiences on higher-risk subreddits. Additionally, keywords suggesting reduced substance use, such as **take half**, **half pill**, and **extended release**, appeared more frequently in de-escalating transitions. This might indicate that individuals transitioning to lower-risk subreddits are sharing their harm reduction strategies and experiences that involve reducing the amount of substance consumed.
Table 5. Top salient n-grams (n=2, 3) associated with escalating and de-escalating transitions in posts, as identified by Sparse Additive Generative Model (SAGE) analysis. A positive SAGE score indicates the terms frequently associated with the beginning of escalation transitions, while a negative SAGE score points to the terms linked with the beginning of de-escalation transitions.

<table>
<thead>
<tr>
<th>n-gram</th>
<th>SAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top salient n-grams in escalating transitions</strong></td>
<td></td>
</tr>
<tr>
<td>legal, state</td>
<td>0.155</td>
</tr>
<tr>
<td>taking, kratom</td>
<td>0.144</td>
</tr>
<tr>
<td>look, like</td>
<td>0.141</td>
</tr>
<tr>
<td>illegal, state</td>
<td>0.141</td>
</tr>
<tr>
<td>coconut, oil</td>
<td>0.136</td>
</tr>
<tr>
<td>started, smoking</td>
<td>0.134</td>
</tr>
<tr>
<td>smoking, weed</td>
<td>0.128</td>
</tr>
<tr>
<td>maeng, da</td>
<td>0.119</td>
</tr>
<tr>
<td>using, kratom</td>
<td>0.106</td>
</tr>
<tr>
<td>smoke, spot</td>
<td>0.101</td>
</tr>
<tr>
<td>dry, herb</td>
<td>0.090</td>
</tr>
<tr>
<td>different, strain</td>
<td>0.087</td>
</tr>
<tr>
<td>mason, jar</td>
<td>0.086</td>
</tr>
<tr>
<td>green, malay</td>
<td>0.086</td>
</tr>
<tr>
<td>like, weed</td>
<td>0.080</td>
</tr>
<tr>
<td>stop, smoking</td>
<td>0.079</td>
</tr>
<tr>
<td>red, bali</td>
<td>0.078</td>
</tr>
<tr>
<td>medical, marijuana</td>
<td>0.078</td>
</tr>
<tr>
<td>want, smoke</td>
<td>0.076</td>
</tr>
<tr>
<td>take, kratom</td>
<td>0.075</td>
</tr>
<tr>
<td>head, shop</td>
<td>0.073</td>
</tr>
<tr>
<td>people, smoke</td>
<td>0.072</td>
</tr>
<tr>
<td>smell, like</td>
<td>0.072</td>
</tr>
<tr>
<td>tolerance, break</td>
<td>0.072</td>
</tr>
<tr>
<td>peanut, butter</td>
<td>0.071</td>
</tr>
<tr>
<td>get, high</td>
<td>0.071</td>
</tr>
<tr>
<td>smoke, weed</td>
<td>0.069</td>
</tr>
<tr>
<td>kratom, use</td>
<td>0.068</td>
</tr>
<tr>
<td>pax, 2</td>
<td>0.067</td>
</tr>
<tr>
<td>drug, test</td>
<td>0.066</td>
</tr>
<tr>
<td>plant, matter</td>
<td>0.065</td>
</tr>
<tr>
<td>time, day</td>
<td>0.064</td>
</tr>
<tr>
<td>arizer, air</td>
<td>0.063</td>
</tr>
<tr>
<td>fellow, ents</td>
<td>0.061</td>
</tr>
<tr>
<td>smoke, much</td>
<td>0.060</td>
</tr>
<tr>
<td>one, hitter</td>
<td>0.060</td>
</tr>
<tr>
<td><strong>Top salient n-grams in de-escalating transitions</strong></td>
<td></td>
</tr>
<tr>
<td>test, kit</td>
<td>-0.130</td>
</tr>
<tr>
<td>get, clean</td>
<td>-0.093</td>
</tr>
<tr>
<td>stay, safe</td>
<td>-0.089</td>
</tr>
</tbody>
</table>
### Discussion

#### Principal Findings

Overall, the results of this study analyzing substance use–related social media behavior on Reddit show preliminary support for the gateway hypothesis; however, these results were mixed and require greater contextualization and examination; that is, while some of the proponents of the gateway hypothesis were supported based on patterns of escalation trajectories (marijuana subreddits to higher-risk subreddits), other patterns noted for escalation and de-escalation trajectories were more complex and were not supported by the original gateway framework. For example, we did observe that engagement in marijuana-related subreddits along with the pattern of specific n-grams such as *kratom*, *weeds*, and *marijuana* were predictive of future risk escalation transitions. This aligned with the gateway hypothesis, suggesting a potential progression from substances perceived as lower risk to ones associated with higher risks. However, most individuals who began in marijuana and other low-risk subreddits did not progress to higher-risk subreddits, a finding that does not support the “causal” notions of the original gateway hypothesis. This finding aligns with the earlier-mentioned research from the National Institute on Drug...
Abuse in the United States, which found that while marijuana use typically precedes the use of “harder” drugs, most marijuana users do not transition to these more potent substances [6]. This consistency between web-based drug use behavior on subreddits and real-world drug consumption patterns is noteworthy. It underscores the relevance and potential utility of social media analysis as a complementary tool to traditional survey research, highlighting its capacity to offer insights into one’s behavior and progression trends. As noted by previous literature [3], there are numerous confounds that could impact the risk for progression from one substance to another, particularly escalation transitions from marijuana to higher-risk substances, given the rapid changes in accessibility, legality, and medical use of marijuana in recent decades. Furthermore, we did note instances of de-escalation transitions from higher-risk subreddits to marijuana subreddits, contrary to the gateway hypothesis and consistent with recent literature on the use of marijuana and other substances (eg, kratom) to mitigate withdrawal from higher-risk substances such as opioids [32,66,67]. Similarly, in line with this literature, we found terms related to harm reduction and recovery methods across high-risk subreddits, such as test kit, getclean, and stay safe, to be predictive of de-escalation transitions. These findings suggested that individuals might leverage a range of strategies and substances to manage and reduce their substance use, a perspective aligned with harm reduction theory.

A significant observation based on the historical behaviors was the relatively short time span within which users tend to escalate or de-escalate transitions to their risk levels. This novel finding is salient for clinicians and health care providers. By understanding these timelines, health care professionals can proactively strategize and implement interventions tailored to these predictive timelines. In addition, providing individuals with insights about the potential expectations can serve as a form of preventative care, making them aware of the potential risks and the rapidity with which transitions can occur. Our study underscored the significant potential of using machine learning and deep learning models, specifically BERT and RoBERTa, in understanding and predicting substance use transitions. The models, leveraging deep learning techniques, outperformed the traditional logistic regression model, indicating their effectiveness in capturing and learning from the complex language patterns found in substance use–related discussions. Furthermore, our findings emphasized the value of linguistic features as predictors of risk escalation and de-escalation transitions. This suggested that individuals’ narratives and discussions around their experiences, including their feelings and perceptions, could provide important clues about their trajectory of substance use. Interestingly, the prevalence of time-related terms among salient features in de-escalating transitions suggested the potential role of timing in substance use management. These findings highlight the need for timely interventions, which could potentially alter an individual’s substance use trajectory if the associated language patterns were detected early.

This study marks an initial step in understanding the role of social media data in substance use research, advocating for cautious interpretation and deeper investigation. By understanding the patterns of language use, more effective policies and therapeutic interventions could be designed to address substance use disorders. Policy makers could specifically use these findings to develop more complex and all-embracing approaches to manage substance use disorders at both the individual and community levels, specifically taking into consideration the prevention approaches among youth as well as treatments that address polysubstance use. However, we should proceed cautiously when considering how to apply these discoveries. While our work shows that machine learning and deep learning models could be used to predict progressions of substance use, they should not be used in isolation. A thorough, multifaceted approach to addressing substance use disorders is required due to human behavior’s complexity and individual experiences’ differences. Consequently, these models should serve as one tool among many, contributing to a better understanding of substance use patterns that could inform more effective intervention strategies.

Limitations and Future Work
We note that our study has some limitations, some of which might suggest future research. This study is limited by selection biases. We have gathered historical data only from individuals who were active on a certain number of substance-related subreddits. For example, some Reddit users might stop using web-based communities to discuss substance-related experiences or seek suggestions. This is especially true, considering the stigma and discrimination experienced by individuals with substance use disorders. The selection biases in our study are also compounded by an uneven distribution of Reddit users across different risk levels, as well as a large number of marijuana-related subreddits in risk level 2. These imbalances may reflect a nonrepresentative sample of the broader population, further skewing our results and influencing the likelihood of associations between marijuana-related discourse and escalation transitions. Thus, our observations might not be generalizable to other web-based communities or population-level trends.

It is important to note that our study adopts a broad definition of risk transitions. We cannot claim that the transitions in which a Reddit user begins posts and comments in different risk level subreddits directly correspond to clinical substance–using status. The predictors and linguistic cues that we have identified are preliminary and require further validation within a clinical setting. Our approach, while informative, cannot be directly translated into definitive assessments of substance use disorders. Future research has the potential to merge these insights with clinically validated assessments and social media data. Our method of modeling transitions might not fully capture the long-term fluctuating escalating or de-escalating behaviors of individuals. It highlights the need for more nuanced methodologies that can accurately reflect the dynamic and potentially fluctuating nature of substance use behavior over time. By doing so, a more comprehensive and generalizable understanding could be developed, providing deeper insights into substance misuse prevention, harm reduction, and recovery strategies.
In addition, our study’s approach to assigning majority labels within time bins based on the most frequent risk level presents a limitation. By focusing solely on the frequency of engagements, this method may overlook critical but less frequent high-risk interactions, potentially misrepresenting a user’s actual risk exposure. For instance, an individual’s occasional but significant high-risk activity within a predominantly low-risk engagement period could be overshadowed by the majority labeling process. Despite this limitation, it is important to underscore that our approach remains exploratory and holds significant value. It provides a foundational understanding of user trajectories in substance-related subreddit activities and offers a novel perspective on risk engagement over time. Future research might benefit from exploring methodologies that better capture the complexity of web-based behaviors, ensuring a more accurate depiction of risk levels in substance-related subreddit activities.

We highlight the limitation of the temporal restriction of our data set. The Reddit data that we used were collected from Google BigQuery spanning from 2015 to 2019. In recent years, collecting such data has become increasingly restricted due to rising privacy concerns and the evolving landscape of data protection regulations. As a consequence, researchers are often constrained to working with historical data when aiming to extract insights from such platforms. This limitation introduces the potential for outdated linguistic and behavioral references in our data set. As web-based communities and their dynamics evolve rapidly, it is possible that patterns and behaviors from 2015 to 2019 do not accurately reflect the current situation. Future studies aiming to capture a more contemporary perspective might need to explore alternative data sources that comply with current privacy standards and ethical policies.

Conclusions

In conclusion, this study successfully used machine and deep learning models to predict future risk transitions in substance use, leveraging a rich longitudinal data set from Reddit. The fine-tuned BERT and RoBERTa models notably outperformed logistic regression models in predicting risk escalation and de-escalation transitions, with specific n-gram features playing a pivotal role in these predictions. Highlighting the critical role of language patterns in understanding substance use trajectories, the findings open avenues for future research, particularly in developing these methods with social media data to inform potential interventions in substance use behaviors.

Acknowledgments

The authors thank Kevin Davet and Nina Kaiser for their contributions to coding the Reddit data and developing the code book, which were essential to the success of this research. The valuable assistance of Sanjana Garg is also acknowledged. The authors appreciate the computational resources provided by the Aalto Science-IT project. This work was partially supported by a grant from the National Institute of Mental Health (R01MH117172) with MDC as the principal investigator and funds from the Centers from Disease Control and Prevention. The content herein is the sole responsibility of the authors and does not necessarily represent the official views of the sponsors. The National Institute on Drug Abuse (grant: K02 DA043657) also provided partial funding for this work. The authors would like to thank the members of the Social Dynamics and Well-Being Lab at Georgia Tech for their feedback on early iterations of the research.

Conflicts of Interest

PCR is a consultant for Rissana, LLC, and PredictView.

Multimedia Appendix 1

Classification of subreddits by risk level with examples and definitions.
[DOCX File, 22 KB - formative_v81e54433_app1.docx ]

References


Abbreviations

ATT: average transition time
BERT: Bidirectional encoder representations from transformers
LIWC: Linguistic Inquiry and Word Count
RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Approach
RQ: research question
SAGE: Sparse Additive Generative Model
Please cite as:
Yuan Y, Kasson E, Taylor J, Cavazos-Rehg P, De Choudhury M, Aledavood T
Examining the Gateway Hypothesis and Mapping Substance Use Pathways on Social Media: Machine Learning Approach
JMIR Form Res 2024;8:e54433
URL: https://formative.jmir.org/2024/1/e54433
doi:10.2196/54433
PMID:38713904

©Yunhao Yuan, Erin Kasson, Jordan Taylor, Patricia Cavazos-Rehg, Munmun De Choudhury, Talayeh Aledavood. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Research Letter

The Relationship Between Changes in Mindfulness and Subsequent Changes in Well-Being Following Psychedelic Use: Prospective Cohort Study

Grant Jones¹, MA; Felipe Herrmann¹, BA; Adam Bear¹, PhD; Robin Carhart-Harris², PhD; Hannes Kettner², MSc

¹Harvard University, Cambridge, MA, United States
²University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Grant Jones, MA
Harvard University
32 Mill St Box 78
Cambridge, MA, 02138
United States
Email: gmj005@g.harvard.edu

Abstract

This study demonstrates that changes in mindfulness predict subsequent changes in well-being in a data set including individuals who recently engaged in psychedelic use.

(JMIR Form Res 2024;8:e54632) doi:10.2196/54632

KEYWORDS
psychedelics; mindfulness; observational; web-based survey; psychedelic; meditation; mental health; anxiety; depression; survey; surveys; drug; drugs; substance use; hallucinogen; hallucinogens

Introduction

Psychedelics have been linked to improvements in depression, anxiety, and overall well-being in clinical trials and cross-sectional research [1-3]. Psychedelic use has also been linked to improvements in mindfulness [4,5], leading researchers to hypothesize that the link between psychedelic use and improvements in well-being may be driven in part by increased mindfulness [4,6,7]. Yet, there is a dearth of studies that directly explore this research domain.

This study uses a prospective data set of individuals who recently used psychedelics to examine, a priori, whether changes in mindfulness are linked to subsequent improvements in well-being, depression, and anxiety. Additionally, this study builds upon and addresses the core limitations of Mans et al [8], which found that psychedelic use was linked to improvements in mindfulness and well-being in the aforementioned data set. As Mans et al [8] did not control for major demographic and substance use confounds, include anxiety as an outcome in the analyses, nor assess how changes in mindfulness may potentially drive subsequent improvements in mental health, we aim to address these critical limitations within this study.

Methods

Overview

This study includes data that were collected in 2017 through a web-based platform called “Psychedelic Survey.” Haijen et al [9] provide further details on the data collection.

We used linear regression to assess whether changes in mindfulness from time 1 (baseline) to time 4 (2 weeks post psychedelic experience) were associated with changes in overall well-being, depression, and anxiety from time 1 to time 5 (4 weeks post psychedelic experience). We controlled for key demographic and substance use variables including age, gender, education level, and prior classic psychedelic use.

We used the Cognitive and Affective Mindfulness Scale–Revised to measure mindfulness, the Warwick-Edinburgh Mental Wellbeing Scale to measure overall well-being, the Quick Inventory of Depressive Symptomatology to measure depression, and the State-Trait Anxiety Inventory–6 (STA6) to measure anxiety. As the STA6 is commonly used to measure state (ie, momentary) anxiety, the scale was adapted to measure trait (ie, persisting) anxiety; these adaptations are included in Multimedia Appendix 1.
Ethical Considerations
The study was approved by the Joint Research Compliance Office and the Imperial College Research Ethics Committee (ICREC reference 18IC4346). Participant data is anonymous, and participants were not compensated for their responses. All participants provided informed consent before participation.

Results
Table 1 presents the demographics of our participants. Table 2 presents the results of our regression models. Changes in mindfulness (T1→T4) were associated with changes in overall well-being, depression, and anxiety (T1→T5). $R^2$ values ranged from 0.19 to 0.35, indicating moderate to strong effect sizes for our models. Multimedia Appendix 1 includes supplementary robustness analyses that assess the inverse relationship between well-being and mindfulness (i.e., are changes in well-being from T1→T4 associated with changes in mindfulness from T1→T5?). Although these associations were significant, $R^2$ values for these models were lower than those for our main models assessing the relationship between mindfulness changes (T1→T4) and well-being changes (T1→T5; $R^2$ range 0.11-0.17). Thus, these additional results offer further evidence supporting the notion that mindfulness changes drove well-being changes in this study.

Table 1. Demographics of the sample.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants (N=163)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>28 (23-38)</td>
</tr>
<tr>
<td>Range</td>
<td>16-71</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Left school before age 16 without qualifications</td>
<td>5 (3)</td>
</tr>
<tr>
<td>Some high school/GCSE(^a) level (in UK)</td>
<td>14 (9)</td>
</tr>
<tr>
<td>High school diploma/A-level education (in UK)</td>
<td>16 (10)</td>
</tr>
<tr>
<td>Some university (or equivalent)</td>
<td>26 (16)</td>
</tr>
<tr>
<td>Bachelor’s degree (or equivalent)</td>
<td>63 (39)</td>
</tr>
<tr>
<td>Post-graduate degree (e.g., masters or doctorate)</td>
<td>39 (24)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>113 (69)</td>
</tr>
<tr>
<td>Female</td>
<td>50 (31)</td>
</tr>
<tr>
<td><strong>Prior psychedelic use, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>14 (9)</td>
</tr>
<tr>
<td>Only once</td>
<td>11 (7)</td>
</tr>
<tr>
<td>2-5 times</td>
<td>43 (26)</td>
</tr>
<tr>
<td>6-10 times</td>
<td>23 (14)</td>
</tr>
<tr>
<td>11-20 times</td>
<td>23 (14)</td>
</tr>
<tr>
<td>21-50 times</td>
<td>30 (18)</td>
</tr>
<tr>
<td>51-100 times</td>
<td>7 (4)</td>
</tr>
<tr>
<td>More than 100 times</td>
<td>12 (7)</td>
</tr>
</tbody>
</table>

\(^a\)GCSE: General Certificate of Secondary Education.
Discussion

Using a sample of individuals who recently used psychedelics, this study demonstrated that changes in mindfulness (T1→T4) were significantly associated with subsequent changes in mental well-being (T1→T5) [8]. Furthermore, models assessing the inverse associations between well-being and mindfulness (ie, assessing whether changes in well-being [T1→T4] predicted subsequent changes in mindfulness [T1→T5]) were weaker, strengthening the possibility that mindfulness drove improvements in well-being in this study.

Hill’s criteria for causation provide guidelines for evaluating the causal power of our results [10]. This study fulfills 7 of the 9 criteria: adequate strength of association, consistency with prior findings, specificity, plausibility, coherence (ie, biological plausibility), analogy (ie, comparability to related phenomena), and temporality (ie, the cause occurring before the effect). Additional research is needed to address the remaining criteria: biological gradient (ie, dose-response) and experimental evidence.

Limitations include the lack of a control group and the likelihood of bidirectional influence between changes in well-being and changes in mindfulness, both of which limit our ability to make definitive causal claims within this study. Another core limitation is our inability to control for important demographic factors in this study (eg, race/ethnicity, income, marital status). Future longitudinal studies and randomized trials can address these limitations. Overall, this study provides preliminary evidence that mindfulness may be a potential driver of the link between psychedelic use and salutary mental health outcomes.

Acknowledgments

This study was supported by the Ad Astra Chandaria Foundation and the funders of the Centre for Psychedelics Research.

Data Availability

Data sets are available upon request.

Authors’ Contributions

GJ and HK conceptualized the study. GJ executed all data analysis and oversaw the writing and drafting of the paper. FH helped write the paper. HK and RCH oversaw data collection. All authors read and edited the manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Results from three linear regression models and the Adapted State-Trait Anxiety Inventory–6. [DOCX File, 10 KB - formative_v8i1e54632_app1.docx]

References


Abbreviations

STA16: State-Trait Anxiety Inventory–6
Research Letter

Smartwatch Versus Routine Tremor Documentation: Descriptive Comparison

Catharina Marie van Alen¹, MSc; Alexander Brenner¹, MSc; Tobias Warnecke², MD; Julian Varghese¹, MSc, MD

¹Institute of Medical Informatics, University of Münster, Münster, Germany
²Department of Neurology and Neurorehabilitation, Klinikum Osnabrück – Academic Teaching Hospital of the University of Münster, Osnabrück, Germany

Corresponding Author:
Julian Varghese, MSc, MD
Institute of Medical Informatics
University of Münster
Albert-Schweitzer-Campus 1, Bldg A11
Münster, 48149
Germany
Phone: 49 0251 8354 714
Email: julian.varghese@uni-muenster.de

Abstract

We addressed the limitations of subjective clinical tremor assessment by comparing routine neurological evaluation with a Tremor Occurrence Score derived from smartwatch sensor data, among 142 participants with Parkinson disease and 77 healthy controls. Our findings highlight the potential of smartwatches for automated tremor detection as a valuable addition to conventional assessments, applicable in both clinical and home settings.

(KEYWORDS: Parkinson disease; tremor; smart wearables; smartwatch; mobile apps; movement disorders; tremor documentation; tremor occurrence; tremor score

Introduction

Clinical assessment of tremor is limited by its subjectivity, making it challenging to detect subtle tremors (<0.05 g) [1]. The need for technology-based evaluations led to extensive research on sensor-based tremor assessments [2-6]. However, the benefits of smart consumer devices for clinical documentation remain unaddressed. We aimed to bridge this research gap by comparing tremor documentation from routine neurological assessments with a Tremor Occurrence Score (TOS) derived from smartwatch sensor recordings. We included 142 participants with Parkinson disease (PD) and 77 healthy controls. During clinical visits, routine neurological assessments were conducted and wrist-motion data were captured using a previously established Smart Device System (SDS) in a controlled setting [1,7]. The SDS is based on consumer smartwatches and smartphones. The smartwatches were validated in a geophysical experiment for tremor analysis with comparison to a gold-standard seismometer [1]. We investigated the potential of complementing routine tremor documentation with tremor capture using smartwatches [8].

Methods

Overview

The SDS parent study [1] recruited 450 participants, including patients with PD and other movement disorders and healthy controls. The SDS implemented data capture through electronic questionnaires and app-guided movement tasks with smartwatch sensing. Varghese et al [7] provide a detailed description of the study. Of the 450 participants, 101 were excluded due to diagnoses other than PD. Moreover, 130 patients with PD without explicit information regarding tremor in routine documentation were excluded, resulting in 219 participants. Table 1 summarizes the age, sex, and years from diagnosis onset distribution.
Table 1. Overview of sex, age, and years to diagnosis onset for the 2 study cohorts.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>PD cohort (n=142)</th>
<th>HC cohort (n=77)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female</td>
<td>42 (29.6)</td>
<td>48 (62.3)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>100 (70.4)</td>
<td>29 (37.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (y), mean (SD)</strong></td>
<td>64.9 (9.4)</td>
<td>62.6 (12.4)</td>
<td>.08</td>
</tr>
<tr>
<td><strong>Years from diagnosis onset</strong></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>6.70 (5.36)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Range, n (%)</td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>0-4</td>
<td>60 (42.2)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td>39 (27.5)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>31 (21.8)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>15-19</td>
<td>6 (4.2)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>≥20</td>
<td>6 (4.2)</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

aPD: Parkinson disease.
bHC: healthy control.
cN/A: not applicable.

Patients with PD were labeled as either (1) *TremorDocumented*: tremor was documented in routine documentation, regardless of type or amplitude (78/142, 54.9%), or (2) *NoTremorDocumented*: no information regarding tremor in routine documentation (64/142, 45.1%). We compared the routine tremor documentation to a Tremor Occurrence Score (TOS) from smartwatch recordings of a 20-second task, during which participants remained seated in an armchair. The TOS is normalized, from 0=no tremor occurrence to 1=tremor occurrence with a clear frequency peak, with 0.5 representing the average TOS of healthy controls. Details on calculations can be found in Multimedia Appendix 1 [9,10].

**Ethical Considerations**

The SDS parent study has been registered (ClinicalTrials.gov NCT03638479) and approved by the ethical board of the University of Münster and the physician’s chamber of Westphalia-Lippe (reference 2018-328-f-S). All participants gave written informed consent, and data were pseudonymized.

**Results**

Figure 1 shows boxplots of the TOSs according to participants’ routine tremor documentation. Overall, 16 (25%) out of 64 patients in the *NoTremorDocumented* group had a TOS ≥0.75 (the median TOS in the *TremorDocumented* group) and an average amplitude of 0.07 g. Particularly, 10 (16%) patients in the *NoTremorDocumented* group had a clear tremor (TOS >0.88, amplitude >0.10 g). The *TremorDocumented* group had a TOS ≥0.75 and an average amplitude of 0.11 g. Overall, patients with a clear but low-amplitude tremor (TOS ≥0.75, amplitude <0.05 g) were 1.59 times less likely to have a documented tremor (*NoTremorDocumented*) than those with a nonsubtle tremor (TOS ≥0.75, amplitude >0.05 g).
Figure 1. Boxplots of Tremor Occurrence Scores (TOSs) for all participants grouped by their routine tremor documentation. Each data point represents a participant of the respective group. Points are colored according to the average wrist amplitude in grams. The middle line is the median. The top and bottom sections are upper and lower quartiles, respectively. The box represents the middle 50% of TOSs for the group. The upper and lower whiskers represent scores outside the middle 50%. PD: Parkinson disease.

Discussion

Smartwatches can capture clear tremors in certain patients with PD, wherein the tremor was not documented during routine neurological assessments. Generally, patients with a clear but low-amplitude tremor (<0.05 g) were 1.59 times less likely to have a documented tremor than those with a strong tremor. We hypothesize that the tremor may not have been documented during the examination because of its small amplitude and difficulty of being seen. Moreover, a few patients (10/64, 16%) in the NoTremorDocumented group showed a clear tremor with visible amplitudes (>0.10 g), possibly due to the tremor being unnoticed or absent during routine examination. Likewise, tremors may not always manifest during smartwatch-based data capture, ultimately affecting the findings’ reproducibility.

Nevertheless, we suggest that integrating smartwatch-based data capture into routine neurological assessments enhances tremor documentation by providing objective numbers on frequency and amplitude for biomarker analysis. Smartwatch integration would further enable improved longitudinal assessment at home. The findings are not suitable for assessing the potential of diagnostic improvements, as routine manual documentation was only compared to smartwatch-based data capture. The parent study is currently evaluating a larger patient cohort, including differential diagnosis, to evaluate the diagnostic accuracy by using machine learning and more detailed tremor features, other movement characteristics, and nonmotor symptoms [1].

Acknowledgments

We acknowledge support from the Open Access Publication Fund of the University of Münster. We thank the Department of Neurology at the University Hospital Münster for integrating the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Tremor Occurrence Score (TOS) and tremor characteristic calculation.

References


**Abbreviations**

PD: Parkinson disease

SDS: Smart Device System

TOS: Tremor Occurrence Score

---

Edited by A Mavragani; submitted 26.07.23; peer-reviewed by D Parry, J Lopez-Rodriguez; comments to author 08.09.23; revised version received 13.10.23; accepted 07.02.24; published 20.03.24.

Please cite as:

van Alen CM, Brenner A, Warnecke T, Varghese J

Smartwatch Versus Routine Tremor Documentation: Descriptive Comparison

JMIR Form Res 2024;8:e51249

URL: https://formative.jmir.org/2024/1/e51249

doi: 10.2196/51249

PMID:38506919

©Catharina Marie van Alen, Alexander Brenner, Tobias Warnecke, Julian Varghese. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Research Letter

A Health Information Technology Protocol to Enhance Colorectal Cancer Screening

Adam Baus¹*, MA, MPH, PhD; Dannell D Boatman²*, MS, EdD; Andrea Calkins¹*, MPH; Cecil Pollard¹*, MA; Mary Ellen Conn²*, MS; Sujha Subramanian³*, MA, PhD; Stephenie Kennedy-Rea²*, MA, EdD

¹Department of Social and Behavioral Sciences, School of Public Health, West Virginia University, Morgantown, WV, United States
²Cancer Prevention and Control, West Virginia University Cancer Institute, Morgantown, WV, United States
³Implenomics, Dover, DE, United States
*all authors contributed equally

Corresponding Author:
Adam Baus, MA, MPH, PhD
Department of Social and Behavioral Sciences
School of Public Health
West Virginia University
64 Medical Center Drive
PO Box 9190
Morgantown, WV, 26506
United States
Phone: 1 304 293 1083
Fax: 1 304 293 6685
Email: abaus@hsc.wvu.edu

Abstract

This study addresses barriers to electronic health records–based colorectal cancer screening and follow-up in primary care through the development and implementation of a health information technology protocol.

(JMIR Form Res 2024;8:e55202) doi:10.2196/55202

KEYWORDS
electronic health record; EHR; colorectal cancer screening; health information technology; cancer; colorectal cancer

Introduction

Cancer is a pressing global public health problem and the second leading cause of death in the United States, accounting for an estimated 1670 deaths daily [1]. Colorectal cancer (CRC) is the third most commonly diagnosed cancer, the second leading cause of cancer death worldwide [2], and the third most common cause of cancer-related deaths in the United States [3]. More effective use of health information technology (HIT), including electronic health records (EHRs), can aid in improving CRC screening and care [4]. Studies from as early as the 1990s have shown that EHRs and associated clinical decision support tools have promise in helping with patient care and population health needs [5]. However, barriers like clinician readiness [6] and clinical workflow integration [7] hinder EHRs’ full benefits. This study aims to address barriers to EHR-based CRC screening and follow-up through the development and implementation of a universally applicable EHR protocol tailored to identify and overcome practice workflow and EHR challenges.

Methods

Overview

This study used a mixed methods approach, involving quantitative and qualitative data collection techniques, conducted across 3 diverse health systems in West Virginia to develop and implement an EHR protocol for CRC screening and follow-up. These health systems were purposefully chosen to encompass diverse sizes, organizational structures, geographic locations, patient demographics, and EHR preferences, thereby supporting the generalizability of the study’s findings. These included a free and charitable clinic, a larger, urban, federally qualified health center, and a smaller, rural, federally qualified health center. Key stakeholders, including health care administrators, clinicians, and information technology personnel, were identified as potential participants. This study was conducted from April 2021 through April 2022. Implementation mapping methodology guided the assessment of current CRC screening practices and the development, implementation, and evaluation of the EHR protocol.
The developed protocol was implemented in Health Systems B and C following its refinement based on feedback from the development site (Health System A). Implementation involved training sessions for clinic staff on protocol utilization and ongoing support from the research team. Eight staff members from the participating health systems completed the protocol. The results demonstrate the successful development and initial implementation of an EHR protocol aimed at enhancing CRC screening in primary care settings. The protocol’s favorable reception by clinic staff, as indicated by high scores on acceptability, appropriateness, and feasibility measures, suggests its potential effectiveness in addressing identified barriers. The diverse representation of health systems and EHR platforms involved in the study enhances the generalizability of findings. Limitations include the small sample size and the focus on a specific geographic region. Future research will assess the protocol’s performance across additional EHR systems and health care settings for enhanced scalability and further evaluate the protocol’s impact on CRC screening outcomes.

Discussion

The results demonstrate the successful development and initial implementation of an EHR protocol aimed at enhancing CRC screening in primary care settings. The protocol’s favorable reception by clinic staff, as indicated by high scores on acceptability, appropriateness, and feasibility measures, suggests its potential effectiveness in addressing identified barriers. The diverse representation of health systems and EHR platforms involved in the study enhances the generalizability of findings. Limitations include the small sample size and the focus on a specific geographic region. Future research will assess the protocol’s performance across additional EHR systems and health care settings for enhanced scalability and further evaluate the protocol’s impact on CRC screening outcomes.

Acknowledgments

The authors acknowledge the funding and support from the Research Triangle Institute (grant 1-312-0216648-66244L).

Conflicts of Interest

None declared.

References


Abbreviations

AIM: Acceptability of Intervention Measure  
CRC: colorectal cancer  
EHR: electronic health record  
FIM: Feasibility of Intervention Measure  
HIT: health information technology  
IAM: Intervention Appropriateness Measure

©Adam Baus, Dannell D Boatman, Andrea Calkins, Cecil Pollard, Mary Ellen Conn, Sujha Subramanian, Stephenie Kennedy-REA. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Case Report

Examining Passively Collected Smartphone-Based Data in the Days Prior to Psychiatric Hospitalization for a Suicidal Crisis: Comparative Case Analysis

Ross Jacobucci¹, PhD; Brooke Ammerman¹, PhD; Nilam Ram², PhD

¹Department of Psychology, University of Notre Dame, Notre Dame, IN, United States
²Departments of Communication and Psychology, Stanford University, Stanford, CA, United States

Corresponding Author:
Ross Jacobucci, PhD
Department of Psychology
University of Notre Dame
390 Corbett Family Hall
Notre Dame, IN, 46556
United States
Phone: 1 574 631 6650
Email: rjacobucci@nd.edu

Abstract

Background: Digital phenotyping has seen a broad increase in application across clinical research; however, little research has implemented passive assessment approaches for suicide risk detection. There is a significant potential for a novel form of digital phenotyping, termed screenomics, which captures smartphone activity via screenshots.

Objective: This paper focuses on a comprehensive case review of 2 participants who reported past 1-month active suicidal ideation, detailing their passive (ie, obtained via screenomics screenshot capture) and active (ie, obtained via ecological momentary assessment [EMA]) risk profiles that culminated in suicidal crises and subsequent psychiatric hospitalizations. Through this analysis, we shed light on the timescale of risk processes as they unfold before hospitalization, as well as introduce the novel application of screenomics within the field of suicide research.

Methods: To underscore the potential benefits of screenomics in comprehending suicide risk, the analysis concentrates on a specific type of data gleaned from screenshots—text—captured prior to hospitalization, alongside self-reported EMA responses. Following a comprehensive baseline assessment, participants completed an intensive time sampling period. During this period, screenshots were collected every 5 seconds while one’s phone was in use for 35 days, and EMA data were collected 6 times a day for 28 days. In our analysis, we focus on the following: suicide-related content (obtained via screenshots and EMA), risk factors theoretically and empirically relevant to suicide risk (obtained via screenshots and EMA), and social content (obtained via screenshots).

Results: Our analysis revealed several key findings. First, there was a notable decrease in EMA compliance during suicidal crises, with both participants completing fewer EMAs in the days prior to hospitalization. This contrasted with an overall increase in phone usage leading up to hospitalization, which was particularly marked by heightened social use. Screenomics also captured prominent precipitating factors in each instance of suicidal crisis that were not well detected via self-report, specifically physical pain and loneliness.

Conclusions: Our preliminary findings underscore the potential of passively collected data in understanding and predicting suicidal crises. The vast number of screenshots from each participant offers a granular look into their daily digital interactions, shedding light on novel risks not captured via self-report alone. When combined with EMA assessments, screenomics provides a more comprehensive view of an individual’s psychological processes in the time leading up to a suicidal crisis.

(JMIR Form Res 2024;8:e55999) doi:10.2196/55999

https://formative.jmir.org/2024/1/e55999

JMIR Form Res 2024 | vol. 8 | e55999 | p.3531
(page number not for citation purposes)
Introduction

Background

In 2021, a total of 48,183 Americans died by suicide, with rates rising in the United States up to 2019 [1] and early indications of a continuing trend into 2022 [2]. While the number of suicides continues to increase, our ability to predict suicidal thoughts and behaviors (STBs) remains stagnant [3], underscoring the potential benefits of shifting focus from identifying who is at risk to when individuals are at risk for suicide. Studies using ecological momentary assessment (EMA) to collect data multiple times per day have demonstrated that key suicide risk factors, including suicidal ideation, change rapidly across the course of the day [4]. However, this research has been limited by the temporal granularity of assessments (ie, every 3-4 hours). It has yet to pay due attention to prospectively predicting suicidal behaviors (ie, suicidal planning and suicide attempts) [5]. Notably, the transition from contemplating suicide to deciding to act on such thoughts can transpire in as little as 10 minutes [6,7], making this decision-making process an ideal candidate for study in temporally intensive designs.

Intensive time sampling studies that use EMA have been shown to capture suicide risk processes effectively [8]; however, these methodologies carry a high participant burden, with the expectation that participants are willing and able to actively report on their experiences consistently throughout and across several days, including during times of distress. This contributes to the high rates of missing data within EMA designs, with compliance rates often around 60% in clinical populations [8]. Despite these limitations, asking self-report items at an intensive time scale has shed light on the dynamic nature of STB [4], with recent work becoming increasingly granular, prompting assessments every 10 minutes for an hour [9]. Furthermore, EMA research has also shed more light on the dynamic influence (both concurrent and prospective) of theoretically identified constructs [10], with a prominent focus on theoretically relevant risk factors, such as thwarted belongingness and perceived burdensomeness [11].

In contrast to active assessment approaches, the reliable detection of experiences within a relatively short window (eg, minutes) can be enhanced by technologically innovative methodologies that can passively and continuously capture the dynamic nature of factors contributing to suicide risk. This has provided, in part, the foundation for the surge in applications of passive sensing—including digital phenotyping [12]—in clinical research. Digital phenotyping is the use of data from smartphones and other personal digital devices to measure behaviors, cognitive functions, mood, and other psychological variables [13,14]. By collecting and analyzing data such as keystrokes, voice patterns, geolocation, social media activity, and other digital traces, digital phenotyping offers a real-time, high-resolution, yet low burden, approach to understanding and monitoring clinical phenomena.

Digital phenotyping has seen an increase in application across clinical research broadly, with the term being previously applied to suicide risk profiles based on active assessments (ie, EMA) [15]. However, little research has specifically targeted passive assessment approaches to suicide risk detection, with existing findings being mixed concerning its predictive validity. For example, in a recent study, Czyz et al [16] found that sensor-based assessments (resting heart rate, heart rate variability, step count, and sleep duration) did not add incremental validity to the prediction of near-term suicidal thoughts above and beyond EMA self-report questions among high-risk adolescents. Alternatively, it was demonstrated that psychological monitoring of electrodermal activity did improve fit upon EMA-only models in the prediction of suicidal ideation severity [17]. The limited work in this realm leaves open the possible use of passive assessment in better understanding suicide risk.

One of the most recently available methods in performing digital phenotyping is screenomics [18,19], wherein participants passively provide temporally intensive records of their psychological and social life as it manifests on the screens of their digital devices [19]. In this case, screenomes, or the record of experiences on digital devices with screens [19], are captured via software that takes screenshots of an individual’s smartphone screens every 5 seconds when the device is in use. Initial work using screenomics has described the timing and patterns of smartphone usage by extracting various textual and graphical features from screenomes, allowing for a mapping of app use and topic of content viewed [20], among other key use features such as the emotion valence of content viewed [21]. More recently, screenomes have informed the examination of digital communications, including differentiation between produced (ie, actively producing smartphone content) and consumed (ie, passive consumption) communication content [22], as well as the communication dynamics within relationships [23].

This Study

Following the possibility of using screenomics to learn about individuals’ emotional experiences and interpersonal dynamics, this study is the first, to our knowledge, to engage screenomics with a clinical population and, more specifically, for the study of suicide risk. This paper focuses on a comprehensive case review of 2 participants, detailing their passive (ie, obtained via screenomics screenshot capture) and active (ie, obtained via EMA) risk profiles that culminated in suicidal crises and subsequent psychiatric hospitalization. In our approach, we intentionally focused on building a qualitative, rather than quantitative, understanding of the data, purposely eschewing statistical modeling to capture the nuanced individual experiences and complexities inherent in suicide risk. Further, we iteratively followed inductive and deductive approaches.
[24,25] in deciding which variables to define (in passively collected data) and focus on (in actively collected data). Through this analysis, we shed light on the timescale of risk processes as they unfold prior to hospitalization, as well as introduce the novel application of screenomics within the field of suicide research. To underscore the potential benefits of this method in comprehending suicide risk, we concentrate on a specific type of data gleaned from screenshots—text—captured prior to hospitalization, alongside self-reported EMA responses.

**Methods**

**Overview**

This case review focuses on 2 participants who had suicidal crises and subsequent psychiatric inpatient treatment, coincident with their participation in an intensive time-sampling study that included both active and passive assessment approaches. To participate in the larger study, participants were required to have an Android smartphone, daily access to the internet (via Wi-Fi or cellular data), a past 1-month history of active suicidal ideation, and no history of being diagnosed with schizophrenia. Participants completed a baseline assessment, 28 days of active assessments obtained via signal-contingent EMA prompts, and 35 days of passive assessments obtained via screenshot collection. Participants were randomly assigned to 2 protocols where they provided the additional 7 days of passive screenshot collection (with no EMA prompts) in the week prior to the 28-day EMA+screenshot collection period (person A) or the week following the 28-day EMA+screenshot collection period (person B). For the 2 participants included in this study, data collection occurred before, during, and after an acute crisis and hospitalization spanning 38 (person A) and 51 (person B) days, respectively. A general overview of the data collected and how the screenshots were processed is depicted in **Figure 1**.

**Figure 1.** General overview of the study design. EMA: ecological momentary assessment; OCR: optical character recognition.

**Passive**

- Screenshots collected every 5 s phone is in use

**Content**

<table>
<thead>
<tr>
<th></th>
<th>Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone use</td>
<td>Screenshot time stamps</td>
</tr>
<tr>
<td>Suicide</td>
<td>OCR text</td>
</tr>
<tr>
<td>Theoretical risk factors</td>
<td>OCR text</td>
</tr>
<tr>
<td>Social content</td>
<td>Image classification and OCR text</td>
</tr>
</tbody>
</table>

**Active**

- EMA prompts 6× per day

**Ethical Considerations**

All procedures were approved by the institutional review board of the University of Notre Dame (protocol number 21-12-6965). Prior to engaging in any study procedures, participants were provided a full description of the study and the opportunity to ask questions, including speaking with study principal investigators (BA and RJ) or the institutional review board if any concerns were expressed. All participants provided written informed consent. They were compensated up to US $215 for participation (ie, US $40 for baseline assessment, US $100 for a 28-day EMA period, US $20 for completing at least 75% of EMA surveys, and US $55 for a 35-day screenomics period). Self-reported data from this study are deidentified; screenshot data are coded with layered IDs, encrypted upon storage upload, processed to remove potentially identifying information (ie, names and addresses), and accessible only to core study team members.

**Active Assessment**

**Overview**

Across the 28 days, participants received 6 prompts each day to complete a short survey on their smartphone (ie, signal-contingent design); prompts were delivered randomly within 2-hour windows across a 12-hour participant-selected block (eg, 9 AM to 9 PM). When participants received a prompt, they had 30 minutes (with one 15-minute reminder) to complete the short survey (expected to take <3-4 minutes). The EMA protocol was administered via the LifeData software on participants’ smartphones. To improve compliance and data quality, participants completed a “practice” EMA signal-contingent prompt during the baseline session and were incentivized for completing ≥75% of EMA signal-contingent responses (determined every week); EMA compliance was calculated based on the number of prompts received versus the number of prompts completed. We identified “high” scores for...
each variable depicted below by calculating whether an individual score was at or above the 90th percentile based on the person-specific empirical distribution.

**Risk Factors**

For this paper, we primarily focus on 2 key risk factors related to suicide: thwarted belongingness and perceived burdensomeness, as explained in the Interpersonal Psychological Theory of Suicide [26]. However, we also examined other important risk factors supported by prior research, including positive and negative emotions, urges, and behaviors related to nonsuicidal self-injury, alcohol and drug use urges, interpersonal conflict, self-reported sleep patterns, and physical pain [10]. The majority of these items were answered on a 5-point Likert scale, except interpersonal conflict, self-injurious and alcohol use behaviors, and nightmares, which were binary. Further, assessments of pain were reported on a 0-100 scale. Multimedia Appendix 1 [27,28] shows details for EMA questions to access risk factors.

**Suicide Risk**

**Suicidal Ideation**

Four questions assessed momentary (ie, “In this moment…”) passive (ie, “Life is not worth living for me”; “There are more reasons to die than to live”) and active (ie, “I want to die”; “I think about taking my life”) suicidal ideation, all items were answered on a 5-point Likert scale. Based on prior work by the study team, which showed that the 4 items form 1 factor (rather than 2 factors) [11], the item responses were summed to obtain a composite momentary suicidal ideation score.

**Suicidal Planning**

Three items assessed suicidal planning since the last prompt (ie, “Considered a specific suicide method”; “Identified how to acquire your suicide method”; “Made other preparations for your death”); these items were summed to create a composite suicidal planning score. All items were answered on a 5-point Likert scale.

**Suicidal Desire**

Two questions assessed momentary suicidal urges (ie, “How strong is your urge to make a suicide attempt?”; “How intense is your desire to kill yourself?”); items were answered on a 5-point Likert scale. These items were summed to create a composite suicidal desire score.

**Passive Assessment**

**Overview**

Continuous smartphone data collection was collected via the ScreenLife Capture [29] app, which is tailored for integration with HIPAA (Health Insurance Portability and Accountability Act of 1996)-compliant servers at the University of Notre Dame. This software passively captures screenshots of participants’ Android smartphone screens at 5-s intervals during smartphone use and stores the screenshots locally on the smartphone before bundling, encrypting, and transmitting them to research servers at intervals that accommodate bandwidth and device memory constraints. The screenshots are obtained continuously as the participant uses their smartphone in everyday life without requiring active engagement from the participant. In postprocessing, the screenshots’ content features are extracted using various image processing tools. In postprocessing, the text from each screenshot was extracted using EasyOCR, an open-source optical character recognition tool for Python, known for its efficiency in processing images containing text, even under less than perfect clarity or with intricate layouts. Optical character recognition technology converts various document types, such as scanned paper documents, PDF files, or images taken by a digital camera, into editable and searchable data. We removed any text extracted with a confidence value of less than 0.70. The text was then normalized by lemmatization, case reduction, and removal of punctuation and numbers and examined using a variety of dictionaries.

**Risk Factor–Related Content**

**Linguistic Inquiry and Word Count**

The first dictionary-based method that we tested is the linguistic inquiry and word count (LIWC) [30], which is implemented as a commercially available software program [31]. LIWC analyzes text according to a dictionary comprised of 6400 words, word stems, and emoticons, primarily related to psychological processes (eg, affective and social processes) [31]. The output contains scores on 90 dictionary components and linguistic or grammar dimensions of their data. For this paper, we only extracted components that we thought could be related to psychological processes or risk factors for suicide (eg, substances). Multimedia Appendix 2 [31] shows an overview of LIWC components. To display the scores, we opted to depict a ratio (score for each day/total) score for each instead of the absolute number of mentions for each category.

**Topic-Specific Dictionaries**

Since the specific words in the LIWC components are proprietary, we also created our own dictionary of common words likely associated with these components, particularly those crucial for understanding the suicide risk process. For example, we created a substances dictionary. Multimedia Appendix 3 contains words used in specific dictionaries.

**Social Content**

We identified which smartphone app was in use through the text portrayed when an app was in use. After examining participants’ screens to identify phone-specific app displays, we used the following words to identify text messaging (SMS): “text message,” “sms,” and “text,” while the following text for Meta: “write a comment,” “what’s on your mind,” “like,” “facebook,” and “Instagram.” We acknowledge that this dictionary will not perfectly distinguish between each app and similar apps, with a more robust method needed when extrapolating beyond these 2 individuals. Additionally, we used external data to train a model to identify whether the screenshot contained indicators of using a social app. We used the Rico data set [32], a repository of 66,261 screen images from Android operating systems. Each screenshot in Rico captures a momentary snapshot of a mobile app’s user interface. For our purposes, one of the crowdsourced annotations is “Is_Social” (4172 positive cases), which classifies whether or not a screen contains social media elements or functionalities. From this,
we randomly split the images into an 80/20 train-test split and fine-tuned ResNet18, which is a variant of the Residual Network family, a type of convolutional neural network architecture. The model was trained for 50 epochs, choosing a final model based on the test set area under the curve score, which was maximized at 0.89. We applied this model to each person’s screenshots, aggregating the predicted probabilities by day.

**Suicide Language**

We defined our dictionary to derive scores specific to suicide-related words (ie, “kill myself” and “end it all”) and a separate dictionary based on theoretically linked risk factors (ie, “trapped” and “crying”). These dictionaries were created through literature review and expert consultation. Multimedia Appendix 3 contains words used in specific dictionaries. Finally, we used the dictionary, termed crisis, which was developed by screening crisis chat messages [33]; this dictionary contains 227 words, ranging from specific states regarding suicidality (ie, “hurt myself” and “I just want this all to end”), suicide-related emotion states (ie, “hate my life” and “hopeless”), and potential suicidal behavior–related words (ie, “drown” and “rope”). To clean the text, we first removed screens that contained EMA prompts, namely, to avoid capturing suicide words that were used as a part of the EMA assessments.

**Results**

**Case Overview**

**Person: A**

**Overview**

Person A was a 37-year-old, straight, White, non-Hispanic individual born female and identifies as a woman. They reported being single, employed full-time, and with an annual household income of US $30,000-US $39,000. This participant’s diagnostic assessment revealed they met the diagnostic criteria for current major depressive disorder, panic disorder, agoraphobia, generalized anxiety disorder, posttraumatic stress disorder, and cannabis use disorder. They reported having engaged in prior individual counseling, as well as spiritual counseling, and that they were currently taking medication for psychiatric and medical reasons, including heart problems and epilepsy. Concerning suicide risk history, they reported chronic suicidal ideation, beginning at age 10 years and attempting suicide on 5 occasions, with the first attempt at age 11 years and the most recent 2 attempts in the year before study enrollment (1 in a month prior). In the week prior to enrolling in the study, they reported thinking about suicide 14 times and denied any suicidal planning.

**Provision of Active and Passive Data**

Person A responded to EMA prompts on 25 days; on these days, they, on average, only completed 4.04 prompts (SD 1.70), resulting in compliance of 60.1% (101 completed prompts out of 168 total prompts) or 67.3% counting only the days in which they responded at least once. They provided 139,276 screenshots. An average day involved 5.92 (SD 2.75) hours of screenshot data collection.

**Person B**

**Overview**

Person B was a 29-year-old, White, non-Hispanic individual who was born female and has a nonbinary gender identity; they reported questioning their sexual identity at the time of baseline assessment. They reported being employed full-time, living with a partner, and having an annual household income of US $40,000-US $49,000. This participant’s diagnostic assessment revealed they met current diagnostic criteria for bipolar I, as well as current panic disorder, social anxiety disorder, and substance use disorders (cannabis, amphetamine, and cocaine). They reported having previously engaged in several forms of outpatient treatment (ie, individual counseling, group counseling, couples counseling, and day treatment) and currently taking medication for medical and psychiatric reasons. Stated medical concerns included heart and respiratory problems.

Concerning suicide risk history, person B reported chronic suicidal ideation, experiencing suicidal thoughts on 95% of days, beginning at the age of 7 years. They also reported having attempted suicide on 4 occasions, with the first being at the age of 13 years, and the most recent 5 years prior to study involvement. In the week prior to enrolling in the study, they reported thinking about suicide 7 times and planning for suicide 3 times.

**Provision of Active and Passive Data**

Person B completed EMA surveys on 23 days, responding to, on average, 2.7 (out of 6; SD 1.63) prompts per day, resulting in a compliance of 37.5% (63 completed prompts out of 168 total prompts) or 45.6% counting only the days in which they responded at least once. Participant B provided 58,098 screenshots. An average day involved 3.59 (SD 2.74) hours of screenshot data collection.

**Comprehensive Suicide Risk Assessments**

To ensure participant safety, EMA responses that indicated elevations in imminent suicide risk triggered comprehensive risk assessments by study staff. Information obtained during these calls is not presented among active assessment results but rather provided in the Results section to provide a more in-depth contextual understanding of each detailed participant. During study enrollment, person A’s responses resulted in a comprehensive suicide risk assessment 9 times. Through these calls, person A expressed feelings of being overwhelmed and hopeless; they also expressed difficulties with having limited social support. They engaged in means restriction pertaining to their most recent suicide attempt method. Suicidal ideation was reported on these calls; however, the participant denied suicidal planning or intent. Day 26 in the study was the last time person A’s responses triggered a suicide risk assessment (study staff unable to reach them after 3 attempts) prior to their choice for hospitalization; however, on day 29 of participation, the participant reported self-admitting for psychiatric hospitalization due to concerns of being unable to keep themselves safe in the context of their suicidal thinking, and study staff spoke with them after they were on the way to the hospital due to EMA responses triggering a suicide risk assessment. They were hospitalized on days 29-32 of study participation.
Comprehensive Suicide Risk Assessments

During study enrollment, person B’s responses to the suicide-specific EMA questions triggered the study risk assessment protocol 8 times. During these calls, the participant expressed their suicidality was prompted by a panic attack, as well as interpersonal conflict that was impacting their housing stability. Suicidal ideation and suicidal planning were expressed in multiple calls; however, intent to act on these plans was low, and confidence in keeping themselves safe was high. They reported the use of several coping strategies, such as connecting with their care provider (resulting in adjustments to their psychiatric medication), means restriction, reviewing reasons for living, future planning, listening to music, sitting with their emotions, and deep breathing. However, on day 18 of participation, the participant attempted suicide and subsequently received inpatient psychiatric treatment from days 18 to 22; person B emailed the study staff to notify them of hospitalization after discharge (ie, day 22).

Comparative Analysis

In the following sections, we describe each person’s active assessment and passive assessment risk profiles surrounding their suicidal crises and subsequent hospital admission. The vertical red line in the visualizations represents the first day of each person’s hospitalization.

General Use Patterns

Active Assessment Compliance

For both individuals, we can see that EMA compliance was lowest surrounding their hospitalization. Particularly for person A (Figure 2), we see a decline in compliance with the active assessment protocol leading up to hospitalization, with a more precipitous decline for person B (Figure 3). For person A, the nonprovision of active assessments presents several problems for assessing suicide risk via EMA, as it suggests that some individuals may be less likely to respond to research survey prompts during periods when they are at the highest risk.

Figure 2. Person A: ecological momentary assessment compliance across the study. The vertical red line denotes the first day of hospitalization.

Figure 3. Person B: ecological momentary assessment compliance across the study. The vertical red line denotes the first day of hospitalization.
Phone Use

We see a different picture regarding phone use leading up to hospitalization, as depicted in Figures 4 and 5. While the number of passively collected screenshots stays relatively consistent leading up to the hospitalization for person A (Figure 4), we see an increase in phone use for person B (Figure 5) the day prior.

Phone use leading up to hospitalization can be further decomposed to look at the timing of when screenshots were collected within a given day; see Figures 6 and 7, where a horizontal red line denotes 50% phone use (ie, 30 minutes=360 five-second intervals) for that hour. For person A (Figure 6), we can see relatively consistent phone use during specific times of the day, along with periods in which their phone is not in use. Most days include at least 1 hour in which they were intensively using their phone (>50% time spent on the phone). In contrast, person B (Figure 7) was more variable in their phone use—some days included periods of intense use, while on others, phone use was sparse. Notably, neither person A nor B’s screenshot patterns indicate a consistent use pattern. The above phone use patterns do not suggest poor sleeping habits as revealed by periods of nonuse; however, we further examine this by considering phone use across hours of the day, aggregated across days. We calculated the percentage of each hour, on average, that each person used their phone (Figure 8: the horizontal red line represents 50% [ie, 30 minutes] average use for that hour). On average, neither person approaches using their phone 50% of the time. As mentioned earlier, person B (Figure 8: right) uses their phone less than person A (Figure 8: left). However, we see similar trends across individuals, with increasing phone use across the day. Notably, we also see some degree of either variable sleep schedules or sleep schedules being interrupted by phone use, as there are few hours of the day with no phone use for either person.
Figure 5. Person B: number of screenshots by day in study. The vertical red line denotes the first day of hospitalization.

Figure 6. Person A: screenshot count across days and hours of each day. The horizontal red line depicts 50% phone use for that hour.
Active Assessment

Risk Factors

Self-reported sleep indicators were assessed each morning. Subjective quality of sleep (sleep quality: bottom panel) and the presence of nightmares (top panel: nightmares yes or no) are depicted in Figure 6 (person A: Figure 9 and person B: Figure 10).

Compliance was particularly low for this set of questions. However, there are clear trends for both. While person A had consistent nightmares leading up to the hospitalization, their sleep quality improved (5=very good night of sleep). This is in contrast to person B, where nightmares increased and sleep quality decreased prior to the hospitalization.

Of theoretically and empirically relevant variables that we assessed via EMA, we do not see any increasing trends in perceived burdensomeness, thwarted belongingness, negative affect, or positive affect in Figures 11 and 12 (notably, however, there are a few spikes in negative affect for person A immediately prior to hospitalization). Figures 11 and 12 depict the empirically supported risk factors that demonstrated change leading up to hospitalization. We can see that both individuals reported consistent interpersonal conflicts across the study. However, there were no clear increases leading up to the hospitalization. For person A (Figure 11), we do see a slight increase in urges to drink alcohol; however, when we look at the responses related to alcohol use, they only reported drinking alcohol on 2 days. Further, both individuals report consistent urges to use drugs, with person B exhibiting a slight increase.
in urges prior to hospitalization. Finally, person A reported moderate urges to self-injury, with reported self-injury engagement on 1 occasion. However, there is no clear pattern of changes in risk factors prior to hospitalization.

**Figure 9.** Person A: ecological momentary assessment sleep responses across days in the study. Nightmares (yes or no) in the top panel, subjective sleep quality in the bottom panel. The vertical red line denotes the first day of hospitalization.

**Figure 10.** Person B: ecological momentary assessment sleep responses across days in the study. Nightmares (yes or no) in the top panel, subjective sleep quality in the bottom panel. The vertical red line denotes the first day of hospitalization.
Figure 11. Person A: risk factor ecological momentary assessment responses across days in the study. The blue line is a loess curve, while the vertical red line denotes the first day of hospitalization.

Figure 12. Person B: risk factor ecological momentary assessment responses across days in the study. The blue line is a loess curve, while the vertical red line denotes the first day of hospitalization.

Suicide Risk

EMA responses related to suicidal ideation, planning, and desire for persons A and B are depicted in Figures 13 and 14, respectively; loess curves depict general trends, with red data points denoting higher relative values (ie, above the 90th percentile). For person A, we can see generally increasing trends for self-reported suicidal ideation and suicidal desire leading up to their hospitalization, with the highest reported values of suicidal ideation on the day of hospitalization. For person B, we can see elevated responses of self-reported suicidal ideation and planning leading up to their hospitalization. However, these elevations dissipate in the days immediately prior. As noted in Figures 2 and 3, we see lower response rates for surveys, including questions assessing suicidal ideation, planning, and desires in the days prior to hospitalization.
Figure 13. Person A: suicide ecological momentary assessment responses across days in the study. The blue line is a loess curve, while the vertical red line denotes the first day of hospitalization.

Figure 14. Person B: suicide ecological momentary assessment responses across days in the study. The blue line is a loess curve, while the vertical red line denotes the first day of hospitalization.

Passive Assessment

Risk Factor–Related Content

To assess passive assessment indicators of sleep, we attempted to link sleep and wake times, obtained via active assessment once per day, to screenshot collection timestamps. However, person A reported sleep and wake times on 16 days, while person B reported on 14 days. Given the small number of responses for each person (ie, approximately 50% missing data) and variability in their reported sleep and wake times, we did not feel confident calculating whether screenshots were collected during reported sleep times. Thus, we opted against reporting and visualizing this.

For person A (Figure 15), we see peaks in several components leading up to their hospitalization, potentially most evident for the death and conflict scores. Probing further into the screenshot content during this day, we found that person A spent most of the day watching the TV show with “murder” in the title, likely contributing to these peaks. We also see a clear peak in the substances score prior to hospitalization. We created a substance-specific dictionary to further examine this (Multimedia Appendix 3). In applying this dictionary, we identified the top 5 words, and plot their use across the time period in Figure 16. To contextualize these findings, we examined a random selection of 1000 screenshots from the day prior to hospitalization. From these screenshots, it is evident that person A was experiencing a high degree of pain; this was corroborated by active responses of self-reported daily physical pain: although there were only 15 responses, their mean response was 74.4 (SD 16.2). They also disclosed through a SMS text message that an over-the-counter painkiller was not enough,
and they could not get into a doctor to receive a prescription until the following month. Further, given the high identification of “weed” in screenshots the day prior to hospitalization, we extracted co-occurring text (i.e., text from screenshots where “weed” was identified), and it became evident that person A was viewing posts about the medicinal use of marijuana on Facebook.

For person B (Figure 17), we see a number of components peak in the day prior to hospitalization. To contextualize these findings, we examined a random selection of 1000 screenshots from the day prior to hospitalization. We saw that they spent a large amount of time on a dating app (increase in sexual score in Figure 17) and conversing with individuals from that app. Further, like person A, person B was struggling with physical pain (specifically, head pain) and spent considerable time looking at health care providers and trying to identify a diagnosis, thus explaining the spike in a number of LIWC categories (i.e., physical and health). Notably, they only completed 10 active responses of self-reported daily physical pain, making it hard to derive any conclusions from their average score of 26.8 (SD 23.5).

We further examined risk factor text content on the day prior to hospitalization, as presented in Figures 18 (person A) and 19 (person B). For person A, a number of spikes in several risk factors were seen during the evening, whereas for person B, similar spikes were seen earlier in the day.

Figure 15. Person A: Linguistic inquiry and word count scores across days in the study. Note that scores are displayed as a ratio, using the daily score divided by the sum of all time points. The vertical black line denotes the first day of hospitalization.

Figure 16. Person A: substance use scores across days in the study. The vertical red line denotes the first day of hospitalization.
Figure 17. Person B: linguistic inquiry and word count scores across days in the study. Note that scores are displayed as a ratio, using the daily score divided by the sum of all time points. The vertical black line denotes the first day of hospitalization.

Figure 18. Person A: Linguistic inquiry and word count scores on the day before hospitalization.
Figure 19. Person B: Linguistic inquiry and word count scores on the day before hospitalization.

Social Content

Given the high amount of social contact for each person, we further examined changes in social behavior prior to hospitalization. Figure 20 shows the SMS text message and Meta (ie, Facebook, Instagram, and WhatsApp) use for person A. While the use of Meta apps does not show a clear change in the pattern of use relative to hospitalization, the use of text messaging peaks on the day of hospitalization. For person B (Figure 21), we see an increased use of Meta apps immediately prior to hospitalization, along with a slight increase in SMS text messaging in the days prior.

A similar conclusion can be drawn from the modeling of whether a screenshot contains social content in Figures 22 (person A) and 23 (person B). In examining a random sample of 1000 screenshots from the day of hospitalization, it appeared that person A was preparing for hospitalization by contacting friends to watch her dog and find a ride. Further, through disclosing her need for hospitalization, she was in frequent SMS text message conversations in which multiple individuals provided support for her decision. Alternatively, for person B, we see an increase in the use of social apps leading up to the day of hospitalization. However, on the day right before hospitalization, we do not see a large predicted total score of a number of screenshots containing social content.

Figure 20. Person A: SMS text message and Meta usage across days in the study. The vertical red line denotes the first day of hospitalization.
Figure 21. Person B: SMS text message and Meta usage across days in the study. The vertical red line denotes the first day of hospitalization.

Suicide Language
In applying the suicide-specific dictionaries, for person A (Figure 24), we only see the clear spike related to viewing the TV show. This is unsurprising given the other information gleaned from this participant’s screenshots, as their hospitalization was often discussed relative to pain, rather than suicidal thoughts or behaviors. Similarly, for person B (Figure 25), we see more of a peak with respect to risk factors and crisis words as opposed to suicide-specific words.

We then examined the top 10 endorsed individual risk words in Figures 26 and 27. For person A (Figure 26), we unsurprisingly see a large spike in pain on the day prior to hospitalization.
hospitalization. This is further seen in counting the word identification within the day prior to hospitalization (Figure 28).

A peak in pain is followed by similar peaks in guilty and alone later in the day. Similar to person A, for person B (Figure 27), we see pain as peaking on the day before hospitalization, in addition to smaller, but noticeable peaks in anxiety, alone, and depression, as well as numb, in the few days prior. When examining the day prior to hospitalization (Figure 29), we see that the identification of pain primarily occurred earlier in the day, followed by smaller peaks in alone, depression, numb, and anxiety.

**Figure 24.** Person A: suicide risk scores across days in the study. Each score was derived from applying each dictionary in Multimedia Appendix 3 to text extracted from the screenshots. The vertical red line denotes the first day of hospitalization.

**Figure 25.** Person B: suicide risk scores across days in the study. Each score was derived from applying each dictionary in Multimedia Appendix 3 to text extracted from the screenshots. The vertical red line denotes the first day of hospitalization.
Figure 26. Person A: individual risk words across day in the study. The vertical red line denotes the first day of hospitalization.

Figure 27. Person B: individual risk words across day in the study. The vertical red line denotes the first day of hospitalization.
Discussion

The goal of this comparative case review was to detail risk profiles obtained from both actively and passively collected smartphone data in the days leading up to a suicidal crisis. Through this, we also aimed to introduce the novel application of screenomics in suicide research. Our preliminary findings underscore the potential of passively collected data in understanding and predicting suicidal crises. The vast number of screenshots from each participant offers a granular look into their daily digital interactions, which, when combined with EMA assessments, provides a more comprehensive view of an individual’s psychological processes in the time leading up to a suicidal crisis. While this variegated picture of each individual’s life, as captured through a phone, makes a clean summarization impossible, it is worth noting some of the common elements evidenced in the data.

The first was decreased EMA compliance leading up to one’s hospitalization, muddying our understanding of an individual’s risk state, including the presence of STB. Perhaps most notably, this highlights the potential downsides of relying solely on one’s ability and willingness to disclose their experiences in times of distress or crisis, having important implications for the prediction of, and intervention during, such risk states. We recognize this is conjecture based on the analysis of only 2 individuals and requires future research as, to our knowledge, this has not been previously assessed or tested in intensive longitudinal studies of STBs. However, if true, it has important consequences for the rising use of just-in-time interventions targeting suicide risk [34]. For example, relying on active (i.e., EMA self-reported responses) may offer sufficient temporal granularity for intervening upon lower level risk states but may not be adequate for detecting crises given the burdensome design, particularly when in crises. Rather, as demonstrated in
our case examples, personalized phone use patterns might be a more reliable indicator of impending suicidal crisis than EMA-assessed suicidal ideation.

A second takeaway from our analysis is that screenomics captured the salient precipitating factors for each suicidal crisis, which were not detected via EMA, due to a variety of reasons (ie, nonresponse and ceiling effects). The first illustrative example is that of physical pain. Self-reports via EMA were able to capture the role of physical pain in suicide risk, to some extent, for person A, who responded to approximately 54% of daily questions about pain. However, for both person A and B, texts extracted from their screenshots were able to more fully portray the significant, and dynamic, nature of this experience leading up to one’s hospitalization. While we recognize this risk process may be specific to the individuals selected for inclusion in this report, an effect with similar implications was demonstrated for the theoretically relevant risk factor of thwarted belongingness. For example, EMA responses with regard to thwarted belongingness were elevated across the majority of the study period, preventing an observation of increases, or spikes, prior to hospitalization. Alternatively, for both person A and person B, we see that alone, a central aspect to the construct of thwarted belongingness (ie, “I am alone” [26]), represented one of the top 10 risk factors words used and also demonstrated a spike in use during the day immediately preceding hospitalization. Together, these examples underscore the potential of passive data collection, such as screenomics, to tap into risk processes that are not evident from, or demonstrated by, self-report assessments. Indeed, the low-burden nature of such passive assessment may allow for a more nuanced detection of a wide array of potential risk factors, lending to meaningful personalized risk detection.

Finally, and possibly the most surprising element was a general increase in phone use leading up to the hospitalization, including the increase in social use. It remains unclear as to whether this was a causal factor—that is, contributed to suicide risk—or was the result of seeking social support for their distress, making arrangements to be hospitalized (pet care or work), or other possible help-seeking behaviors, to cope with their escalating suicide risk. Further, the function of social phone use at or near times of suicidal crises may be person or situation specific (ie, idiographic). For example, person A appeared to be engaging in help-seeking behavior via SMS text message (eg, arranging rides) prior to hospitalization, while person B largely engaged in social and dating app usage, potentially contributing to acute interpersonal stressors. It will be useful to further develop our data processing pipeline to further explore the relationships between phone use, particularly social use, and suicide risk.

Of note, we purposely did not report any results from statistical models. While it is becoming more commonplace to apply linear models at the individual level [35], we opted against this. Our main focus for this comparative case report was on identifying risk patterns preceding hospitalization; the forgoing statistical models allowed us to explore a greater number of risk pathways in a more in-depth manner. Further, each participant had a limited number of assessments, with an increasing number of missed prompts leading up to the hospitalization (Figures 2 and 3), making the application of linear models difficult. Moreover, applying a longitudinal model to components derived from screenshots presents a number of challenges due to the high dimensionality, and uncertainty regarding forms of validity (eg, in naming components).

Within suicide research, and particularly intensive longitudinal research, ethical consideration of participant safety is a primary concern [36]. The incorporation of active assessments has the clear benefits of face validity to specific suicide questions (ie, “Are you planning to attempt suicide today?”), along with straightforward processing to derive risk score cutoffs that indicate required follow-up by study staff and the ability to readily access such data (ie, manually examining responses). As noted above, there are potential issues with relying on active assessment responses (ie, noncompliance [37,38]). However, the use of passively collected data (broadly defined) also presents a number of challenges for use in ascertaining risk. The first is the clear lack of published research on which data sources are valid markers of risk. Second, while some data points are easily calculated (ie, the number of screenshots and steps walked), others would require complex and expensive processing pipelines to process data in real time. For instance, in screenshots presented in this case study, this would require automatic decryption of uploaded images to a server that is integrated with a graphics processing unit (or central processing unit) architecture that can apply pretrained models to generate new data sources (eg, SMS text messages) or predictions (eg, suicide risk scores based on the text). Similarly, most smartwatches require manual syncing with software before data are uploaded to the cloud, thus precluding real-time analysis. Thus, a large degree of additional research is required before incorporating various types of passive assessment into real-time participant safety protocols.

A number of limitations to this paper should be mentioned. First, as this is the first comprehensive investigation of smartphone screenshot data to improve our understanding of suicide risk, aspects of our approach are preliminary in nature, including the use of specific (ie, suicide, risk factor, and substance) dictionaries. It will be important for further validation of these dictionaries prior to wide-scale applications. Our focus on studying 2 individuals with a complex case profile limits the generalizability of our findings across the wider spectrum of severity of suicidal ideation, planning, and behaviors. Consequently, our results may not fully represent the experiences of those at lower levels of suicide risk (ie, not experiencing a suicidal crisis). The findings may also not generalize to individuals at a similar risk level but do not consider psychiatric hospitalization as a treatment option. Further research with larger and more diverse samples is necessary to validate and extend our findings to a broader population. Finally, we only included one form of passive sensing, screenomics, while a number of other modalities, such as geo-location or watch-based assessments (ie, sleep, steps, and heart rate), have seen wide application in clinical research [39].

In conclusion, the primary aim of this paper was to provide an in-depth examination of both passive and active data modalities from 2 individuals who were hospitalized for suicidal crises. We hope this paper has introduced the value of collecting
passive data, specifically screenshots, at an intensive timescale, as well as highlighting the potential of its ability to incrementally add to our understanding of suicide risk timescale and underlying processes.

Acknowledgments
This project was funded by the National Institute of Mental Health (1R21MH129688) awarded to BA and RJ.

Data Availability
The screenshot data set collected and analyzed during this study is not available from the authors at this time. However, it will eventually be accessible through the NIMH National Data Archive. Derivative data sets used to create each figure are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Ecological Momentary Assessment (EMA) questions to assess risk factors.
[DOCX File, 13 KB - formative_v8i1e55999_app1.docx]

Multimedia Appendix 2
Linguistic Inquiry and Word Count (LIWC) components [30].
[DOCX File, 14 KB - formative_v8i1e55999_app2.docx]

Multimedia Appendix 3
Words used in specific dictionaries.
[DOCX File, 14 KB - formative_v8i1e55999_app3.docx]

References


Abbreviations

EMA: ecological momentary assessment
HIPAA: Health Insurance Portability and Accountability Act of 1996
LIWC: linguistic inquiry and word count
STB: suicidal thoughts and behavior
Tailoring of Health-Promotion Video Messaging for Reproductive-Aged Women at Risk for Developing Cardiometabolic Disease: Qualitative Focus-Groups Study

Jacqueline Kent-Marvick1*, BSN; Bryan Gibson2*, DPT, PhD; Alycia A Bristol1*, RN, AGCNS-BC, PhD; Stephanie St Clair1*, BSN; Sara E Simonsen1*, CNM, MSPH, PhD

1College of Nursing, University of Utah, Salt Lake City, UT, United States
2School of Medicine, University of Utah, Salt Lake City, UT, United States
*all authors contributed equally

Corresponding Author:
Jacqueline Kent-Marvick, BSN
College of Nursing
University of Utah
10 South 2000 East
Salt Lake City, UT, 84112
United States
Phone: 1 4356686932
Email: jacqueline.kent-marvick@utah.edu

Abstract

Background: Targeting reproductive-aged women at high risk for type 2 diabetes (T2D) provides an opportunity for prevention earlier in the life course. A woman’s experiences during her reproductive years may have a large impact on her future risk of T2D. Her risk is 7 to 10 times higher if she has had gestational diabetes (GDM). Despite these risks, T2D is preventable. Evidence-based programs, such as the National Diabetes Prevention Program (DPP), can reduce the risk of developing T2D by nearly 60%. However, only 0.4% of adults with prediabetes have participated in the DPP to date and reproductive-aged women are 50% less likely to participate than older women. In prior work, our team developed a mobile 360° video to address diabetes risk awareness and promote DPP enrollment among at-risk adults; this video was not designed, however, for reproductive-aged women.

Objective: This study aims to obtain feedback from reproductive-aged women with cardiometabolic disease risk about a 360° video designed to promote enrollment in the DPP, and to gather suggestions about tailoring video messages to reproductive-aged women.

Methods: Focus groups and a qualitative descriptive approach were used. Women with at least 1 previous pregnancy, aged 18 to 40 years, participated in one of three focus groups stratified by the following health risks: (1) a history of GDM or a hypertensive disorder of pregnancy, (2) a diagnosis of prediabetes, or (3) a BMI classified as obese. Focus-group questions addressed several topics; this report shared findings regarding video feedback. The 3 focus-group discussions were conducted via Zoom and were recorded and transcribed for analysis. Deductive codes were used to identify concepts related to the research question and inductive codes were created for novel insights shared by participants. The codes were then organized into categories and themes.

Results: The main themes identified were positive feedback, negative feedback, centering motherhood, and the importance of storytelling. While some participants said the video produced a sense of urgency for health-behavior change, all participants agreed that design changes could improve the video’s motivating effect on health-behavior change in reproductive-aged women. Participants felt a tailored video should recognize the complexities of being a mother and how these dynamics contribute to women’s difficulty engaging in healthy behaviors without stirring feelings of guilt. Women desired a video with a positive, problem-solving perspective, and recommended live links as clickable resources for practical solutions promoting health behavior change. Women suggested using storytelling, both to describe how complications experienced during pregnancy impact long-term health and to motivate health behavior change.

Conclusions: Reproductive-aged women require tailored lifestyle-change messaging that addresses barriers commonly encountered by this population (eg, parenting or work responsibilities). Moreover, messaging should prioritize a positive tone that harnesses storytelling and human connection while offering realistic solutions.
Introduction

Targeting reproductive-aged women at high risk for type 2 diabetes (T2D) provides an opportunity for prevention earlier in the life course [1]. This is especially important as a woman’s health status during her reproductive years may have a large impact on her future risk of T2D [2-5]. Her risk is 7 to 10 times higher if she has had gestational diabetes (GDM) [2,5]. Despite these risks, T2D is preventable. Evidence-based programs, such as the National Diabetes Prevention Program (DPP), can reduce the risk of developing T2D by nearly 60% [6,7]. However, very few adults with prediabetes participate in the DPP and reproductive-aged women are 50% less likely to participate than older women [8]. Our team earlier developed a 360° video to increase risk perception of T2D. The video helps participants conceptualize T2D complications throughout the lifespan and is designed to promote DPP enrollment among at-risk adults. It was developed with input from Hispanic community members and was not designed specifically for reproductive-aged women [9]. Because few studies have attempted to understand methods motivating reproductive-aged women to engage in healthy lifestyle change, this study aims to obtain feedback from reproductive-aged women with cardiometabolic disease risk factors in order to tailor messaging for this population.

Methods

Overview

Designed for middle-aged Hispanic men and women with prediabetes, the 360° video can be viewed with a man’s or woman’s voice and in English or Spanish. Viewers can move their phones around to look at the video’s world [10]. The video depicts an individual who has made unhealthy eating choices and has forgone dental care, thereby increasing their risk for T2D. It shows progression from prediabetes through diabetes leading to a heart attack. The video ends with a statement that diabetes can be prevented through the DPP.

Eligible participants included women who have had at least 1 previous pregnancy, were 18 to 40 years of age, and had one of the following three health risks: (1) history of GDM or hypertensive disorder of pregnancy, (2) diagnosis of prediabetes, or (3) BMI classified as obese. Those with GDM or prediabetes were eligible regardless of BMI. The 3 focus groups were conducted via Zoom by an interviewer from the University of Utah’s Center for Clinical and Translational Science Community Collaboration and Engagement Team (CCET) who had been trained in focus-group best practices and community-engaged research. Focus-group questions enabled key themes relating to the topic to be identified [11]. Focus-group participants were recruited by a CCET study coordinator using flyers. These were posted in the University of Utah cardiology and obstetrics and gynecology offices, and in Facebook and LinkedIn groups within the CCET’s network. The coordinator screened for eligibility and reviewed the study’s consent cover letter with potential participants. The CCET supports community-engaged research that addresses investigator- and community-identified priorities with the goal of promoting community health needs [12].

Data collection included audio-recording the 1-hour-long focus-group discussions between February and April of 2022. Participants were asked to share their feedback on ways to modify the story, the voice-over content, and the imagery to make the video more impactful and appropriate for reproductive-aged women. The focus-group interviewer prompted participants to provide additional details about their perspectives and experiences. Additionally, the interviewer noted commonalities and differences between participant experiences to foster further insight. Members of the research team were present at the focus groups to answer questions and guide follow-up. The research team had no prior relationship with participants. Audio recordings of the focus groups were transcribed by a research assistant and verified for accuracy by additional team members.

Analysis was completed by 4 team members trained in qualitative methods and women’s health. Qualitative content analysis was used to examine the data. Deductive codes were used to identify concepts related to the research questions (eg, positive and negative feedback). Inductive codes were created for novel insights shared by participants (eg, importance of using storytelling and centering the video’s focus on family relationships). The codes were then organized into categories and themes. Lincoln and Guba’s trustworthiness criteria [13] were used, including weekly coding meetings with authors (AAB, SES, JK-M, and SSC) to discuss nuances in the data to ensure consistency in the application of codes. Coding across transcripts was compared to assess similarities and differences between participant experiences not only as women who had given birth, but also as women with differences in health risks. To address reflexivity during analysis, a group memo document was maintained and reviewed weekly. A member of the team audited group notes and coding to assess for consistency and accuracy. No members had a personal history of hypertensive disorder of pregnancy, GDM, or T2D, but all had clinical and research expertise. Our assumption was that participants would provide critical feedback on the video.

Ethical Considerations

Before the study’s initiation, waiver-of-consent documentation was received from the University of Utah Institutional Review Board (IRB 00146243). Participants received a consent cover letter containing information about the study’s purpose, and potential risks associated with participation; this information was reviewed with participants before focus-group sessions began, and the voluntary nature of participation was emphasized. Participants received a US $75 gift card for participation in a
demographic survey, pre-event Zoom technical session, and focus-group discussion. This amount was suggested by the CCET and was approved and deemed noncoercive by the IRB.

**Results**

**Overview**

In total, 20 women participated in the 3 focus groups (see Table 1 for demographics).

The main themes were positive video feedback, negative video feedback, centering motherhood, and the importance of storytelling. See Textbox 1 for exemplar quotes.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>34.2 (6.46)</td>
</tr>
<tr>
<td>Age (years), Range</td>
<td>26-48</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
</tr>
<tr>
<td>18-28</td>
<td>6 (30)</td>
</tr>
<tr>
<td>29-39</td>
<td>10 (50)</td>
</tr>
<tr>
<td>40-48</td>
<td>4 (20)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>2 (10)</td>
</tr>
<tr>
<td>White or European</td>
<td>15 (75)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Unreported</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latina</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Non-Hispanic or Latina</td>
<td>17 (85)</td>
</tr>
<tr>
<td>Partnership status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Married or living with a partner</td>
<td>15 (75)</td>
</tr>
<tr>
<td>Divorced</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Separated</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Single</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Associate degree</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>14 (70)</td>
</tr>
<tr>
<td>Master degree</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Setting, n (%)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Suburban</td>
<td>13 (65)</td>
</tr>
<tr>
<td>Urban</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Number of live births, n (%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5 (25)</td>
</tr>
<tr>
<td>2</td>
<td>6 (30)</td>
</tr>
<tr>
<td>3</td>
<td>8 (40)</td>
</tr>
<tr>
<td>4</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Number of children living in home, n (%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5 (25)</td>
</tr>
<tr>
<td>2</td>
<td>8 (40)</td>
</tr>
<tr>
<td>3</td>
<td>6 (30)</td>
</tr>
<tr>
<td>4</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Prior formal lifestyle program attendance, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8 (40)</td>
</tr>
<tr>
<td>No</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Variable</td>
<td>Participants</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Weight (lbs), mean (SD)</td>
<td>179.025 (40.6740)</td>
</tr>
<tr>
<td>Weight, range</td>
<td>125-280</td>
</tr>
</tbody>
</table>

Textbox 1. Themes and exemplar quotes.

**Positive video feedback**

- I thought the information was interesting. I knew that, like, eye doctors often catch diabetes, but I wasn’t aware about the dental connection. So, it was information that I… I personally didn’t realize that dentists would give that kind of information about diabetes, or potential risks.
  
  [Participant #5; obesity focus group; Hispanic or Latino]

- When I watched it, I was like, ‘Oh, I need to get my crap together and I don’t want to be that.’ … I felt somewhat motivated to want to, like, improve my life. Like, I don’t know, but it’s more my personality just to be very direct.
  
  [Participant #3, pregnancy complication focus group; White, Asian, non-Hispanic]

**Negative video feedback**

- I don’t do well with scare tactics. And, so, it’s just… That kind of style is really off-putting to me. It kind of scares me away from it more.
  
  [Participant #1; obesity focus group; White, non-Hispanic]

- I think it lacks empathy. … While I was watching it, I felt a lot of guilt and shame, um, which is something I think if you’re trying to reach women specifically of childbearing age that that should not be something that I think should be the first thing that ignites in someone.
  
  [Participant #7; prediabetes focus group; Hispanic or Latino]

**Centering motherhood**

- I would much prefer something that was like super fun and encouraging and saying, you know, like, about being excited about having energy to play with my kids, and being around, you know?
  
  [Participant #1; obesity focus group; White, non-Hispanic]

- I don’t think the fear tactic is gonna work. Um, like, that’s like, I feel like that’s somebody that’s like yelling at me to go to the gym, ‘You’re just being lazy.’ Like, ‘Just get up and go.’ Like, ‘Disregard your feelings and your postpartum depression and your postpartum anxiety and all of those other things’.
  
  [Participant #7, pregnancy complication focus group; White, non-Hispanic]

**Importance of storytelling**

- I think our just our innate human beings, like we really love connection. … What really would have got me, is if I saw like a mother. … and how she was given that news at the dentist. And saw her progress with her family.
  
  [Participant #5; prediabetes focus group; Native Hawaiian or Pacific Islander]

- Maybe a mother and daughter, where the mother is struggling with the beginnings of these issues and shares the information with her daughter to help prevent from developing issues in the future.
  
  [Participant #2, pregnancy complication focus group; White, non-Hispanic]

**Positive and Negative Video Feedback**

Feedback about the videos was generally negative. While some participants said the video produced a sense of urgency for healthy behavior change, all participants agreed that it did not motivate health-behavior change in reproductive-aged women.

**Centering Motherhood**

The participants felt that a tailored video should convey empathy for the complexities of motherhood and recognize how these dynamics contribute to women’s difficulty engaging in healthy behaviors. Participants emphasized that this tailored messaging should convey urgency. They recommended messages that avoided stirring feelings of guilt. Participants desired a video with a positive, problem-solving perspective. They suggested live links could be embedded within the video as clickable resources for practical solutions, helping to promote health behavior change (eg, links to the DPP website, food-preparation tips, and ways to incorporate children into physical activity).

**Importance of Storytelling**

One woman suggested that an optimal approach to the video would summarize a woman’s personal risk from her doctor’s point of view (eg, “you’ve had gestational diabetes, so this is what that means for your long-term health risk”), and then quickly pivot to a problem-solving tone: “Now here’s how we’re going to help you get to where you need to be.” Women across the 3 focus groups agreed that storytelling would be effective in describing how pregnancy complications impact long-term health. For example, a participant suggested the video could include a mother talking with her adult daughter about developing T2D and how she wishes she had acted earlier for
prevention. The mother could also encourage her daughter to begin finding solutions while still young and able to maintain health. Other participants agreed that a focus on storytelling and human connection would improve the video’s effectiveness. Finally, the women suggested that it should end by illustrating the benefits of healthy behavior change (eg, being able to keep up with their kids).

**Discussion**

In focus group interviews with reproductive women at risk for T2D, participants were enthusiastic about the use of 360° videos to promote healthy-lifestyle change and DPP enrollment, but many found the video to be negative or guilt-inducing. Suggestions for change included positive messages and avoidance of guilt-based messaging. Additionally, participants felt that messaging should acknowledge the time demands and challenges faced by reproductive-aged women, including parenting and work demands, and should convey urgency without creating feelings of guilt.

Few efforts have been made at tailoring diabetes-prevention messaging for reproductive-aged women. An example is the Health-e mums study, which used smartphone app–based messaging to target women with a history of GDM. The app prompted women to improve their health behaviors through goal-setting, and personalized responses to the recording of body weights, diet, and activity. Women provided feedback on messages received through the app. As was true of our study, some Health-e mums participants disliked the use of “negative storytelling,” instead preferring positive, health-coaching messages [14]. Recently, the Centers for Disease Control and Prevention (CDC) released a 1-minute-long promotional DPP video entitled “More Energy Means More Family Fun” [15] that is tailored to reproductive-aged women. It is unclear if the video has been successful in motivating their target population to enroll in the DPP.

Study strengths include participant characteristics (eg, racial or ethnic diversity) and a range of cardiometabolic risk factors. Limitations include the small sample size and the homogeneity of the participants, including the fact that all were mothers. Thus, findings may not apply to women who are not mothers.

In sum, participants highlighted the importance of acknowledging shared experiences and struggles of motherhood, while also offering realistic solutions and a positive tone through storytelling. These findings underscore the importance of tailoring health behavior change messaging to address unique needs of reproductive-aged women and to motivate enrollment in lifestyle change programs. Such messaging is essential for engaging this population in lifestyle change so as to reduce disease burden.

**Acknowledgments**

The authors of this paper give thanks to the University of Utah’s Center for Clinical and Translational Science Community Collaboration and Engagement Team for their collaboration on this project and to Roger Altizer and Greg Stoddard for their assistance with developing the video. This study was supported by funding from the Educational Resource Development Council at the University of Utah. Additionally, author JK-M received support from the National Institute of Nursing Research of the National Institutes of Health under award number (F31NR020431). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. These funding sources had no impact on the study’s design and data collection, or the analysis, interpretation, and writing of the study’s results.

**Authors’ Contributions**

The conceptualization of the study is attributed to SES and BG. SES, BG, and JK-M created the interview guide for the focus groups. The design and implementation of the qualitative analysis was done by AAB. Data analysis was done by AAB, JK-M, SES, and SSC. The manuscript was written by AAB, JK-M, SES, BG, and SSC. All authors approved the final version for publication.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

CCET: Community Collaboration and Engagement Team
CDC: Centers for Disease Control and Prevention
DPP: Diabetes Prevention Program
GDM: gestational diabetes
IRB: institutional review board
T2D: type 2 diabetes

©Jacqueline Kent-Marvick, Bryan Gibson, Alycia A Bristol, Stephanie St Clair, Sara E Simonsen. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.